# Workload-Aware Voltage Regulator Optimization for Power Efficient Multi-Core Processors

Abhishek A. Sinkar, Hao Wang, and Nam Sung Kim
Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706, USA
{sinkar, hwang223, nskim3}@wisc.edu

*Abstract*— **Modern multi-core processors use power management techniques such as dynamic voltage and frequency scaling (DVFS) and clock gating (CG) which cause the processor to operate in various performance and power states depending on runtime workload characteristics. A voltage regulator (VR), which is designed to provide power to the processor at its highest performance level, can significantly degrade in efficiency when the processor operates in the deep power saving states. In this paper, we propose VR optimization techniques to improve the energy efficiency of the processor + VR system by using the workload dependent P- and C-state residency of real processors. Our experimental results for static VR optimization show up to 19%, 20%, and 4% reduction in energy consumption for workstation, mobile and server multi-core processors. We also investigate the effect of dynamically changing VR parameters on the energy efficiency compared to the static optimization.**

***Keywords-DVFS; switching voltage regulator; P-state; C-state;***

## I. Introduction

Many leading CPU manufacturers today sell multi-core processors which leverage technology scaling to pack multiple processing units or *cores* in a single die. The maximum performance that can be obtained from these multiple cores is often limited by their power consumption which in turn is constrained by the cost of packaging and cooling solutions in high performance servers and by battery capacity in mobile processors. Most modern multi-core processors use power saving techniques like dynamic voltage and frequency scaling (DVFS), clock gating (CG), and per-core power gating (PG) which reduce processor power consumption during low activity periods. In recent multi-core processors (e.g., Intel®'s Core™ i7), these power saving methods are achieved by enforcing various performance and power states (P and C States). P-states are implemented by DVFS mechanism, tuning the core voltage and frequencies for different performance levels. C-states are implemented by various degrees of clock and power gating.

Voltage scaling for DVFS in state-of-the-art processors is achieved by a switching voltage regulator (VR) which is located on the motherboard and communicates with the processor through a digital bus. The voltage scaling time of such off-chip VRs is relatively large (of the order of tens of microseconds) due to the interconnect impedance on the power delivery path. A single VR is typically used to supply a single voltage domain which is shared by all the cores on die to limit the platform and packaging cost. Integration of the VR with core can enable fast, nanosecond scale DVFS while lowering platform and packaging costs. Several recent works indicate the feasibility of fully monolithic VRs in commercial processors [1] [2].

A VR is designed to provide the rated thermal design power (TDP) of the processor with high efficiency. However, our experiments with commercial workloads show that a processor spends more than 50% of the run time in one or more of the low power/performance states. During these states, the power consumption of the CPU drops significantly from its rated value causing the VR to operate at a lower efficiency. This reduces the power efficiency of the system consisting of the VR and processor. In this paper, we propose an optimization of the integrated VR design considering the power consumption characteristics of the processor workload. Although many works in the area of switching regulators have addressed the problem of efficiency optimization, we believe this is the first study to optimize VR design taking into account the workload dependent power consumption trend of processors. Our specific contributions include:

- An analysis of the impact of different processor P- and C-states on the efficiency of an integrated VR design.
- Measurement and analysis of P- and C-state residency of workloads running on real, commercial multi-core processors from three different market computing segments.
- Optimization of VR design taking into account the residency statistics of P- and C-states of multi-core processors. We first present the results of workload residency aware design optimization. Next, for processors showing variable P- and C-state residencies such as servers, we investigate the effect of dynamically changing the VR parameters to maximize energy efficiency.

## II. Integrated VR Efficiency Analysis

Switching voltage regulators such as shown in Fig. 1 are commonly used to supply power to processors due to their high power conversion efficiency over a wide output voltage range compared to linear regulators. Voltage conversion is achieved using switching MOSFETs and passive devices to filter the switching ripple. A multi-core processor can draw a maximum average current of 15~20A per core. For supplying such high current, multiple phases of the circuit in Fig. 1 can be connected in parallel and operated in an interleaved fashion. Integration of VR on the processor die in CMOS technology can enable high switching frequency which reduces the size of the filter L and C components compared to an off-chip VR.

978-3-9810801-8-6/DATE12/©2012 EDAA

The main sources of power loss in a switching VR are the capacitive and resistive losses in the switching MOSFETs and their drivers and the inductor losses. The total power loss, $P_{VR}$, in an integrated VR is described in [3] and can be given by equation (1):

$$P_{VR} = \left(W_p \cdot C_p + W_n \cdot C_n\right) \cdot V_{in}^2 \cdot f_s + \left(\frac{R_p}{W_p} \cdot D + \frac{R_n}{W_n} \cdot (1-D)\right) \cdot i_{rms}^2 + R_{ind} \cdot i_{rms}^2 \tag{1}$$

where $W_p$ and $W_n$ are the effective widths of the PMOS and NMOS switches, $C_p$ and $C_n$ are the effective capacitances of the PMOS and NMOS switches including the driver capacitance, $R_p$ and $R_n$ are the on-state resistances per unit width of the switches, $i_{rms}$ is the RMS value of the current through the inductor, $R_{ind}$ is the effective series resistance of the inductor, $f_s$ is the switching frequency and $D$ is the duty ratio of the PMOS. The above model captures the inductor loss for an inductor technology in the effective resistance $R_{ind}$ which can be calculated as

$$R_{ind} = \frac{\omega \cdot L}{Q} = \frac{2 \cdot \pi \cdot f_s \cdot L}{Q} \tag{2}$$

where $Q$ is the quality factor of the inductor which varies with frequency $f_s$. The optimization of the VR efficiency involves choosing the values of $f_s$, $L$, $W_p$ and $W_n$ such that $P_{VR}$ is minimized. In this study we assume air core inductors mounted on package similar to [2]. The inductance and $Q$ factor data for this inductor technology were obtained from Coilcraft (www.coilcraft.com).

For fixed VR parameters, the efficiency of a VR varies with output voltage and current. Fig. 2-(a) shows the efficiency of a single phase of a VR optimized to provide TDP of a typical 4-core server at 0.9V (For processor configuration, see section V). At light and heavy load the efficiency degrades due to increased capacitive switching loss and resistive loss respectively. The efficiency reaches a maximum value at a current which produces resistive loss equal to the capacitive loss. Further, as output voltage is decreased, the efficiency falls due to increased ripple current. It can be inferred that a fixed VR design, optimized to operate at a given output voltage and current, can suffer from significant efficiency degradation when operated over a wide dynamic output range; e.g. the design in Fig. 2-(a) shows ~24% less efficiency at (0.7V, 0.3A) compared to (0.9V, 2.25A).

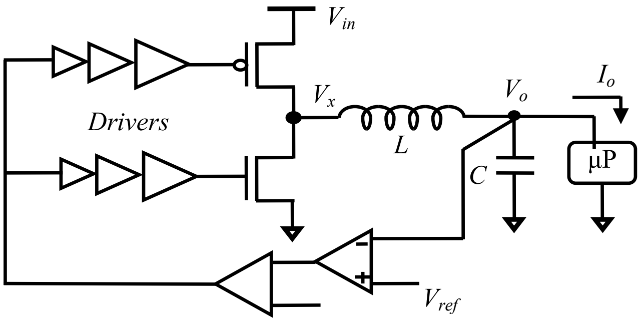


**Figure. 1. Buck converter schematic with feedback control**

Multi-core processors with sophisticated power management techniques, often operate in low power/performance states which cause the VR to deviate from its peak efficiency operation. Fig. 2-(b) shows the efficiency variation of two VR designs, one optimized for supplying TDP to the processor and the other optimized for operation in low performance state with CG ($P_3C_1$). The efficiency of the TDP design falls as the processor goes into higher P-states with low voltages and currents. In the absence of clock gating (solid curves), the TDP design is more efficient than the low power design but degrades in higher P-states and has the same efficiency as the low power design in $P_3$ state. However, the low power design remains 4%, 10%, 14% and 29% more efficient than the TDP design in $P_0$, $P_1$, $P_2$ and $P_3$ states respectively when CG (dashed curves) is applied. CG prevents any dynamic current consumption which often forms 60~ 90% of the total current in processors.

In summary, a VR designed for providing TDP output may not be the best design for workloads which cause the processor to operate in low power/performance states. Such design can reduce the energy savings of runtime power management schemes when total system energy consumption (processor + VR) is considered.

## III. Multi-core P- And C-State Residency Analysis

In a multi-core processor with dynamic power management, higher P-states correspond to operation at a reduced voltage, $V_{DD}$, and maximum operating frequency, $F_{MAX}$, with $F_{MAX}$ in state $P_0$ corresponding to $F_{MAXTDP}$. A P- and C-state combination can exist concurrently in a processor. E.g. a core in $C_0$ state runs at the $F_{MAX}$ dictated by the existing P-state without any CG or PG applied. The $C_1$ state corresponds to CG applied to a part of or the entire core which prevents any dynamic power consumption. Note that a core in $C_1$ state still consumes leakage power $P_{LKG}$ at the $V_{DD}$ corresponding to the existing P-state.

Fig. 3 shows the average P-C state residency of each core in an AMD Opteron 6-core processor targeting a workstation computing segment. In Fig. 3 $P_0$, $P_1$, $P_2$, and $P_3$ correspond to 2.4GHz/0.9V, 1.6GHz/0.8V, 1.2GHz/0.75V and 0.8GHz/0.7V, and each bar represents the residency of each core in a P-state. The bottom part of each bar represents the fraction of time the

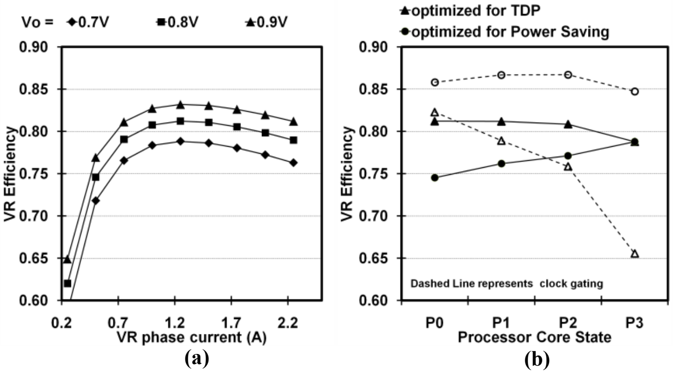
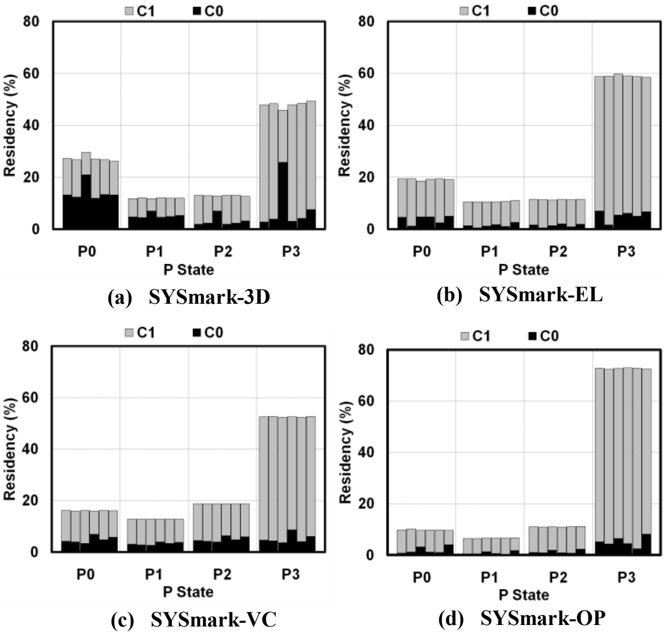


**Figure. 2. VR efficiency variation with output voltage and current in (a) and for two designs across different processor P- and C-states in (b)**



(a) SYSmark-3D

(b) SYSmark-EL

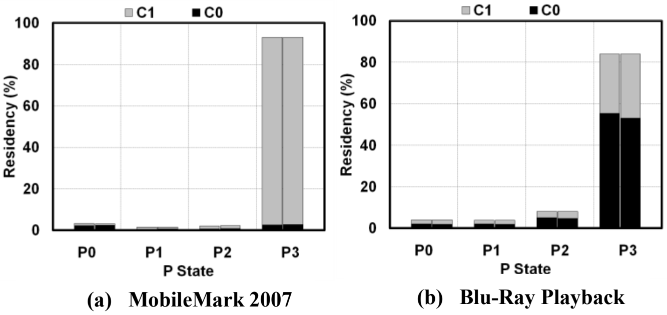
(c) SYSmark-VC

(d) SYSmark-OP

**Figure. 3. Average P-C state residency of cores in a 6-core AMD Opteron workstation. *$P_0$, $P_1$, $P_2$,* and *$P_3$* correspond to 2.4, 1.6, 1.2, and 0.8GHz.**

core is in $C_0$ state while operating at the $V_{DD}$ and $F_{MAX}$ of that P- state. The top portion represents the fraction of time the core is in $C_1$ (clock gated) state. These data were gathered by running SYSmark® 2007 on a real AMD processor. SYSmark® 2007 family of benchmarks mimics usage patterns of business users in the areas of Video creation, E-learning, 3D Modeling and Office Productivity (www.bapco.com). Fig. 3 shows that on average, 73%, 81%, 84%, and 90% of run-time is spent in $P_1$~$P_3$ states (including $C_0$ and $C_1$ time) for SYSmark-3D, EL, VC, and OP respectively. Furthermore, the processor spends 56%, 73%, 71%, and 83% of the run-time in $C_1$ state for these benchmarks.

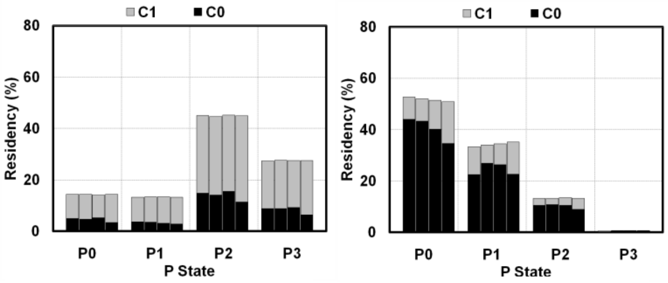
Fig. 4 shows the average P-C state residency of each core in an AMD Athlon II 2-core processor targeting a mobile computing segment. The residency statistics were collected by running the MobileMark 2007 benchmark suite and Blue-Ray Playback which represent typical mobile applications (www.bapco.com). MobileMark 2007 causes both cores to be in the CG condition for more than 90% of its run-time. Meanwhile, Blu-Ray shows only 30% total residency in the $C_1$ state due to its continuous video stream decoding function. Total $P_3$ state residency of Blu-Ray is found to be 84%.



(a) MobileMark 2007

(b) Blu-Ray Playback

**Figure. 4. Average P-C state residency of cores in a 2-core AMD Athlon mobile processor. *$P_0$*, *$P_1$*, *$P_2$,* and *$P_3$* correspond to 2.0, 1.5, 1.2, and 0.8GHz .**



(a) SPECPower-20%

(b) SPECPower-80%

**Figure. 5. Average P-C state residency of cores in a 4-core AMD Opteron server. *$P_0$, $P_1$, $P_2$,* and *$P_3$* correspond to 3.0, 2.3, 1.8, and 0.8GHz.**

Fig. 5 shows the average P-C state residency of each core in an AMD Opteron 4-core processor targeting a server computing segment. The residency statistics were collected by running the SPECPower Server Side Java (SSJ) 2008 benchmark suite (http://www.spec.org/power_ssj2008) which represents typical server applications. When the load is low (~20%), the server spends a substantial fraction of time in the $C_1$ state (60~47%) in the $P_1$ ~ $P_3$ states. Such usage pattern is common in servers from midnight to dawn. At increased load level (~80%) representative of day time usage of data centers, residency in the $P_0$ and $P_1$ states rises with less degree of CG. It can be inferred from this data that a VR design targeting peak performance in $P_0C_0$ state will suffer from reduced energy efficiency for these processors.

## IV. Workload Aware VR Optimization

The objective of the optimization is to minimize the total energy consumption of the VR for the given residency values by optimizing switching frequency $f_s$, inductor value $L$, and the effective widths of the PMOS and NMOS bridge devices, $W_p$ and $W_n$. We consider the losses in the inductor, MOSFETs and drivers. Output capacitor losses are small in multiphase integrated VRs due to ripple cancellation and are not considered although they can be easily included in our optimization framework. The optimization problem can be formulated as follows:

$$\textbf{\textit{Minimize}} \sum_{i=1}^{N} \sum_{j=1}^{M} (P_{VR,i,j} + P_{core,i,j}) \cdot R_{i,j} \qquad (3)$$

***Constraints:***

$f_{s,min} \le f_s \le f_{s,max}$ , $L_{min} \le L \le L_{max}$ , $W_{min} \le W_{p(n)} \le W_{max}$

$0 < Duty_j < 1, j = 1, 2, \ldots, M$

where $P_{VR,i,j}$ is the power loss in the VR when the processor core $i$ is operating in the $j^{th}$ processor state ($P$- and $C$-state combination), $P_{core,i,j}$ is the power consumption and $R_{i,j}$ is the residency of the $i^{th}$ processor core in the $j^{th}$ processor state. $N$ is the total number of cores and $M$ is the number of possible states. In this work we assume M = 8 (4 P-states and 2 C-states), $f_{smin}$ = 50MHz, $f_{smax}$ = 500MHz, $L_{min}$=0.5nH, $L_{max}$ = 10nH, $W_{min}$ = 500µm, and $W_{max}$ = 50e+03µm. We set the constraints such that the VR operates in continuous conduction mode in all processor states. The input voltage was assumed to be 1.2V in all designs. The VR designed for TDP without considering the P- and C-state residency is used as a baseline for comparing the energy reduction obtained with the optimized designs.

### *A. Static optimization*

The VR parameters were optimized for each of the benchmarks shown in Section III. The energy consumption of VR + processor was computed by applying each set of VR parameters to every benchmark within a computing segment. Table I shows the average percentage energy reduction for each core within a segment. Energy savings are more for benchmarks which cause the processor to run in the high P-states with CG. On the contrary, benchmarks which cause a high fraction of $C_0$ state show less benefit since they lead to VR parameters that are close to TDP operation (baseline design).

TABLE I. AVERAGE PERCENTAGE ENERGY REDUCTION BY WORKLOAD-AWARE VR DESIGN

| | Benchmark | Core ID | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Work-station | SYS-3D | 6.2 | 6.4 | 2.0 | 6.7 | 6.0 | 5.7 |
| | SYS-EL | 12.0 | 16.2 | 11.3 | 11.1 | 13.8 | 10.8 |
| | SYS-VC | 12.1 | 12.5 | 13.1 | 9.3 | 11.5 | 10.3 |
| | SYS-OP | 19.2 | 18.7 | 15.5 | 18.8 | 19.6 | 14.1 |
| Mobile | MobileMk | 20.1 | 19.7 | - | - | - | - |
| | Blu-Ray | 8.1 | 8.6 | - | - | - | - |
| Server | SPEC-20% | 3.2 | 3.3 | 3.2 | 4.0 | - | - |
| | SPEC-80% | -0.4 | -0.4 | -0.3 | 0.0 | - | - |

### *B. Runtime phase count and frequency modulation*

We investigated the effect of operating the VR with different switching frequencies and number of phases in the $C_0$ and $C_1$ states. For the results shown in this section, these additional variables were included in the optimization framework. Table II shows the average energy reduction achieved by using two separate frequencies and phase counts for $C_0$ and $C_1$ states. Reducing the switching frequency and phase count lowers the switching losses in $C_1$ state and improves efficiency. Assigning optimal frequency and phase count for individual processor states can further increase the energy savings and can be accomplished with phase locked loop circuits.

## V. EXPERIMENTAL METHODOLOGY

The P- and C-state residency data were obtained by running the benchmarks on commercial processors and Windows operating systems described in Table III. The *xperf* built-in tool of Windows O/S was used to log the P- and C-state transitions.

TABLE II. AVERAGE PERCENTAGE ENERGY REDUCTION BY DYNAMIC PHASE AND FREQUENCY MODULATION

| | Benchmark | Core ID | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Work-station | SYS-3D | 13.3 | 13.4 | 9.8 | 13.5 | 13.1 | 12.8 |
| | SYS-EL | 16.5 | 19.2 | 16.5 | 16.2 | 17.7 | 15.9 |
| | SYS-VC | 16.9 | 17.2 | 17.6 | 15.1 | 16.7 | 15.8 |
| | SYS-OP | 22.8 | 22.5 | 20.4 | 22.7 | 23.4 | 19.4 |
| Mobile | MobileMk | 34.3 | 33.9 | - | - | - | - |
| | Blu-Ray | 17.9 | 18.3 | - | - | - | - |
| Server | SPEC-20% | 7.1 | 7.1 | 7.1 | 7.3 | - | - |
| | SPEC-80% | 6.2 | 6.1 | 6.1 | 6.0 | - | - |

The processor power consumption and voltage, ($P_{TDP}$ and $V_{DDTDP}$) in the $P_0C_0$ state is obtained from the processor specifications. Next, we assume $P_{LKG}$ in $P_0C_0$ state as 30%, 20%, and 40% of $P_{TDP}$ for the workstation, mobile, and server computing segments respectively. Then, assuming the total power in $x = P_0C_0$ state as 1, the normalized dynamic and leakage power in any state $x$ can be written as

$$P_{DYN}(x) = D \cdot f(V_{DD}(x)) \cdot v(V_{DD}(x))^2 \tag{4}$$

$$P_{LKG}(x) = L \cdot l(V_{DD}(x)) \cdot v(V_{DD}(x)) \tag{5}$$

where $D$ and $L$ are fractions of $P_{TDP}$ in x= $P_0C_0$ state representing dynamic and leakage power respectively. $f$, $l$, and $v$ are scaling factors for $F_{MAX}$, $I_{LKG}$, and $V_{DD}$. $f = l = v = 1$ in $P_0C_0$ state and, for other $P$- and $C$- states, can be obtained from the HSPICE simulations as follows. To obtain $V_{DD}$ for a $F_{MAX}(x)$, we simulated a 24 stage FO4 inverter chain in a 32nm predictive technology [4] using HSPICE and recorded $V_{DD}$ corresponding to $F_{MAX}$ for each state. $P_{LKG}$ was modeled by performing HSPICE simulation on a circuit consisting of a large number of NOT, NAND, and NOR gates with effective widths of 50%, 30%, and 20% respectively, representative of a typical processor. The leakage current as a function of $V_{DD}$ is measured by applying a large number of static input combinations to the gates. The MOSFET related capacitances and resistances for the VR designs are extracted from HSPICE simulations by following the procedure outlined in [3].

## VI. CONCLUSION

Multi-core processors with DVFS and CG operate in low power/performance states for a significant fraction of their run-time. Workload- aware VR design can lead to energy savings of up to 19%, 20%, and 4% in workstation, mobile and server computing segments. Further, by choosing lower VR switching frequency and phase count during CG states, energy reduction of up to 23%, 34%, and 7% can be obtained for these computing segments. Workloads with higher residency in higher P-states with CG offer the most opportunity for energy saving.

TABLE III. PROCESSOR CONFIGURATIONS FOR RESIDENCY MEASUREMENT

| Segment | Processor | TDP | Memory | O/S |
|---|---|---|---|---|
| Work-station | 6 core AMD Opteron (SR56x0) | 110W | 8GB, 2- ch., DDR3 1333MHz | Windows 7 Ultimate 64bit |
| Mobile | 2 core AMD Athlon (Tigris) | 35W | 8GB, 2- ch., DDR3 1066MHz | |
| Server | 4 core AMD Opteron (SR56x0) | 130W | 8GB/skt, 2-ch, DDR2 1066 MHz | Win Server 2008R2 |

## VII. REFERENCES

[1] J Ted DiBene II et al., "A 400 Amp Fully Integrated Silicon Voltage with In-die Magnetically Coupled Embedded Inductors," in *Applied Power Electronics Conference*, Palm Springs, CA, 2010.

[2] P. Hazucha et al., "A 233MHz, 80-87% efficient, integrated, 4-phase DC-DC converter in 90nm CMOS," in *Symposium on VLSI Circuits*, 2004.

[3] G. Schrom et al., "Optimal Design of Monolithic Integrated DC-DC Converters," in *International Conference on Integrated Circuit Design and Technology*, 2006.

[4] Predictive Technology Model (PTM). [Online]. http://ptm.asu.edu