Approximate Voltage Regulation for Energy Efficient Error Tolerable Applications

Longfei Wang
Qualcomm Corporation
San Diego, California 92121
longwang@qti.qualcomm.com

Selçuk Köse
Electrical and Computer Engineering
University of Rochester
Rochester, New York 14627
selcuk.kose@rochester.edu

Abstract—The primary objective of on-chip voltage regulation is to achieve the best possible output quality. In a centralized on-chip voltage regulation scheme with a single voltage regulator, the transient voltage droop can be reduced faster transient response as compared to off-chip voltage regulation. To further improve the quality of the generated supply voltage, a distributed on-chip voltage regulation scheme with many localized voltage regulators can be implemented to improve the transient performance and reduce static IR drop. In this work, approximate voltage regulation is investigated to investigate a trade-off between designing a voltage regulator that can achieve the best possible transient and steady state performance or one with just enough output quality. Approximate voltage regulation can fill the niche design space for energy efficient error tolerable applications for both centralized and distribution on-chip voltage regulation cases. Leveraging the error tolerance of the load circuitry, more than 26% current efficiency savings have been achieved with a digital low-dropout regulator.

I. INTRODUCTION

On-chip voltage regulation has become an essential component in modern electronic devices to provide different voltage levels at high quality as required by various functional blocks. The power conversion efficiency of on-chip voltage regulators is a significant design concern as it largely affects the battery life and on-chip thermal profile [1], [2]. Meanwhile, steady state and transient voltage noise profile directly affect the execution speed and error rate of the underlying load circuit [3] within both centralized and distributed on-chip voltage regulation schemes [4]. However, power conversion efficiency of on-chip voltage regulators has typically been compromised to achieve better transient performance and lower static voltage noise, as demonstrated in [1], [2], [4]–[6]. Transient performance includes transient voltage droop and response speed while static voltage noise profile includes steady state output voltage ripple of the voltage regulator and static IR drop within a power grid.

Various techniques have been proposed to improve the output quality, such as the transient voltage droop, output voltage ripple, and transient response speed of on-chip voltage regulators. A better steady state output voltage ripple can be achieved with larger number of phases for a buck converter, but at the cost of power conversion efficiency degradation at light load current conditions [1]. A faster transient response for switched-capacitor DC-DC converters can be achieved with a higher switching frequency [5]. However, power efficiency

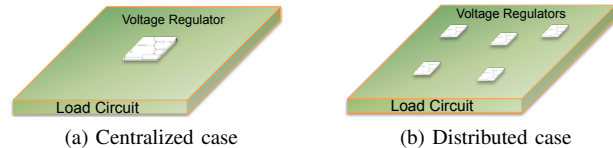(a) Centralized case   (b) Distributed case

Fig. 1. Different types of on-chip voltage regulation

may be lowered with increased frequency. Similarly, a faster settling speed can be obtained with a higher switching frequency, leading to reduced current efficiency for a digital low-dropout regulator (DLDO) [6]. Furthermore, considering distributed on-chip voltage regulation where many individual on-chip voltage regulators are scattered across the chip, more number of voltage regulators are necessary to reach the desired IR drop profile. In this case, the added control circuits incur additional power loss.

Achieving a better trade-off between power conversion efficiency and output quality of on-chip voltage regulators is quite important as both are important design considerations. Circuits with error detection and correction capability can tolerate larger transient voltage droop or static voltage noise [7]. Furthermore, different applications may also have different degrees of error tolerance capability. For example, larger rate of error can be tolerated in the data flow than in control [9]. Instead of designing a voltage regulator to achieve the best possible output quality, approximate voltage regulators with *just enough* output quality can be implemented for error tolerable applications to save energy. For error tolerable applications, a better power conversion efficiency and output quality trade-off for on-chip voltage regulators is explored in this work to realize energy efficient design.

The main contributions of this work are threefold. First, the error tolerance capability of certain applications is investigated to relax the output quality requirement of on-chip voltage regulators to save energy. Second, approximate voltage regulation concept is explored in both centralized and distributed on-chip voltage regulation schemes. Third, the benefits of approximate voltage regulation within error tolerable applications are validated through practical case studies.

The rest of this paper is organized as follows. The trade-off between voltage regulator power efficiency and output quality is explained in Section II. Leveraging approximate voltage regulation for energy efficient error tolerable applications is
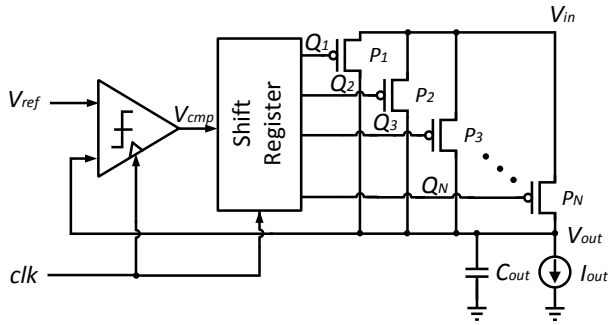
Fig. 2. Schematic of an on-chip DLDO.

Fig. 3. Illustration of voltage regulator transient voltage waveform.

explored in Section III. Evaluation results based on practical case studies are presented in Section IV. Conclusions are offered in Section V.

II. POWER EFFICIENCY AND OUTPUT QUALITY TRADE-OFF FOR ON-CHIP VOLTAGE REGULATORS

Based on the number of voltage regulators deployed, on-chip voltage regulation can be typically categorized into centralized and case as illustrated, respectively, in Figs. 1 (a) and 1 (b). For centralized on-chip voltage regulation, there is only one voltage regulator deployed on chip to supply all the required load current. Alternatively, multiple tiny voltage regulators are implemented at different locations of the chip distributed on-chip voltage regulation. Each voltage regulator only supplies a small portion of the total load current. A mesh power delivery network is inserted between on-chip voltage regulators and load circuits to serve as local power grids for current distribution [4].

To better illustrate voltage regulator power efficiency and output quality trade-off, a specific type of on-chip voltage regulator, DLDO, is adopted in this work. DLDO regulators have been widely investigated in both industry and academia in recent years [8], [9]. Although a DLDO is considered in this paper, the presented analysis presented is general for most voltage regulator types. The schematic of an on-chip DLDO is shown in Fig. 2. Parallel connected power transistors are controlled by a shift register and a clocked comparator to supply the output current I_{out}. Output voltage V_{out} is compared with a reference voltage V_{ref} at the rising edge of each clock clk cycle to determine the number of active power transistors for on-chip voltage regulation. As the power conversion efficiency of an LDO is limited by the dropout voltage between input voltage V_{in} and output voltage V_{out}, current efficiency is typically adopted for evaluation. Current efficiency η_c^{DLDO} for a DLDO can be express as

$$\eta_c^{DLDO} = \frac{I_{out}}{I_{in}} = \frac{I_{out}}{I_{in\_pmos} + I_{in\_sr} + I_{in\_ccmp}} \quad (1)$$

where I_{in}, I_{in_pmos}, I_{in_sr}, and I_{in_ccmp} are, respectively, the total input current of a DLDO, input current of the power transistor array, input current of the shift register, and input current of the clocked comparator.

A. Centralized On-Chip Voltage Regulation

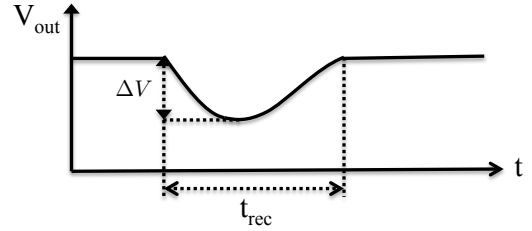For a centralized on-chip voltage regulation scheme with a single voltage regulator, in the case of a DLDO, under a certain load current condition, η_c^{DLDO} is largely affected by I_{in_sr} and I_{in_ccmp} as $I_{in\_pmos} \approx I_{out}$. Meanwhile, the magnitude of a voltage regulator transient voltage droop ΔV as shown in Fig. 3 is a critical design parameter. When a load current step ΔI occurs, V_{out} temporarily drops to $V_{out} - \Delta V$ before it recovers. A droop magnitude that is larger than a certain threshold can lead to voltage emergencies [3], which is not desirable. The magnitude of the transient voltage droop can be estimated as [10]

$$\Delta V = R\Delta I - I_{pmos}f_{clk}R^2 C_{out} ln(1 + \frac{\Delta I}{I_{pmos}f_{clk}RC_{out}}). \quad (2)$$

where R, I_{pmos}, and f_{clk} are the average output resistance of the DLDO before and after load current change, current that is provided by a single power transistor, and clock frequency, respectively. ΔV reduces with increased f_{clk} as detailed in [9]. However, with increased f_{clk}, the switching losses of the shift register and clocked comparator also increase as they are proportional to fCV^2. Accordingly, I_{in_sr} and I_{in_ccmp} will also increase, which can significantly degrade η_c^{DLDO}.

B. Distributed On-Chip Voltage Regulation

For the case of distributed on-chip voltage regulation shown in Fig. 1 (b), a resistive mesh model can be leveraged to investigate the characteristics of on-chip power delivery network [11]. The horizontal and vertical unit resistance can be denoted, respectively, as r_h and r_v. The effective resistance $R_{x,y}$ between two different nodes $N_1(x_1, y_1)$ and $N_2(x_2, y_2)$ can be expressed as [11]

$$R_{x,y}/r = \frac{\sqrt{k}}{2\pi}ln(x^2 + ky^2) + 3.44388 - 0.0033425k$$
$$- \frac{0.1975k(k-1)}{\pi} \quad (3)$$

where $x = |x_1 - x_2|$, $y = |y_1 - y_2|$, $r_v = r$, $r_h = kr$, and k is the ratio between r_h and r_v. A larger distance between two different nodes leads to larger effective resistance as noted in (3). For a certain on-chip power delivery network with increased number of on-chip voltage regulators, the maximum distance between an arbitrary node and a voltage regulator decreases. The reduced maximum distance translates into the reduced maximum static IR drop.

In order to reduce the maximum chip-wide static IR drop, more number of on-chip voltage regulators can be utilized. However, if we consider a case of N distributed on-chip voltage regulators, according to (1), the current efficiency $\eta_c^{disDLDO}$ for all of the N distributed on-chip voltage regulators becomes the ratio of the total distributed output current I_{out_dis} and the sum of total distributed power transistor

727

input current $I_{in\_dispmos}$, total distributed shift register input current I_{in_dissr}, and total distributed clocked comparator input current $I_{in\_disccmp}$. Considering the same f_{clk} for all of the N distributed on-chip DLDOs, $I_{out} = I_{out\_dis}$, the same total number of power transistors and power transistor size for each individual power transistor for both centralized and distributed cases, $I_{in\_dispmos} \approx I_{in\_pmos}$, $I_{in\_dissr} > I_{in\_sr}$, and $I_{in\_disccmp} \approx N I_{in\_ccmp}$. When N is increased, $\eta_c^{disDLDO}$ drops due to additional power loss introduced by the control circuits.

III. ENERGY EFFICIENT ERROR TOLERABLE APPLICATIONS

Optimizing an on-chip voltage regulation scheme to achieve the best possible output quality may degrade power efficiency. For certain applications or circuits that can tolerate error, approximate voltage regulation with *just enough* output quality can be utilized to balance power consumption and voltage quality. Approximate voltage regulation concept is explained with the following centralized and distributed on-chip voltage regulation schemes.

A. Centralized On-Chip Voltage Regulation

Approximate on-chip voltage regulators can be implemented either at the design stage or during run time. With a single voltage regulator, if the underlying load circuit can tolerate more transient voltage noise, the switching frequency of DLDO can be lowered to save energy. If the load circuit has error detection and correction capability or the application is data-centric, the DLDO switching frequency can be lower than a control-centric application or load circuits that are more sensitive to errors.

B. Distributed On-Chip Voltage Regulation

For the case of distributed on-chip voltage regulation with more than one voltage regulator, with increased number of on-chip voltage regulators, the distance between an arbitrary node and an adjacent voltage regulator reduces. This essentially reduces the maximum effective resistance between the load circuits and on-chip voltage regulators. Under the condition of the same load current distribution, the maximum IR drop also decreases. Accordingly, to improve the chip-wide IR drop performance, increasing the number of on-chip voltage regulators provides better results. However, as analyzed in Section II-B, increasing the number of on-chip voltage regulators will lead to degraded $\eta_c^{disDLDO}$. To mitigate $\eta_c^{disDLDO}$ degradation, if the distributed load circuits can tolerate higher level of IR drop, the number of necessary on-chip voltage regulators can be reduced to mitigate the power loss due to additional control circuits.

For both centralized and distributed on-chip voltage regulation schemes, voltage noise tolerance capability of the underlying load circuits can be leveraged to achieve a better voltage regulator output quality and power efficiency trade-off. Such a trade-off relaxes the design requirements for on-chip voltage regulators, leading to reduced power loss and design complexity.

TABLE I
MAGNITUDE OF TRANSIENT VOLTAGE DROOP ΔV, TOTAL RECOVERY TIME t_{rec}, AND CURRENT EFFICIENCY η_c^{DLDO} FOR A DLDO UNDER DIFFERENT SWITCHING FREQUENCY f_{clk}

f_{clk}	10 MHz	100 MHz	500 MHz	1 GHz	2 GHz
ΔV	121.08mV	121.82mV	109.42mV	94.52mV	74.01mV
t_{rec}	12.62us	1.28us	256.51ns	126.89ns	63.46ns
η_c^{DLDO}	98.05%	91.76%	72.34%	57.69%	42.01%

IV. EVALUATION RESULTS

To demonstrate the effectiveness of approximate voltage regulation for energy efficient error tolerable applications, both centralized and distributed on-chip voltage regulation schemes are investigated as detailed below regarding a better voltage quality and efficiency trade-off.

A. Centralized On-Chip Voltage Regulation

For the case of a centralized on-chip voltage regulation scheme, a DLDO with input and output voltages of, respectively, 1.2 V and 1.1 V is designed and simulated in SPICE using 32 nm CMOS process. There are 256 power transistors in the power transistor array and a maximum output current supply capability of around 100 mA is achieved with an output capacitance of 1 nF to mimic the load circuits. With a load current transition from 50 mA to 100 mA, the simulated magnitude of transient voltage droop ΔV, total recovery time t_{rec} as shown in Fig. 3, and current efficiency η_c^{DLDO} are summarized in Table I under different switching frequency f_{clk}.

1) Magnitude of Transient Voltage Droop: As listed in Table I, there is small ΔV change when f_{clk} changes from 10 MHz to 100 MHz. ΔV reduction is more observable when f_{clk} is further increased to the GHz frequency range. However, with increased f_{clk}, η_c^{DLDO} significantly drops from 98.05% to 42.01%. Note that a much higher f_{clk} is typically used during load transient of a DLDO while a lower f_{clk} is adopted during steady state operation [6]. For a load circuit that can tolerate a maximum of 125 mV transient voltage noise, there is no need to implement a 500 MHz f_{clk} for fast load transient conditions. Based on the frequency of load transient occurrence, up to 26% current efficiency saving can be achieved. Further savings in the current efficiency can be achieved when a higher f_{clk} is utilized.

2) Number of Voltage Emergencies: The total recovery time t_{rec} reduces significantly as f_{clk} increases as listed in Table I. An improved t_{rec} can be achieved at the cost of reduced current efficiency η_c^{DLDO}. When the transient voltage droop ΔV is larger than $a\%$ of nominal V_{dd}, voltage emergency happens, leading to increased error rate [3]. Here, a is a constant that is affected by the error tolerance capability of the load circuits. Suppose the supply voltage of the load circuit is slightly above the voltage level where no voltage emergencies occur, for a certain CPU clock frequency f_{CPU}, the number of voltage emergencies is proportional to the duration when the instantaneous supply voltage is below $(1-a\%)V_{dd}$. V_{dd} is the same as V_{out} in the case of a voltage regulator. For an error tolerable application or load circuit that can tolerate N_1

728

f_{clk}	10 MHz	100 MHz	500 MHz	1 GHz	2 GHz
1 DLDO	121.08mV	121.82mV	109.42mV	94.52mV	74.01mV
2 DLDO	125.69mV	122.87mV	98.16mV	81.05mV	62.26mV
4 DLDO	124.34mV	117.76mV	92.8mV	65.24mV	25.84mV
8 DLDO	137.11mV	102.59mV	50.22mV	50.98mV	13.51mV

f_{clk}	10 MHz	100 MHz	500 MHz	1 GHz	2 GHz
1 DLDO	12.62us	1.28us	256.51ns	126.89ns	63.46ns
2 DLDO	8.02us	807.65ns	130.79ns	70.81ns	35.18ns
4 DLDO	4.44us	333.76ns	68.84ns	36.48ns	18.62ns
8 DLDO	2.16us	169.01ns	34.63ns	24.96ns	8.73ns

voltage emergencies during t_{rec} under a f_{clk} of 10 MHz, a constant f_{clk} of 10 MHz can be used even during load transient conditions. Similar to the case described in Section IV-A1, more than 26% savings in the current efficiency can be realized.

B. Distributed On-Chip Voltage Regulation

A distributed on-chip voltage regulation scheme can have varying number of on-chip DLDOs. For a fair comparison, the same input and output voltages as in Section IV-A are considered. The unit power transistor size and total number of power transistors are the same as in Section IV-A, thus the maximum current supply capability of different distributed DLDO cases is the same. For a case of 1, 2, 4, and 8 distributed DLDOs, the number of power transistors within each individual DLDO is, respectively, 256, 128, 64, and 32. Phase interleaving among different individual DLDOs is adopted to improve the transient performance. A 100 by 100 power grid is utilized with a unit grid resistance of 45 mΩ. The same amount of load current and output capacitance as in Section IV-A are uniformly distributed across the power grid.

1) Transient Performance: With the same load current transition from 50 mA to 100 mA, the transient simulation results regarding transient voltage droop and total recovery time for different number of distributed on-chip voltage regulators are summarized in Tables II and III, respectively. As listed in Table II, ΔV does not significantly improve for f_{clk} below 500 MHz as the number of DLDOs increases. For a higher f_{clk} larger than 500 MHz, the ΔV improvement can be considerable as the number of DLDOs increases. To achieve a superior ΔV performance, a larger number of DLDOs operating at GHz frequency range may be necessary. Similarly, as shown from Table III, under a certain f_{clk}, t_{rec} reduces with increased number of DLDOs. With increased f_{clk}, t_{rec} also improves. A superb t_{rec} performance can also be realized with a large number of DLDOs operating at GHz frequency range. However, as the number of distributed DLDOs and f_{clk} increase, η_c^{DLDO} significantly drops. For example, for distributed on-chip voltage regulation with 2 DLDOs, η_c^{DLDO} with f_{clk} of 10 MHz, 100 MHz, 500 MHz, 1 GHz, and

2 GHz is, respectively, 97.64%, 88.77%, 66.09%, 52.02%, and 37.94%, which is about 5% drop as compared to the case of a single DLDO. For a larger number of DLDOs, the degradation in η_c^{DLDO} can be more significant. For error tolerable applications, the potential η_c^{DLDO} savings can be more noticeable.

2) Static IR Drop Profile: With larger number of DLDOs, the maximum effective resistance as well as the current that is provided by each individual DLDO reduce. For evenly distributed of DLDOs and load current, each individual DLDO provideS approximately the same amount of load current. Furthermore, the maximum distance between an arbitrary node and the closest DLDO reduces, leading to reduced effective resistance. The maximum IR drop is determined by the total number of DLDOs and the reduced effective resistance, which can be beneficial. However, this IR drop improvement also comes at the cost of reduced η_c^{DLDO}. For applications that can tolerate certain amount of IR drop, significant η_c^{DLDO} savings can also be observed.

V. CONCLUSION

Approximate on-chip voltage regulation achieving just enough output quality is investigated in this work for energy efficient error tolerable applications. More than 26% savings in the current efficiency for on-chip DLDOs can be realized. Furthermore, voltage regulation and load circuit co-design considering not only the load current behavior, but also the error tolerance capability of the load circuit needs to be considered to achieve a better quality and efficiency trade-off.

REFERENCES

[1] E. A. Burton *et al.*, "FIVR - Fully Integrated Voltage Regulators on 4th Generation Intel Core^{TM} SoCs," *Proceedings of IEEE APEC*, pp. 432-439, March 2014.
[2] S. K. Khatamifard, L. Wang, W. Yu, S. Köse, and U. R. Karpuzcu, "ThermoGater: Thermally-Aware On-Chip Voltage Regulation," *Proceedings of the IEEE ISCA*, pp. 120-132, June 2017.
[3] R. Zhang, K. Wang, B. H. Meyer, M. R. Stan, and K. Skadron, "Architecture Implications of Pads as a Scarce Resource," *Proceedings of the IEEE ISCA*, pp. 373-384, June 2014.
[4] S. Köse and E. G. Friedman, "Distributed On-Chip Power Delivery," *IEEE JETCAS*, vol. 2, no. 4, pp. 704-713, December 2012.
[5] Orhun Aras Uzun and S. Köse, "Converter-Gating: A Power Efficient and Secure On-Chip Power Delivery System," *IEEE JETCAS*, vol. 4, no. 2, pp. 169-179, June 2014.
[6] S. B. Nasir, S. Gangopadhyay, and A. Raychowdhury, "All-Digital Low-Dropout Regulator With Adaptive Control and Reduced Dynamic Stability for Digital Load Circuits," *IEEE Transactions on Power Electronics*, vol. 31, no. 12, pp. 8293-8302, December 2016.
[7] D. Ernst, N. S. Kim, S. Das, S. Pant, R. Rao, and T. Pham, "Razor: A Low-Power Pipeline Based on Circuit-Level Timing Speculation," *Proceedings of MICRO*, pp. 1-12, December 2003.
[8] Z. Toprak-Deniz *et al.*, "Distributed System of Digitally Controlled Microregulators Enabling Per-core DVFS for the POWER8TM Microprocessor," *Proceedings of ISSCC*, February 2014.
[9] L. Wang, S. K. Khatamifard, U. R. Karpuzcu, and S. Köse, "Exploiting Algorithmic Noise Tolerance for Scalable On-Chip Voltage Regulation," *IEEE TVLSI*, vol. 27, no. 1, pp. 229-241, January 2019.
[10] S. Leitner, P. West, C. Lu, and H. Wang, "Digital LDO Modeling for Early Design Space Exploration," *Proceedings of the IEEE SOCC*, pp. 7-12, September 2016.
[11] S. Köse and E. G. Friedman, "Efficient Algorithms for Fast IR Drop Analysis Exploiting Locality," *Integration, the VLSI Journal*, vol. 45, no. 2, pp. 149-161, March 2012.