Regulator-Gating: Adaptive Management of On-Chip Voltage Regulators

Selçuk Köse
Department of Electrical Engineering
University of South Florida
Tampa, Florida
kose@usf.edu

ABSTRACT

Design-for-power has become one of the primary objectives with the continuous demand to improve the battery life of mobile devices or minimize the cooling costs of servers. To save power and mitigate thermal emergencies, circuits typically enter reduced power states when the workload is light. Voltage regulators, however, operate indifferently under varying workload conditions due to the lack of different operating modes. When a voltage regulator is optimized for a particular load current, significant power is dissipated during voltage conversion while delivering a different load current. Adaptive activity management of on-chip voltage regulators based upon the workload information is exploited in this paper to force each on-chip regulator to operate in its most power-efficient load current. In the proposed regulator-gating (ReGa) technique, regulators are adaptively turned on (off) when the current demand is high (low) to improve the voltage conversion efficiency. With the proposed ReGa technique, the overall voltage conversion efficiency from the battery or off-chip power supply to the output of on-chip voltage regulators is improved $\sim 3\times$.

Categories and Subject Descriptors

B.7.1 [**Hardware**]: INTEGRATED CIRCUITS

Keywords

On-chip voltage regulation; parallel voltage regulation; power management

1. INTRODUCTION

With continuous advancements in the semiconductor industry, transistors with smaller than 20 nm feature size have enabled the integration of multi-billion transistors on a single die [13, 28, 34]. With the failure of Dennard's scaling [1], however, only a fraction of the transistors on a die can operate at full voltage/frequency to not exceed the thermal design power (TDP) [8]. A large proportion of the circuit blocks is either inactive (dark silicon) or in a reduced-power state (dim silicon) at any given time to satisfy the power and thermal constraints [4, 5]. Despite the significant amount of research and growing necessity for a holistic power optimization technique, existing efforts to minimize power dissipation are typically not coherent. The existing research efforts are disjointed into two pieces: i) the dynamic and static power loss at the load circuits is minimized or ii) the power loss during power-conversion is minimized.

There is a growing trend for integrating the voltage regulators fully on-chip [10, 18, 19, 27] to improve the quality of voltage delivered to the load circuits. Voltage regulators are typically designed to provide the highest power-conversion efficiency when delivering a particular output current regime (*i.e.*, typically the maximum current for LDO and SC regulators). Since dynamically changing the design parameters of a voltage regulator under different workloads is difficult, existing power management techniques suffer from increased voltage conversion losses during idle states when the current demand is low [16, 30] and regulator efficiency is reduced. A new parallel voltage regulation architecture and regulator management technique are proposed in this paper to improve the voltage conversion efficiency at different utilization levels.

The related background and motivation for the proposed voltage regulator management technique are presented in the next section. The proposed regulator-gating technique is explained in Section 3 with a sample system of parallel voltage regulators. The overall power efficiency improvement and the proposed management of regulator-gating are also offered in this section. The paper is concluded in Section 4.

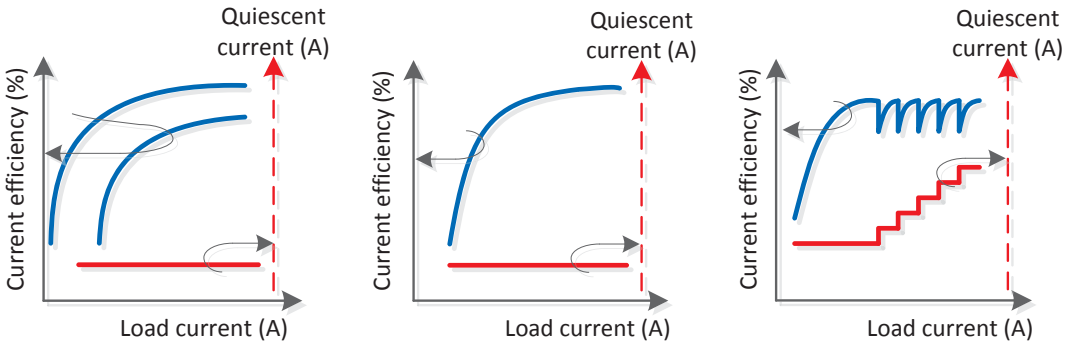
2. BACKGROUND AND MOTIVATION

More than 32% of the overall battery power is dissipated during high-to-low voltage conversion before even reaching the load circuits in modern mobile platforms [22]. The primary reason for this huge power loss is that power delivery networks are designed to satisfy the stringent noise requirements under *worst-case* loading conditions, which is typically the full utilization of the overall chip computing and memory resources when the current demand is the highest.

Parallel voltage regulation has been widely used for buck and SC regulators to reduce the output voltage ripple by interleaving multiple regulators with phase shifted switching frequencies [31, 32]. Advantages of interleaved regulation include reduced filter size for buck converters, improved load response, and higher efficiency [26]. The interleaved architectures, however, have not been exploited until recently to regulate voltage close to the load circuits to minimize

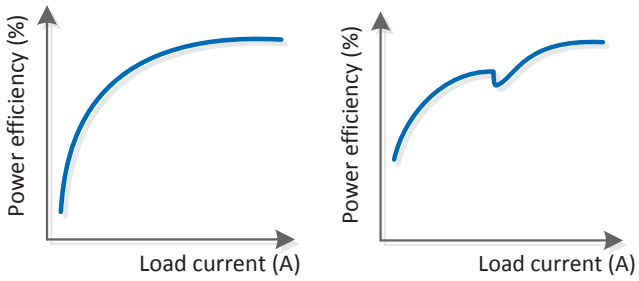
GLSVLSI'14, May 21–23, 2014, Houston, Texas, USA.

http://dx.doi.org/10.1145/2591513.2591524.



(a) (b) (c)

Figure 1: Current efficiency of different LDO regulators. a) Current efficiency of an LDO regulator is significantly degraded when the quiescent current increases at light load currents [20]. b) Current efficiency of an LDO regulator increases monotonically with the load current when the quiescent current is constant [12]. c) Adaptively controlling the quiescent current based upon the load current can improve the current efficiency [23].



(a) (b)

Figure 2: Power efficiency of different SC regulators. a) The power efficiency of a SC regulator is not necessarily monotonic but the maximum efficiency is typically obtained while delivering maximum output current [33]. b) Different techniques can be used to improve the efficiency at light load currents [27].

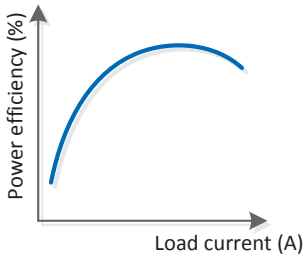
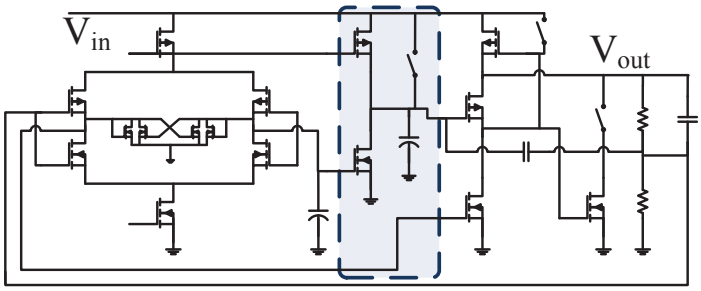


Figure 3: Power efficiency of buck converters. The efficiency graph exhibits a non-monotonic behavior and the maximum power efficiency is not obtained at the full output current [11, 25].

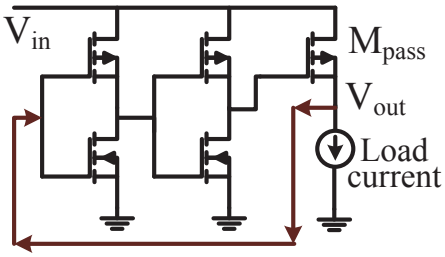
noise [7, 35]. Distributed on-chip voltage regulation is an emerging research area where multiple voltage regulators are connected in parallel, delivering current to the same power network close to the load circuits [2, 3, 6, 7, 17, 21, 35]. Although challenges such as device mismatch, offset voltages among parallel regulators, overall system stability, and balanced current sharing need to be considered, distributed voltage regulation can provide sub-nanosecond load regulation to attain high performance under increased temporal and spatial workload variations in modern ICs. Bulzacchelli *et al.* achieved 500 ps transient response time with a system of eight distributed LDO regulators [6]. Ten LDO regulators and a buck converter are connected in parallel to provide 1.4 amps maximum current in a commercial cellular handset chip [2]. Recently, Lai *et al.* provided detailed guidelines to ensure stability of a distributed voltage regulation system composed of LDO regulators [21] based on a hybrid stability theory. Publications from ST-Ericsson [2] and IBM [6] clearly demonstrate that there is a growing interest not only from academia but also from industry to realize distributed on-chip voltage regulation.

One very important observation that is exploited in this paper is that regulators are optimized at a particular output current, assuming that the regulator typically operates at that particular current regime [29]. The voltage conversion efficiency during the idle periods is therefore significantly degraded since the regulators are almost always designed to provide the highest power efficiency at higher load current [30]. Aggressive power saving mechanisms are currently implemented as a result of the modern ICs exhibiting frequent idle periods [24]. It is projected that more than half of the circuit needs to be idle at 8 nm technology node [5] to satisfy the TDP requirement in server processors or to improve the battery life in mobile processors. Sinkar *et al.* analyzed the potential power savings from workload-aware voltage regulator optimization for a single buck converter and proposed an optimization based on frequency and phase count of the regulator, achieving 34% lower overall power consumption [30].

The power efficiency of low-dropout (LDO) regulators, switched capacitor (SC) regulators, and buck converters are illustrated in Figs. 1, 2, and 3, respectively. The current efficiency of an LDO regulator depends on the quiescent current consumption within the regulator. Although the efficiency of an LDO regulator can be improved by adaptively changing the the quiescent current, current efficiency is significantly



(a) LDO



(b) DLDO

Figure 4: Voltage regulators used in the analysis. a) LDO proposed by Lai *et al.* [20] and b) modified DLDO regulator used in this paper to achieve fast load regulation.

degraded at light load currents, as shown in Fig. 1 [12,20,23]. The power efficiency of an SC regulator typically increases with the output current as shown in Fig. 2a [33]. Although advanced techniques can be used to improve the efficiency at light load currents, as illustrated in Fig. 2b, the efficiency is typically significantly degraded while providing light output current [27]. The power efficiency of a buck converter exhibits a non-monotonic behavior and efficiency is degraded when the load current exceed a certain value. Similar to the other regulator types, the power efficiency of a buck converter is minimized while delivering light load current, as shown in Fig. 3 [11, 25].

As compared to the conventional schemes where the power network is designed targeting the full utilization of the overall chip area, the proposed technique will provide an adaptive power delivery infrastructure that is tailored to provide high voltage conversion efficiency during both fully-utilized and under-utilized modes of operation. One of the primary challenges is to realize a voltage regulator with fast (*a couple of nanoseconds*) turn on and off capability. Voltage regulators with fast turn on and off capability tailored to achieve an adaptive *regulator-gating* methodology are investigated in this paper.

3. REGULATOR-GATING

3.1 Proof of Concept

A distributed power delivery network is constructed with parallel LDO and digital-LDO (DLDO) regulators to provide a *proof of concept* for the proposed ReGa methodology. Although similar results have been obtained with parallel SC and DLDO regulators, in the interest of limited space,

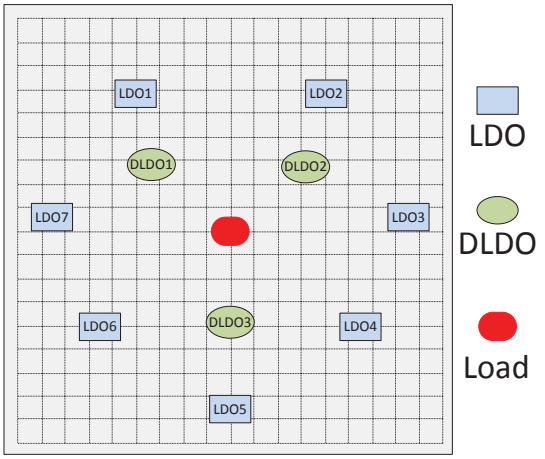


Figure 5: Illustration of a distributed power network with 7 LDO and 3 DLDO voltage regulators connected in parallel.

preliminary results from the system designed with LDO and DLDO regulators are reported in this paper since parallel LDO regulators tend to have more stability problems and are considered to be more difficult to realize.

A DLDO regulator with two skewed inverters to sense the changes at the output voltage and to generate a transient signal to control the gate voltage of a pass transistor M_{pass}, permitting an instant response to transient changes is shown in Fig. 4b. A drop at the output voltage V_{out} causes the pass transistor to provide higher current due to the increased gate voltage. This DLDO regulator is similar to the circuit proposed in [9] with certain differences. The voltage sense portion of the circuit is simpler and a single pass transistor is used without a pull-down NMOS transistor. Due to the smaller area of the sense transistors, multiple copies of this modified DLDO regulator can be distributed across the die in parallel with LDO regulators and can provide a fast response time of ∼400 ps (see Fig. 6). Multiple copies of these DLDO regulators are connected in parallel with the LDO regulator proposed by Lai *et al.* in [20]. The inverting amplifier stage of the LDO regulator, shown in dotted box in Fig. 4a, has been modified to enhance the dynamic response while minimizing the quiescent current consumption at this stage.

Seven LDO and three DLDO regulators are connected to a small power network with 400 nodes, as depicted in Fig. 5. The current contribution from individual regulators to the power grid is shown in Fig. 6 when the load current demand increases from 11 mA to 80 mA. While only one LDO regulator (LDO 7) is sufficient to provide a robust 11 mA current to the load, the rest of the LDO regulators turn on and start providing current to the power grid when the load current demand increases to 80 mA. DLDO regulators turn on immediately after sensing a voltage drop at the power grid and provide instant current to the grid while the LDO regulators are turning on, as shown in Fig. 6. The DLDO regulators remain active only for a couple of nanoseconds (∼4 ns) until the LDO regulators turn on. The DLDO regulators are self-activated, whereas the LDO regulators are controlled by the system-level (global) controller, as explained in the next subsections.

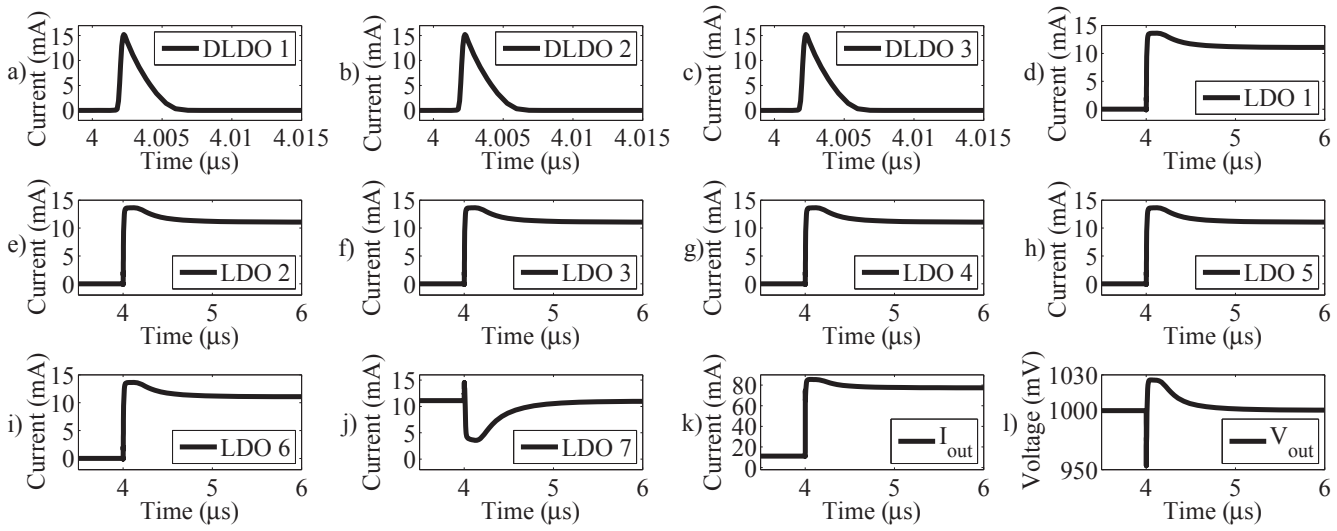


Figure 6: Response time of multiple LDO and DLDO regulators connected in parallel to the same power grid when the load current is increased from 11 mA to 80 mA.

3.2 Regulator Gating to Improve Overall Power Efficiency

The current efficiency of the LDO regulators is around 98% (∼300 μA quiescent current while providing 11 mA current). The quiescent current increases while providing a lower output current, as discussed in [20], and doubles to ∼600 μA when the output current is lower than 2 mA. In this case, the current efficiency of the LDO regulator becomes 2/2.6=∼77%. The increase in the quiescent current is typically observed when an LDO regulator has an AB amplifier type output stage. In the case study where the load current is 11 mA, a single LDO regulator provides the required current with 98% current efficiency. If all of the seven LDO regulators were active while providing 11 mA load current, each LDO regulator would contribute less than 2 mA current to the load with a current efficiency of less than 77%. Without regulator-gating, the total power dissipated during voltage conversion while providing 11 mA load current is $V_{in} * I_{in} - V_{out} * I_{out}$ = 1.2 V * (11 mA + 7 * 0.6 mA) - 1 V * 11 mA = 7.24 mW. Alternatively, with regulator-gating, the total power dissipated during voltage conversion is $V_{in} * I_{in} - V_{out} * I_{out}$ = 1.2 V * 11.3 mA - 1 V * 11 mA = 2.56 mW. This preliminary study demonstrates that power delivery system is ∼3× more power-efficient with ReGa when certain regulators are gated during the idle periods of time. If the idle period lasts 1 ms, the energy savings will be greater than 8 μj for this sample circuit.

On-chip voltage regulation introduces certain overheads, such as area and reduced power efficiencies. In spite of these overheads, on-chip voltage regulation can enable per-core-DVFS, lower the on-chip noise, and reduce the number of dedicated I/O pins [15]. The primary overheads of ReGa assuming that the system already has on-chip voltage regulation are summarized below.

Speed of ReGa: With the utilization of DLDO regulators, the turn on time is decreased to sub-nanosecond range (400 ps in our example). For most of the applications, this turn on time does not degrade system performance.

Area overhead of ReGa Assuming that the power delivery network already has control circuitry for power/clock gating, on-chip voltage or current sensors, and performance counters, the area overhead of ReGa will be the additional area requirement for the DLDO regulators. Please note that DLDO regulators already exist in certain designs without ReGa [9]. When a firmware is used, there is no additional area overhead for ReGa power management.

Power overhead of ReGa The additional power overhead of the proposed ReGa methodology occurs during turning on and off voltage regulators. The power dissipation to turn on an LDO is less than ∼0.1 mW, and the power dissipated by the DLDO is negligible (*i.e.* ∼0.02 mW) when providing 15 mA output current.

3.3 Proposed Regulator-Gating Management

The proposed ReGa control methodology is based on two control loops: i) local control and ii) global control, as illustrated in Fig. 7. A local control provides a sub-nanosecond response to the transient changes in the supply voltage. DLDO regulators are immediately activated based on a simple voltage feedback and provide instant current to the power grid in ∼400 ps. When the voltage emergency is over (*i.e.*, the transient spike is mitigated), these burst mode DLDO regulators are self-deactivated and wait for another interrupt to be activated. Alternatively, a global control loop continuously monitors the overall power consumption of the distributed regulators and compares this information with the available power budget dictated by the system-level controller. If the power consumption exceeds the available power budget limit, certain regulators are turned off. Global and local control loops are fundamentally separate during normal operation, however, the global control loop can override the local control at any time during the operation and can permanently turn off the regulators which are actually controlled locally. When the global controller turns a voltage regulator off, the local control loop cannot turn the regulator on unless the global controller asserts the turn on signal. The proposed power management system is partially imple-

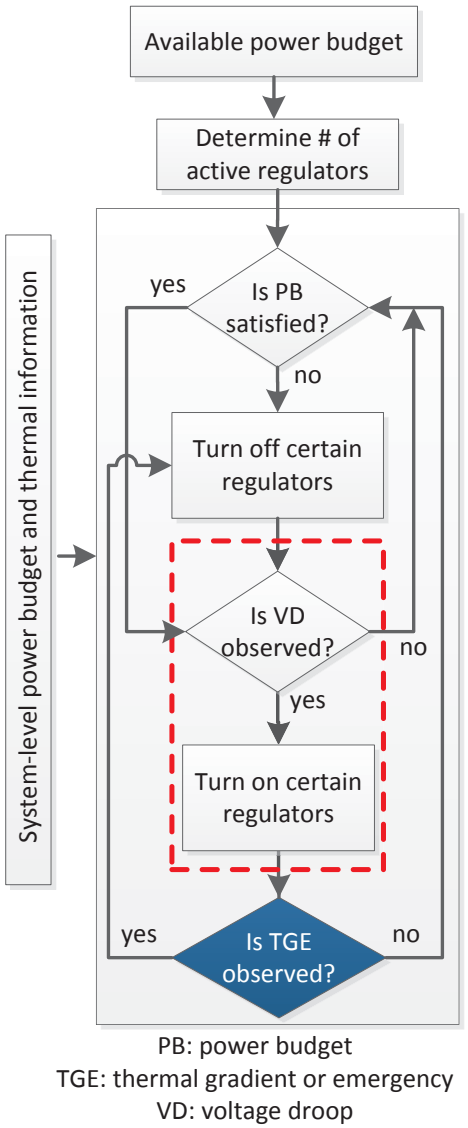


Figure 7: Proposed regulator gating (ReGa) technique. The local control, which is shown with the dotted box, has been implemented in this paper.

mented at the circuit level as demonstrated in the previous section that includes the local control and a portion of the global control which is illustrated in the dotted box in Fig. 7.

4. FUTURE WORK

Since the ReGa technique is orthogonal to other power management techniques, ReGa can be implemented within an existing power management flow with negligible area and power overhead. The system-level workload information to control the DVFS, power-gating, and clock-gating will be leveraged in the proposed ReGa methodology and there will be no additional overhead for workload prediction. A requirement for the proposed methodology is the design of a flexible system-level controller to balance the competing objectives of power, noise, and temperature based upon the data obtained from distributed sensors at runtime. The system-level controller will employ power allocation decisions based on multiple objective optimal control [14]. Distributed sensors within the power management system will report the power, temperature, and noise requirements to the system-level controller, which also considers workload estimations while allocating available power to local regulators through the local controllers. This hierarchical power management organization will be implemented to control the activity of the individual voltage regulators. The stability and feasibility of the overall power management system will also be evaluated.

5. CONCLUSIONS

More than 32% of the overall power is dissipated during voltage conversion in modern integrated circuits. A new power management technique, regulator-gating, is proposed in this paper to improve the voltage conversion efficiency by adaptively controlling the activity of individual voltage regulators within a system of parallel on-chip voltage regulators. The proposed regulator-gating technique achieves $\sim$3$\times$ lower power consumption during voltage conversion in a sample circuit with seven LDO and three DLDO regulators which are connected in parallel. A modified DLDO regulator is utilized to achieve fast turn on capability A power management technique to control the regulator-gating is proposed and partially implemented. The related overheads of the proposed technique are addressed.

6. REFERENCES

[1] R. H. e. Dennard. Design of ion-implanted mosfet's with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9(5):256–268, October 1974.

[2] A. J. D. *et al.* A fully integrated power-management solution for a 65nm cmos cellular handset chip. In *Proceedings of the IEEE International Solid-State Circuits Conference*, pages 382–384, February 2011.

[3] F. L. *et al.* Embedded cmos distributed voltage regulator for large core loads. In *Proceedings of the IEEE European Solid-State Circuits Conference*, pages 521–524, September 2003.

[4] G. V. *et al.* Conservation Cores: Reducing the Energy of Mature Computations. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 205–218, March 2010.

[5] H. E. *et al.* Dark silicon and the end of multicore scaling. In *Proceedings of the International Symposium on Computer Architecture*, pages 365–376, June 2011.

[6] J. F. B. *et al.* Dual-loop system of distributed microregulators with high dc accuracy, load response time below 500 ps, and 85-mv dropout voltage. *IEEE Journal of Solid-State Circuits*, 47(4):863–874, April 2012.

[7] P. Z. *et al.* 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*, pages 263–270, November 2012.

[8] W. H. *et al.* Many-core design from a thermal perspective. In *Proceedings of the IEEE/ACM Design Automation Conference*, pages 746–749, June 2008.

[9] Y. O. *et al.* 0.5-v input digital low-dropout regulator (ldo) with 98.7% current efficiency in 65 nm cmos.

IEICE Transactions on Electronics, 94(6):938–944, June 2011.

[10] J. Guo and K. N. Leung. A 6-μW Chip-Area-Efficient Output-Capacitorless LDO in 90-nm CMOS Technology. *IEEE Journal of Solid-State Circuits*, 45(9):1896–1905, September 2010.

[11] P. e. Hazucha. A 233-MHz 80%-87% Efficient Four-Phase DC-DC Converter Utilizing Air-Core Inductors on Package. *IEEE Journal of Solid-State Circuits*, 40(4):838–845, April 2005.

[12] P. e. Hazucha. Area-Efficient Linear Regulator with Ultra-Fast Load Regulation. *IEEE Journal of Solid-State Circuits*, 40(4):933–940, April 2005.

[13] R. Jakushokas, M. Popovich, A. V. Mezhiba, S. Kose, and E. G. Friedman. *Power Distribution Networks with On-Chip Decoupling Capacitors, Second Edition*. Springer, 2011.

[14] P. P. Khargonekar and M. A. Rotea. Multiple objective optimal control of linear systems: the quadratic norm case. *IEEE Transactions on Automatic Control*, 36(1):14–24, January 1991.

[15] W. Kim, M. S. Gupta, G.-Y. Wei, and D. Brooks. System level analysis of fast, per-core dvfs using on-chip switching regulators. In *Proceedings of the IEEE International Symposium on High Performance Computer Architecture*, pages 123–134, February 2008.

[16] S. Kose. Thermal Implications of On Chip Voltage Regulation Upcoming Challenges and Possible Solutions. In *Proceedings of the IEEE/ACM Design Automation Conference*, June 2014.

[17] S. Kose and E. G. Friedman. Distributed on-chip power delivery. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2(4):704–713, December 2012.

[18] S. Kose, S. Tam, S. Pinzon, B. McDermott, and E. G. Friedman. An Area Efficient On-Chip Hybrid Voltage Regulator. In *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pages 398–403, March 2012.

[19] S. Kose, S. Tam, S. Pinzon, B. McDermott, and E. G. Friedman. Active Filter Based Hybrid On-Chip DC-DC Converters for Point-of-Load Voltage Regulation. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 21(4):680–691, April 2013.

[20] S. Lai and P. Li. A fully on-chip area-efficient cmos low-dropout regulator with fast load regulation. *Analog Integrated Circuits and Signal Processing*, 72(2):925–1030, February 2012.

[21] S. Lai, B. Yan, and P. Li. Stability assurance and design optimization of large power delivery networks with multiple on-chip voltage regulators. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pages 247–254, November 2012.

[22] W. e. Lee. Power conversion efficiency characterization and optimization for smartphones. In *Proceedings of the IEEE/ACM International Symposium on Low Power Electronics and Design*, pages 103–108, July-August 2012.

[23] M. e. Luders. A fully-integrated system power aware ldo for energy harvesting applications. In *Proceedings of the IEEE Symposium on VLSI Circuits*, pages 244–245, June 2011.

[24] D. Meisner, B. T. Gold, and T. F. Wenisch. PowerNap: Eliminating Server Idle Power. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 205–216, March 2009.

[25] M. D. Mulligan, B. Broach, and T. H. Lee. A constant-frequency method for improving light-load efficiency in synchronous buck converters. *IEEE Power Electronics Letters*, 3(1):24–29, March 2005.

[26] X. Peng, Y.-C. Ren, Y. Mao, and F. C. Lee. A family of novel interleaved dc/dc converters for low-voltage high-current voltage regulator module applications. In *Proceedings of the Annual IEEE International Power Electronics Specialists Conference*, pages 1507–1511, June 2001.

[27] Y. Ramadass, A. Fayed, B. Haroun, and A. Chandrakasan. A 0.16 mm^2 Completely On-Chip Switched-Capacitor DC-DC Converter Using Digital Capacitance Modulation for LDO Replacement in 45 nm CMOS. In *Proceedings of the IEEE International Solid-State Circuits Conference*, pages 208–209, February 2010.

[28] R. J. e. Riedlinger. A 32 nm 3.1 billion transistor 12-wide-issue itanium processor for mission-critical servers. In *Proceedings of the IEEE International Solid-State Circuits Conference*, pages 84–86, February 2011.

[29] G. e. Schrom. Optimal design of monolithic integrated dc-dc converters. In *Proceedings of the IEEE International Conference on Integrated Circuit Design and Technology*, pages 1–3, May 2006.

[30] A. A. Sinkar, H. Wang, and N. S. Kim. Workload-aware voltage regulator optimization for power efficient multi-core processors. In *Proceedings of the Conference on Design, Automation and Test in Europe*, pages 1134–1137, March 2012.

[31] N. e. Sturcken. A switched-inductor integrated voltage regulator with nonlinear feedback and network-on-chip load in 45 nm soi. *IEEE Journal of Solid-State Circuits*, 47(8):1935–1945, August 2012.

[32] F. Su, W.-H. Ki, and C.-Y. Tsui. Regulated switched-capacitor doubler with interleaving control for continuous output regulation. *IEEE Journal of Solid-State Circuits*, 44(4):1112–1120, April 2009.

[33] T. e. Umeno. New switched-capacitor dc-dc converter with low input current ripple and its hybridization. In *Proceedings of the IEEE International Midwest Symposium on Symposium on Circuits and Systems*, pages 1091–1094, August 1990.

[34] J. e. Warnock. 5.5 ghz system z microprocessor and multi-chip module. In *Proceedings of the IEEE International Solid-State Circuits Conference*, pages 46–47, February 2013.

[35] P. Zhou, D. Jiao, C. H. Kim, and S. S. Sapatnekar. Exploration of on-chip switched-capacitor dc-dc converter for multicore processors using a distributed power delivery network. In *Proceedings of the IEEE Custom Integrated Circuits Conference*, pages 1–4, September 2011.