This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Low-Cost Per-Core Voltage Domain Support for Power-Constrained High-Performance Processors

Abhishek A. Sinkar, Hamid Reza Ghasemi, Michael J. Schulte, Ulya R. Karpuzcu, and Nam Sung Kim, *Senior Member, IEEE*

***Abstract*— Per-core voltage domains can improve performance under a power constraint. Most commercial processors, however, only have a single voltage domain for all processor cores. This is because splitting the single voltage domain into per-core voltage domains and powering them with multiple off-chip voltage regulators (VRs) incur a high cost for the platform and package designs. Although using on-chip switching VRs can be an alternative solution, integrating high-quality inductors for VRs with cores has been a technical challenge. In this paper, we propose a cost-effective power delivery technique to support per-core voltage domains. Our technique is based on the observations that: 1) core-to-core (C2C) voltage variations are relatively small for most execution intervals when the voltages/frequencies are optimized to maximize performance under a power constraint and 2) per-core power-gating devices augmented with feedback control circuitry can serve as low-cost VRs that can provide high efficiency in situations like 1). Our experimental results show that processors using our technique can achieve power efficiency as high as those using the per-core on-chip switching VRs at a much lower cost.**

***Index Terms*— On-chip voltage regulators (VRs), per-core voltage domains, multicore processors.**

I. Introduction

THE maximum performance of multicore processors operating all available cores is typically limited by a power constraint. When: 1) a processor runs multiple threads or applications and 2) the instructions per cycle (IPC) of each core varies notably within each execution interval, adjusting the voltage/frequency (V/F) of each core can improve performance under a power constraint. This approach, however, requires a voltage domain (and thus a voltage regulator (VR)) for each core. In contrast, most commercial processors have only a single voltage domain for all processor cores. This is because splitting the single voltage domain into per-core domains and powering them using multiple off-chip VRs incur a high cost for the platform and package designs. Alternatively, on-chip switching VRs (i.e., buck converters) integrated with cores can lower the cost associated with the platform and package designs. Yet, no commercial high-performance multicore processors have adopted the on-chip switching VRs because integrating high-quality inductors for the switching VRs with cores is expensive and technically challenging.

In this paper, we propose a cost-effective power delivery technique to support per-core voltage domains using on-chip low-drop-output (LDO) VRs. The on-chip LDO VRs have not previously been considered for high-performance multicore processors, because they cannot power-efficiently provide a wide voltage range for cores; an LDO VR suffers from a high power loss when the voltage difference between its input and output voltages (V_I and V_O) is large. On the other hand, platform architects must assume that the core voltage can be of any value within a given voltage range at which the processor can operate. However, our experiments reveal that the maximum voltage difference between cores at each dynamic V/F scaling (DVFS) interval is small for most intervals, when the V/F of each core is optimized to maximize the performance under a power constraint. In other words, the V_O values of multiple LDO VRs can be close to the shared V_I value. In such a case, an LDO VR, which can be implemented at a low cost by sharing its largest component with an existing on-chip per-core power-gating (PCPG) device, can be more power-efficient than a switching VR.

Note that commercial high-performance multicore processors from Intel and AMD have supported PCPG [1], [2]. This is because PCPG can considerably improve not only power efficiency but also the performance of power-constrained multicore processors [3], [4] with a small investment of chip area (5%–10% of each core's area [5]). Consequently, we envision that future high-performance multicore processors, which have not supported PCPG yet, will adopt PCPG as the performance improvement is limited by a power constraint that barely scales with technology. The key contributions of this paper, which extends our previous paper presented at [6], are as follows.

- We demonstrate that a high-performance multicore processors using per-core voltage domains can deliver higher performance than using a single-voltage domain under a power constraint. Then, we discuss the challenges associated with supporting per-core voltage domains (Section II).
- We demonstrate that C2C voltage variations are relatively small at each DVFS interval, when the V/F of each core is optimized to maximize the performance of a processor under a power constraint. Then, we propose a lost-cost power delivery technique that exploits

Manuscript received May 30, 2012; revised December 31, 2012; accepted March 14, 2013. This work was supported in part by generous grants from AMD, IBM, the Wisconsin Alumni Research Foundation, the National Science Foundation (CCF-095360), and the DARPA Cooperative Agreement HR0011-12-2-0019.

A. A. Sinkar, H. R. Ghasemi, M. J. Schulte, and N. S. Kim are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: sinkar@wisc.edu; hamid@cs.wisc.edu; schulte@engr.wisc.edu; nskim@engr.wisc.edu).

U. R. Karpuzcu is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: ukarpuzc@umn.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.2013.2257900

1063-8210/$31.00 © 2013 IEEE

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

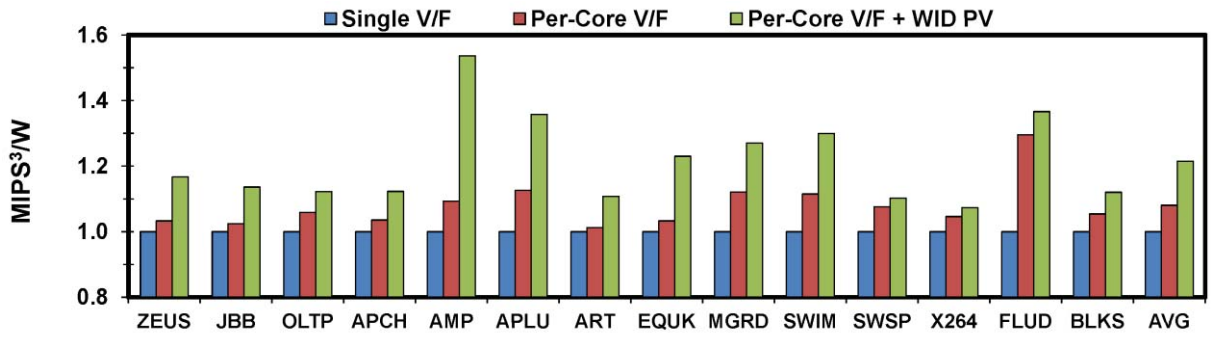

Fig. 1. MIPS^3/W comparison of eight-core processors supported by a single V/F domain, per-core V/F domains, and per-core V/F domains exploiting WID PV. Each DVFS interval is composed of 10 million instructions.

1) small C2C voltage variations and 2) PCPG devices (Section III).

- We demonstrate that using our proposed technique is as effective as using the on-chip per-core switching VRs for running both single- and multithreaded applications (Section IV).
- We discuss a cost-effective power delivery solution for many-core processors (Section V).

The remainder of this paper is organized as follows. Sections II–V describe our key contributions. Section VI discusses related work. Section VII concludes this paper.

II. Per-Core Voltage Domain Support: Opportunities and Challenges

A. Performance Improvement Opportunities Using Per-Core Voltage Domain Support

Providing per-core voltage domains enables per-core DVFS. This allows a multicore processor to effectively exploit runtime performance variations across cores running single- or multithreaded applications in a given execution interval. For example, some cores running threads in memory-intensive phases can operate at lower V/F without impacting the performance while other cores executing threads in compute-intensive phases must operate at higher V/F to maximize the performance. The C2C performance variations increase as more cores are integrated on a chip, which increases the opportunity for improving power efficiency using per-core DVFS. Consequently, many researchers have investigated various per-core DVFS algorithms and their physical implementations to either minimize power consumption under a performance constraint or maximize performance under a power constraint (e.g., [7], [8]).

Fig. 1 compares the million instructions per second cubed per Watt (MIPS^3/W) of eight-core processors that are supported by a single V/F domain for all processor cores, per-core V/F domains, and per-core V/F domains exploiting within-die (WID) process variations (PVs). WID PVs lead to C2C frequency and power variations for the same voltage applied to cores. All the results are normalized to MIPS^3/W of an eight-core processor with a single V/F domain (without considering the power efficiency of on- and off-chip VRs). We use four commercial applications (Apache, JBB, OLTP, and Zeus denoted by APCH, JBB, OLTP, and ZEUS) [9], six SPEC OMP V3.2 benchmarks (ammp, applu, art, equake, mgrid, and swim denoted by AMP, APLU, ART, EQUK, MGRD, and SWIM), and four PARSEC benchmarks (swaptions, x264, fluidanimate, and blackscholes denoted by SWSP, X264, FLUD, and BLKS) [10] running on a GEMS multicore simulator modified to support per-core frequency domains [11]. An oracle DVFS algorithm [7] is modified to maximize MIPS^3/W for a given power constraint. Because we focus on high-performance multicore processors, we choose MIPS^3/W to emphasize the performance aspect of processors [12]. See Section IV for a more detailed description of our experimental methodology.

Supporting per-core V/F domains increases a geometric mean of MIPS^3/W by 8% (22% when C2C frequency and power variations are exploited) over the single V/F domain. As we increase the number of cores per processor, we observe that processors with per-core V/F domains achieve even more MIPS^3/W improvement than a single V/F domain. For example, the MIPS^3/W improvement of 12- and 16-core processors with per-core V/F domains is nearly 3× and 4× higher than that of an eight-core processor with per-core V/F domains. The increase becomes even larger when the C2C PVs are exploited. This signifies the growing importance of providing per-core V/F domains to maximize performance under a power constraint. Furthermore, supporting per-core V/F domains can allow multicore processors to more effectively exploit C2C frequency and power variations, thereby significantly increasing the performance and power efficiency.

B. Technical Challenges for Supporting Per-Core Voltage Domains

Regardless of performance improvement opportunities using per-core V/F domains, most commercial multicore processors have only a single V/F domain for all processor cores. This is because splitting the single-voltage domain into per-core voltage domains and powering them with multiple off-chip VRs incur a high cost for the platform and package designs. Fig. 2 illustrates one of the negative impacts of splitting the single-voltage domain to provide per-core voltage domains, i.e., increasing the overall VR capacity required. Assume that the maximum power consumption of the processor is limited to 120 W and there are four cores. When all four cores are running, each core can consume up to 30 W. Thus, it seems that each per-core VR needs only to support up to 30 W. However, for example, when only two out of four cores are

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

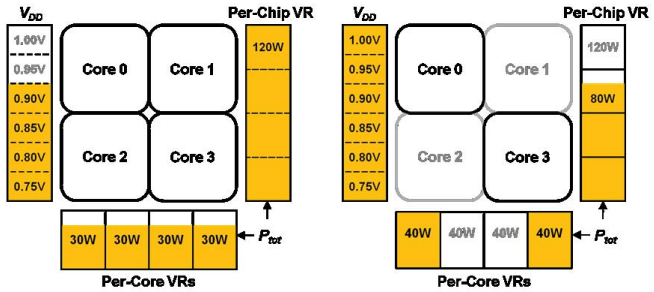

Fig. 2. Impact of splitting the chip-wide voltage domain into per-core voltage domains on the overall VR capacity. (a) All cores are active and consume a total of 120 W. (b) Only two cores are active.

active due to limited parallelism, the two active cores can run at higher V/F (e.g., Intel Turbo Boost Technology [3]) without violating their thermal and power constraints. If the two active cores consume 40 W at such an operating V/F, the capacity of each VR needs to be increased to 40 W and the total combined capacity of all the VRs becomes 160 W.

When the voltage domain is shared, however, a 120 W VR is still sufficient for such a case; the total power consumption of two cores running at the turbo mode is a total of 80 W, which is below the maximum capacity of the VR. Although it is feasible for only a subset of cores to run in turbo mode, we cannot increase the VR capacity for only a subset of the cores. This is because cores are put into turbo mode in a round-robin fashion to prevent excessive aging of a specific core or subset of cores, requiring us to provide the capacity for turbo mode for all the cores. Finally, increasing WID PVs lead to substantial C2C frequency and power variations [13]. In other words, some cores consume notably more power than others due to a high fraction of leakage power in total power (e.g., ~30% [14]) and a large variation of the leakage power across cores. Thus, the per-core VR capacity is determined by such cores, increasing the overall VR capacity even further.

The increased total power capacity requires larger components for VRs and more package pins for power delivery. Note that form-factor is critical even for server platforms to maintain high integration density in data centers, and VRs are the second largest components next to DRAM modules; VRs occupy 63% more platform area than the CPU, the third largest component [15]. Furthermore, many commercial chips are heavily constrained by the available pins; nearly half of all pins are already dedicated for power delivery and the increased overall VR capacity requires more pins. Although the platform and package cost associated with multiple off-chip VRs can be lowered by using the on-chip switching VRs [16], integrating the cores and high-quality inductors for the VRs on the same chip has been also a major technical challenge for manufacturers, potentially impacting both the efficiency of the VRs and the yield of dies [17].

III. LDO VRs Exploiting Small C2C Voltage Variations and PCPG Devices

A. C2C Voltage Variations

Fig. 3 shows the maximum voltage difference between cores for the "Per-Core V/F" case in Fig. 1 is less than or equal to 100 mV for at least 90% of the execution intervals in most applications where the voltage change is restricted to increments of 50 mV. We also observe similar statistics for the "Per-Core V/F + WID PV" case. In other words, the maximum voltage difference between cores in a processor with per-core V/F domains is not large at each execution interval. In such a case, the power loss by LDO VRs can be lower than the switching VRs when a proper V_I value for LDO VRs is selected to minimize the difference between the V_I value and the V_O values across cores, as briefly discussed in Section I. Furthermore, an LDO VR can be implemented very cost-effectively since: 1) it does not require a large inductor and/or capacitor [18] and 2) it can share its largest component (i.e., the output device) with a PCPG device.

B. PCPG-Based LDO VRs

PCPG devices are provided for most commercial multicore processors to reduce standby leakage power of idle cores. In active state, a PCPG device incurs a slight voltage drop across it (i.e., between the supply voltage and the actual voltage applied to the core). The voltage drop is inversely proportional to the size (i.e., total transistor width) of the PCPG device for a given amount of total current (dynamic + leakage) drawn by the core. In fact, the voltage applied to the core can be modulated by controlling the effective width (i.e., resistance) of the PCPG device [19].

A PCPG device, which is implemented with many parallel transistors and on/off signal buffers, is similar to the largest component (i.e., the output device between V_I and V_O) of a typical LDO VR, as illustrated in Fig. 4(a). In other words, an LDO VR can be implemented by a PCPG device augmented with feedback control circuitry composed of an error amplifier, an analog-to-digital converter, and a reference voltage generator; it was reported that the output device and its buffers, both of which can be shared with a PCPG device, accounted for 83% of the total LDO VR area [18]. Since a PCPG device consumes 5%–10% of a core's area [5], we estimate that the extra overhead due to the feedback control circuitry to implement the LDO VR is less than 2% of the core's area. In contrast, the on-chip switching VRs require a large inductor and/or capacitor. As a result, a switching VR has at least four times larger chip area than a comparable LDO VR [20]. Furthermore, LDO VRs can provide faster transient responses than the switching VRs [21] and, unlike the switching VRs, they do not inject the switching noise in the substrate. This is desirable for the operation of highly sensitive mixed signal circuits.

Fig. 4(b) shows two different approaches to distribute supply voltages to an eight-core processor with per-core V/F domains. Both approaches use a first-stage off-chip VR to convert 5 V to an intermediate voltage level, V_I of the on-chip per-core VRs; we cannot supply 5 V for the on-chip switching VRs directly due to the oxide-reliability of nanoscale transistors implementing both VRs and cores. This voltage is further down converted using the on-chip per-core VRs to the voltage ($V_O[i]$) required by core i. The arrangement on the left uses

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

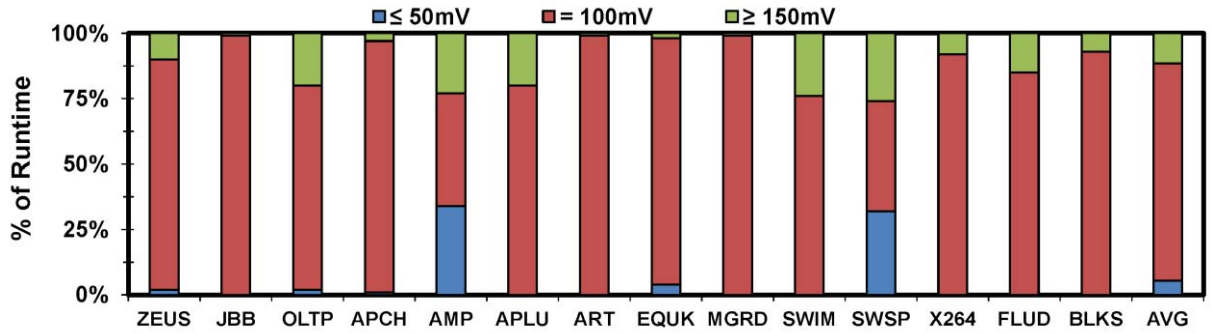

Fig. 3. Fraction of execution intervals exhibiting equal to or less than 50 mV, equal to 100 mV, and equal to or larger than 150-mV voltage differences between cores for the "Per-Core V/F" case presented in Fig. 1.

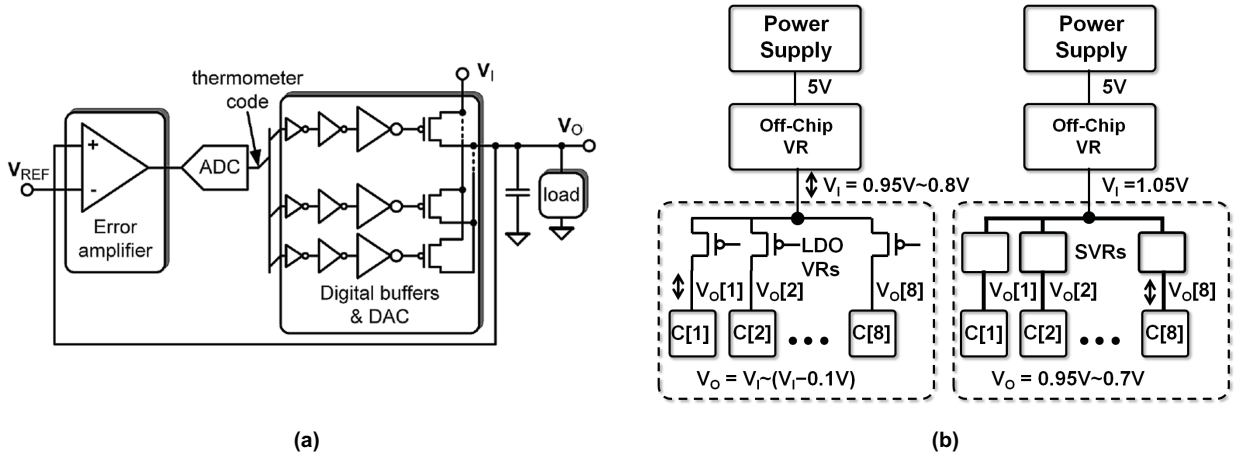

Fig. 4. (a) A typical LDO VR architecture; the image is reproduced from [18]. (b) An example of V_I and V_O ranges of LDO VRs in the left and switching VRs denoted by SVRs in the right for supporting per-core voltage domains; for illustration purposes, we ignore the default voltage drop of the LDO VRs due to the small resistance of the fully turned-on PCPG devices in the figure. "$C[i]$" in (b) denotes core i.

LDO VRs (i.e., PCPG devices augmented with the feedback control circuitry to implement LDO VRs). The efficiency of an LDO VR is a function of its V_O/V_I ratio.

When the voltages demanded by individual cores are restricted to a limited range (e.g., within 100 mV of one another), a high V_O/V_I ratio can be achieved for all the cores by adjusting the V_O of the first stage (i.e., V_I of the second stage) such that it is sufficient to provide the highest V_O demanded by any of cores. Thus, a processor adopting per-core LDO VRs can be tuned to achieve high efficiency by jointly optimizing both their V_I and V_O. The arrangement on the right uses the per-core on-chip switching VRs to provide the necessary core voltage. A switching VR uses two active devices, an inductor and a capacitor, to provide high voltage conversion efficiency across a wide range of values of V_O. This efficiency is primarily determined by the switching losses in the active devices and their conduction losses. The V_I value for the switching VRs is fixed to 1.05 V in this example.

C. Efficiency Comparison: LDO Versus Switching VRs

Fig. 5 compares the efficiency of a switching VR with that of an LDO VR (the on-chip second stage only in (a) and both the off- and on-chip stages in (b), respectively). The efficiency of the LDO VRs is higher than that of the switching VRs when V_I–V_O is small (or V_O/V_I is high), but it becomes lower as V_I–V_O increases (or V_O/V_I decreases). If V_I–V_O is more than 100 mV, the efficiency of LDO VRs usually is lower than that of switching VRs, as shown in Fig. 5(a). We model the efficiencies of both switching and LDO VRs assuming that each core consumes the maximum allowed current for each operating voltage. To measure the maximum efficiency of the switching VR at each operating point (i.e., voltage/current), we search for and activate the optimal number of phases out of eight available phases for a given voltage/current; Table I summarizes the key design parameters of various VR stages described in this paper.

The off-chip switching VR efficiency computation is based on [22] with V_I fixed at 5 V. Off-chip switching VR designs built with off-the-shelf components typically have very high efficiencies ($>$ 90%) due to a low-loss high-quality inductor and capacitor. Their efficiency reaches a maximum value for a certain load current and then drops with further increase in current due to an increase in conduction losses. Consequently, as the on-chip regulator voltage V_O for LDO VRs decreases, the efficiency degrades, and thus the overall efficiency of LDO VRs becomes slightly lower than that of the switching VRs,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

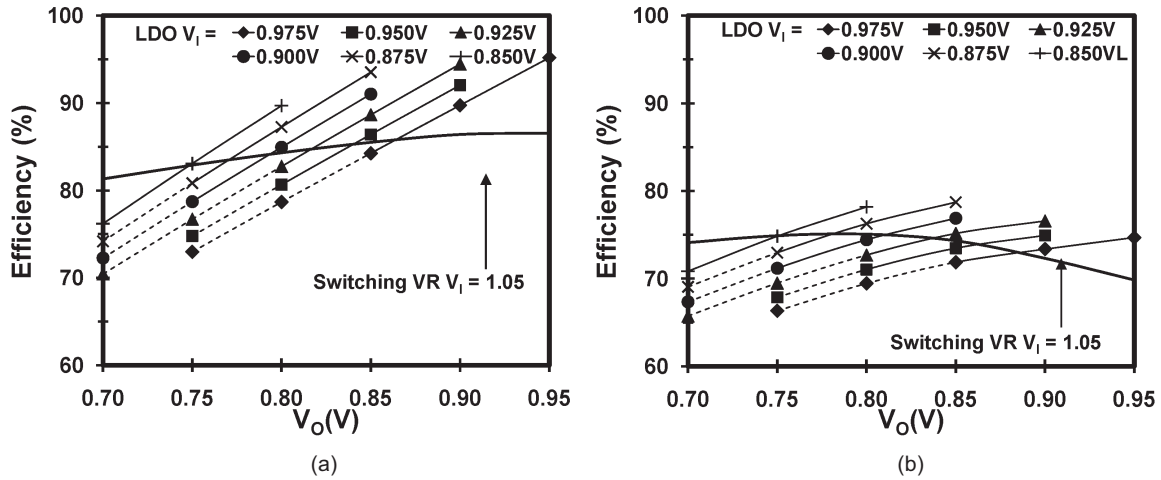

Fig. 5. Efficiency comparison between switching and LDO VRs. The efficiency of (a) the on-chip (the second stage) only and (b) both the off- and on-chip (the first and second stages) are considered.

TABLE I
SUMMARY OF VR DESIGN PARAMETERS

	Off-Chip Switching VR	On-Chip Switching VR	On-Chip LDO VR
V_I/V_O	5 V/1.05 V to 5 V/0.85 V	1.05 V/0.95 V to 1.05/0.7 V	0.95 V/0.7 V to 0.7 V/0.95 V
Technology	N/A	32 nm	32 nm
f_{sw}	300 KHz	100 MHz	N/A
L/phase	360 nH ($r_L = 0.5$ mΩ)	63.5 nH ($Q = 20$ @100 MHz)	N/A
No. of Phases	6	8	N/A

as plotted in Fig. 5(b). The efficiency of an LDO can be calculated as

$$\eta_{LDO} = \frac{V_O \cdot I_O}{V_O \cdot I_O + (V_I - V_O) \cdot I_O + V_{bias} \cdot I_Q} \quad (1)$$

where I_Q is the quiescent current of the LDO and V_{bias} is the biasing voltage for the reference and feedback control circuitry. A steady analog $V_{bias} = 0.9$ V generated on chip from the variable V_I is assumed in this paper. The current efficiency of a typical LDO VR is defined by [23] as

$$\eta_I = \frac{I_O}{I_O + I_Q} \quad (2)$$

where η_I is a measure of the power loss in the control and biasing circuitry of the LDO. On-chip LDO designs with current efficiencies in the range of 95%–99% have been reported. The LDO efficiency is computed assuming a current efficiency of 97% at I_O corresponding to 120 W/0.9 V [18].

The efficiency of the switching VRs with an integrated inductor was modeled for different CMOS technology generations in [24]. The efficiency is mainly a function of the inductor Q factor. An inductor in a CMOS process is made from the available metal layers, and it attains low Q values for realistic dimensions due to the substrate losses and frequency-dependent conduction losses. It was shown that a fully monolithic switching VR achieves ∼62% efficiency with on-die inductors in 90-nm CMOS [24]. Such low efficiency is not acceptable considering the performance benefit that can be brought by per-core voltage domains under a power constraint. The efficiency can be improved by using alternate inductor technologies with high Q. This may include inductors with magnetic materials compatible with a CMOS process or inductors mounted externally on the package while only the active devices are integrated on die [16]. A switching VR with 80%–87% efficiency with integrated active devices and on-package inductors ($Q = 20$) was demonstrated [16].

Our efficiency analysis assumes a 32-nm CMOS process with inductor ($Q = 20$ @ 100 MHz) similar to [16], since switching VR with an on-die inductor exhibits poor efficiency; on-package inductors incur packaging design and integration issues, but we will not discuss them in detail in this paper. The design is optimized to achieve a conversion ratio of 1.05 V/0.9 V at a load current of 16.67 A per core (corresponding to total of 120 W for eight cores at 0.9 V) with an efficiency of 88%. An eight-phase topology is used with 63.5 nH inductance per phase. As V_O and load current are reduced, the efficiency of the switching VRs decreases monotonically. This is because the switching loss constitutes a higher percentage of the output power as the V_O value reduces. The efficiency is strongly dependent on the operating point at which the switching VR design is optimized. For a design optimized for higher V_O, the efficiency at low output voltage drops more rapidly compared to a design optimized for lower V_O [24]. In Fig. 5(a), the on-chip switching VRs are optimally designed for $V_O = 0.8$ V.

IV. EVALUATION

A. DVFS Algorithms for MIPS^3/W Comparison

The key objective of this section is to evaluate the effectiveness of the LDO VRs derived from PCPG devices.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II
SUMMARY OF DVFS ALGORITHMS EXPLORED IN THIS STUDY

Algorithms	Voltage Domain	Frequency Domain	Process Variation Aware	Thread Migration	Off-Chip VR V_O	On-Chip VR	Constraint
ShV/F	Shared	Shared	No	No	Varying	N/A	$V_{O1} = V_{O2} = \cdots = V_{ON}$
SeV/F	Separate	Separate	No	No	Fixed	SVR	
SeV/F(PV)			Yes	No			
SeV/F(PV/TM)			Yes	Yes			
LDOSeV/F	Virtually Separate	Separate	No	No	Varying	LDO VR	$V_I - V_{Oi} \leq 100$ mV
LDOSeV/F(PV)			Yes	No			
LDOSeV/F(PV/TM)			Yes	Yes			

Notes: "Sh," "Se," "PV," and "TM" denote "Shared," "Separate," "Process Variation," and "Thread Migration," respectively.

Thus, we can use various per-core DVFS algorithms optimized for high-performance multicore processors including the algorithms exploiting: 1) C2C frequency and power variations and 2) thread migrations (TMs) [8]. For the evaluation, we adopt an integer linear programming (ILP) method for the DVFS algorithms. The ILP formulation is similar to the one used in [7], which aims to minimize the power consumption of a multicore processor for a given performance constraint. We modify the formulation to search for the optimal V_O for each core to maximize MIPS under a power constraint at each DVFS interval, which comprised ten million instructions, as follows.

Objective:

$$\text{maximize} \left(\sum_{i=1}^{N} MIPS_i = \sum_{i=1}^{N} \sum_{j=1}^{M} IPC_i, \cdot F_{ij} \cdot x_{ij} \right). \quad (3)$$

Constraints:

$$\sum_{i=1}^{N} \sum_{j=1}^{M} P_{ij} \cdot x_{ij} \leq P_{totmax} \quad \text{and} \quad \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij} \leq N$$
$$\text{s.t. } \exists i : \sum_{j=1}^{M} x_{ij} = 1 \quad (4)$$

where N is the number of cores; M is the number of V_O steps supported by a DVFS algorithm; $MIPS_i$ and IPC_i are the MIPS and IPC of core i; F_{ij} is the frequency of core i at voltage level j; x_{ij} corresponds to one bit of an M-bit binary variable for core i that is guaranteed to assign core i to only one of M possible V/F states; P_{ij} is the power consumption of core i, which is a function of $V_O[i]$; P_{totmax} is the allowed total power consumption of the processor; and (4) is the constraint, respectively. In (4), the second constraint is to enforce one V_O selection for each core. The V_I for all LDO VRs is determined by taking the maximum value among $V_O[1], V_O[2], \ldots,$ and $V_O[N]$.

As discussed in [8], this algorithm requires manufacturers to store per-core frequency and power values at each voltage level for DVFS algorithms to exploit C2C frequency and power variations. These values can be characterized by the manufacturer and stored, along with many other processor-tuning parameters, in a nonvolatile memory of the processor. Like other DVFS algorithms, we also need to predict workload characteristics like the IPC of each thread to assign a proper V/F to each core for the next DVFS interval. Although we can use various methods to predict the IPC of the next interval based on the current IPC, we assume that the IPC value of each thread at every interval is known in advance (as with an oracle method). This is to isolate the impact of the IPC prediction from the MIPS^3/W results so that we can fairly compare the efficacy of the two different VR schemes. Finally, we adopt a simple scheme for the TM technique; we assign threads to cores one-to-one in the order of IPC and frequency values. For example, the thread with the highest IPC is assigned to the core with the highest frequency at a given voltage (i.e., the fastest core considering C2C frequency variations). Table II summarizes the DVFS algorithms explored in this paper and constraints for specific algorithms. Our baseline processor has a single V/F domain using an off-chip VR (i.e., ShV/F).

B. Architecture Simulation Environment

Our processor configuration contains eight cores. Each core is four-wide with 32-kB private L1 caches and a shared 512-kB L2 cache. The cores are connected to each other using crossbar switches. We evaluate different DVFS algorithms using a full-system cycle-level simulator, GEMS [11], after we modify GEMS to support per-core frequency domains and TM that requires L1 cache flushing. We evaluate four commercial applications (Apache, JBB, OLTP, and Zeus), six SPEC OMP V3.2 benchmarks (ammp, applu, art, equake, mgrid, and swim), and four PARSEC benchmarks (swaptions, x264, fluidanimate, and blackscholes) [10]. In addition, we also evaluate five mixes of compute- and memory-bound SPEC2006 benchmarks (eight copies of bzip2, six copies of bzip2 and two copies of libquantum, four copies of bzip2 and four copies of libquantum, two copies of bzip2 and six copies of libquantum, and eight copies of libquantum denoted by B8L0, B6L2, B4L4, B2L6, and B0L8, respectively) with the processor simulation parameters summarized in Table III.

C. Core Frequency and Power Models

Our objective is to maximize the performance for a given maximum power consumption constraint. Typically, an operating system (OS) determines V/F of cores based on a given power management algorithm, but both the OS and VRs

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
SUMMARY OF PROCESSOR SIMULATION PARAMETERS

Fetch/Issue/Retire	4/4/4	# of Cores	8
IL1	32kB/4-way/64B 3 cycles	Branch Predictor/BTB/RAS	YAGS/1K/32
L2	512kB/8-way/64B 10 cycles	DL1	32kB/4-Way/64B 3cycles
Cache Coherency Protocol	Directory-based MESI	Main Memory (size/block/page/latency)	DDR3-1.6 GHz 4 GB/64 B/4 kB/7-7-7 -20 ns
# of MSHRs	8	Write-buffer entries	16

cannot track and respond to instantaneous changes of power consumption. Thus, the OS must conservatively assume the power consumption of cores at each given operating V/F and guarantee that the entire chip does not exceed its maximum power consumption, if it aims to optimize performance without violating a power constraint at any given moment.

To model the maximum power consumption of cores, we assume that: 1) the total maximum power consumption of eight-cores is 120 W and 2) 30% of the total power is active leakage [14], [25] at 0.9 V. Each core has its own shared L2 cache that shares the V/F domain with the core. Thus, we assume that the L2 power scales with the core power consumption. The power consumption of I/O and other peripheral components including on-chip interconnects, which are tied to other separate fixed V/F domains, is not included in our analysis since it can be regarded as a fixed power cost for all the cases we explore in this paper; I/O and on-chip interconnects are responsible for $\sim$15% of the total power in Niagara 2 [25].

Due to WID C2C frequency and leakage power variations, the power consumption of each core differs. To analyze the impact of WID PVs on the frequency and leakage power consumption of each core, we first generate 100 variation maps for the threshold voltage (V_{th}) and effective channel length (L_{eff}) of transistors in a die and characterize frequency and power consumption by following the methodology presented in [13]: WID correlation distance coefficient $\phi = 0.5$ and WID V_{th} and L_{eff} variations $\sigma_{\text{sys}} = 6.4\%$ and 3.2% of the nominal V_{th} and L_{eff} values, respectively. We apply the V_{th} and L_{eff} values of each grid point to an FO4 inverter chain and a dummy circuit—which is composed of 50% inverters, 30% NAND gates, and 20% NOR gates—to obtain the frequency and leakage power scaling factors of each grid, respectively; NAND and NOR gates in a dummy circuit can have up to four inputs and their inputs are assigned randomly with either one or zero.

Second, we measure the frequency and leakage scaling factors of each grid point at 0.95 to 0.7 V using a 32-nm technology model and SPICE. We assume that the frequency of each core is determined by the slowest grid point in the core [13] and the frequency of the slowest core is 3.2 GHz at 0.9 V. Then, each core's maximum dynamic power consumption at 0.9 V is $(F_i / \sum_{j=1}^{N} F_j) \cdot 120\ W \cdot 0.7$ where F_i and F_j are the frequencies of cores i and j, and N is the number of cores. With the known frequency, voltage, and dynamic power values, we can calculate the maximum core-switching capacitance (i.e., C_{dyn}). This allows us to calculate the dynamic power at any given V/F. The leakage power of each grid point is scaled such that the sum of the leakage power from all grid points in a die is equal to 30% of 120 W at 0.9 V. The sum of the scaled leakage power from all the grid points belonging to a particular core becomes the core's leakage power.

Finally, we allow some cores to run at V/F higher than 0.9 V/3.2 GHz as long as the total power constraint is satisfied; this is possible when other cores run at V/F lower than 0.9 V/3.2 GHz. Since all cores in our baseline processor run at the same frequency, the dynamic power consumption of the processor is lower than when other processors use per-core V/F domains. Thus, we increase the V/F of the processor until 120 W is fully used (i.e., 0.9125 V and 3.3 GHz).

Note that C2C frequency and power variations change across different dies. However, for our analyses, we pick a typical die map from the 100 generated maps because a large amount of simulation time is required to repeat the same experiment for hundreds of die maps. Thus, the MIPS^3/W results, which exploit C2C frequency and leakage power variations, represent the value close to the median value of the 100 die maps. Table IV tabulates the frequency and power consumption of each core as function of $V_O[i]$. For each core the frequency (GHz) and power (Watts) are given in the left and right columns, respectively.

D. MIPS^3/W Comparison

1) Impact of Limiting V_I–V_O Range on MIPS^3/W: Fig. 6 compares MIPS^3/W of eight-core processors using an on-chip LDO and switching VRs. In this experiment, we do not include the power consumption (i.e., power loss) by both on- and off-chip VRs when we calculate MIPS^3/W to observe the impact of constraining the V_O range of LDO VRs. Although WID C2C PV is not exploited and the TM technique is not applied, the MIPS^3/W difference between LDOSeV/F and SeV/F is around 2% (3% versus 5% improvement over ShV/F) on average (i.e., geometric mean). However, when WID C2C PV is exploited, the MIPS^3/W difference between the processors using the LDO and switching VRs becomes 1% (14% versus 15% improvement over ShV/F) on average. Finally, the MIPS^3/W difference between the two schemes leads to a 3% difference (22% versus 25% improvement over ShV/F) on average when both the TM technique is applied and WID C2C PVs are incorporated with the DVFS algorithms. Finally, exploiting C2C frequency/power variations and TMs can mitigate the potential limitation of LDO VRs

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
SUMMARY OF FREQUENCY AND POWER CONSUMPTION OF EACH CORE AS A FUNCTION OF $V_O[i]$

$V_O[i]$	Core 1		Core 2		Core 3		Core 4		Core 5		Core 6		Core 7		Core 8	
0.95 V	3.6	15.9	3.8	17.9	4.4	18.2	3.9	16.9	3.8	17.2	4.1	19.3	4.4	29.1	4.1	19.7
0.90 V	3.2	12.5	3.4	14.0	4.0	14.5	3.5	13.3	3.4	13.5	3.7	15.1	4.0	22.0	3.7	15.3
0.85 V	2.8	9.6	3.0	10.8	3.5	11.3	3.0	10.2	3.0	10.4	3.3	11.7	3.5	16.6	3.3	11.8
0.80 V	2.4	7.2	2.5	8.1	3.1	8.6	2.6	7.7	2.4	7.8	2.8	8.8	3.1	12.3	2.8	8.9
0.75 V	2.0	5.3	2.1	6.0	2.6	6.4	2.2	5.7	2.1	5.7	2.3	6.5	2.6	9.0	2.4	6.6
0.70 V	1.6	3.7	1.7	4.2	2.1	4.6	1.7	4.0	1.7	4.1	1.9	4.6	2.1	6.4	1.9	4.7

Note: For each core, the frequency (GHz) and power (Watts) are given in the left and right columns, respectively.

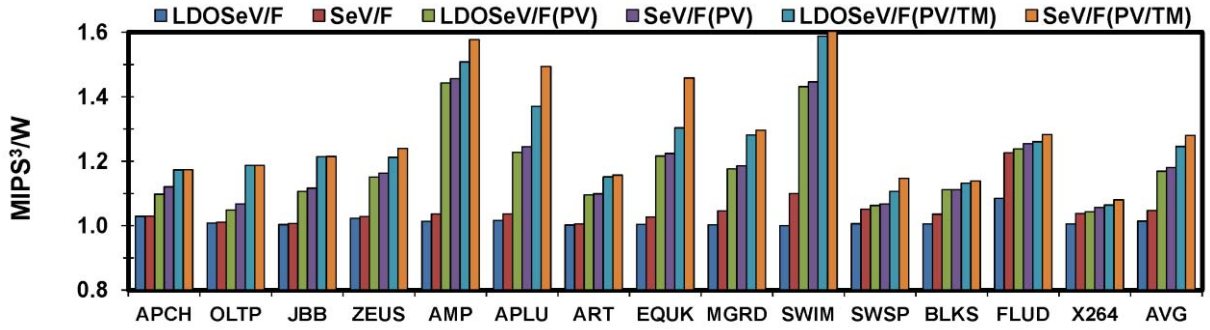

Fig. 6. MIPS^3/W comparison of eight-core processors supported by LDO (algorithms beginning with the LDOSeV/F prefix) and switching VRs (algorithms beginning with SeV/F). All results are normalized to a processor with ShV/F and *do not* include the power loss by both the on- and off-chip VRs. Each interval is comprised of ten million executed instructions.

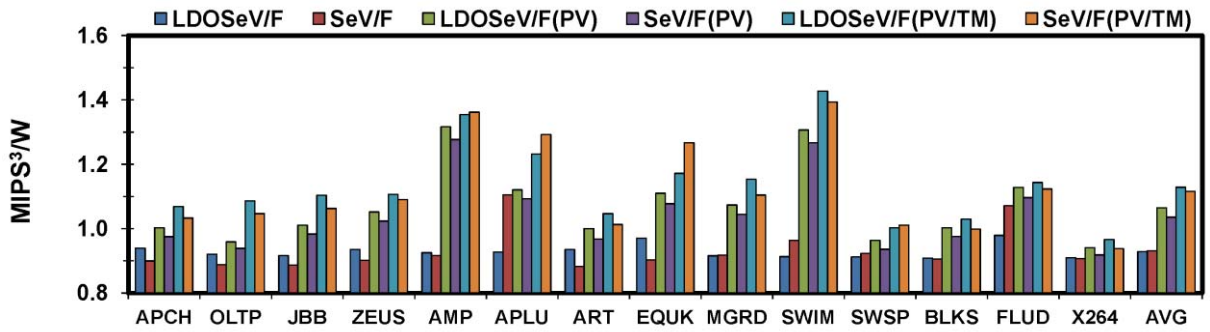

Fig. 7. MIPS^3/W comparison of eight-core processors supported by LDO (algorithms beginning with the LDOSeV/F prefix) and switching VRs (algorithms beginning with SeV/F) including the power loss by both on- and off-chip VRs. All results are normalized to a processor with ShV/F and include the power loss by the off-chip VR. Each interval is composed of ten million executed instructions.

and its relative benefit can be higher for processors using LDO VRs.

2) Impact of VR Efficiency on MIPS^3/W: Fig. 7 compares MIPS^3/W of eight-core processors using the on-chip LDO and switching VRs when the power loss by both on- and off-chip VRs are accounted for. When the processors with LDOSeV/F and SeV/F do not exploit WID PVs, they exhibit worse MIPS^3/W than the processor with ShV/F, which uses only an off-chip VR for all processor cores. This is because the power loss by the on-chip VRs completely negates the benefit of supporting per-core V/F domains for multithreaded applications. However, per-core V/F domains allow processors to more effectively exploit WID PVs and VSeV/F(PV) and SeV/F(PV) can provide 6% and 4% higher MIPS^3/W than ShV/F on average. Furthermore, when the TM technique is also applied, LDOSeV/F(PV/TM) and SeV/F(PV/TM) can provide 13% and 12% higher MIPS^3/W than ShV/F on average.

We observe that the processor using LDO VRs exhibits higher MIPS^3/W than the one using switching VR in Fig. 7. This is the opposite of the trend shown in Fig. 6, where the power loss by VRs was not considered in computing MIPS^3/W and the MIPS^3/W of the processor using LDO VRs was lower than the one using the switching VRs. This is mainly due to small C2C voltage variations in multithreaded applications, which allows LDO VRs to provide voltages with higher efficiency than that by the switching VRs, as shown in Fig. 5(b).

Note that LDOSeV/F and SeV/F (i.e., per-core V/F domains) lead to a much higher MIPS^3/W improvement for

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

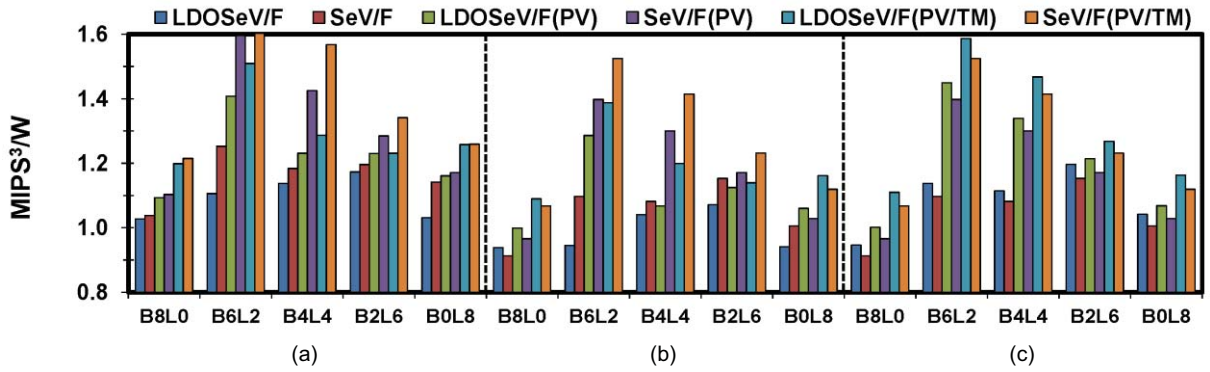

Fig. 8. MIPS^3/W comparison of eight-core processors supported by LDO (algorithms beginning with the LDOSeV/F prefix) and switching VRs (algorithms beginning with SeV/F). The power loss by on- and off-chip VRs (a) is not included and (b) is included. The V_I–V_O constraint is removed for LDO VRs in (c) while the power loss by on- and off-chip VRs is accounted for.

most SPEC OMP benchmarks than most commercial applications and PARSEC benchmarks. This is because all the threads of the commercial applications and PARSEC benchmarks exhibit very similar behavior in both computations and synchronizations. As a result, cores running these threads require similar performance at each execution interval. For example, most SPEC OMP benchmark have much longer parallel loops than most PARSEC benchmarks, leading to more diverse behavior in each execution interval because they are more likely running different sections of parallel loops.

3) Impact of the DVFS Interval Period on MIPS^3/W: The interval period for applying a DVFS algorithm also impacts the efficacy of DVFS. In theory, a shorter DVFS interval can capture more C2C IPC variations, resulting in more C2C voltage variations. This may negatively affect the efficacy of the proposed technique that exploits small C2C voltage variations. To analyze the impact of the DVFS interval period on MIPS^3/W, we reduce the DVFS interval period to every five million instructions. The MIPS^3/W values for both LDOSeV/F(PV/TM) and SeV/F(PV/TM) increase, but the relative difference between them remains almost the same. Note that we do not evaluate more aggressive DVFS interval periods because the following factors prohibit the use of a very short DVFS interval: 1) the computational overhead of the DVFS algorithm; 2) the phase-locked loop (PLL) re-locking time for a frequency change [26]; and 3) the VR efficiency degradation during V_O transitions [27].

4) Multiprogram Environment: A processor executing multiple applications can exhibit more substantial C2C IPC variations than one running multithreaded applications, depending on the mix and characteristics of applications. Consequently, supporting a wider range of V_O values using the switching VRs may lead to higher MIPS^3/W than using LDO VRs under a specified power constraint. Fig. 8 shows the MIPS^3/W comparison between two processors using on-chip switching and LDO VRs when running five mixes of memory- and compute-bound applications; we run the mixes of applications (not in isolation) using the multicore simulator to accurately model the interaction between applications.

When the power loss by both off- and on-chip VRs is not considered, a processor with per-core V/F domains using either on-chip switching or LDO VRs exhibits much higher MIPS^3/W than one using a single V/F domain for B6L2, B4L4, and B2L6 in Fig. 8(a), thereby signifying the importance of supporting per-core V/F domains for such an environment. This is because these mixes of applications present much higher C2C IPC variations than the multithreaded applications in general. For example, LDOSeV/F(PV/TM) and SeV/F(PV/TM) can provide 34% and 55% higher MIPS^3/W than ShV/F on average.

The processor that supports per-core V/F domains using LDO VRs provides substantially higher MIPS^3/W than the one that provides a single V/F domain using an off-chip VR, but 16% lower MIPS^3/W than the one using the switching VRs. This is a notable MIPS^3/W difference compared to the multithreaded environment evaluated in Fig. 6. However, this does not imply that the processor using LDO VRs is inferior to the one using the switching VRs. When the power loss by both the on- and off-chip VRs is considered, as shown in Fig. 8(b), LDOSeV/F(PV/TM) and SeV/F(PV/TM) can yield 24% and 39% higher MIPS^3/W than ShV/F on average. LDOSeV/F(PV/TM) still results in lower MIPS^3/W than SeV/F(PV/TM), yet the difference between LDOSeV/F(PV/TM) and SeV/F(PV/TM) is reduced from 16% to 12%. This is because the power loss by LDO VRs is lower than that with the switching VRs in many DVFS intervals.

In the previous experiments, we limited the difference between V_I–V_O to 100 mV. This is because the efficiency of LDO VRs becomes lower than that of switching VRs once the voltage difference becomes larger than 100 mV. However, we observed that forcing such a constraint misses potential power reduction opportunities that can be achieved by operating cores at lower V/F. In other words, the benefit of reducing V/F of cores more can outweigh the lower power efficiency of LDO VRs operating at V_I–V_O larger than 100 mV for all processor cores. Thus, we remove the V_I–V_O constraint for the processor using LDO VRs in Fig. 8(c). When V_I–V_O is larger than 100 mV, the power loss by LDO VRs is higher than that of the switching VR. Nonetheless, the power loss by LDO VRs becomes lower than that of the switching VRs for the DVFS intervals exhibiting V_I–V_O less than 100 mV.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

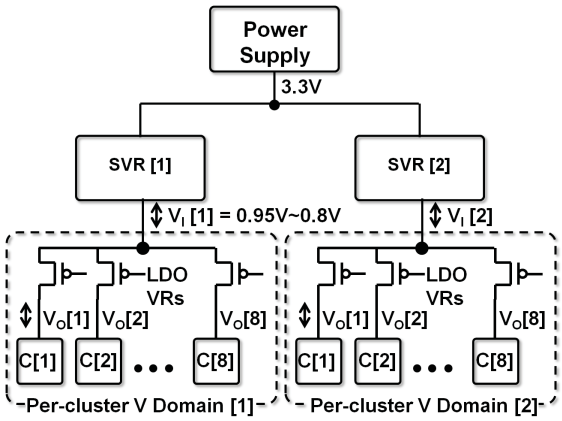

Fig. 9. Hierarchical VR scheme.

Consequently, as long as we have more DVFS intervals with V_I–V_O less than 100 mV, the processor using LDO VRs can lead to higher MIPS^3/W than the one using the switching VRs. Fig. 8(c) shows that LDOSeV/F(PV/TM) exhibits 4% higher MIPS^3/W than SeV/F(PV/TM).

To validate this result, we analyze the fraction of DVFS intervals exhibiting V_I–V_O more than 100 mV. For B4L4, we measure the fraction of DVFS intervals in which LDO VRs have lower efficiency than the switching VRs using the V/F and core power consumption traces from SeV/F(PV/TM). This reveals that the LDO VRs show higher efficiency than the switching VRs for nearly 60% of the total DVFS intervals that are experienced by individual cores.

5) Impact of Removing V_I–V_O Range Constraint on MIPS^3/W of Multithreaded Applications: After we discover that limiting V_I–V_O can negatively impact MIPS^3/W, we re-evaluate the MIPS^3/W for multithreaded applications after removing the V_I–V_O constraint. Although WID C2C PVs are not exploited and the TM technique is not applied, LDOSeV/F provides 4% higher average MIPS^3/W than SeV/F, while LDOSeV/F led to slightly lower average MIPS^3/W than SeV/F before removing the V_I–V_O constraint. When WID C2C PVs (and TM) are exploited, removing the V_I–V_O constraint for LDOSeV/F increases the average MIPS^3/W improvement over ShV/F from 6% to 8% (from 13% to 16%). This suggests that we should not enforce the V_I–V_O constraint.

V. Discussion

In this paper, we demonstrated that using LDO VRs can be effective for multithreaded applications that exhibit a narrow V_I–V_O range due to relatively small IPC variations across cores. However, we also showed that adopting LDO VRs can provide a significant performance improvement even for the mixes of applications exhibiting large IPC variations across cores. In contrast, the cost of on-chip per-core LDO VRs is much less than that of the on-chip per-core switching VRs. This is because LDO VRs can share their most expensive component with PCPG devices, if PCPG devices are already implemented in the processor, as described in Section III.

However, future processors will have more cores (i.e., many-core processors) and they can potentially exhibit higher C2C IPC variations. This may result in a wider V_I–V_O range and thus worse MIPS^3/W due to poor power conversion efficiency of LDO VRs for such a case. On the other hand, it may not be practical to provide a large number of switching VRs for manycore processors due to the cost; integrating a high-quality on-chip inductor becomes more challenging with technology scaling while integrating an on-package inductor will not be a scalable solution for a larger number of cores due to physical constraints. In such a case, we propose a hierarchical VR scheme in which a switching VR provides a shared voltage domain for a subset (or a cluster) of cores and LDO VRs provide per-core voltage domains within each cluster, as illustrated in Fig. 9. This hierarchical power delivery architecture allows us to support cost-effective per-core voltage domains, which can minimize the number of switching VRs while maximizing the power efficiency of individual LDO VRs. The detailed evaluation for a hierarchical power delivery architecture is left as future work.

VI. Related Work

Several researchers have investigated the benefits of the per-core DVFS for multicore processor. Li *et al.* [4] analyzed the performance of per-core DVFS combined with dynamic core scaling for multicore processors running parallel applications. They exploit the observation that parallel applications with limited problem size do not use all cores efficiently. Thus, they jointly adjust the number of active cores and perform per-core DVFS to maximize the performance under a power constraint. Kim *et al.* [7] demonstrated the potential power reduction opportunities using the on-chip switching VRs for embedded processors. They also provided detailed background on the switching VRs using air-core inductors and an analysis on their efficiency. Recently, Eyerman *et al.* [28] also evaluated the benefit of fine-grain applications of DVFS and proposed a fine-grain microarchitecture-driven DVFS mechanism. Many researchers also studied the impact of WID PVs, which lead to C2C frequency and power variations [8], [29], [30], on the performance of multicore processors, and proposed DVFS algorithms that can exploit the C2C frequency and power variations. Teodorescu *et al.* [8] investigated a DVFS algorithm based on linear programming to maximize the performance of multicore processors under a power constraint. Their DVFS algorithm also exploits WID C2C PVs for workload scheduling, as well as power management. Rangan *et al.* [31] proposed a thread migration technique to minimize the cost of the transition time for the VR output voltage. They introduced voltage domains in which each core operates at a fixed but different voltage level. If threads require different V/F levels for power-efficient operations, they migrate to the cores that can provide an appropriate performance level, instead of changing the V/F of cores. Dighe *et al.* [30] analyzed energy reduction using per-core DVFS and core scaling in an 80-core processor and design trade-offs of implementing multiple V/F domains. Their observations clearly support and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

motivate our proposed methods. They also demonstrated that energy reduction increases as the number of voltage domains increases with fewer cores per domain and as the voltage step resolution increases; they briefly compared three methods to vary the core voltage. While an independent collapsible voltage rail controlled by an off-chip VR avoids chip design complexity and energy overhead of a sleep transistor, its benefits diminish rapidly for high-performance targets as shown in [30]. Truong *et al.* [32] demonstrated a 167-core processor with per-core V/F control. Their processor is limited to two discrete voltage levels implemented as two separate global power grids and utilizes power switches to connect a core to one of the grids. This method trades voltage scaling granularity with the complexity of implementing multiple global power grids.

VII. Conclusion

In this paper, we propose a cost-effective technique to support per-core voltage domains for high-performance multicore processors. We demonstrate that existing PCPG devices augmented with a small amount of circuitry can operate as low-cost LDO VRs. Unlike on-chip switching VRs, LDO VRs do not require inductors that pose a major technical challenge for the on-chip integration. However, the power efficiency of LDO VRs degrades as the difference between V_I and V_O increases. Consequently, multicore processors using LDO VRs may have lower performance/power efficiency than multicore processors using the switching VRs, when cores require quite different voltage levels for performance optimization. However, our experiments show that C2C voltages variations are relatively small when the V/F of each core is optimized to maximize performance under a power constraint. After modeling the power efficiency of both the LDO and switching VRs using a 32-nm technology, we show that the MIPS^3/W of an eight-core processor using LDO VRs is slightly higher than that of a processor using the switching VRs. This is because the efficiency of LDO VRs is higher than that of the switching VRs for small C2C voltage variations exhibited in most DVFS intervals.

References

[1] S. Rusu, T. Simon, H. Muljono, J. Stinson, D. Ayers, J. Chang, R. Varada, M. Ratta, S. Kottapalli, and S. Vora, "A 45 nm eight-core enterprise Xeon processor," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 7–14, Jan. 2010.

[2] R. Jotwani, S. Sundaram, S. Kosonocky, A. Schaefer, V. F. Andrade, A. Novak, and S. Naffziger, "An x86–64 core in 32 nm SOI CMOS," *IEEE. J. Solid-State Circuits*, vol. 46, no. 1, pp. 162–172, Jan. 2011.

[3] Intel Corporation. (2011, Mar.). *Intel Turbo Boost Technology 2.0.*, Santa Clara, CA, USA [Online]. Available: http://www.intel.com/technology/turboboost/index.htm

[4] J. Li and J. Martinez, "Dynamic power-performance adaptation of parallel computation on chip multiprocessors," in *Proc. IEEE/ACM Int. Symp. High-Perf. Comput. Archit.*, 2006, pp. 77–87.

[5] Y. Hoskote, S. Vangal, A. Singh, N. Borkar, and S. Borkar, "A 5-GHz mesh interconnect for a teraflops processor," *IEEE Micro*, vol. 27, no. 5, pp. 51–61, Sep.–Oct. 2007.

[6] H. R. Ghasemi, A. A. Sinkar, M. J. Schulte, and N. S. Kim, "Cost-effective power delivery to support per-core voltage domains for power-constrained processors," in *Proc. IEEE/ACM Design Autom. Conf.*, 2012, pp. 56–61.

[7] W. Kim, M. S. Gupta, G.-Y. Wei, and D. Brooks, "System level analysis of fast, per-core DVFS using on-chip switching regulators," in *Proc. IEEE/ACM Int. Symp. High-Perform. Comput. Archit.*, 2008, pp. 123–134.

[8] R. Teodorescu and J. Torrellas, "Variation-aware application scheduling and power management for chip multiprocessors," in *Proc. IEEE/ACM Int. Symp. Comput. Archit.*, 2008, pp. 363–374.

[9] A. R. Alameldeen, C. J. Mauer, M. Xu, P. J. Harper, M. M. K. Martin, D. J. Sorin, M. D. Hill, and D. A. Wood, "Evaluating non-deterministic multi-threaded commercial workloads," in *Proc. Comp. Archit. Evaluation Using Commercial Workloads*, 2002, pp. 30–38.

[10] Univ. Princeton. (2008, Jan.). *PARSEC Benchmark Suite*, Princeton, NJ, USA [Online]. Available: http://parsec.cs.princeton.edu/

[11] M. M. K. Martin, D. J. Sorin, B. M. Beckmann, M. R. Marty, M. Xu, A. R. Alameldeen, K. E. Moore, M. D. Hill, and D. A. Wood, "Multifacet's general execution-driven multiprocessor simulator (GEMS) toolset," *ACM SIGARCH Comput. Archit. News*, vol. 33, no. 4, pp. 92–99, Nov. 2005.

[12] D. M. Brooks, P. Bose, S. E. Schuster, H. Jacobson, P. N. Kudva, A. Buyuktosunoglu, J.-D. Wellman, V. Zyuban, M. Gupta, and P. W. Cook, "Power-aware microarchitecture: Design and modeling challenges for next-generation microprocessors," *IEEE Micro*, vol. 8, no. 6, pp. 26–44, Nov.–Dec. 2000.

[13] S. R. Sarangi, B. Greskamp, R. Teodorescu, J. Nakano, A. Tiwari, and J. Torrellas, "VARIUS: A model of process variation and resulting timing errors for microarchitects," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 1, pp. 3–13, Feb. 2008.

[14] K. Aygun, M. J. Hill, K. Eilert, K. Radhakrishnan, and A. Levin, "Power delivery for high-performance microprocessor," *Intel Technol. J.*, vol. 9, no. 4, pp. 273–283, Nov. 2005.

[15] *Intel Workstation Board S975XBX2 Technical Product Specification*, Intel Corporation, Santa Clara, CA, USA, Oct. 2006.

[16] P. Hazucha, G. Schrom, J. Hahn, B. A. Bloechel, P. Hack, G. E. Dermer, S. Narendra, D. Gardner, T. Karnik, V. De, and S. Borkar, "A 233-MHz 80%–87% efficient four-phase DC-DC converter utilizing air-core inductors on package," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 838–845, Apr. 2005.

[17] L. E. Mosley, "Power delivery challenges for multi-core microprocessors," in *Proc. Capacitor Resistor Technol. Symp.*, 2008, pp. 445–579.

[18] P. Hazucha, T. M. Sung, G. Schrom, F. Paillet, D. Gardner, S. Rajapandian, and T. Karnik, "High voltage tolerant linear regulator with fast digital control for biasing integrated DC-DC converters," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 66–73, Jan. 2007.

[19] N. S. Kim, J. Seomun, A. Sinkar, J. Lee, T. H. Han, K. Choi, and Y. Shin, "Frequency and yield optimization using power gates in power-constrained designs," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design*, 2009, pp. 121–126.

[20] W. Fu and A. Fayed, "A feasibility study of high-frequency buck regulators in nanometer CMOS technologies," in *Proc. IEEE Dallas Circuits Syst. Workshop*, 2009, pp. 1–4.

[21] P Hazucha et al., "Area-efficient linear regulator with ultra-fast load regulation," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 933–940, Apr. 2005.

[22] J. Klein. (2006). *Fairchild Semiconductors* [Online]. Available: http://www.fairchildsemi.com/an/AN/AN-6005.pdf

[23] J. Gjanci and M. H. Chowdhury, "A hybrid scheme for on-chip voltage regulation in system-on-a-chip (SOC)," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 99, no. 10, pp. 1–11, Oct. 2010.

[24] J. Lee, G. Hatcher, L. Vandenberghe, and C. K. Yang, "Evaluation of fully-integrated switching regulators for CMOS process technologies," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 15, no. 9, pp. 1017–1027, Sep. 2007.

[25] HP Labs. (2009, Dec.). *McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures* [Online]. Available: http://www.hpl.hp.com/research/mcpat

[26] J. Park, D. Shin, N. Chang, and M. Pedram, "Accurate modeling and calculation of delay and energy overheads of dynamic voltage scaling in modern high-performance microprocessors," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design*, 2010, pp. 419–424.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[27] W. Kim, M.S. Gupta, G.-Y. Wei, and D. Brooks, "System level analysis of fast, per-core DVFS using on-chip switching regulators," in *Proc. IEEE/ACM Int. Symp. High-Perform. Comput. Archit.*, 2008, pp. 123–134.

[28] S. Eyerman and L. Eeckhout, "Fine-grained DVFS using on-chip regulators," *ACM Trans. Archit. Code Optim.*, vol. 8, no. 1, pp. 1–24, Apr. 2011.

[29] S. Herbert and D. Marculescu, "Variation-aware dynamic voltage/frequency scaling," in *Proc. IEEE/ACM Int. Symp. High Perform. Comput. Archit.*, 2009, pp. 301–312.

[30] S. Dighe, S. R. Vangal, P. Aseron, S. Kumar, T. Jacob, K. A. Bowman, J. Howard, J. Tschanz, V. Erraguntla, N. Borkar, V. K. De, and S. Borkar, "Within-die variation-aware dynamic-voltage-frequency-scaling with optimal core allocation and thread hopping for the 80-core TeraFLOPS processor," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 184–193, Jan. 2011.

[31] K. Rangan, G.-Y. Wei, and D. Brooks, "Thread motion: Fine-grained power management for multi-core systems," in *Proc. IEEE/ACM Int. Symp. Comput. Archit.*, 2009, pp. 302–313.

[32] D. N. Truong, W. H. Cheng, T. Mohsenin, Z. Yu, A. T. Jacobson, G. Landge, M. J. Meeuwsen, C. Watnik, A. T. Tran, Z. Xiao, E. W. Work, J. W. Webb, P. V. Mejia, and B. M. Baas, "A 167-processor computational platform in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1130–1144, Apr. 2009.

Abhishek A. Sinkar received the B.E. degree in electrical engineering from the Victoria Jubilee Technical Institute, Mumbai, India, in 2001, the M.S. degree in electrical engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 2005, and the Ph.D. degree in electrical and computer engineering from the University of Wisconsin-Madison, WI, USA, in 2012.

He is currently a Senior CPU Design Engineer with Oracle America Inc., Redwood, CA, USA. Prior to joining Oracle, he was a Co-Op Engineer with Intel Corporation, Santa Clara, CA, USA, where he was involved in power modeling and estimation for Intel's next generation of many-core processors. His current research interests include the areas of low-power circuit design, circuit and micro-architectural power management, circuit reliability, and VLSI testing.

Hamid Reza Ghasemi received the B.S. and M.S. degree from Electrical and Computer Engineering Department, University of Tehran, Tehran, Iran. He is currently pursuing the Ph.D. degree with the Computer Sciences Department, University of Wisconsin-Madison, Madison, WI, USA.

His current research interests include computer architecture, power efficient high-performance processing, and low power architecture.

Michael J. Schulte received the B.S. degree in electrical engineering from the University of Wisconsin-Madison, Madison, WI, USA, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas at Austin, Austin, TX, USA.

He is a Senior Fellow with AMD Research, Austin, TX, USA, where he conducts research in the areas of power-efficient processor design, heterogeneous processor architectures, and high-performance computing. He is currently a Principal Investigator and a Technical Lead on AMD's FastForward Extreme-Scale Computing Research and Development efforts. He is an Adjunct Faculty Member with the Electrical and Computer Engineering Department, University of Wisconsin-Madison, Madison, WI, USA, where he was a Tenured Faculty Member and directed the Madison Embedded Systems and Architectures Laboratory. He is an inventor on ten patents and has numerous patents pending. He has published over 200 research papers. His current research interests include high-performance computing, domain-specific processing, heterogeneous computing, power-efficient processor design, computer architecture, reconfigurable computing, and computer arithmetic.

Dr. Schulte served as an Associate Editor for the IEEE TRANSACTIONS ON COMPUTERS and the *Journal of VLSI Signal Processing*. He is a recipient of the National Science Foundation CAREER Award, the Alfred Noble Robinson Award, the Frank Hook Assistant Professorship, and service awards from the National Society of Black Engineers and IEEE.

Ulya R. Karpuzcu received the Ph.D. degree in electrical and computer engineering from the University of Illinois, Urbana-Champaign, IL, USA, in 2012.

She is an Assistant Professor of electrical and computer engineering with the University of Minnesota, Twin Cities, MN, USA. Her current research interests include process technology on computer architecture.

Nam Sung Kim (S'96–M'03–SM'09) received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, and the Ph.D. degree in computer science and engineering from the University of Michigan-Ann Arbor, Ann Arbor, MI, USA.

He is an Associate Professor with the University of Wisconsin-Madison, Madison, WI, USA, where he has been conducting interdisciplinary research that cuts across device, circuit, and architecture for power-efficient computing. Prior to joining the University of Wisconsin-Madison, he was a Senior Research Scientist with Intel from 2004 to 2008, where he conducted research in power-efficient digital circuit and processor architecture. He has more than 2000 combined citations and a combined citation of all the papers exceeds 3400. He has published over 90 refereed articles to highly-selective conferences and journals in digital circuit, processor architecture, and computer-aided design. His current research interests include designing power-efficient and robust computing systems.

Dr. Kim has served several prominent international conferences as a Technical Program Committee Member. He received the IEEE Design Automation Conference Student Design Contest Award in 2001, Intel Fellowship in 2002, the IEEE International Conference on Microarchitecture Best Paper Award in 2003, the National Science Foundation CAREER Award in 2010, and the IBM Faculty Award in 2011 and 2012.