ThermoGater: Thermally-Aware On-Chip Voltage Regulation

S. Karen Khatamifard* Longfei Wang† Weize Yu† Selçuk Köse† Ulya R. Karpuzcu*

* University of Minnesota † University of South Florida
{khatami,ukarpuzc}@umn.edu {longfei, weizeyu}@mail.usf.edu kose@usf.edu

ABSTRACT

Tailoring the operating voltage to fine-grain temporal changes in the power and performance needs of the workload can effectively enhance power efficiency. Therefore, power-limited computing platforms of today widely deploy integrated (i.e., on-chip) voltage regulation which enables fast fine-grain voltage control. Voltage regulators convert and distribute power from an external energy source to the processor. Unfortunately, power conversion loss is inevitable and projected integrated regulator designs are unlikely to eliminate this loss even asymptotically. Reconfigurable power delivery by selective shut-down, i.e., *gating*, of distributed on-chip regulators in response to spatio-temporal changes in power demand can sustain operation at the minimum conversion loss. However, even the minimum conversion loss is sizable, and as conversion loss gets dissipated as heat, on-chip regulators can easily cause thermal emergencies due to their small footprint.

Although reconfigurable distributed on-chip power delivery is emerging as a new design paradigm to enforce sustained operation at minimum possible power conversion loss, thermal implications have been overlooked at the architectural level. This paper hence provides a thermal characterization. We introduce *ThermoGater*, an architectural governor for a collection of practical, thermally-aware regulator gating policies to mitigate (if not prevent) regulator-induced thermal emergencies, which also consider potential implications for voltage noise. Practical ThermoGater policies can not only sustain minimum power conversion loss throughout execution effectively, but also keep the maximum temperature (thermal gradient) across chip within 0.6°C (0.3°C) on average in comparison to thermally-optimal oracular regulator gating, while the maximum voltage noise stays within 1.0% of the best case voltage noise profile.

CCS CONCEPTS

• **Hardware** → **Power and energy**; **Thermal issues**;

KEYWORDS

Power distribution, on-chip voltage regulation, thermal emergencies

ACM Reference format:
S. Karen Khatamifard Longfei Wang Weize Yu Selçuk Köse Ulya R. Karpuzcu University of Minnesota University of South Florida. 2017. ThermoGater: Thermally-Aware On-Chip Voltage Regulation . In *Proceedings of ISCA '17, Toronto, ON, Canada, June 24-28, 2017,* 13 pages.
http://dx.doi.org/10.1145/3079856.3080250

1 MOTIVATION

Due to gradual (if not stagnated) voltage scaling, chip power density (power per chip area) has been growing over technology generations [11]. At the same time, cooling limitations prevent a proportional expansion of the chip power budget. In this power-limited environment, the only way to avoid performance degradation becomes to enhance the power efficiency, i.e., performance improvement per unit power consumed. Both, power and performance strongly depend on the operating voltage V_{dd}. Therefore, tailoring V_{dd} to fine-grain temporal changes in the power and performance needs of the workload can effectively enhance power efficiency. As a result, to enable fast fine-grain voltage control, power-limited computing platforms of today widely deploy *integrated* – be it partially *on-package* [4, 21] or entirely *on-chip* [38] – voltage regulation.

In supplying a fixed or time-varying V_{dd} to logic and memory blocks, voltage regulators convert and distribute power from an external energy source to the processor. Regulator *power conversion efficiency*, η, is defined as the ratio of output power to the input power (of the regulator). Due to inevitable losses in power conversion, the best-known and projected integrated regulators fail to reach 100% efficiency even asymptotically. Conversion efficiency η changes as a function of the load current at the regulator output as depicted in Fig. 1 for a representative subset of highly-optimized, recent regulator designs presented in ISSCC 2015 [1, 14, 15, 26, 29, 31, 36, 37]. Output load current I_{out} (as captured by the x-axis) represents a proxy for microarchitectural activity.

Under stringent (voltage) noise requirements, conventional integrated regulators are calibrated to deliver the peak efficiency, η_{peak}, at a specific I_{out} which typically matches maximum microarchitectural activity. However, once deployed, regulators spend most of the time under much lighter load conditions [33]. This is the primary reason behind further power conversion losses.

An off-chip voltage converter supplies the input to on-chip voltage regulators over the *global* power grid. Each on-chip regulator, in turn, delivers power over a *local* grid to its V_{dd}*-domain*, i.e., logic or memory blocks connected to the output of the voltage regulator over the local power grid. Power managers can control the V_{dd} of each domain separately.

An emerging practice is distributing many small regulators (which can be homogeneous or heterogeneous in their topology and electrical characteristics), across each V_{dd}-domain. Connected in parallel, these small regulators can cumulatively supply an I_{out} corresponding to the sum of the $I_{out,r}$ of each component regulator r. In this setting,

ISCA '17, June 24-28, 2017, Toronto, ON, Canada

ACM ISBN 978-1-4503-4892-8/17/06…$15.00
http://dx.doi.org/10.1145/3079856.3080250

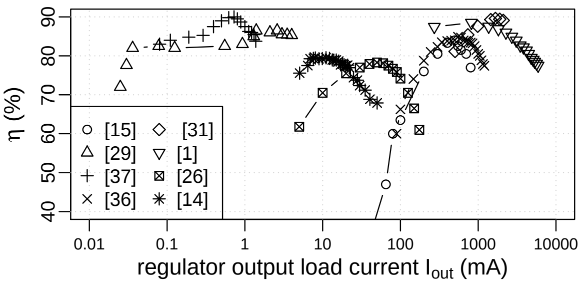


Figure 1: Reported power conversion efficiency η of a representative subset of recent, highly optimized regulators from ISSCC 2015 [1, 14, 15, 26, 29, 31, 36, 37].

each component regulator can operate at its peak efficiency $\eta_{peak,r}$ at a specific value (usually maximum) of $I_{out,r}$. If each component regulator is always enforced to operate at its respective $\eta_{peak,r}$, by modulating the number of active component regulators within the V_{dd}-domain, n_{on}, a wide range of I_{out} can be supplied to the respective V_{dd}-domain at the minimum possible power conversion loss. Such reconfigurable power delivery by selective shut-down, i.e., *gating*, of distributed on-chip regulators in response to spatio-temporal changes in I_{out} can sustain operation at $\eta_{peak(,r)}$ throughout the execution.

However, even at η_{peak}, a notable power conversion loss of around 10% is the case for the best-known industrial integrated regulators of today [4, 38]. At the same time, integrated regulators are miniaturized to minimize the area overhead. As power conversion loss gets dissipated as heat, integrated regulators can easily cause thermal emergencies due to their small footprint. Deviation from the peak efficiency η_{peak} as a result of fluctuations in microarchitectural activity can only exacerbate the thermal profile by generating even more heat due to higher conversion loss.

Integrated regulators close to high-activity functional blocks can increase the span and the temperature of hotspots, or give rise to new hotspots. Integrated regulators close to low-activity regions such as caches, on the other hand, can raise the local temperature. Consequently, the *maximum temperature* observed across chip, T_{max}, can rise, which in turn can lead to a higher *maximum spatial difference in temperature*, i.e., *thermal gradient*. A higher T_{max} may exceed the permissible maximum and degrade performance by triggering thermal throttling. A higher thermal gradient may enforce operation at a lower speed to mask potential timing violations due to temperature-induced spatial timing variations [27]. Mean time to failure (MTTF) for silicon wear-out mechanisms can also decrease notably with increasing temperature [17]. Worse, static power consumption skyrockets at higher temperatures.

Although reconfigurable distributed on-chip power delivery is promising as a new design paradigm to enforce sustained operation at η_{peak} throughout the execution, thermal implications have been overlooked at the architectural level. *How to address thermal emergencies due to integrated voltage regulation* hence forms the focus of this study, considering multiple V_{dd}-domains (each featuring many small regulators) dispersed across chip. This paper

- investigates thermal implications of distributed integrated voltage regulation;
- provides an architectural exploration of how spatio-temporal selective gating of regulators can mitigate regulator-induced thermal emergencies;
- introduces *ThermoGater*, a collection of practical runtime policies to orchestrate thermally-aware regulator gating.

Thermally-aware regulator gating policies from ThermoGater spatio-temporally manage a parallel network of many small regulators dispersed across a V_{dd}-domain, by closely tracking microarchitectural activity. Circuit blocks serviced by a hot regulator do not starve if the respective regulator is gated. Rather, cooler regulators in the same V_{dd}-domain take over the load of the gated regulator. Therefore, spatio-temporal regulator gating can effectively spread the heat concentrated around a hot regulator to neighboring circuit blocks and cap local peak temperatures [12] without compromising performance, as opposed to the vast majority of conventional dynamic thermal management techniques [17]. The end effect is a reduction in both, T_{max} and the thermal gradient across chip.

Be it thermally-aware or not, the goal of regulator gating is to sustain operation at η_{peak} over a wide I_{out} range, which imposes a stringent constraint on the maximum number of regulators that can be gated at a given time. Hence, ThermoGater first determines the number of active regulators within a V_{dd}-domain, n_{on}, required to sustain operation at η_{peak}. ThermoGater next identifies a subset of the regulators of size n_{on} to turn on. Therefore, under thermally-aware regulator gating, circuit blocks may not always be supplied by the closest regulator. The potential adverse affect is higher voltage noise, but ThermoGater takes the implications into account.

To our knowledge, this paper represents the first architectural study of thermally-aware regulator gating, which has been only very recently explored at the circuit-level [19] in broad terms where quite generic guidelines are provided for the voltage regulator physical placement to mitigate thermal emergencies. In the following, Section 2 details thermal implications of integrated voltage regulation; Section 3 provides the background; Section 4 covers how regulator gating can help mitigate (if not avoid) regulator-induced thermal emergencies; Sections 5 and 6 provide the evaluation; and Section 7 concludes the paper.

2 INTEGRATED VOLTAGE REGULATION: THERMAL IMPACT

The primary design objectives for integrated voltage regulators are i) to maximize the power conversion efficiency η, ii) to minimize the on-chip area overhead, and iii) to minimize the response time to transient changes in I_{out} due to microarchitectural activity. Regulator power conversion efficiency is defined as $\eta = P_{out}/P_{in}$ where P_{in} and P_{out} are, respectively, the input and output power of the regulator. As depicted in Fig. 1, η may degrade significantly off the peak (which typically corresponds to maximum anticipated microarchitectural activity).

Voltage regulators convert and distribute power from an external energy source to the processor. The power dissipated during conversion, P_{loss}, is the difference $P_{in} - P_{out}$. $P_{in} = V_{in} \times I_{in}$ and $P_{out} = V_{out} \times I_{out}$ where V_{in}, I_{in}, V_{out}, and I_{out} are, input voltage, input current, output voltage, and output (load) current of the regulator, respectively. Hence,

$$P_{loss} = P_{out} \times (1/\eta - 1) = V_{out} \times I_{out} \times (1/\eta - 1) \qquad (1)$$

applies. As Fig. 1 reveals, η is a function of I_{out}.

In a reconfigurable distributed power delivery network, the η_r of each component regulator r typically monotonically increases with $I_{out,r}$, reaches a peak, and degrades past the peak, similar to the the curves from Fig. 1. However, modulating the number of active component regulators (as explained in Section 1, as a function of microarchitectural activity) can make the effective cumulative η barely change with I_{out} (which represents the sum of component $I_{out,r}$) over a wide I_{out} range [4].

Coming back to Eqn. 1, at a given V_{out} (=V_{dd}), how P_{loss} changes with I_{out} closely tracks how η evolves with I_{out}. P_{loss} is dissipated as heat, and therefore can increase the local temperature by the regulator significantly, as a function of the physical footprint of the regulator. The physical footprint of regulators from Fig. 1 varies, as well, due to differences in regulator topologies. From different regulator designs featuring similar η_{peak}, the smaller ones are more likely to cause thermal problems. For each regulator design from Fig. 1, P_{loss} per area tends to dip at η_{peak} which incurs the minimum P_{loss} by construction.

A Motivating Case Study: The reported regulator P_{out} per area in Intel's Haswell processor is 33.6W/mm^2 [21], with η_{peak}=90%. In this case, according to Eqn. 1, P_{loss} per area becomes 3.7W/mm^2 at η_{peak} [4]. As air (microchannel) cooling limit is around 1.5W/mm^2 (7.9W/mm^2) [13], this P_{loss} per area of 3.7W/mm^2 can result in thermal emergencies, depending on where a regulator resides on-chip[1].

3 BACKGROUND & RELATED WORK

3.1 Distributed Voltage Regulation

Spatio-temporal selective shutdown of integrated regulators, *regulator gating*, applies to any distributed power delivery network, where a number of regulators are connected in parallel and dispersed across a V_{dd}-domain to maximize the physical proximity to their respective load (circuit blocks). The increased proximity provides very fast response time in tailoring V_{dd} to varying load conditions, i.e., microarchitectural activity. At the same time, regulating V_{dd} at close proximity to the load (circuit blocks) minimizes voltage noise [20, 46]. The component regulators can be *homogeneous* or *heterogeneous* [40] in terms of circuit topology and other electrical characteristics.

Modern processors widely deploy three types of integrated regulators: buck, switched capacitor (SC), and linear low-dropout (LDO). The conversion efficiency of buck regulators can reach over 90%, but usually at a high area penalty. Intel Haswell's fully integrated voltage regulator (FIVR) represents a buck regulator, which mitigates the area overhead due to bulky inductors by keeping these on package while the remaining components of the regulator are placed on chip [4]. While regulator components are distributed between the package and the chip, the regulation happens on-chip in this case. η_{peak}=90% applies where the reported P_{out} per area is 33.6W/mm^2 [21]. For SC regulators, η can reach over 90%, as well, and the footprint is usually smaller than comparable buck regulators [2]. LDO regulators are also area efficient, but both buck and SC regulators can deliver a higher η over a wide V_{out} range when compared to LDO regulators. This is because η of LDO regulators closely tracks V_{out}/V_{in}, and hence degrades significantly as the difference between V_{in} and V_{out} increases (e.g., while supplying a low V_{out} =V_{dd} to the processor). IBM's POWER8 features LDO regulators with a P_{out} per area of 34.5W/mm^2, and η_{peak} of 90.5% [8, 38].

All three types of regulators can represent component regulators in a distributed power delivery network. In this setting, close physical proximity of component regulators to their respective circuit blocks enables sub-nanosecond response times while tracking microarchitectural activity [20, 46]. For example, POWER8 features a network of 64 uniformly distributed LDO regulators per V_{dd}-domain [8, 38]. Each regulator has a bypass mode (i.e., the equivalent of gating) to eliminate P_{loss} when the respective output load is low. POWER8 design corresponds to a *homogeneous* distributed network as component LDO regulators (i.e., *microregulators* in IBM terminology) are electrically identical.

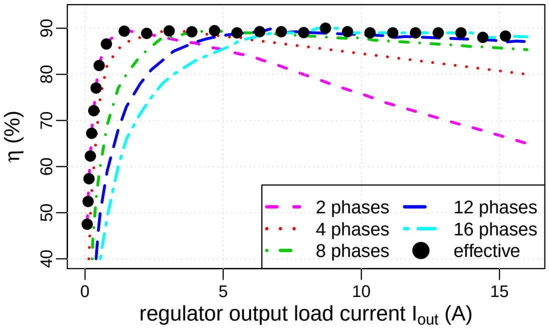


Figure 2: η of a 16-phase regulator from Intel [21].

Multi-phase regulators are among the most efficient buck and SC regulators. In this case, the regulator itself comprises a parallel network of multiple component regulators which are electrically identical by design, but operate at different phases [15, 16, 26, 28, 36]. Throughout execution, a varying number of the phase-shifted component regulators (i.e., *phases*) are activated to feed the load circuit blocks. For example, Haswell's fully integrated voltage regulator (FIVR) design features phase interleaved buck converters [4, 10, 21]. Different numbers of active phases give rise to different η vs. I_{out} characteristics, each reaching η_{peak} at a different I_{out} (Fig. 2). Hence, by tailoring the number of active phases to the instantaneous I_{out} demand (as dictated by microarchitectural activity), a multi-phase regulator can effectively sustain η_{peak} throughout execution. In this case, distributing phases across a V_{dd}-domain gives rise to an alternative *homogeneous* distributed network design [20, 46].

Although reconfigurable distributed power delivery is emerging as a new design paradigm to increase the performance [9], η [23], and battery life [4, 21], to reduce the energy consumption [33], or to minimize voltage noise [8, 38], thermal implications have been overlooked at the (micro)architectural level.

[1] In fact, IBM POWER8 sensors table from [7] features a "VRHOT" signal, the description of which points to "regulator overheating".

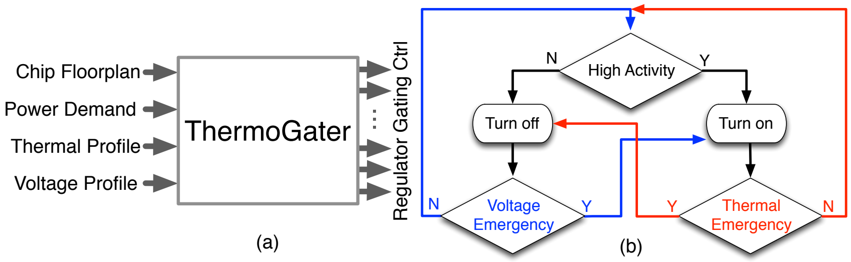
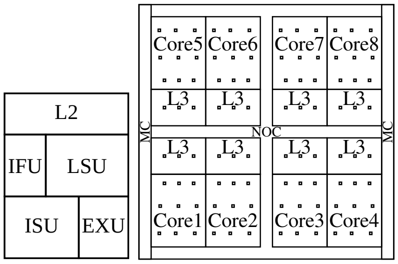


Figure 3: Thermally-aware regulator gating: macroscopic (a) and microscopic (b) view.



(a) core **(b) chip**

Figure 4: Simplified floorplan.

3.2 (Thermally-Oblivious) Regulator Gating

Regulator power conversion efficiency η strongly depends on the output load current I_{out}, as shown in Fig. 1. Any time I_{out} deviates from $I_{out,peak}$, at which the regulator operates at peak power conversion efficiency η_{peak}, conversion efficiency degrades. It is possible to reconfigure regulators in response to varying I_{out} demand, such that η vs. I_{out} curve shifts to enforce peak efficiency η_{peak} at the instantaneous I_{out} [22, 33]. In a distributed power delivery network (Section 3.1), modulating the number of active regulators can alter the *effective* η vs. I_{out} characteristics. *Regulator gating* refers to such demand-driven turning-on/off of individual component regulators with the goal of sustained operation at η_{peak}. Regulator-gating applies to distributed multi-phase SC [39], distributed multi-phase buck [4], or distributed LDO regulators [38]. Recent representative examples include modulation of bypass modes in IBM POWER8 [8, 38]; and active phases, in Intel Haswell [4, 10, 21] (Section 3.1).

As an example, Fig. 2 illustrates how gating component regulators (i.e., phases in this case) on demand can sustain operation at the peak efficiency over a wide I_{out} range for the 16-phase Intel buck regulator [21]. Each curve corresponds to a different regulator configuration, as dictated by the number of active phases. Therefore, each curve reaches the peak efficiency at a different I_{out}. Consequently, adaptive gating of active phases can shift the curve to match the instantaneous I_{out} demand (as governed by instantaneous microarchitectural activity) at the peak efficiency, for the entire duration of execution. The *effective* curve, as a result, takes the form of the black dotted trend-line in Fig. 2: a practically constant conversion efficiency over a wide I_{out} window, closely tracking the peak η_{peak}. When the I_{out} demand increases under high microarchitectural activity, additional phases become active to deliver a higher total output current. When the I_{out} demand decreases under low microarchitectural activity, gating applies to a subset of the phases.

4 THERMOGATER: THERMALLY-AWARE REGULATOR GATING

As shown in Section 2, due to inevitable power conversion losses and the small physical footprint, integrated voltage regulators can easily challenge cooling limits. Regulators may become hotspots themselves or cause significant temperature increase in their close proximity, possibly giving rise to new hotspots or raising the thermal gradient. Both the regulator area and physical placement [19] determine the severity of the regulator incurred thermal problems.

In its thermally-oblivious form as covered in Section 3.2, regulator gating can effectively sustain operation at the peak efficiency (Fig. 2). However, even the peak efficiency η_{peak} remains around 90% for recent, highly optimized designs, hence incurs a notable power conversion loss. Therefore, sustained operation at the peak efficiency cannot eliminate regulator induced thermal emergencies.

In a distributed power delivery network, regulator gating can be leveraged to mitigate regulator induced thermal problems. Selectively turning off *hot* regulators can keep the local temperature under control. Gating a regulator does not imply (power-)gating its respective load circuit. Cooler regulators (within the same V_{dd}-domain) in close proximity take over the load of the gated regulator. This does not necessarily translate into operation at a lower power conversion efficiency than the peak, i.e., into more conversion efficiency loss. This is because, *be it thermally-aware or not, the goal of regulator gating is to sustain a practically constant peak efficiency over a wide I_{out} range*.

By *spatio-temporally* modulating the location of the active regulators, regulator-gating can help reduce the number and temperature of thermal hotspots, and smoothen the thermal gradient. Spatially changing (the location of) active regulators spreads the heat concentrated around a hot regulator to neighboring circuit blocks. Temporally changing (the set of) active regulators can reduce the average power dissipated by each regulator (i.e., P_{loss}) over time, and therefore, cap the maximum local temperature.

Be it thermally-aware or not, the goal of regulator gating is to sustain operation at η_{peak} over a wide I_{out} range, which imposes a stringent constraint on the number of active regulators at any point in time during execution. This is because, under regulator gating, only a specific number of active regulators (within a V_{dd}-domain), n_{on}, can supply the demanded I_{out} while operating at η_{peak}. Therefore, a viable regulator gating policy first needs to determine the number of active regulators (within a V_{dd}-domain), n_{on}, required to sustain operation at η_{peak}. Thermal awareness comes only at the next step, in identifying a subset of the regulators of size n_{on} to turn on.

Under thermally-aware regulator gating, logic or memory blocks may not always be supplied by the regulator in closest proximity. The potential adverse affect is higher voltage noise – particularly IR drop, as the effective impedance observed by the load circuit block increases with distance to the supplying regulator. Accordingly, thermally-aware regulator gating is subject to the potential onset of voltage emergencies.

Putting it all together: Three critical factors drive thermally-aware regulator gating decisions: the first factor **(I)** sets the number of active regulators, n_{on}, while the second **(II)** and third **(III)**, determine which n_{on} from the entire set of component regulators within a V_{dd}-domain to activate:

(I) **The instantaneous I_{out} demand, as determined by microarchitectural activity**: Thermally-aware regulator gating is only legal if n_{on} active regulators can collectively supply the required I_{out} at η_{peak}. As the maximum current a component regulator can supply is limited, the instantaneous I_{out} demand can easily restrict thermally-aware regulator gating.

(II) **Thermal emergencies**: The n_{on} regulators selected to be turned on should not trigger thermal emergencies.

(III) **Voltage emergencies**: The n_{on} regulators selected to be turned on should not trigger voltage emergencies.

Fig. 3 provides the macroscopic (a) and microscopic (b) view of ThermoGater, an architectural framework which orchestrates thermally-aware regulator gating subject to the constraints imposed by **(I)**, **(II)**, and **(III)**. As depicted in Fig. 3(a), in achieving this, ThermoGater has to monitor the instantaneous power demand along with the thermal and voltage profiles per V_{dd}-domain across the chip. To this end, ThermoGater can deploy thermal or voltage sensors [6, 18, 32] and/or emergency predictors [30] to proactively alter thermally-aware regulator gating decisions. The resulting ThermoGater control loop is depicted in Fig. 3(b). When the I_{out} demand increases (decreases) under high (low) microarchitectural activity, ThermoGater turns on (off) the required number of component regulators to sustain operation at η_{peak}, in a thermally- and voltage-noise-aware manner. We will next introduce and evaluate a collection of practical policies to implement ThermoGater's control loop from Fig. 3(b).

5 EVALUATION SETUP

In the rest of the paper, we will refer to *phases* (in Intel terminology) or *microregulators* (in IBM terminology) – as covered in Section 3.1 – as *(component) voltage regulators (VR)*.

Benchmarks: We experiment with all benchmarks from SPLASH2x [42] to cover a representative range of application domains and characteristics. We restrict our analysis to the region-of-interest (ROI) of the benchmarks where the actual computation takes place. Our simulations involve 8 threads.

Architectural, thermal, power simulation: To quantitatively characterize thermally-aware regulator gating, without loss of generality, we model an 8-core processor similar to IBM POWER8 [8]. Table 1 captures the technology and architecture parameters. Fig. 4a depicts the floorplan of a core which comprises an IFU (instruction fetch unit), an ISU (instruction scheduling unit), an EXU (execution unit), an LSU (load store unit) and a private L2. L1 data cache (not shown in the figure) resides inside LSU; L1 instruction cache, inside IFU. Fig. 4b demonstrates the floorplan for the entire chip of 8 cores, including the memory controller (MC), network-on-chip (NOC), and 96 integrated voltage regulators, shown as squares.

We experiment with 16 V_{dd}-domains: a separate V_{dd}-domain for each core (+ private L2), and for each L3 bank, mimicking the IBM POWER8 design [8]. Each per core V_{dd}-domain incorporates 9; each per L3 bank V_{dd}-domain, 3 (component) VRs. Over 8 cores and 8 L3 banks (Fig. 4b), this totals up to 96 on-chip VRs distributed among 16 V_{dd}-domains[2].

We integrated MR2 [43] version of McPAT [24] into SNIPER6.0 [5] microarchitectural simulator to collect power traces. The model is calibrated such that the share of static power does not exceed 30% of the total consumption (of the entire chip) at 80°C. We use Hotspot6.0 [35] to model temperature. Temperature (the output of Hotspot) is used to calculate the static power (an input to Hotspot), therefore Hotspot is invoked each time in a closed feedback loop until convergence. We adapt HotSpot6.0's default cooling package (which mimics POWER7+). We expect our observations to generally hold under better cooling, because: (i) cooling solutions usually uniformly affect the chip; (ii) on-chip regulators have much smaller footprint than logic or memory blocks; (iii) regulator power efficiency loss is inevitable.

Technology Parameters
Technology node: 22nm, Frequency: 4.0GHz TDP: 150W, Area: 441mm^2, Vdd: 1.03V
Architecture Parameters
cores: 8, issue width: 8 # architectural floating point registers: 64 # architectural integer registers: 32
L1-I cache: 32KB, 8-way, 64B, LRU, 1-cycle hit L1-D cache: 64KB, 8-way, 64B, LRU, 1-cycle hit L2 cache: 512KB, 8-way, 128B, LRU, 11-cycle hit L3 cache: 64MB, 8-way, 128B, LRU, 30-cycle hit

Table 1: Technology and architecture parameters.

Voltage noise simulation: We deploy an extended version of VoltSpot [44] to capture the impact of thermally aware regulator gating on voltage noise. The extended version, validated against SPICE, models all critical (component) VR parameters. VoltSpot requires cycle-accurate power traces. Since generating them for the entire duration of execution is too expensive, we rely on sampling, following the suggested methodology for VoltSpot [45]: 200 equally distant samples are captured from the entire execution of each application. Each sample contains 2K cycles. The first 1K is used to warm up VoltSpot; the rest, for actual analysis.

Voltage regulator properties: In the simulations, we use 96 voltage regulators (VRs) distributed across the chip over 16 V_{dd}-domains, as captured by the little squares in Fig 4b. Area of each VR is 0.04mm^2. Without loss of generality, we calibrate these VRs to match the conversion efficiency η vs. I_{out} characteristics of Intel's Haswell design [10]. Recall that this design keeps the inductors off-chip (on the package), however, the actual voltage regulation is performed on-chip. We picked these curves just for calibration purposes, as this design represents one of the most efficient regulators from industry. Fig. 5 captures the η vs. I_{out} characteristics for the (component) VRs in each per-core V_{dd}-domain. In this case, each component VR (i.e.,

[2] The high computational complexity of our thermal and voltage noise simulators prevented experimentation with larger number of component on-chip regulators. A lower regulator count worsens both the thermal and the voltage noise profile, therefore, we selected the maximum possible (component) regulator count permissible by our simulation infrastructure. IBM design features a similar number of V_{dd}-domains, however, many more (component) regulators per domain.

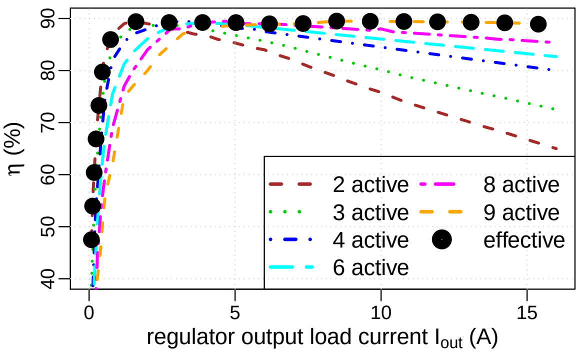


Figure 5: η vs. I_{out} used for calibration.

phase) provides around I_{out}=1.5A load current at the highest conversion efficiency η_{peak}=90%. In each per-core V_{dd}-domain, while all 9 (component) VRs should be active at the highest performance point, lower number of active VRs can still provide operation at η_{peak} at lower processor utilization (Section 3.2).

Voltage regulator placement: On-chip regulator placement can affect the voltage noise profile significantly. In order to eliminate any adverse bias from our analysis, we obtain the (voltage-noise) optimal placement following the methodology from [41], which finds the optimal C4[3] pad locations in order to minimize the maximum (both transient and steady-state) voltage noise. We mimic the devised algorithm (i.e., *Deep Optimization*) to find the optimal location of on-chip component VRs since similar to C4 pads, they act as inputs to the power delivery network[4]. Starting with the VRs in immediate vicinity of where the voltage noise peaks, we attempt to move VRs away one by one in each iteration. We accept a change of position only if it decreases the maximum voltage noise. We continue until the placement converges. The resulting optimized placement slightly deviates from the uniformly distributed placement from Fig. 4b (which results in an increase in the maximum voltage noise by less than 0.4%). We find that the voltage noise profiles of the two placements are very similar otherwise. The uniform placement is more convenient to model due to its regularity. Therefore, in the following, we will stick to the more regular uniform placement.

6 EVALUATION

Any viable thermally-aware regulator gating policy needs to carefully determine two critical parameters: *number* of active regulators and their respective *location*, on a per V_{dd}-domain basis. Section 6.1 focuses on the first; Section 6.2, on the second, parameter. In Section 6.3 we devise practical ThermoGater policies which can effectively orchestrate thermally-aware spatio-temporal regulator gating. Section 6.4 concludes the evaluation with a design space exploration.

6.1 Setting the Number of Active Regulators

Regulator power conversion efficiency η is a strong function of the load current I_{out} (Section 3.1). To be able to sustain operation at the peak efficiency η_{peak} throughout the execution, only as many VRs should be active as necessary to supply the instantaneous current demand I_{out} at η_{peak} as demonstrated in Fig. 5. If more or less VRs remain active than needed to operate at η_{peak}, the degradation in power conversion efficiency causes higher power conversion loss, P_{loss} (Eqn. 1), to be dissipated as heat, which can easily exacerbate the thermal profile and cause thermal emergencies.

As a representative example, Fig. 6 shows how the cumulative number of active regulators over all V_{dd}-domains to enforce operation at η_{peak} (i.e., cumulative n_{on}) changes over the execution time for an 8-threaded run of *lu_ncb*. Time is shown on the x-axis while the left y-axis represents the total power demand and the right one shows the active regulator count. As Fig. 5 reveals, operation at η_{peak} under higher (lower) I_{out} demand requires more (less) active regulators. Accordingly, we observe in Fig. 6 that regulator activity closely tracks temporal changes in total power demand, which represents $V_{dd} \times I_{out}$ with V_{dd} being constant. Recall that thermally-aware regulator gating is only viable if the set of active regulators can collectively supply the required I_{out} at η_{peak} (Section 4).

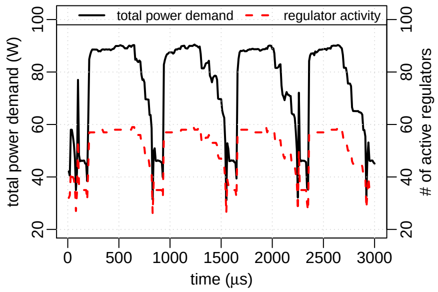


Figure 6: Evolution of #active regulators with time.

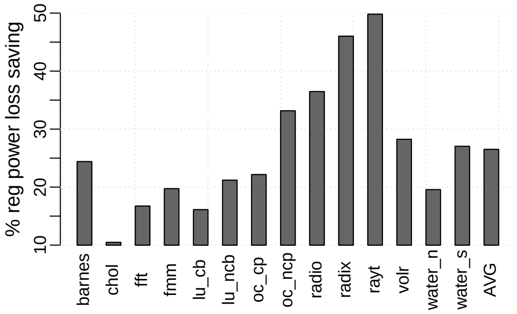


Figure 7: P_{loss} improvement under optimal gating.

We next analyze the impact of active regulator count on the power conversion loss, P_{loss}. We compare two cases: keeping all 96 regulators active (*all-on*) vs. keeping only as many regulators active as necessary to sustain η_{peak} (i.e., cumulative n_{on} over all V_{dd}-domains, as it was the case for Fig. 6, by regulator gating). We find the average (over time and space) P_{loss} of the regulators for each case, and report the difference in Fig. 7. The y-axis captures % P_{loss} saving in comparison to *all-on* under regulator gating. We observe a wide range of savings from 10.4% for *cholesky* up to 49.8% for *raytrace*. % saving strongly depends on the total power consumption of the application:

[3] Controlled Collapse Chip Connection (C4) pads connect the off-chip voltage converter to the global power grid.
[4] C4 pads feed the global; on-chip VRs, the local power grids, respectively.

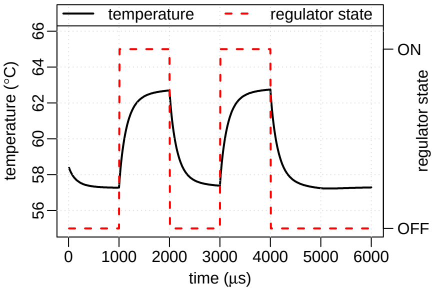


Figure 8: A representative thermal profile under Naïve.

If power consumption stays high (*cholesky*) throughout execution, many more active regulators are required to operate at η_{peak} (as Figs. 5 and 6 reveal), hence the difference to *all-on* becomes much less pronounced where we keep all regulators active all the time. As a result, regulator gating (if at all) can only save a very little fraction of P_{loss}. On the other hand, for low power applications (*raytrace*), savings become significant as only a notably lower number of active regulators is necessary to operate at η_{peak}. Overall, we observe that for most of the applications, regulator gating can significantly reduce P_{loss}, by ≈26.5% on average.

6.2 Setting the Location of Active Regulators

Be it thermally-aware or not, the goal of regulator gating is to sustain operation at η_{peak} over a wide I_{out} range. Once we determine the number of active regulators, n_{on} (on a per V_{dd}-domain basis), required to sustain operation at η_{peak}, the question becomes *which subset of the regulators of size n_{on} to select to turn on*. Thermally-oblivious regulator gating policies (Section 3.2) typically focus on n_{on} calculation only and do not consider potential thermal implications at all, as opposed to ThermoGater.

In selecting n_{on} regulators (within a V_{dd}-domain) to turn on, omitting the hottest regulators can decrease the local temperature, which in turn can help reduce the number or temperature of thermal hotspots, and keep the thermal gradient under control. Such a selection, however, may cause on-chip logic and memory blocks not to be supplied by the regulators in closest physical proximity, and hence, may accelerate the onset of voltage emergencies. In the following, we analyze these adverse effects by starting with a thermally-aware greedy gating policy (Section 6.2.1) and continuing with predictive voltage-noise oblivious thermally-aware (Section 6.2.2) and thermally-oblivious voltage-noise-aware (Section 6.2.3) gating policies, before concluding with an oracular policy which considers the thermal and voltage noise implications together (Section 6.2.4). These policies only differ in the selection of n_{on} regulators on a per V_{dd}-domain basis.

6.2.1 Thermally-Aware *Naïve* Gating

Spatio-temporal thermal information can reveal which regulators are more likely to become potential hotspots. Based on this information, we designate a greedy gating policy, *Naïve*, which picks the n_{on} coolest of the regulators to turn on (in each V_{dd}-domain). This gives the hottest regulators time (until the next gating decision takes place) to cool down. Practically, *Naïve* keeps the maximum possible number of hottest regulators off at each decision point. All of the thermally-aware gating policies in this study, including *Naïve* draw gating decisions every 1ms[5].

Fig. 8 depicts how the temperature T around a representative regulator changes under *Naïve* during the execution of *lu_ncb*. The x-axis captures the time; the y-axis on left (right), T in °C (regulator state). A similar trend applies to other applications or regulators across different V_{dd}-domains. At the first decision point at 1ms, this specific regulator is not one of the hottest regulators in its respective V_{dd}-domain, so *Naïve* turns it on. However, at the second decision point at 2ms, the regulator becomes hot enough (with respect to the rest of the regulators) that *Naïve* turns it off and lets it cool down at least until the next decision point. Regulator T changes by more than 5°C during this process.

Considering implications for reliability, keeping the thermal gradient (the maximum spatial difference in T) under control can become more critical than capping the maximum temperature across chip, T_{max} [27]. Throughout this study, we will report the impact of different gating policies on both, the thermal gradient and T_{max}. Specifically, we will report the *temporal maximum* of both, the thermal gradient and T_{max}.

Fig. 10 compares the maximum thermal gradient under different gating policies with two baselines: *all-on* where all 96 on-chip regulators are active all the time, and *off-chip* which excludes on-chip regulation. When compared to *off-chip*, *all-on* increases the thermal gradient significantly, by 79.4% on average. *Naïve* enforces η_{peak} throughout the execution, however, exacerbates the thermal gradient further, by 12.5% on average, over *all-on*. Similarly, as Fig. 9 reveals, *all-on* increases T_{max} by 5.4°C over *off-chip*; and *Naïve* by 1.1°C over *all-on*. Clearly, *Naïve* does not represent a feasible thermally-aware gating policy.

6.2.2 Thermally-Aware Oracular Gating

At each temporal decision point, *Naïve* swaps a hot regulator (from each V_{dd}-domain) which was on, say *hot*, with a cooler one which was off, say *cool*. At the decision point t, the temperature of *hot*, $T_{t,hot}$ is higher than $T_{t,cool}$. However, depending on the instantaneous power demand, once *cool* is turned on, T_{cool} can easily exceed the temperature *hot* would assume if *hot* was kept on. Therefore, keeping *hot* on until the next decision point can turn out to be a better option, although there are cooler regulators such as *cool* to turn on. *Naïve*'s poor performance hence stems from not considering the impact of gating to the future thermal profile. This motivates a policy of predictive nature. We start with an oracular policy, $Orac_T$, which turns off the *hottest-to-be* (as opposed to the instantaneous *hottest* under *Naïve*) regulators at each decision point in selecting n_{on} regulators to turn on within each V_{dd}-domain. We assume that $Orac_T$ has full-fledged oracular knowledge about the output power demand and temperature of each regulator at each decision point in time under all possible gating decisions. Later in Section 6.3 we will factor in practical limitations.

Fig. 10 reveals that $Orac_T$ can improve the maximum thermal gradient by 10.9% on average over *all-on*. At the same time, recall that $Orac_T$ enforces operation at η_{peak} throughout execution (as

[5] According to our experiments, choosing a 100× shorter period improves the accuracy only by less than 1%.

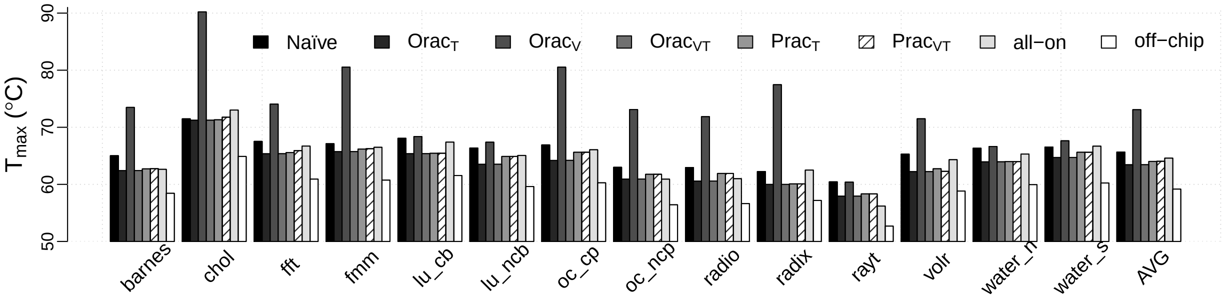


Figure 9: Maximum chip-wide temperature under different regulator gating policies.

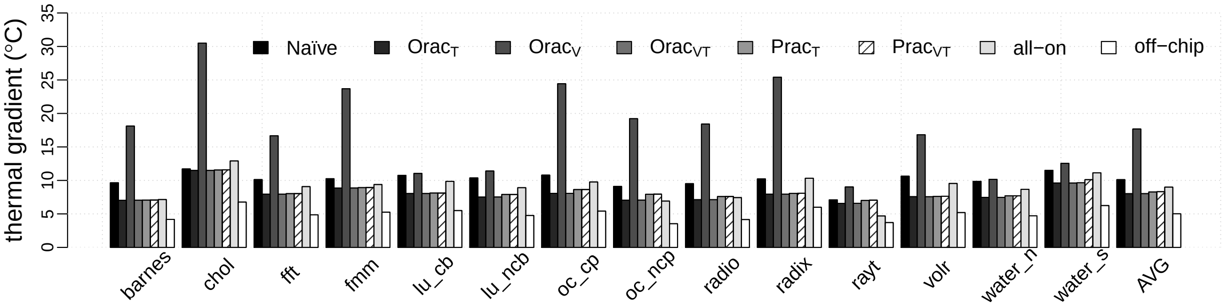


Figure 10: Maximum thermal gradient under different regulator gating policies.

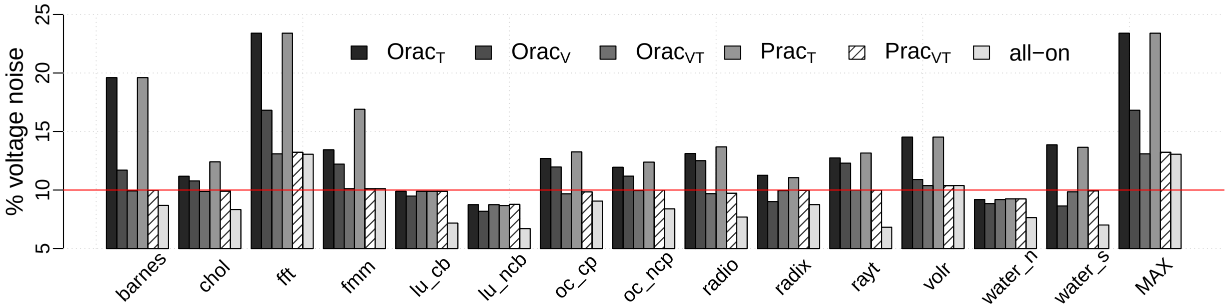
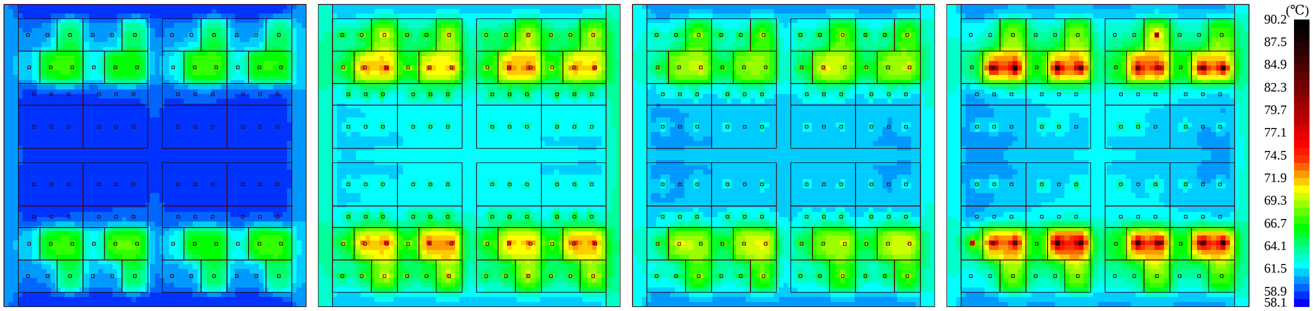


Figure 11: Maximum voltage noise under different regulator gating policies.

the rest of the gating policies we study in this paper), while power conversion loss is inevitable under *all-on* (Fig. 5). The decrease in maximum thermal gradient over *Naïve* is over 20.7%. Similarly, as Fig. 9 reveals, $Orac_T$ decreases T_{max} by 1.2°C over *all-on*; and by 2.2°C over *Naïve*.

We next analyze the spatial impact of regulator gating using heat maps. Fig. 12 shows a representative thermal frame of *cholesky* without loss of generality, where T_{max} peaks during execution (under different gating policies we report, the frames provided differ by less than 100μs). If we do not use on-chip regulators at all (*off-chip*), T_{max} does not exceed 66°C (Fig. 12a). Keeping all on-chip regulators on all the time (*all-on*) triggers hotspots on some of the LSUs and EXUs, increasing T_{max} to 73°C (Fig. 12b). As Fig. 12c reveals, by predictive thermally-aware gating, $Orac_T$ can eliminate these hotspots and decrease the instantaneous T_{max} to nearly 71.2°C, while operating at η_{peak} as opposed to *all-on*.

On-chip memory blocks are usually cooler and less power hungry than logic units. In picking n_{on} regulators to sustain operation at η_{peak} (on a per V_{dd}-domain basis), $Orac_T$ effectively moves active regulators farther from logic units and closer to the memory blocks. This reduces the maximum thermal gradient and T_{max}, however, keeping supplying regulators further away from logic units worsens the voltage noise profile. Fig. 11 captures the impact of different gating policies on voltage noise. $Orac_T$ exacerbates the voltage noise significantly for all applications: On average, the maximum voltage noise becomes 23.4% of the nominal V_{dd}, which is 79.3% larger than *all-on*. Safe operation under $Orac_T$ therefore would demand an excessive V_{dd} guard-band, which implies operation at a much higher nominal V_{dd}. This in turn can increase the power consumption noticeably and is more than likely to wipe out the savings in P_{loss} under regulator gating. Accordingly, a feasible thermally-aware gating policy cannot be voltage-noise-oblivious.

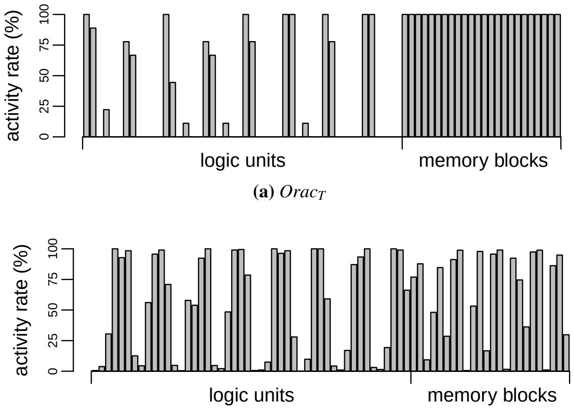


(a) off-chip **(b) all-on** **(c)** $Orac_T$ **(d)** $Orac_V$

Figure 12: Representative heat maps under different regulator gating policies.

6.2.3 Voltage-Noise-Aware Oracular Gating

We next analyze the voltage-noise-aware dual policy of $Orac_T$: $Orac_V$. In picking n_{on} regulators to sustain operation at η_{peak} (on a per V_{dd}-domain basis), $Orac_V$ chooses regulators based on the spatial voltage noise profile only, in a thermally-oblivious way. Therefore, $Orac_V$ tends to keep the regulators physically closest to high voltage noise regions on. Similar to $Orac_T$, we assume that $Orac_V$ has full-fledged oracular knowledge about the output power demand per regulator and domain-wide voltage noise profile across chip at each decision point in time under all possible gating decisions. Later in Section 6.3 we will factor in practical limitations.



(a) $Orac_T$

(b) $Orac_V$

Figure 13: Regulator activity under $Orac_T$ **vs.** $Orac_V$**.**

Typically, voltage noise is worse near logic units since they are more power hungry than on-chip memory blocks. Fig. 13 compares spatial regulator activity under $Orac_T$ (a) and $Orac_V$ (b) for *lu_ncb* as an example, without loss of generality. Y-axis represents % of execution time during which a regulator stays on. On the x-axis, we have 72 bars, each representing a regulator from a (per-core) V_{dd}-domain (over all domains), binned into two groups according to the location within the V_{dd}-domain: regulators supplying logic units on the left; on-chip memory blocks, on the right. Recall that each per core V_{dd}-domain incorporates a core along with its private L1 and private L2 (Section 5). We observe that while $Orac_T$ tends to turn more regulators off in the immediate proximity of logic units, $Orac_V$ does the opposite to keep the voltage noise under control.

Fig. 14 shows a critical spatio-temporal voltage noise sample from *fft* which causes the worst voltage noise profile under $Orac_T$ across all the applications as Fig. 11 reveals. Gating according to spatial voltage noise information helps a lot in this case: $Orac_V$ decreases the maximum voltage noise significantly by 28.2% over $Orac_T$.

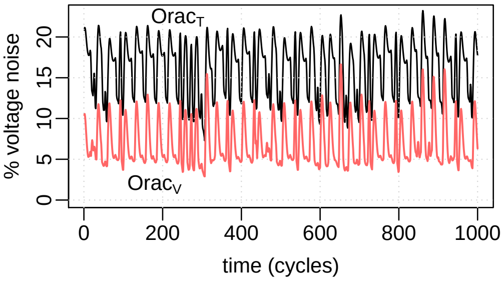


Figure 14: Impact on voltage noise: $Orac_T$ **vs.** $Orac_V$**.**

Fig. 11 reports the maximum voltage noise for all applications under different gating policies. We report the maximum across all V_{dd}-domains. We observe that under $Orac_V$ maximum voltage noise remains within 28.4% of the maximum voltage noise under *all-on* (which represents the best case for voltage noise, as each logic or memory block is guaranteed to be supplied by the regulator in closest physical proximity). However, $Orac_V$ can sustain operation at η_{peak}, while the effective conversion efficiency η under *all-on* ($\leq \eta_{peak}$) fluctuates throughout the execution. Voltage noise profile under $Orac_V$ is worse than under *all-on* for most of the applications. For these cases, turning on more regulators than n_{on} can help at the expense of operating at a degraded η much below η_{peak}, but maximizing conversion efficiency criterion prevents $Orac_V$ from turning on more than n_{on} regulators (as necessary to operate at η_{peak}) on a per V_{dd}-domain basis.

By turning on the regulators mostly near logic units (Fig. 13b), $Orac_V$ can severely worsen the thermal profile as the heat map from Fig. 12d reveals: T_{max} increases to more than 90° in this case. The poor thermal profile under $Orac_V$ is summarized for all applications in Fig.s 10 and 9. $Orac_V$ increases the maximum thermal gradient by nearly 96.3% (120.9%); T_{max}, by 8.5°C (9.6°C), in comparison to *all-on* ($Orac_T$), respectively.

6.2.4 Putting It All Together: $Orac_{VT}$

We next devise a thermally-aware gating policy which also takes the implications for voltage-noise into account: $Orac_{VT}$. Similar to $Orac_T$ and $Orac_V$, $Orac_{VT}$ is of oracular nature (in Section 6.3 we will factor in practical limitations). $Orac_{VT}$ captures the impact on voltage noise by tracking the frequency (of occurrence) of voltage emergencies. We define a *voltage emergency* as the maximum voltage noise exceeding a threshold of 10% (of the nominal V_{dd}), on a per V_{dd}-domain basis. The horizontal line in Fig. 11 marks this threshold.

App. names	barnes	chol	fft	fmm	oc_cp	oc_ncp	radio	radix	rayt	volr	water_s	**AVG**
% exec. time	0.67	0.001	0.49	0.024	0.50	0.002	0.008	0.06	0.032	0.002	0.11	**0.13**

Table 2: % execution time spent in voltage emergencies under $Orac_T$ (reported are non-zero values only).

Table 2 summarizes % of cycles spent in voltage emergencies throughout the execution across all benchmarks under $Orac_T$. We observe that for all of the applications % cycles spend in voltage emergencies stays under 1%. Moreover, changes in temperature take much longer than the duration of a voltage emergency (which is in the order of cycles). In other words, the time constant for temperature changes is higher, therefore, a change in gating schedule of very short duration is usually not long enough to affect the temperature.

Based on these observations, $Orac_{VT}$ mimics $Orac_T$ by default and switches to *all-on*, on a per V_{dd}-domain basis, only upon the prediction of a voltage emergency (the prediction accuracy is perfect under $Orac_{VT}$ due to the oracular nature). Therefore, for applications which do not experience any voltage emergencies per our definition (such as *lu_ncb*), $Orac_{VT}$ becomes equivalent to $Orac_T$. Fig.s 10, 9, and 11 reveal the maximum thermal gradient, T_{max}, and maximum voltage noise under $Orac_{VT}$. We observe that under $Orac_{VT}$ the thermal profile effectively converges to the thermal profile under $Orac_T$; and the voltage noise profile, to *all-on*. Higher temperature under $Orac_V$ can increase the static power significantly, and thereby, the overall power consumption. This is how $Orac_V$ can render higher voltage noise than the less power-hungry $Orac_{VT}$ for some benchmarks.

6.3 Practical ThermoGater Policies

We next devise practical ThermoGater policies to orchestrate thermally aware regulator gating following the closed control loop from Fig. 3. We closely mimic oracular policies from Section 6.2 and factor in the practical limitations.

We start with $Prac_T$, which corresponds to $Orac_T$ and gates regulators based on spatial thermal information only. Mimicking $Orac_T$, in selecting n_{on} regulators to ensure operation at η_{peak} (on a per V_{dd}-domain basis), $Prac_T$ ranks regulators by their anticipated temperature, turns on n_{on} of the regulators with coolest anticipated temperature, and turns the rest off. To this end, $Prac_T$ should be able to predict the anticipated temperature of each regulator as a function of changes in the power demand throughout execution.

In predicting the anticipated temperature of each regulator, $Prac_T$ assumes a linear relationship between changes in power dissipation (ΔP) and changes in temperature (ΔT) between two decision points, on a per regulator basis:

$$\Delta T_i = \theta_i \Delta P_i \quad i = 1,2,...,96 \tag{2}$$

where θ_i represents a constant of proportionality (or model fitting parameter) that we extract for each regulator i using (temporal) power and thermal traces from a profiling pass. As pointed out in Skadron et. al. [34] such linear models often fail short of accurately capturing the on-chip thermal profile. However, using HotSpot, we validated that *confined* deployment of this model only to predict the anticipated temperature of on-chip regulators (considering their small footprint) provides highly accurate results. ΔP_i captures the anticipated change (between two decision points) in the power dissipation (i.e., in $P_{loss,i}$) of regulator i, which is dictated by the anticipated change in the power demand of load circuit blocks (i.e., $P_{out,i}$) per Eqn. 1. Accordingly, $Prac_T$ tracks $\Delta P_{out,i}$ in determining ΔP_i.

To find the accuracy of prediction for Eqn. 2, we use *coefficient of determination*, R^2, a common statistical metric to quantify predictability [25]:

$$R^2 = 1 - \frac{\sum_i (T_{i,HotSpot} - T_{i,Prediction})^2}{\sum_i (T_{i,HotSpot} - T_{avg,HotSpot})^2} \tag{3}$$

applies where $T_{i,HotSpot}$ and $T_{i,Prediction}$ denote the T of regulator i obtained from Hotspot simulation and predicted using Eqn. 2, respectively. $T_{avg,HotSpot}$ captures the average regulator T obtained from HotSpot. If the prediction error is absolutely zero, the numerator of Eqn. 3 goes to zero, leading to $R^2 = 1$. So, a more accurate prediction implies a closer-to-one R^2. We calibrate θ_i values to keep R^2 around 0.99.

Thermal sensors are widely used in modern commercial processors. For instance, IBM POWER7 employs 44 digital thermal sensors to detect chip wide hotspots [18]. Thermal sensors capable of providing up to 10K thermal readings per second exist [6]. For this type of sensor, in the worst case, thermal readings $Prac_T$ uses at each decision point would be 100μs old. We place such a thermal sensor at immediate vicinity of each regulator, and assume that the overhead of gathering and sorting sensor readings is comparable to the sensor delay (100μs in this case) by relying on microcontroller firmware such as POWER8 On-Chip Controller (OCC) [8].

To be able to predict the anticipated temperature by using Eqn. 2, $Prac_T$ needs a prediction for the anticipated power demand of the load circuit blocks (i.e., P_{out}), as well. To this end, we use the Weighted Moving Average (WMA) based model from [3], which can derive the anticipated power consumption from the power consumption history spanning the last three decision points.

Putting it all together: $Prac_T$ deploys Eqn. 2 at each decision point as follows: θ_i values are extracted from a profiling pass and do not change if the floorplan is fixed. $Prac_T$ first collects the instantaneous readings from the thermal sensors. Let the reading be $T_{t,i}$ for regulator i at decision point t. $Prac_T$ also estimates the anticipated power demand from the power demand history of the last three decision points, and calculates ΔP_i from the difference between the anticipated demand and the demand at the previous decision point. The next step is deploying Eqn. 2 to derive the anticipated temperature for each regulator i from $T_{t,i} + \theta_i \Delta P_i$. Recall that the anticipated temperature corresponds to the temperature the regulator would assume if it was turned on (until the next decision point). Finally, $Prac_T$ sorts the anticipated temperatures and picks the n_{on} of the regulators with the lowest anticipated temperatures to turn on, on a per V_{dd}-domain basis.

Factoring in all sensor and control related overheads, Fig.s 10 and 9 report the maximum thermal gradient and T_{max} under $Prac_T$. We observe that the thermal profile under $Prac_T$ slightly degrades in comparison to $Orac_T$ mainly due to the thermal sensing delay and prediction error from Eqn. 2: on average, the maximum thermal gradient increases by ≈3%; T_{max}, by 0.5°C over $Orac_T$. As it is the case for $Orac_T$, gating only based on thermal information renders a poor voltage noise profile under $Prac_T$, as Fig. 11 reveals, increasing the overall maximum by 79.9% in comparison to *all-on*.

We add voltage-noise awareness to $Prac_T$ by mimicking $Orac_{VT}$, and call the resulting policy $Prac_{VT}$, which needs to predict the onset of voltage emergencies. As demonstrated in [30], voltage emergencies are predictable with more than 90% accuracy. $Prac_{VT}$ deploys a voltage emergency detector per core similar to [30] instead of alternative spatial voltage emergency sensors [32]. Further, $Prac_{VT}$ turns all regulators of the affected domain on upon a voltage emergency alert. This policy relaxes the power conversion efficiency constraint by possibly turning on more than n_{on} regulators, where n_{on} regulators would be necessary to operate at η_{peak} (on a per V_{dd}-domain basis). However, since emergency events are rare (Table 2), and $Prac_{VT}$ only turns on all regulators within a few critical domains upon an alert, the power conversion efficiency η degrades negligibly, by less than 0.1% on average, with the maximum degradation reaching 0.5% for *barnes*. When compared to $Prac_T$, $Prac_{VT}$ improves the voltage noise profile significantly as Fig. 11 reveals: overall maximum voltage noise stays at 13.22% of the nominal V_{dd} under $Prac_{VT}$; at 13.05%, under *all-on*.

In conclusion, $Prac_{VT}$ can effectively sustain operation within 0.5% of the peak efficiency η_{peak}, while degrading the maximum thermal gradient by around 3%; T_{max}, by 0.5°C over $Orac_T$, and the maximum voltage noise by less than 1.3% over *all-on*, subject to the accuracy of sensors and the predictive model from Eqn. 2, including calibration. Parametric variation due to manufacturing imperfections may render per-chip calibration necessary and can increase manufacturing testing overhead, but $Prac_{VT}$ is ranking-based and can tolerate calibration errors as long as inaccuracies keep relative ranking intact (where absolute parameter values may fluctuate significantly).

6.4 Design Space Exploration

The evaluation so far assumed an Intel FIVR [4, 10, 21] like regulator design in calibrating the regulator power conversion efficiency curves according to Fig 5. In this case, each distributed component regulator (as depicted by small squares, 96 in total, in Fig. 4b) corresponds to a phase. Our observations and ThermoGater policies, however, are equally applicable to different types of regulators, as well.

We repeat our analysis for an LDO based design similar to IBM's POWER8 [8, 38]. In this case, each distributed component regulator represents a digital LDO (micro)regulator. The reported η_{peak} and P_{out} per area assume very similar values in comparison to FIVR[6]. For an apples-to-apples comparison, we calibrate this LDO based design to follow the efficiency curves from Fig. 5 ([8, 38] do not report the efficiency curves).

Due to very similar P_{out} per area values (and because we calibrate both designs to render the very same power conversion efficiency curves under gating), the LDO based design's thermal profile closely tracks the FIVR based design. The main difference comes from the faster response time of LDO regulators, which is anticipated to lead to lower voltage noise.

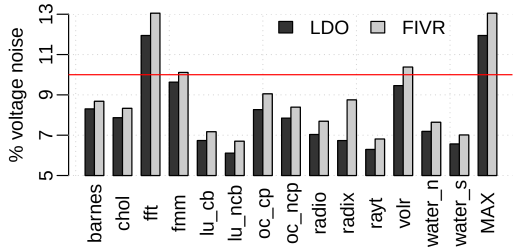


Figure 15: Maximum voltage noise: LDO vs. FIVR.

Fig. 15 compares the maximum voltage noise for both of these designs across all benchmarks, if all component regulators stay on all the time (*all-on*). The LDO based design decreases the voltage noise by around 0.7% on average, while the overall maximum (for *ftt*) decreases by around 1.1% over the FIVR based design. This small improvement in voltage noise did not render any notable deviation from our results or observations in Section 6.

For FIVR, the power loss on the on-package inductor does not directly contribute to on-chip heating. Recall that we rely on FIVR only in calibrating the regulator power conversion efficiency curves. The only reason why we used FIVR for calibration was the public disclosure of the corresponding efficiency curves, as best-known representatives from industry. Otherwise, ThermoGater decisions are governed by the *effective* curve from Fig. 5 which takes a very similar form for both (FIVR like) buck- and LDO-based designs. Note that a given (η, I_{out}) point on this curve can result in a different number of active regulators for buck- vs. LDO-based designs. This is where the inaccuracy in our modeling comes from, which is unlikely to change our fundamental observations.

7 CONCLUSION & DISCUSSION

Even at the peak conversion efficiency, η_{peak}, a significant power conversion loss applies to the best-known and projected integrated regulators. As power conversion loss gets dissipated as heat, integrated regulators can easily cause thermal emergencies due to their small footprint. Deviation from η_{peak} as a result of changes

[6] η_{peak}=90.5%, P_{out} per area = 34.5W/mm^2 for this case; η_{peak}=90%, P_{out} per area = 33.6W/mm^2 for FIVR.

in microarchitectural activity can only make regulator-induced thermal problems worse. While regulator gating in response to spatio-temporal changes in the processor power demand can sustain operation at η_{peak}, thermal implications have been overlooked at the architectural level. This paper introduces *ThermoGater*, an architectural governor for practical, thermally-aware regulator gating to mitigate regulator-induced thermal emergencies, which also considers the impact on voltage noise. Practical ThermoGater policies can not only sustain operation within 1% of η_{peak}, but also keep the maximum temperature (thermal gradient) across chip within 0.6°C (0.3°C) on average in comparison to thermally-optimal oracular regulator gating, while the maximum voltage noise stays within 1.0% of the best case voltage noise profile. The goal for ThermoGater policies is not directly minimizing power consumption, but sustaining operation at the peak power conversion efficiency throughout execution in a thermally- and voltage-noise-aware fashion.

ThermoGater policies are likely to affect aging because utilization per regulator does not necessarily stay uniform throughout the execution, as Fig 13 suggests. Under $Prac_{VT}$ the utilization profile is likely to be similar to Fig. 13a. This is because $Prac_{VT}$'s periodic gating decision interval is based on temperature, while gating based on voltage-noise is only the case when a voltage-emergency happens. Voltage-noise-based gating is not periodic, but rather event-driven. Therefore, highly-utilized regulators are more likely to reside at cooler regions (closer to memory). This may balance out aging, particularly considering wear-out paradigms where aging rate increases exponentially with temperature.

This study characterized the worst-case corner under parametric variation, which affects both static power (a strong function of temperature) and vulnerability to voltage noise. Therefore, a yield impact is inevitable, and we leave the exploration to future work.

ThermoGater controls each voltage-domain independently and accounts for the evolution of the power conversion efficiency with the workload. Therefore, ThermoGater policies can accommodate heterogeneity in the workload, including multi-programming.

The physical placement of component regulators also has a thermal impact. In this study, we started with a (close-to) optimal placement which minimizes voltage noise (Section 5). Thermally-aware placement [19] can exploit potential lateral heat transfer between hotter (e.g., core logic) and cooler regions (e.g., memory blocks). However, placing regulators further away from logic units (and closer to on-chip memory blocks) is very likely to boost voltage noise due to the increased distance between the respective regulators and their load (logic units).

8 ACKNOWLEDGEMENTS

This work was supported in part by NSF (CCF-1421988 and CAREER CCF-1350451).

REFERENCES

[1] Toke Meyer Andersen, Florian Krismer, Johann Walter Kolar, Thomas Toifl, Christian Menolfi, Lukas Kull, Thomas Morf, Marcel Kossel, Matthias Brândii, and Pier Andrea Francese. 2015. A feedforward controlled on-chip switched-capacitor voltage regulator delivering 10W in 32nm SOI CMOS. In *Solid-State Circuits Conference-(ISSCC), 2015 IEEE International*. IEEE, 1–3.

[2] Toke M Andersen, Florian Krismer, Johann W Kolar, Thomas Toifl, Christian Menolfi, Lukas Kull, Thomas Morf, Marcel Kossel, Matthias Brändli, Peter Buchmann, and others. 2013. A 4.6 W/mm 2 power density 86% efficiency on-chip switched capacitor DC-DC converter in 32 nm SOI CMOS. In *Applied Power Electronics Conference and Exposition (APEC), 2013 Twenty-Eighth Annual IEEE*. IEEE, 692–699.

[3] Ehsan K Ardestani, Francisco J Mesa-Martinez, Gabriel Southern, Elnaz Ebrahimi, and Jose Renau. 2013. Sampling in Thermal Simulation of Processors: Measurement, Characterization, and Evaluation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 32, 8 (2013).

[4] Edward A Burton, Gerhard Schrom, Fabrice Paillet, Jonathan Douglas, William J Lambert, Kaladhar Radhakrishnan, and Michael J Hill. 2014. FIVR – Fully Integrated Voltage Regulators on 4th Generation Intel Core SoCs. In *Applied Power Electronics Conference and Exposition (APEC), 2014 Twenty-Ninth Annual IEEE*. IEEE, 432–439.

[5] Trevor E. Carlson, Wim Heirman, and Lieven Eeckhout. 2011. Sniper: Exploring the Level of Abstraction for Scalable and Accurate Parallel Multi-core Simulation. In *International Conference for High Performance Computing, Networking, Storage and Analysis*.

[6] Poki Chen, Chun-Chi Chen, Chin-Chung Tsai, and Wen-Fu Lu. 2005. A Time-to-Digital-Converter-Based CMOS Smart Temperature Sensor. *IEEE Journal of Solid-State Circuits* 40, 8 (August 2005).

[7] Wael El-Essawy. Version 0.4 (2016). IPMItoolRaw Command Interface to Open-POWER POWER8 On Chip Controller: Sensor Reading Commands (https://github.com/open-power/docs/blob/master/occ/OCC_ipmitool_sensors.pdf). (Version 0.4 (2016)).

[8] Eric J Fluhr, Joshua Friedrich, Daniel Dreps, Victor Zyuban, Gregory Still, Christopher Gonzalez, Allen Hall, David Hogenmiller, Frank Malgioglio, Ryan Nett, and others. 2014. POWER8: A 12-Core Server-Class Processor in 22nm SOI with 7.6Tb/s Off-Chip Bandwidth. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[9] Waclaw Godycki, Christopher Torng, Ivan Bukreyev, Alyssa Apsel, and Christopher Batten. 2014. Enabling Realistic Fine-Grain Voltage Scaling with Reconfigurable Power Distribution Networks. In *Proceedings of the Annual IEEE/ACM International Symposium on Microarchitecture*.

[10] Per Hammarlund, Alberto J Martinez, Atiq A Bajwa, David L Hill, Erik Hallnor, Hong Jiang, Martin Dixon, Michael Derr, Mikal Hunsaker, Rajesh Kumar, and others. 2014. Haswell: The Fourth-Generation Intel Core Processor. *IEEE Micro* 34, 2 (March 2014).

[11] Mark Horowitz. 2014. Computing's Energy Problem (and what we can do about it). *Keynote at IEEE International Solid-State Circuits Conference* (2014). http://isscc.org/media/2014/plenary/Mark_Horowitz/NewStandardPlayer.html?plugin=HTML5&mimetype=video%2Fmp4

[12] Wei Huang, Shougata Ghosh, Sivakumar Velusamy, Karthik Sankaranarayanan, Kevin Skadron, and Mircea R Stan. 2006. HotSpot: A Compact Thermal Modeling Methodology for Early-Stage VLSI Design. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 14, 5 (May 2006).

[13] Wei Huang, Mircea R Stan, Sudhanva Gurumurthi, Robert J Ribando, and Kevin Skadron. 2010. Interaction of Scaling Trends in Processor Architecture and Cooling. In *IEEE Semiconductor Thermal Measurement and Management Symposium*.

[14] Junmin Jiang, Yan Lu, Cheng Huang, Wing-Hung Ki, and Philip KT Mok. 2015. A 2-3-Phase Fully Integrated Switched-Capacitor DC-DC Converter in Bulk CMOS for Energy-Efficient Digital Circuits with 14% Efficiency Improvement. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[15] Seong Joong Kim, Romesh Kumar Nandwana, Qadeer Khan, Robert Pilawa-Podgurski, and Pavan Kumar Hanumolu. 2015. A 1.8V 30-to-70MHz 87% Peak-Efficiency 0.32mm^2 4-Phase Time-Based Buck Converter Consuming 3μA/MHz Quiescent Current in 65nm CMOS. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[16] Wonyoung Kim, Meeta S Gupta, Gu-Yeon Wei, and David Brooks. 2008. System Level Analysis of Fast, Per-Core DVFS Using On-Chip Switching Regulators. In *Proceedings of the IEEE International Symposium on High Performance Computer Architecture*.

[17] Joonho Kong, Sung Woo Chung, and Kevin Skadron. 2012. Recent Thermal Management Techniques for Microprocessors. *ACM Computing Survey* 44, 3 (June 2012).

[18] Joonho Kong, Sung Woo Chung, and Kevin Skadron. 2012. Recent Thermal Management Techniques for Microprocessors. *ACM Computing Surveys (CSUR)* 44, 3 (2012).

[19] Selçuk Köse. 2014. Thermal Implications of On Chip Voltage Regulation: Upcoming Challenges and Possible Solutions. In *Proceedings of the IEEE/ACM Design Automation Conference*.

[20] Selçuk Köse and Eby G Friedman. 2012. Distributed On-Chip Power Delivery. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 2, 4 (December 2012).

[21] Nasser Kurd, Muntaquim Chowdhury, Edward Burton, Thomas P Thomas, Christopher Mozak, Brent Boswell, Praveen Mosalikanti, Mark Neidengard, Anant Deval, Ashish Khanna, and others. 2014. Haswell: A Family of IA 22nm Processors. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[22] Woojoo Lee, Yanzhi Wang, and Massoud Pedram. 2014. VRCon: Dynamic Reconfiguration of Voltage Regulators in a Multicore Platform. In *Proceedings of the Conference on Design, Automation and Test in Europe*.

[23] Woojoo Lee, Yanzhi Wang, and Massoud Pedram. 2015. Optimizing a Reconfigurable Power Distribution Network in Multicore Platform. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 34, 7 (July 2015).

[24] Sheng Li, Jung Ho Ahn, Richard D. Strong, Jay B. Brockman, Dean M. Tullsen, and Norman P. Jouppi. 2009. McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures. In *Proceedings of the Annual IEEE/ACM International Symposium on Microarchitecture*.

[25] David J Lilja. 2005. *Measuring Computer Performance: a Practitioner's Guide*. Cambridge University Press.

[26] Yan Lu, Junmin Jiang, Wing-Hung Ki, C Patrick Yue, Sai-Weng Sin, U Seng-Pan, and Rui Paulo Martins. 2015. A 123-phase DC-DC Converter-Ring with Fast-DVS for Microprocessors. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[27] Francisco Javier Mesa-Martinez, Ehsan K. Ardestani, and Jose Renau. 2010. Characterizing Processor Thermal Behavior. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems*.

[28] Hans Meyvaert, Gerard Villar Piqué, Ravi Karadi, Henk Jan Bergveld, and Michiel SJ Steyaert. 2015. A Light-Load-Efficient 111 Switched-Capacitor DC-DC Converter with 94.7% Efficiency While Delivering 100mW at 3.3V. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[29] Sung-Yun Park, Jihyun Cho, Kyuseok Lee, and Euisik Yoon. 2015. PWM Buck Converter with >80% PCE in 45μA-to-4mA Loads Using Analog-Digital Hybrid Control for Implantable Biomedical Systems. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[30] Vijay Janapa Reddi, Meeta Gupta, Glenn Holloway, Michael D Smith, Gu-Yeon Wei, and David Brooks. 2010. Predicting Voltage Droops Using Recurring Program and Microarchitectural Event Activity. *IEEE Micro* 30, 1 (January 2010).

[31] Christopher Schaef, Kapil Kesarwani, and Jason T Stauth. 2015. A Variable-Conversion-Ratio 3-Phase Resonant Switched Capacitor Converter with 85% Efficiency at 0.91Wmm^2 using 1.1nH PCB-Trace Inductors. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[32] Anuja Sehgal, Peilin Song, and Keith A Jenkins. 2006. On-chip Real-Time Power Supply Noise Detector. In *Proceedings of the 32nd European Solid-State Circuits Conference*.

[33] Abhishek A Sinkar, Hao Wang, and Nam Sung Kim. 2012. Workload-Aware Voltage Regulator Optimization for Power Efficient Multi-Core Processors. In *Proceedings of the Conference on Design, Automation and Test in Europe*.

[34] Kevin Skadron, Mircea R Stan, Wei Huang, Sivakumar Velusamy, Karthik Sankaranarayanan, and David Tarjan. 2003. Temperature-aware Microarchitecture. In *ACM SIGARCH Computer Architecture News*, Vol. 31.

[35] Kevin Skadron, Mircea R. Stan, Karthik Sankaranarayanan, Wei Huang, Sivakumar Velusamy, and David Tarjan. 2004. Temperature-aware Microarchitecture: Modeling and Implementation. *ACM Transactions on Architecture and Code Optimization* 1, 1 (March 2004).

[36] Min Kyu Song, Lei Chen, Joseph Sankman, Stephen Terry, and Dongsheng Ma. 2015. A 20V 8.4W 20MHz Four-Phase GaN DC-DC Converter with Fully On-Chip Dual-SR Bootstrapped GaN FET Driver Achieving 4ns Constant Propagation Delay and 1ns Switching Rise Time. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[37] Yi-Ping Su, Chiun-He Lin, Shen-Yu Peng, Ru-Yu Huang, Te-Fu Yang, Shin-Hao Chen, Ting-Jung Lo, Ke-Homg Chen, Chin-Long Wey, Ying-Hsi Lin, and others. 2015. 90% Peak Efficiency Single-Inductor-Multiple-Output DC-DC Buck Converter with Output Independent Gate Drive Control. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[38] Zeynep Toprak-Deniz, Michael Sperling, John Bulzacchelli, Gregory Still, Ryan Kruse, Seongwon Kim, David Boerstler, Tilman Gloekler, Raphael Robertazzi, Kevin Stawiasz, and others. 2014. Distributed System of Digitally Controlled Microregulators Enabling Per-Core DVFS for the POWER8 Microprocessor. In *Proceedings of the IEEE International Solid-State Circuits Conference*.

[39] Orhun Aras Uzun and Selçuk Köse. 2014. Converter-Gating: A Power Efficient and Secure On-Chip Power Delivery System. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 4, 2 (June 2014).

[40] Inna Vaisband and Eby G Friedman. 2013. Heterogeneous Methodology for Energy Efficient Distribution of On-Chip Power Supplies. *IEEE Transactions on Power Electronics* 28, 9 (September 2013).

[41] Ke Wang, Brett H Meyer, Runjie Zhang, Micrea Stan, and Kevin Skadron. 2014. Walking Pads: Managing C4 Placement for Transient Voltage Noise Minimization. In *Proceedings of the IEEE/ACM Design Automation Conference*.

[42] Steven Cameron Woo, Moriyoshi Ohara, Evan Torrie, Jaswinder Pal Singh, and Anoop Gupta. 1995. The SPLASH-2 Programs: Characterization and Methodological Considerations. In *Proceedings of the 22nd Annual International Symposium on Computer Architecture (ISCA '95)*.

[43] Sam Likun Xi, Hans Jacobson, Pradip Bose, Gu-Yeon Wei, and David Brooks. 2015. Quantifying Sources of Error in McPAT and Potential Impacts on Architectural Studies. In *Proceedings of the IEEE International Symposium on High Performance Computer Architecture*.

[44] Runjie Zhang, Kaushik Mazumdar, Brett H Meyer, Ke Wang, Kevin Skadron, and Mircea R Stan. 2015. Transient Voltage Noise in Charge-recycled Power Delivery Networks for Many-layer 3D-IC. In *Proceedings of the IEEE/ACM International Symposium on Low Power Electronics and Design*.

[45] Runjie Zhang, Ke Wang, Brett H. Meyer, Mircea R. Stan, and Kevin Skadron. 2014. Architecture Implications of Pads As a Scarce Resource. In *Proceedings of the International Symposium on Computer Architecture*.

[46] Pingqiang Zhou, Won Ho Choi, Bongjin Kim, Chris H Kim, and Sachin S Sapatnekar. 2012. Optimization of On-Chip Switched-Capacitor DC-DC Converters for High-Performance Applications. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*.