System-Level Power Analysis of a Multicore Multipower Domain Processor With ON-Chip Voltage Regulators

Ayan Paul, Sang Phill Park, Dinesh Somasekhar, Young Moon Kim, Nitin Borkar, Ulya R. Karpuzcu, and Chris H. Kim, *Senior Member, IEEE*

Abstract—In this paper, we study two different ON-chip power delivery schemes, namely, fully integrated voltage regulator (FIVR) and low-dropout regulator (LDO), and analyze their effect on total system power under process variation, assuming a realistic dynamic voltage–frequency scaling (DVFS) system. The impact of different task scheduling algorithms on the overall system power was also analyzed. We find that in a hypothetical 256-core processor, under a per-core DVFS assumption, the FIVR-based power delivery consumes 20% less power than the LDO-based one for a 50% throughput. However, as the number of cores in the processor reduces, the difference in power consumption between the FIVR-based and LDO-based power delivery schemes becomes smaller. For example, in the case of a 16-core processor with per-core DVFS capability, FIVR-based design was found to consume about the same power as the LDO-based design.

Index Terms—Circuit simulation, dynamic voltage scaling, integrated circuit modeling, multicore processing, power dissipation, regulators, switching converters.

I. INTRODUCTION

POWER consumption of multicore processors can be reduced by individually controlling the supply voltage of each core based on the processor workload. A prerequisite of such a per-core dynamic voltage–frequency scaling (DVFS) scheme is the integration of voltage regulator modules into the processor die. Hence, the design of integrated voltage regulators has gained momentum over the past few years. Switching regulators that use on-die thick-metal inductors are not suitable for integration because of the low quality factor and large area overhead [1]. On the other hand, even with novel

high density capacitor technology, switched-capacitor-based ON-chip dc–dc converters have shown to suffer from relatively low output power density, especially when the $V_{\text{OUT}}/V_{\text{IN}}$ ratio deviates from the target [2]. Intel's Haswell processors use air-core package inductors as the inductors of the switching regulators, and they integrate voltage regulator on-die [3]. This kind of switching regulators, which use package inductors in lieu of OFF-chip inductors, is termed fully integrated voltage regulators (FIVR). Similarly, IBM introduced a distributed low-dropout regulator (LDO) for controlling supply voltage on a per-core basis in its POWER8 processor [4]. With various state-of-the-art ON-chip power delivery solutions reported thus far, it remains to be seen whether switching regulators or linear regulators will result in lower overall system-level power consumption. In this paper, we compare the power consumption of a multicore system with either FIVR or LDO as the ON-chip power delivery unit, while considering different core count, power domain count, scheduling algorithm, and process variation.

In order to estimate power savings, it is important to take power loss of the voltage regulators into account. Several previous works have attempted to evaluate power/energy benefits of ON-chip voltage regulators. For instance, [5] discusses that workload-aware voltage regulator designs can result in system-level energy saving. Reference [6] presents a dynamic reconfiguration of networks that connect voltage regulators to the cores, resulting in system-wide energy saving. Reference [7] shows that per-core DVFS using the ON-chip voltage regulation scheme can provide significant system energy reduction. With this knowledge, it becomes necessary to find out which of the ON-chip power delivery solutions would result in maximum reduction in system energy/power.

In this paper, we do not present a new circuit-level power delivery solution, nor do we aim to put forward a CAD methodology for efficient power delivery. We rather explore the power-performance design space of a many-core processor system, when the processors are powered by different types of voltage regulators. The choice of FIVR and LDO as the ON-chip power regulator in our power-performance exploration study was obvious, given that they are the current state of the art and are being used by industry leaders as the ON-chip power delivery units. Performance metric for this paper has been assumed to be normalized throughput,

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

2   IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS

which we define as the ratio of the average throughput and the maximum possible throughput of the system. Since server applications are generally limited by throughput, it makes sense to use throughput as the performance parameter in our power-performance analysis of a many-core processor designed for server applications.

The contribution of this paper is twofold. First, it presents a systematic framework to compute system power consumption of a many-core processor by incorporating power dissipated in the cores, voltage regulators, and power grids for various workload profiles. Second and most importantly, it compares two state-of-the-art power delivery solutions (FIVR-based and LDO-based) from a system perspective. If there are more cores than the number of ON-chip regulators, then a number of cores have to share the same power domain. In this scenario, total system power consumption is minimized when the cores with equal supply voltage requirement are grouped in the same power domain. However, such homogeneous grouping of cores may not always be possible due to limitations in scheduling algorithms, and furthermore, it will result in an overly optimistic estimate in terms of total system power consumption. Hence, our initial analysis assumes that both homogeneous and inhomogeneous grouping of cores are equally probable, and we perform a Monte Carlo analysis to find out the range of power consumptions of FIVR-based and LDO-based power delivery techniques. Later, we show results based on a minimum power scheduler, which assigns cores to different power domains in a homogeneous fashion.

II. Efficiency Models of Switching Regulators and LDO

Switching voltage regulators are integral components of power delivery systems. Traditional OFF-chip buck converters typically down convert OFF-chip supply voltage to logic voltage to be used by microprocessor cores. In this paper, in order to support DVFS on a cluster of cores, one more level of voltage conversion has been assumed to take place ON-chip, between OFF-chip buck converter and microprocessor cores. We assume that this ON-chip power delivery module can be either FIVR or LDO.

A. Overview of FIVR

FIVR is a synchronous buck converter built ON-chip. It can have up to 16 phases. In order keep filter passives small, FIVR has to be operated at relatively high frequencies (e.g., 140 MHz according to [3]). Cascode nMOS and pMOS are used as the power switches of this switching regulator. Built in 22-nm Intel's logic process, these switches can handle an input voltage of 1.8 V and are distributed across the die. They are placed right above the connections of the package inductors in order to minimize routing cost. Because of the close proximity of the regulator and the circuits, extra bumps can be placed on the circuit, and routing can be done using a thick metal layer, which effectively increases power density provided by FIVR. Bottom of the package and the die of Intel's Haswell processors along with FIVR inductors have been shown in Fig. 1 (top). Very fast voltage ramp times of the order of submicroseconds can be achieved using an FIVR-based
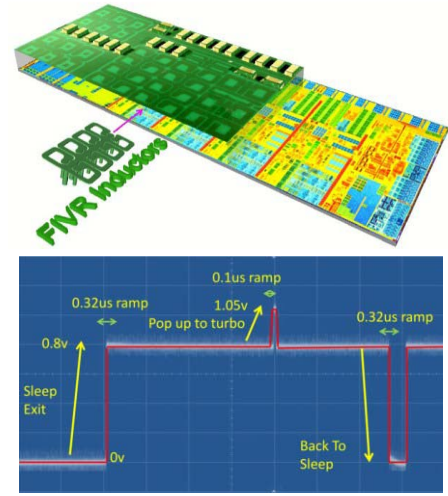

Fig. 1. Top: bird's eye view of Intel's Haswell processor die with package inductors for FIVR [3]. Bottom: fast DVFS transients enabled by FIVR [8].
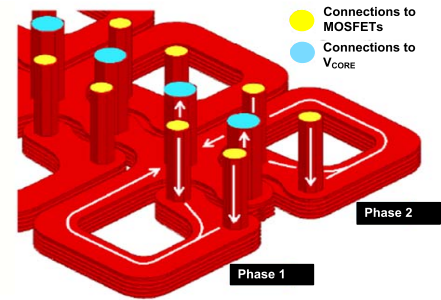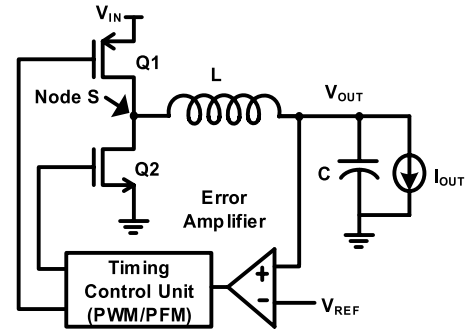

Fig. 2. 3-D view of two FIVR inductors [3].

Fig. 3. Simplified schematic of a step-down switching voltage regulator.

DVFS system, as shown in Fig. 1 (bottom). FIVR inductors have an air core and, hence, are nonmagnetic. A 3-D view of the FIVR inductor with two phases has been shown in Fig. 2. For decoupling purpose, ON-chip metal–insulator–metal (MIM) capacitors and package ceramic capacitors are used. MIM capacitors provide decoupling from output rail and show good transient characteristics. On the other hand, both package ceramic capacitors and ON-chip MIM capacitors are used to provide decoupling from the input rail.

B. Switching Voltage Regulator Model

Schematic of a generic step-down switching voltage regulator, which can be an OFF-chip buck converter or an ON-chip FIVR, is shown in Fig. 3. It consists of MOSFET
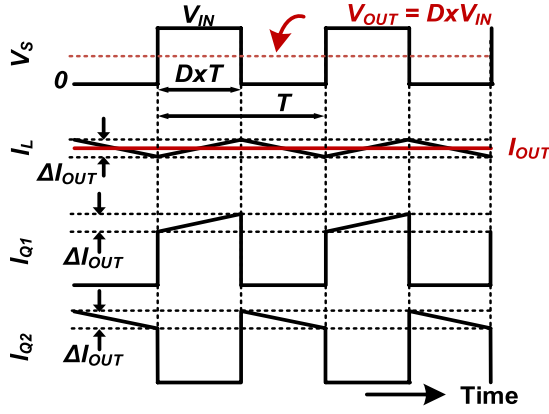
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

PAUL *et al.*: SYSTEM-LEVEL POWER ANALYSIS OF A MULTICORE MULTIPOWER DOMAIN PROCESSOR

3

Fig. 4. Current and voltage waveforms of a switching voltage regulator in steady state.

Fig. 5. Efficiency versus I_{OUT} for single-phase OFF-chip buck regulator, and single-phase ON-chip FIVR.

switches $Q1$ and $Q2$, filter network comprising of a filter inductor and a capacitor, and a feedback control loop. Voltage level required by the microprocessor core sets the voltage of the inverting input of a hysteretic comparator. Other input of the comparator is driven by the output of the switching converter. The comparator generates error voltage, which in turn, drives a pulsewidth modulated (PWM) or pulse frequency modulated (PFM) controller to generate precise turn-ON and turn-OFF timings of the upper/lower switching MOSFETs, $Q1$ and $Q2$. The voltage at the output node of the MOSFETs then drives a low-pass filter formed by L and C.

Fig. 4 shows the current waveforms of the switching converter through $Q1$, $Q2$, and L along with voltage at node S. $Q1$ is ON for a time $D \times T$ during which $Q2$ should be OFF, in which T is the time period of the clock generated by the timing control unit, and D is the duty cycle of the clock. $Q2$ is kept ON for the remaining of the time period, which is $(1-D) \times T$. From the current waveforms shown in Fig. 4, we can see that I_{OUT} is the average output current and ΔI_{OUT} is the inductor current ripple. The rms values of I_L, I_{Q1}, and I_{Q2} can be written as $I_{L,\mathrm{rms}} = (I_{\mathrm{OUT}}^2 + (\Delta I_{\mathrm{OUT}}^2/12))^{1/2}$, $I_{Q1,\mathrm{rms}} = ((V_{\mathrm{OUT}}/V_{\mathrm{IN}}) \cdot (I_{\mathrm{OUT}}^2 + (\Delta I_{\mathrm{OUT}}^2/12)))^{1/2}$, and $I_{Q2,\mathrm{rms}} = ((1-(V_{\mathrm{OUT}}/V_{\mathrm{IN}})) \cdot (I_{\mathrm{OUT}}^2 + (\Delta I_{\mathrm{OUT}}^2/12)))^{1/2}$, respectively. Hence, the conduction losses in the switches $Q1$ ($P_{\mathrm{COND\_}Q1}$) and $Q2$ ($P_{\mathrm{COND\_}Q2}$), and in the parasitic resistance of the inductor ($P_{\mathrm{PAR\_}L}$) can be written as $P_{\mathrm{COND\_}Q1} = I_{Q1,\mathrm{rms}}^2 \cdot R_{\mathrm{SW\_}Q1}$, $P_{\mathrm{COND\_}Q2} = I_{Q2,\mathrm{rms}}^2 \cdot R_{\mathrm{SW\_}Q2}$, and $P_{\mathrm{PAR\_}L} = I_{L,\mathrm{rms}}^2 \cdot R_{\mathrm{PAR\_}L}$, respectively, where $R_{\mathrm{SW\_}Q1}$ and $R_{\mathrm{SW\_}Q2}$ are the average ON-resistances of switches $Q1$ and $Q2$, and $R_{\mathrm{PAR\_}L}$ is the inductor parasitic resistance. Apart from the conduction loss, another important loss component is the MOSFET gate drive loss, which can be given as $P_{\mathrm{GATE}} = C_{\mathrm{GATE}} V_{\mathrm{GATE}}^2 f$, where C_{GATE} is the total gate capacitance of $Q1$ and $Q2$. The final power loss component comes from control circuitry (P_{CTRL}), which consists of an error amplifier,
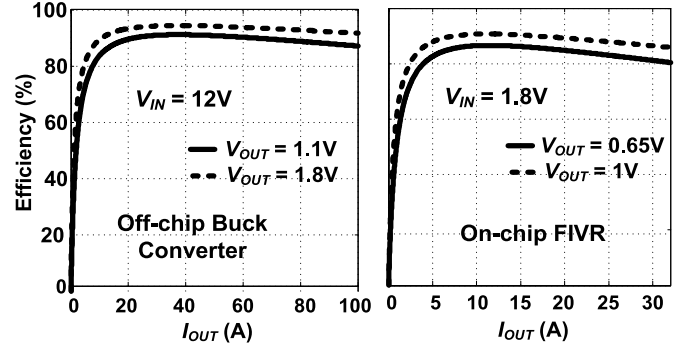
a compensation circuit, and a digital controller. Function of the control loop is to generate PFM or PWM control signals. Power loss in the control loop can be represented as $P_{\mathrm{CTRL}} = V_{\mathrm{IN}} \cdot I_{\mathrm{sub}} + K_c \cdot V_{\mathrm{IN}}^2 \cdot f$, where I_{sub} is the static power loss in the control loop of the converter, K_c is proportional to the gate capacitance of the devices in the control loop, and f is the switching frequency of the converter. Major portion of this power loss is due to the quiescent current in the control loop. Absolute control loop power loss of the converter does not depend on converter size or load condition. In a per-core DVFS scenario, converter size will be much smaller compared with the case when the converter is delivering power to many cores. Hence, in terms of total power loss in the control loop of all converters, a per-core DVFS scheme will be worse.

FIVR model has been built assuming $R_{\mathrm{PAR\_}L} = 16$ mΩ/per phase, $R_{\mathrm{SW\_}Q1} = 64$ mΩ/per phase, and $R_{\mathrm{SW\_}Q2} = 48$ mΩ/per phase. We assume that when only one phase of a 16-phase FIVR is operating, FIVR can achieve an efficiency of 86% while delivering 0.5 A/per phase at an output voltage $V_{\mathrm{OUT}} = 1$ V. Switching frequency of 140 MHz and 30% inductor current ripple has been assumed.

With these loss components taken into account, power efficiency (η) of a switching regulator can be written as shown at the bottom of this page, where η reaches a peak value for a certain load condition. Below that load current, efficiency suffers because of load-independent gate drive and control circuit loss, and above this load current efficiency drops due to excessive conduction loss. Efficiency versus load characteristics of an OFF-chip buck converter and an ON-chip FIVR are shown in Fig. 5. An OFF-chip buck converter that sits on a motherboard can typically take 12 V from the power supply unit, and down convert it to the voltage level to be used by either the FIVR or LDO [9]. On the other hand, FIVR uses 1.8 V as input voltage and generates different voltage levels based on the requirement of the cores, to which the converter is delivering power [3], [8]. As the output voltage of the converter reduces at constant load current, converter efficiency reduces. It can be verified from Fig. 5.

$$\eta = \frac{V_{\mathrm{OUT}} \cdot I_{\mathrm{OUT}}}{V_{\mathrm{OUT}} \cdot I_{\mathrm{OUT}} + P_{\mathrm{COND\_}Q1} + P_{\mathrm{COND\_}Q2} + P_{\mathrm{PAR\_}L} + P_{\mathrm{GATE}} + P_{\mathrm{CTRL}}}$$
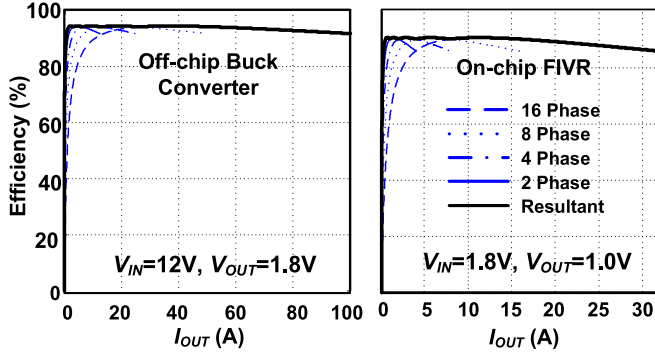
Fig. 6. Efficiency versus I_{OUT} for OFF-chip buck regulator, and ON-chip FIVR.
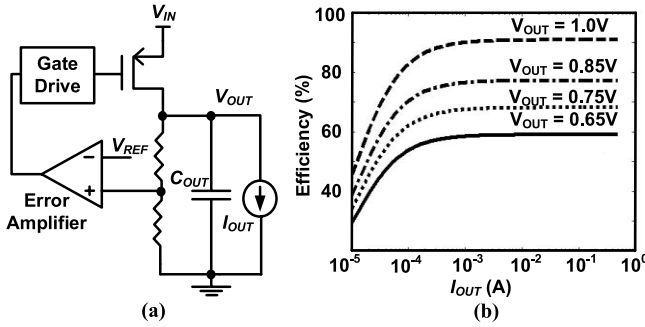

Fig. 7. (a) Block diagram and (b) efficiency versus I_{OUT} of LDO.

In order to improve light load efficiency and reduce output voltage ripple, instead of building a single converter, smaller converter modules are built. Running these converter modules in a phase-interleaved fashion ensures smaller output ripple, and phase dropping at light load ensures improvement in light load efficiency. Typical efficiency versus I_{OUT} characteristics of a 16-phase OFF-chip converter and a 16-phase FIVR are shown in Fig. 6 (left) and (right), respectively.

C. LDO Model

Block diagram of an LDO is shown in Fig. 7(a). An LDO has an n-type/p-type pass element, which generates a regulated output voltage (V_{OUT}) by dropping a portion of the input voltage (V_{IN}) across it. As V_{IN} reduces, or load (I_{OUT}) increases, V_{OUT} starts to drop and is sensed by the error amplifier. The error amplifier then generates a larger gate drive to regulate the output voltage. In order for the output to be regulated at a proper level, a minimum voltage, known as the dropout voltage of the regulator, has to be maintained across the pass gate. Efficiency of an LDO can be given as

$$\eta = \frac{V_{\text{OUT}} \cdot I_{\text{OUT}}}{V_{\text{IN}} \cdot (I_{\text{OUT}} + I_q)}$$

in which I_q is quiescent current of the LDO circuitry [10]. As output voltage deviates further from the input voltage, loss in the pass element increases, and efficiency of the regulator reduces. Efficiency of an LDO is also limited by I_q. At light load condition, I_q dominates over the load current, and hence, LDO efficiency drops at light load. Efficiency versus I_{OUT}
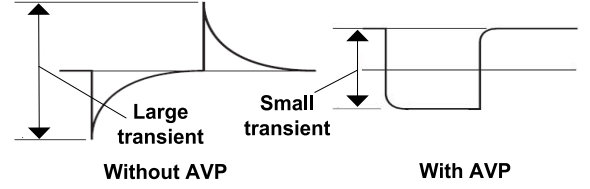

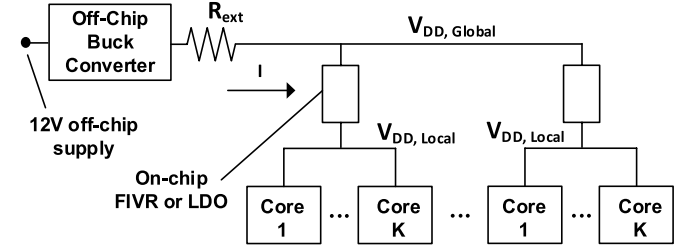Fig. 8. Transient response improves with AVP [11].

Fig. 9. Block diagram of power delivery scheme under consideration.

characteristics of the modeled LDO for a range of output voltages are shown in Fig. 7(b).

D. Active Voltage Positioning

In order to reduce the output ripple during voltage transients, regulation at the output of the converter is not made perfect by design [11]. At minimum load, V_{OUT} is set at a slightly higher voltage than its nominal value. Regulation is done in such a way that at full-load condition, V_{OUT} attains its nominal value. This technique is known as active voltage positioning (AVP) and is commonly used in voltage regulators in order to reduce transient microprocessor power at the expense of reduced output regulation. Simple waveforms in Fig. 8 show that AVP reduces the peak-to-peak output excursion. Although this paper concentrates on the steady-state power consumption of the system, and power consumed during microprocessor transient is out of scope of this paper, we still incorporate AVP into our steady-state system power analysis as AVP modulates steady-state load characteristics of the voltage regulators.

III. POWER DELIVERY SYSTEM ARCHITECTURE

Fig. 9 shows the block diagram of the power delivery system used in this paper. There is one 16-phase buck regulator sitting outside of the chip on the motherboard. It uses 12 V supply and generates output voltage levels to be used by subsequent converter stages. Efficiency versus load characteristics of this buck converter are shown in Fig. 6 (left). Inside the chip, there are ON-chip regulators (FIVR or LDO) and processor cores. IR noise due to external wire and package resistances is accounted for with a single lumped resistor, R_{ext}, in our model of the power delivery network. We assume that due to R_{ext}, the worst case power loss is ∼5%, which is typical of the current state-of-the-art power delivery networks.

In our analysis, we have assumed a processor with 256 cores. This assumption is in line with the number of cores in several recently developed processors, including NVIDIA's GPU accelerator Tesla K80 that has 4992 CUDA cores [12], and Intel's Xeon Phi processor that can have up to

61 cores [13]. Reference [14] presents an 80-tile TeraFLOPS processors built in Intel's 65-nm process. TILE-Mx is a 100-core processor from Tilera, and it is targeted toward high compute workloads [15]. Our system-level power-performance analysis aims to explore the design space of a 256-core server processor. However, our approach is generic and can be applied to processors with any number of cores without any modifications. Although we present our analysis based on a future 256-core processor, for completeness, we also include the results from 16-core and 64-core processors toward the end of this paper.

Like in any exploratory research, we had to make assumptions at various stages of the analysis. For example, the cores in our hypothetical processor are assumed to be homogeneous in nature, and they can be power gated individually. Supply line of these cores ($V_{\text{DD,Local}}$) is driven by an LDO in the case of the LDO-based power delivery scheme or by an FIVR in the case of the FIVR-based power delivery scheme. Each of these cores has DVFS capability with maximum and minimum operating frequencies of f and $f/2$ for corresponding logic $V_{\text{DD\_min}}$ values of 1 V ($V_{\text{DD\_HI}}$) and 0.65 V ($V_{\text{DD\_LO}}$), respectively. These cores are also equipped with a power-down mode for idle state. Although a continuous DVFS scheme would be more useful in terms of power savings, its implementation in a 256-core processor might be limited because of synchronization overhead across cores. Furthermore, DVFS p-states of processors are typically quantized and only a few of these states are frequently accessed, as can be seen from Fig. 1 (bottom). Hence, our assumption of a two-level DVFS operation is an acceptable compromise for keeping the analysis insightful and practical. Please note that 6T static random access memory-based caches are usually operated under a separate nominal voltage due to read and write margin constraints, and they will require separate voltage regulators. Our analysis is focused on power delivery to the core logic only.

In FIVR-based power delivery, we perform the analysis assuming various number of FIVRs (i.e., 16, 128, and 256) present ON-chip. For a 256-core processor, it translates to 16-, 2-, and 1-core per power domain, respectively. Current FIVR technology can support 59 inductors on an land grid array package with an area of 20 mm × 8 mm [3]. Die size of future 256-core processor is likely to be bigger. In addition, according to the package design rules, air-core inductors of FIVRs can be densely placed on the package. Although the feasibility of a 256 FIVR inductor on package is unknown, for the purpose of comparison with the LDO-based per-core DVFS scheme, we assume that the future FIVR technology will be able to support 256 inductors on a single package. The FIVR output voltage is determined by the activity of the cores powered up by that FIVR. Unless all the cores inside an FIVR domain run at a frequency of $f/2$, voltage of that domain has to be maintained at $V_{\text{DD\_HI}}$ in order to maintain the required throughput. However, if all the cores in a power domain can run at a frequency of $f/2$, then that power domain voltage can be set to $V_{\text{DD\_LO}}$. Per-core DVFS is possible when the number of FIVR increases to 256. Here, in our analysis, we assume that the no-load and full-load input voltages of
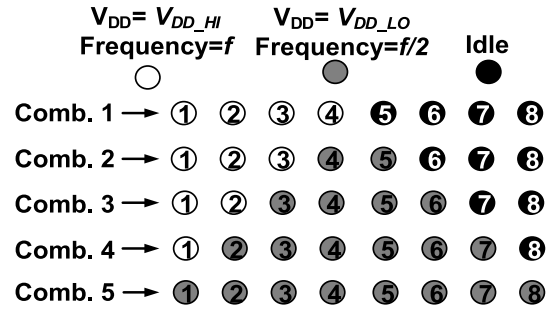

Fig. 10. Example showing an eight-core processor with the same throughput of 0.5 but different power consumptions.

FIVR are 1.8 and 1.7 V, respectively [3]. Efficiency versus load characteristics of FIVR are shown in Fig. 6 (right).

On the other hand, LDO-based power delivery can use 256 LDOs to supply power to 256 cores. This is due to the fact that the LDOs are inexpensive to build and usually occupies very small area. The LDO architecture with 16 and 128 power domains is inferior to the LDO architecture with 256 power domains in terms of total power consumptions. Hence, we did not include results for 16 and 128 power domain cases for the LDO. Because of the presence of 256 LDOs ON-chip, per-core DVFS is possible. However, it does not guarantee lower total system power than FIVR-based architectures because of conversion loss at low output voltages. In our analysis, we assume an LDO whose efficiency versus load characteristics are shown in Fig. 7(b).

IV. SYSTEM POWER ANALYSIS METHODOLOGY

For our processor, we assume a throughput-oriented architecture, in which the processor has a lot of inherent parallelism. Because of our choice of throughput as system performance metric, we used power consumption instead of energy consumption as the comparison metric of FIVR-based and LDO-based power delivery schemes. In case all the cores run at maximum frequency, we assume a normalized throughput of 1. However, the same throughput can result in different powers consumed by the cores. Fig. 10 shows various core configurations for the same normalized throughput of 0.5 in an eight-core configuration. From Fig. 10, we find that, in order to obtain normalized throughput of 0.5, four cores can run at frequency f, whereas other four cores can remain idle. However, this particular combination results in maximum power consumption, equal to $P_{\text{8Core}} = 4 \cdot C_{\text{EFF}} V_{\text{DD\_HI}}^2 f + 4 \cdot P_{\text{Leak}} + 4 \cdot P_{\text{Static}}$. Here, C_{EFF} is the effective dynamic capacitance of each core, including activity factor, P_{Leak} is the leakage power of an active core, and P_{Static} is the static power of an idle core and is due to the power gate leakage of the core. The lowest possible core power consumption corresponds to the case when all the cores run at a frequency of $f/2$, and is equal to $P_{\text{8Core}} = 8 \cdot C_{\text{EFF}} V_{\text{DD\_LO}}^2 ((f/2)) + 8 \cdot P_{\text{Leak}}$. Since at normal operating condition, P_{Leak} and P_{Static} are smaller than dynamic power, we find that the latter core combination consumes smaller core power under isothroughput condition, albeit at the expense of a longer execution time. However, we assume, in a power budget-constrained isothroughput scenario, the processor might have to sacrifice latency in lieu of smaller power.
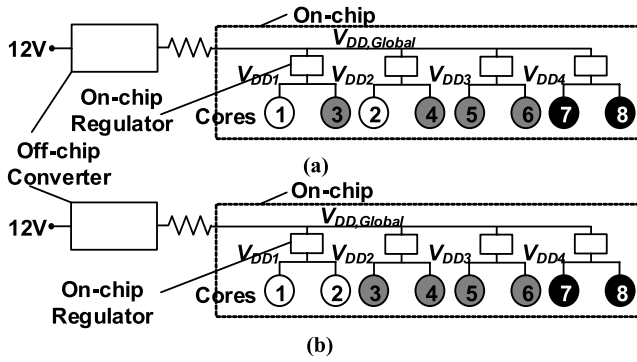
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS

Fig. 11.   Core assignment across power domains. (a) Inhomogeneous V_{DD}. (b) Homogeneous V_{DD}.
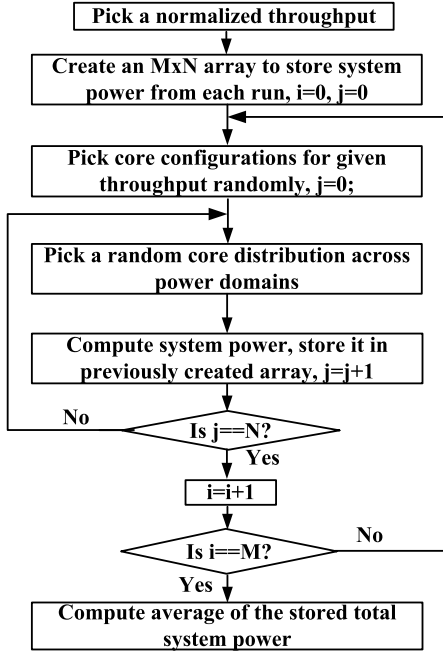

Fig. 12.   Flowchart showing average power computation steps. M is the number of MC runs for different core combinations and N is the number of MC runs for different core distributions across power domains.

In case per-core DVFS is not a viable option, total power consumed by the cores may vary greatly depending on how the cores are distributed among different power domains. In order to explain this point, we pick combination 3 from Fig. 10, and distribute the cores across four power domains in two different ways, as shown in Fig. 11(a) and (b). Fig. 11(b) shows that all the cores with equal supply voltage requirement have been grouped together in the same power domain. However, this is not the case in Fig. 11(a). Because of inhomogeneous grouping of the cores in Fig. 11(a), total power consumed by all cores will be larger compared with the case shown in Fig. 11(b).

To explore the entire design space for a random scheduler, we use a two-step Monte Carlo simulation, as shown in Fig. 12. In order to understand how the average system power computation is done with this technique, let us go back to our eight-core processor example in Fig. 10. For the given throughput, we assume that the scheduler randomly picks any

combination from Fig. 10. We further assume that the cores corresponding to that combination can be distributed across different power domains in a random fashion. Now, if we run Monte Carlo simulation for this two-step randomization process, and compute an average of the system powers obtained from all the occurrences, we will obtain average system power for that particular normalized throughput. At the same time, we can find out the minimum and maximum system power. For example, the minimum power scheduler will maximize the number of homogeneous power domains [16] and pick combination 5 from Fig. 10.

Once configurations of the cores across power domains are decided, total system power can be obtained by adding the power consumed by cores (P_{core}), ON-chip voltage regulators ($P_{ON-chip\_reg}$), power distribution network (P_{supply_net}), and OFF-chip voltage regulator ($P_{OFF-chip\_reg}$), i.e., $P_{total} = P_{core} + P_{ON-chip\_reg} + P_{OFF-chip\_reg} + P_{supply\_net}$. P_{core} includes dynamic and leakage power of an active core, and static power of an idle core, as described in Section III. $P_{ON-chip\_reg}$ and $P_{OFF-chip\_reg}$ are the power lost in the ON-chip voltage regulator (FIVR or LDO), and OFF-chip buck converter, respectively. Depending on the supply voltage and frequency of the cores, and the AVP requirement of the ON-chip regulator for better dynamics, efficiency of the ON-chip regulator and the power lost in it change. Input current–voltage profile of the ON-chip regulator determines the IR drop in the supply network between the ON-chip and OFF-chip regulator, and the output voltage requirement from the OFF-chip buck converter. This variation in the output voltage and current of an OFF-chip converter, and the AVP requirement of the buck converter determines the power loss in the OFF-chip buck converter.

In our analysis, we take output voltage and current-dependent power loss of the voltage regulators into account to find out total system power consumption. To find out load-dependent power loss of LDO, FIVR, and OFF-chip switching regulator, we use the models mentioned in Section II.

V. SIMULATION RESULTS

In our analysis, we assume a hypothetical 256-core processor with: 1) 256 power domains for LDO-based power delivery and 2) 16, 128, and 256 power domains for FIVR-based power delivery. Monte Carlo simulations have been performed at a constant normalized throughput for all power delivery schemes under consideration. Fig. 13 shows the system power versus normalized throughput assuming no process variation (i.e., each core operates at the same V_{DD_HI} or V_{DD_LO} voltage). Fig. 13 indicates that the range of power consumption is maximum for a normalized throughput of 0.5, and it tapers down gradually as the throughput increases or decreases. It is due to the fact that toward midthroughput region, the number of combinations (as shown in Fig. 10, we can have five combinations for a normalized throughput of 0.5 in an eight-core processor) to obtain same throughput increases, thereby increasing the power consumption range. For a random scheduler, in order to get the average system power for a given throughput, we compute the average of the power consumption values for that particular

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

PAUL *et al.*: SYSTEM-LEVEL POWER ANALYSIS OF A MULTICORE MULTIPOWER DOMAIN PROCESSOR 7
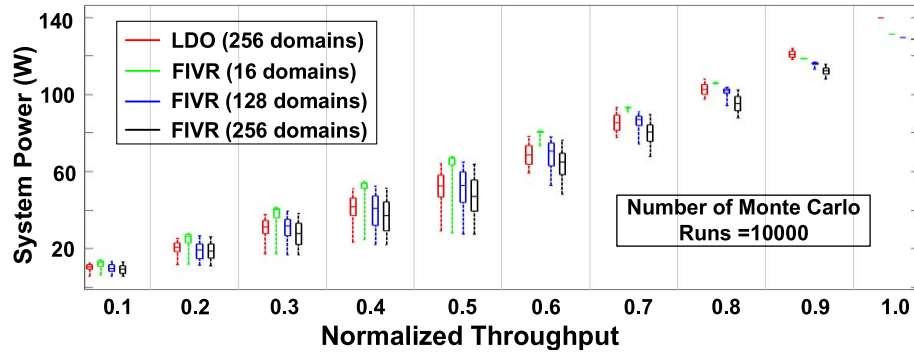

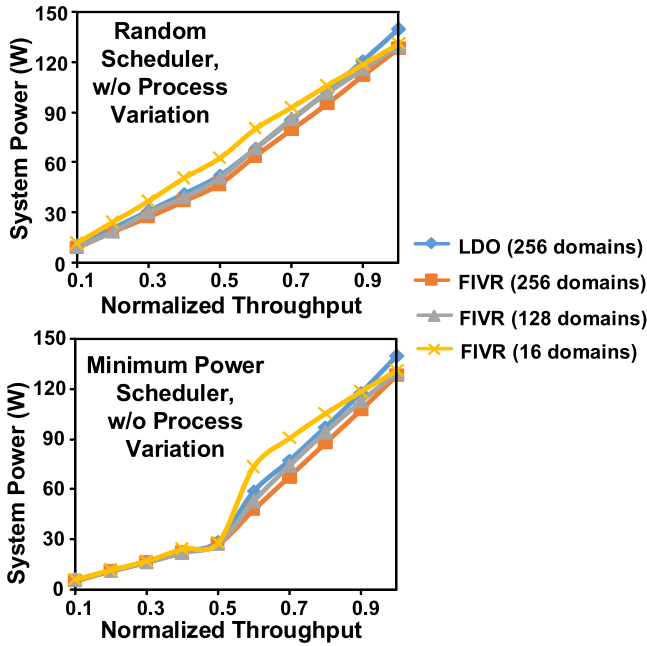Fig. 13.   Power consumption versus throughput for various power delivery options (without process variation).

Fig. 14.   System power versus throughput for various power delivery options (without process variation).

Fig. 15.   Process variation-induced V_{DD_LO} and V_{DD_HI} variation for 256 cores.

throughput from Fig. 13. As for a minimum power scheduler, power consumption corresponds to the minimum power point corresponding to each throughput values of Fig. 13.

Fig. 14 shows system power plotted against normalized throughput for the random scheduler (top) and the minimum power scheduler (bottom). From Fig. 14, we see that the average system power with the LDO is smaller than that with 16 FIVRs because of per-core DVFS capability with LDO (24% less power consumption at normalized throughput of 0.5). However, if the number of power domains using FIVR increases to 128, FIVR-based design becomes comparable to that of the LDO because of better FIVR efficiency. Eventually, power consumption with 256 FIVR domains becomes less than that of the LDO-based design by 12% at a throughput of 0.5 for the random scheduling technique. Note that perfect homogeneous grouping of cores is always possible when normalized throughput $\leqslant 0.5$. Hence, from Fig. 14 (bottom), we see that the minimum system power is independent of the DVFS configuration for normalized throughput $\leqslant 0.5$. When normalized throughput approaches 1, all the cores in all
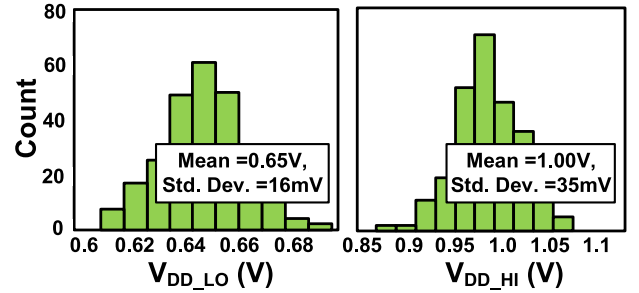
the power domains operate at the maximum frequency. Hence, the system power consumption of FIVR architectures with 16, 128, and 256 power domains becomes almost equal. However, for the same condition, the system power consumption of LDO-based design is higher than that of FIVR designs. This can be attributed to the difference in the OFF-chip regulator efficiency. The LDO requires an input voltage of ~1.05 V, whereas FIVR requires an input voltage of ~1.8 V. Due to the higher power loss in the OFF-chip converter when generating a 1.05 V, the overall system power is higher for the LDO case.

In real world, the systematic and random process variations will cause threshold voltage of transistors to shift. This variation will cause the voltage–frequency relation of cores to differ from one another. As a result, if more than one cores share the same supply voltage, that supply voltage will be determined by the supply voltage requirement of the slowest core. Consequently, total dynamic power consumption will increase. To a first-order, the supply voltage required to meet a target frequency can be approximated as a linear function of threshold voltage [17]. For the sake of analysis, we assume V_{DD_LO} is normally distributed with a mean value of 0.65 V and a standard deviation of 16 mV, and V_{DD_HI} is normally distributed with a mean value of 1.0 V and a standard deviation of 35 mV (Fig. 15). Fig. 16 shows the total system power versus normalized throughput using a random scheduler (top) and a minimum power scheduler (bottom), taking process variation into account. Process variation makes LDO-based DVFS approach less attractive than FIVR-based one due to the following reason: V_{IN} for the LDOs, which is also $V_{DD,Global}$ in Fig. 9, has to be determined by the slowest core under process variation, and hence, must be slightly increased to meet the same performance. Since this will result in a larger

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

8

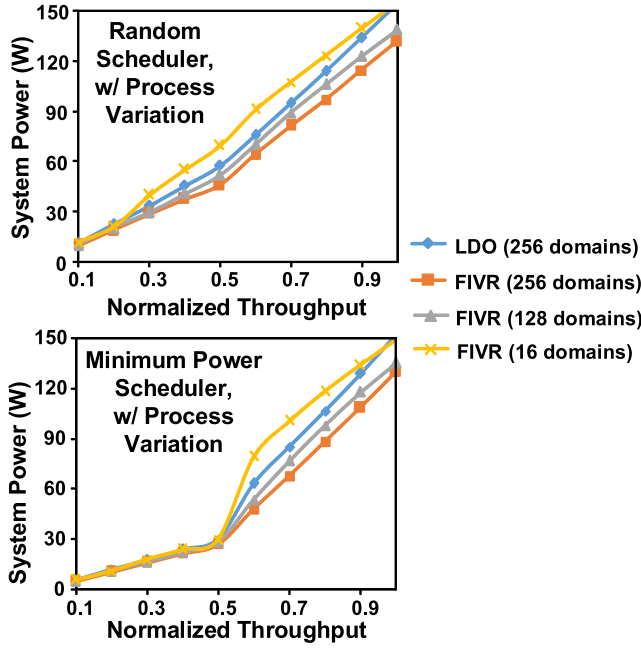IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS

Fig. 16. System power versus throughput for various power delivery options (with process variation).
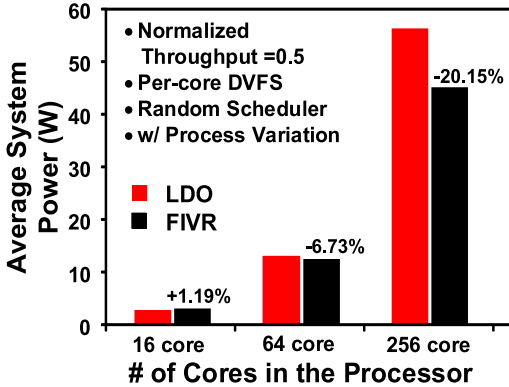

Fig. 17. Average system power versus number of cores for per-core LDO and per-core FIVR.

voltage drop in the LDO circuit, the power will be higher as compared with the no process variation case. As a result, for a throughput of 0.5 and for a random scheduling algorithm, FIVR with 128 domains consumes 10% less power than the per-core LDO, while FIVR with 256 domains consumes 20% less power than the per-core LDO.

Finally, in Fig. 17, we plot average system power versus number of cores in a processor for a throughput of 0.5. The plot shows that as the number of cores increases, FIVR becomes more attractive. This trend can be explained as follows. In both the LDO-based and FIVR-based designs, the input voltage of the ON-chip converter, which is also the output voltage of the OFF-chip switching converter, is determined by the highest core voltage. The highest core voltage under process variation is higher with more number of cores in the processor (and similarly, the lowest core voltage is lower), so the difference between the shared input voltage and the output voltages of the individual ON-chip converters increases. This causes the LDO efficiency to drop whereas the FIVR efficiency remains relatively constant.

VI. CONCLUSION AND KNOWN LIMITATIONS

In this paper, we compare an FIVR-based power delivery solution with an LDO-based one, in terms of system power consumption of a multicore, multipower domain processor. Our analysis shows that under random scheduling and process variation, for a normalized throughput of 0.5, LDO and FIVR-based per-core DVFS systems consume almost similar amount of power for a 16-core system. The advantage of using FIVR as the ON-chip voltage regulator becomes more prominent when the number of cores in the processor increases (e.g., 64 or 256 cores).

Although our estimation methodology is sufficient to gain insight into the overall benefits of FIVR and LDO, it could be refined to study the specific aspects of the two power delivery methods. For instance, a more detailed power distribution network, including ON-chip parasitic, can be used to incorporate the impact of local supply noise at the expense of further complexity of the model. In addition, the impact of LDO-FIVR-based hybrid power delivery solutions on the total system power consumption can be evaluated. Our analysis focuses on the steady-state power consumption, but the methodology can be extended to capture the effect of system transient on instantaneous system power. Finally, since these power-performance characteristics are dependent on the type of scheduler, further research can be carried out based on the application-specific schedulers.

REFERENCES

[1] J. Lee, G. Hatcher, L. Vandenberghe, and C.-K. 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.

[2] 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.

[3] E. A. Burton *et al.*, "FIVR—Fully integrated voltage regulators on 4th generation Intel Core SoCs," in *Proc. Appl. Power Electron. Conf. (APEC)*, Mar. 2014, pp. 432–439.

[4] E. J. Fluhr *et al.*, "5.1 POWER8: A 12-core server-class processor in 22 nm SOI with 7.6 Tb/s off-chip bandwidth," in *Proc. Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2014, pp. 96–97.

[5] A. A. Sinkar, H. Wang, and N. S. Kim, "Workload-aware voltage regulator optimization for power efficient multi-core processors," in *Proc. Design Autom. Test Eur. (DATE)*, Mar. 2012, pp. 1134–1137.

[6] W. Lee, Y. Wang, and M. Pedram, "VRCon: Dynamic reconfiguration of voltage regulators in a multicore platform," in *Proc. Design Autom. Test Eur. (DATE)*, Mar. 2014, pp. 1–6.

[7] W. Kim, M. S. Gupta, and G.-Y. Wei, "System level analysis of fast, per-core DVFS using on-chip switching regulators," in *Proc. High Perform. Comput. Archit. (HPCA)*, Feb. 2008, pp. 123–124.

[8] N. Kurd *et al.*, "Haswell: A family of IA 22 nm processors," in *Proc. Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2014, pp. 112–113.

[9] K. Yao, M. Ye, M. Xu, and F. C. Lee, "Tapped-inductor buck converter for high-step-down DC-DC conversion," *IEEE Trans. Power Electron.*, vol. 20, no. 4, pp. 775–780, Jul. 2005.

[10] G. A. Rincón-Mora, "Current efficient, low voltage, low dropout regulators," Ph.D. dissertation, School Elect. Comput. Eng., Georgia Inst. Technol., Atlanta, GA, USA, 1996.

[11] R. Sheehan. (Nov. 1999). *Active Voltage Positioning Reduces Output Capacitors Linear Technology*. [Online]. Available: http://www.linear.com/docs/5600

[12] NVIDIA Corp. *Tesla Server Solutions*, accessed on Aug. 8, 2015. [Online]. Available: http://www.nvidia.com/object/tesla-servers.html

[13] Intel Corp. *Intel Xeon Phi Product Family: Product Brief*, accessed on Aug. 8, 2015. [Online]. Available: http://www.intel.com/content/www/us/en/high-performance-computing/high-performance-xeon-phi-coprocessor-brief.html

[14] S. R. Vangal *et al.*, "An 80-tile sub-100-W TeraFLOPS processor in 65-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 29–41, Jan. 2008.

[15] Tilera Corp. *TILE-Mx Multicore Processor*, accessed on Aug. 8, 2015. [Online]. Available: http://www.tilera.com/products/?ezchip=585&spage=686

[16] G. Yan, Y. Li, Y. Han, X. Li, M. Guo, and X. Liang, "AgileRegulator: A hybrid voltage regulator scheme redeeming dark silicon for power efficiency in a multicore architecture," in *Proc. High Perform. Comput. Archit. (HPCA)*, Feb. 2012, pp. 1–12.

[17] K. J. Kuhn *et al.*, "Process technology variation," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, Aug. 2011.

Ayan Paul received the B.E. degree in electronics and telecommunication engineering from Jadavpur University, Kolkata, India, in 2005, the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2008, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA. His Ph.D. research was focused on circuit design for power management which included resonant supply noise reduction technique using circuit/architectural approach and the design of on-chip switched capacitor based step-up and step-down dc/dc converters.

He was with PricewaterhouseCoopers, Kolkata, India, in 2005 and 2006, as a Consultant, and Atrenta, Noida, India, as a Corporate Applications Engineer, in 2006 and 2007. He is currently involved with the CPU Memory Design Team, Qualcomm, San Diego, CA, USA. He is also involved in transistor leakage modeling and spin-transfer torque-RAM scaling analysis.

Sang Phill Park received the B.E. degree in architecture engineering from Hongik University, Seoul, South Korea, in 2000, the B.S. degree in computer engineering from the University of Arizona, Tucson, AZ, USA, in 2004, and the Ph.D. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, in 2011.

He was a Software Engineer with Language Bank Inc., Seoul, from 2000 to 2002. He was with the Exploratory VLSI Design Group, IBM Austin Research Laboratory, Austin, TX, USA, in 2008, as a Research Intern. He is currently with the Advanced Path-Finding Research Team, Graphics Architecture Group, Intel Corporation, Hillsboro, OR, USA, where he developed advanced design methodologies for multiple Intel's graphics products. He has authored 28 technical papers and holds three patents in the field of VLSI and computer science.

Dinesh Somasekhar received the B.E. degree in electronics engineering from the Maharaja Sayajirao University of Baroda, Vadodara, India, in 1989, the M.E. degree in electrical communications engineering from the Indian Institute of Science, Bangalore, India, in 1990, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA, in 1999.

He was an IC Design Engineer with Texas Instruments, Bangalore, from 1991 to 1994, where he designed application-specified integrated circuit compiler memories and interface ICs. From 1999 to 2011, he was with the Circuits Research Laboratory, Intel Corporation, Hillsboro, OR, USA, where he was responsible for research on memory technologies. From 2011 to 2012, he was with GlobalFoundries, Sunnyvale, CA, USA, where he was responsible for defining the memory bit-cell menu for 14-nm class technologies. Since 2012, he has been part of the Data-Center Group of Intel Corporation. He is currently a Principal Engineer with Intel Corporation. He is responsible for the memory strategy on the exascale computing initiative—part of the Data-Center Group path-finding. He has authored 35 papers, three book chapters, and holds over 80 patents in the field of VLSI.

Dr. Somasekhar served as a Mentor at the Semiconductor Research Consortium, and has participated on the Technical Program Committee of International Symposium on Low Power Electronics and Design (ISLPED), International Symposium on Quality Electronic Design, Design, Automation and Test in Europe, Great Lakes Symposium on VLSI, and Custom Integrated Circuits Conference.

Young Moon Kim received the B.E. (*summa cum laude*) degree in electronics engineering from Seoul National University, Seoul, South Korea, in 2005, the M.S. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2007, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2013.

He was with IBM, Essex Junction, VT, USA, Intel Corporation, Hillsboro, OR, USA, NEC, Sagamihara, Japan, and Samsung, Kihung, South Korea, as an Intern while pursuing the Ph.D. degree. Since 2013, he has been with Intel Corporation. His current research interests include on-chip memory systems and CMOS circuit reliability.

Dr. Kim was a recipient of the Samsung Scholarship in 2007.

Nitin Borkar received the M.S. degree in physics from the University of Mumbai, Mumbai, India, in 1982, and the M.S. degree in electrical engineering from Louisiana State University, Baton Rouge, LA, USA, in 1985.

He joined Intel Corporation, Hillsboro, OR, USA, in 1986. He has held several technical and senior management positions with the Micro-Processor Design Group, the Supercomputer Systems Group, and the Corporate Technology Research Group, Intel Corporation. He has participated and led development and delivery of number of leadership projects, including Intel 80960 embedded processor, Intel 80486DX2 processor, ASCI-Red Tera-FLOPS system, TCP/IP hardware accelerator, a single chip Tera-FLOPS research processor, and a single chip cloud computer. He is currently a member of the Platform Engineering Group, Hillsboro, and manages the Advanced Development Team for Visual and Parallel Processing Group and the Exascale Prototype Processor Design Teams. He holds ten patents, with nine patents pending in the areas of high performance circuits, low-power VLSI circuit design, on-die communication circuits, and special purpose hardware designs. He has authored or co-authored over 18 papers in the above areas.

Ulya R. Karpuzcu received the Ph.D. degree in computer engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA.

She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA. Her current research interests include the impact of process technology on computing systems.

Chris H. Kim (M'04–SM'10) received the B.S. and M.S. degrees from Seoul National University, Seoul, South Korea, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA.

He joined the University of Minnesota, Minneapolis, MN, USA, in 2004, where he is currently a Professor. His current research interests include digital, mixed-signal, and memory circuit design in silicon and non-silicon (organic thin film transistor and spin) technologies.

Prof. Kim is a recipient of a Council of Graduate Students Outstanding Faculty Award, an NSF CAREER Award, a Mcknight Land-Grant Professorship, a 3M Non-Tenured Faculty Award, DAC/ISSCC Student Design Contest Awards, IBM Faculty Partnership Awards, an IEEE Circuits and Systems Society Outstanding Young Author Award, and ISLPED Low Power Design Contest Awards.