ABSTRACT

Integrating a large number of on-chip voltage regulators holds the promise of solving many power delivery challenges through strong local load regulation and facilitates system-level power management. The quantitative understanding of such complex power delivery networks (PDNs) is hampered by the large network complexity and interactions between passive on-die/package-level circuits and a multitude of nonlinear active regulators. We develop a fast combined GPU-CPU analysis engine encompassing several simulation strategies, optimized for various subcomponents of the network. Using accurate quantitative analysis, we demonstrate the significant performance improvement brought by on-chip low-dropout regulators (LDOs) in terms of suppressing high-frequency local voltage droops and avoiding the mid-frequency resonance caused by off-chip inductive parasitics. We perform comprehensive analysis on the tradeoffs among overhead of on-chip LDOs, maximum voltage droop and overall power efficiency. We conduct systematic design optimization by developing a simulation-based nonlinear optimization strategy that determines the optimal number of on-chip LDOs required and on-board input voltage, and the corresponding voltage droop and power efficiency for PDNs with multiple power domains.

Categories and Subject Descriptors
B.7.2 [Integrated Circuits]: Design Aids.

General Terms

Keywords
Power delivery network, on-chip voltage regulation, power efficiency, low-dropout regulator.

1. INTRODUCTION

Power delivery networks (PDNs) are key subsystems that are responsible for supplying power to all on-chip devices. Voltage regulators are essential components in power delivery. Traditionally, voltage regulators are contained in board-level voltage regulator modules with large inductors or capacitors. The costs and sizes of such bulky modules severely limit their use for multiple power domain regulation. As such, there is a significant interest in developing fully integrated on-chip voltage regulators. Initial studies have shown that on-chip regulators can significantly improve load regulation, reduce crosstalk, eliminate load-transient spikes caused by bondwire and package inductances, and save board space as well as external pins [11]. In addition, thanks to the small size, it is promising to integrate many voltage regulators on chip to facilitate fine-grained multiple power domains [6] [1].

Among existing voltage regulators [7] [15] [8] [11] [6] [12], low-dropout regulators (LDO) are more amenable for on-chip integration due to their small area overhead, low standby current, low dropout voltage, and improved power efficiency. Fully integrated LDOs with no external capacitors are very attractive for regulating large and fast local voltage fluctuations and providing multiple levels of supply voltage. [11] [6] [12]. Despite the design challenges associated with on-chip voltage regulators and the progress has been made, little work has been geared towards understanding the detailed electrical characteristics of a multitude of on-chip voltage regulators operating as a part of a large power delivery network. In addition, there is a strong need to derive basic design tradeoffs and perform whole system optimization for sophisticated delivery networks with on-chip voltage regulation. Along this line, the goals of this paper are severalfold:

1. First, we address the significant challenge in simulating complex PDNs with a large number of integrated LDOs with SPICE-level accuracy by developing an integrated GPU-CPU analysis engine; our engine achieves its efficiency through circuit partitioning and integration of linear iterative, linear direct and nonlinear solvers running on the GPU (graphics processing unit) or CPU hardware, and optimized for large on-die power grids, package models and transistor-level LDO models, respectively.

2. Second, we demonstrate the significant benefits brought by on-chip LDOs and show how localized load regulation can noticeably suppress both the high-frequency switching noises and mid-frequency droops caused by the resonance with off-chip inductive parasitics; we perform comprehensive analysis trading off the number of on-chip LDOs, maximum voltage droop and power efficiency.

3. Finally, we develop a simulation-based nonlinear optimization approach that systematically maximizes power efficiency under maximum voltage droop and LDO overhead constraints; we extract key observations on the optimization of PDNs with multiple power domains.
2. ON-CHIP LDO BACKGROUND

An external capacitorless low-dropout voltage regulator [12], shown in Fig. 1, is comprised of a pass element (M_p), sampling resistors (R_{i1} and R_{i2}), reference voltage (V_{ref}), an error amplifier, and a differentiator. Several important terms and definitions of LDOs are introduced to help understanding the operation and performance of LDOs [10].

Figure 1: An external capacitorless low-dropout voltage regulator.

- **Dropout voltage** V_{drop}: the input-to-output differential voltage at which the circuit ceases to regulate against further reductions in input voltage.
- **Dropout region and regulation region**: if input-to-output voltage difference is less than dropout voltage, the regulator is in the dropout region and the output voltage degrades in proportion to the decreasing input voltage. In contrast, if input-to-output voltage difference is larger than dropout voltage, the regulator is in regulation region and the output voltage maintains a stable level.
- **Power efficiency** ε: the efficiency of an LDO is limited by the quiescent current and input/output voltage, and is defined as,

\[ \varepsilon = \frac{I_{out}V_{out}}{I_{in}V_{in}} = \frac{I_{out}V_{out}}{(I_{out} + I_q)V_{in}} \]  

where I_q is the quiescent current flowing to the ground. Increasing input-to-output voltage difference would reduce power efficiency.
- **Line regulation** (∆V_{out}/V_{in}) and **load regulation** (∆V_{out}/I_{out}): they are used to measure the change of LDO’s output voltage due to the changes of input voltage and loading current respectively.
- **Maximum current loading** I_{max}.

The LDO topology from [12] is adopted in our experimental studies. To mimic the characteristics of projected future on-chip LDO designs and their impacts on power delivery, design alternations are introduced such that the resulting LDO has a quiescent currents of less than 0.05mA and output voltage levels of 1.4V, 1.2V and 1V.

3. POWER DELIVERY NETWORK: MODELING AND SIMULATION

3.1 Power Delivery Network Modeling

A detailed model of multiple-domain power delivery network with on-chip low-dropout regulators is presented in Fig. 2. A ladder RLC network [5] is utilized to capture the parasitics of off chip network which consists of PC board, socket, package, and off chip decoupling capacitors. C4 bumps are modeled as parallel RL pairs. For simplicity, the on-board power supply is modeled as a fixed voltage source. The on-chip PDN has the following components: a global VDD grid, on-chip LDOs, local grids, a global GND grid, decoupling capacitors, and switching transistors. The global VDD grid is a sparse grid that distributes input voltage to on-chip LDOs. Each local grid corresponds to a power domain, and its voltage is provided by LDOs. In the on-chip PDN, each grid is modeled as an RC network with decoupling capacitors. Switching transistors in functional modules are modeled as dynamic current sources. The dynamic current sources are generally approximated with triangular waveforms, whose amplitudes and periods are determined by the power consumption and active rate of the corresponding functional modules.

Figure 2: Model of power delivery network with on-chip voltage regulators.

In this paper, the on-chip power grids of PDNs are generated according to the typical current loadings and wire conductance of the IBM power grid benchmarks [13], while the package level model parameters, such as inductance and capacitance values, as well as total on-die capacitance are adopted from [5]. It should also be noted that the current loadings for our PDNs are assumed to be given.

3.2 Simulation of PDN with On-Chip LDOs

To meaningfully capture the important dynamic behavior of large power delivery networks with on-chip LDOs, thousands of time steps are needed in the transient analysis. However, analyzing such a PDN has two major obstacles:

- The on-chip grids are very large meshes with millions of nodes. Analyzing such huge systems with traditional SPICE simulator is very time and memory consuming.
- On-chip LDOs have nonlinear devices, and off-chip model has inductances. Therefore, the system matrix is no longer a symmetric positive definite matrix. Existing fast power grid simulators can not be utilized to solve such a PDN circuit.

In this paper, we address the simulation challenges by developing a GPU accelerated simulation engine, Gsim, which solves extremely large PDNs (having on-chip LDOs) with good runtime and memory efficiency.

Figure 3: PDN partitioning and simulation.

Gsim utilizes an efficient iterative partitioning relaxation method to analyzes LDOs and off-chip circuit (solved on
CPU) as well as extremely large on-chip power grids (solved on GPU). As presented in the left of Fig. 3, the entire PDN can be partitioned into five major circuit blocks: (1) off-chip circuits, (2) LDOs, (3) global VDD grid, (4) local grids, and (5) global GND grid. For the transient simulation, at each time point, Gsim solves each circuit block individually by fixing the boundary voltages and updates the solutions through the partition boundary until the convergence is reached.

As presented in Fig. 3, to solve each block in the most efficient way: off-chip circuits are solved by a passive network solver on CPU; transistor-level LDO models are analyzed by a SPICE solver on CPU; and all the power grids (VDD grid, local grids and GND grid) are tackled by a GPU multigrid solver [3] which is over 50X faster than the state-of-the-art direct solver CHOLMOD [2]. The updated results for partition boundaries are exchanged through PCI-E between CPU solvers and the GPU solver. It will be shown in experimental results, most of the simulation time is spent on solving power grids. Thus our partitioning simulation scheme which puts power grid simulations on fast GPU engine is very efficient in terms of runtime.

The convergence is examined by checking the average and maximum voltage changes at partition boundaries. Although LDO is a nonlinear device, due to its property that automatically maintains the output voltage, the voltage change at the boundary between LDO and local grids is small for consecutive time steps. Other partition boundaries have small voltage change from time step to step as well. Therefore, the convergence can be quickly reached, which is consistent with our experimental results. Further convergence improvement can be introduced in the form of multi-level Newton method.

4. ON-CHIP VOLTAGE REGULATION: BENEFITS AND TRADEOFFS

4.1 Benefits of On-Chip Voltage Regulation

High-frequency local voltage droops due to the fast switching load currents and mid-frequency global resonance caused by off-chip inductive parasitics are two major contributors to the voltage fluctuation [5] [9]. Suppressing or remediating these effects would significantly improve the performance of PDN.

![Figure 4: Voltage droops for a power domain with LDOs and without LDOs.](image)

With the accurate and powerful analysis engine Gsim, we give a quantitative analysis to the effects of having on-chip voltage regulation. A random node’s voltage droops before and after integrating on-chip LDOs are examined, and the voltage droop waveforms are shown in Fig. 4. When the PDN is working without LDOs, local grids are connected to the VDD grid through vias, otherwise they are connected to LDOs as shown in Fig. 2.

Figure 4 shows that without the regulation of on-chip LDOs, on top of the high-frequency voltage droop, there is a large mid-frequency swing due to chip-package resonance. Even though the voltage droop caused by current loadings is about 140mV, the resonance introduces another 40mV droop, and makes the total voltage droop much larger. However, in contrast, by having on-chip voltage regulators, the benefits are twofold:

1. Suppressing high-frequency local droop. On-chip LDOs provide strong local regulation. They respond quickly to local current fluctuations and automatically maintain the output voltage level. Hence large local voltage swings are suppressed significantly.

2. Remediating mid-frequency global resonance. On-chip LDOs do not suffer the global resonance since the resonance is blocked at the input of LDOs. LDOs have weak transfer functions so that as long as working in the regulation region they are not sensitive to the input changes. Therefore, large voltage fluctuations at the off-chip circuits and VDD grids can not propagate to local grids.

4.2 Design Aspects of On-Chip Voltage Regulation

In order to utilize on-chip voltage regulation for a PDN with high performance and low cost, three important aspects need to be considered: maximum voltage droop, overall power efficiency, and LDO overhead. Obviously, LDO overhead is associated with the number of LDOs, since having more LDOs would occupy more chip area as well as routing resources. The locations of LDOs have an impact on the voltage droop, but since we assume the LDOs are evenly distributed, the location effect is not considered under this context. To better understand the other two design aspects, they are analyzed in detail in the following.

4.2.1 Voltage Droop

For a PDN design with on-chip regulation, there are two factors that affect the maximum voltage droop: the number of on-chip LDOs and the on-board input voltage. We run two sets of simulations to explore the relationships between the voltage droop and the number of LDOs as well as the input voltage. In the first set, PDNs having various numbers of LDOs (all the LDOs are evenly distributed) but the same fixed on-board supply (1.6V) are examined. While in the second set, the number of LDOs is fixed (N=9), but the input voltages are varying. All the testing PDNs have the same package and on-chip models. They only vary in the number of LDOs or input voltage. In order to capture sufficient dynamics, hundreds of clock cycles are simulated. A single power domain with 1.4V Vdd level is examined. The simulation results are shown in Fig. 5, and the relationships between maximum voltage droop and the number of LDOs as well as input voltage are analyzed below. It should be noted that the voltage droop includes voltage droop on the local grid and voltage overshoot in the GND grid.

![Figure 5: Voltage droop vs. number of LDOs and input voltage.](image)

- Maximum voltage droop vs. number of LDOs: As shown in (a) of Fig. 5, integrating more LDOs to
the PDN strengthens the local voltage regulation since more nodes would be covered by the effects of LDOs. In addition, less amount of dynamic current fluctuation is presented to each LDO, which visibly reduces the swings at the output of LDOs. Moreover, having more LDOs would make them a stronger and more reliable regulation system, since they can share work loads and help each other to encounter large load current transitions. Therefore, increasing the number of LDOs would lower maximum voltage droop.

- **Maximum voltage droop vs. input voltage:** As shown in (b) of Fig. 5, when the difference between input voltage and output voltage is less than the dropout voltage, LDOs are working in the dropout region, and they have weaker regulation on output voltage with respect to input voltage change. Thus the local grids are exposed to the mid-frequency resonance, reflected by large voltage droops. However, if the input-to-output voltage difference is above dropout voltage, LDOs are working in the regulation region, and the voltage droop remains a stable level with a small shift due to line regulation. In summary, significantly lowering the input voltage would harm the power integrity of the entire PDN.

### 4.2.2 Power Efficiency

Similar to maximum voltage droop, overall power efficiency is also impacted by the number of LDOs and on-board input voltage. Assume there are \( N \) LDOs, utilizing simulation engine Gsim illustrated in Section 3.2, for time period \( t_1 \) to \( t_2 \), the overall power efficiency can be calculated as (2).

\[
\varepsilon = \frac{\sum_{k=1}^{N} P_{o}^{k}}{\sum_{k=1}^{N} P_{i}^{k}}
\]

where \( P_{o}^{k} = \int_{t_1}^{t_2} V_{out}^k I_{out}^k dt \) and \( P_{i}^{k} = \int_{t_1}^{t_2} V_{in}^k I_{in}^k dt \) are input and output power of the \( k \)th LDO.

We also run the same two sets of simulations as in Section 4.2.1 to explore the relationships between the power efficiency and the number of LDOs as well as input voltage. The simulation results are shown in Figure 6, and the observations are the following:

![Figure 6: Power efficiency vs. number of LDOs and input voltage.](image)

- **Overall power efficiency vs. number of LDOs:** As presented in the left of Figure 6, increasing the number of LDO would bring down power efficiency, but only with a small amount. The efficiency reduction is due to the reason that more LDOs would result in less dynamic current loading on each LDO. According to (1), less \( I_{out} \) would lower efficiency \( \varepsilon \). However, there are two reasons why the reduction is small. The first reason is that due to load regulation, decrease of \( I_{out} \) would increase \( V_{out} \), which makes the overall \( V_{out}I_{out} \) only has a small decrease. Another reason is that the quiescent current \( I_q \) of the LDO we use is very small. Even though \( I_{out} \) gets reduced, it is still large compared with \( I_q \). Thus the overall efficiency is still at a high level.

- **Overall power efficiency vs. input voltage:** On the other hand, power efficiency gets improved by lowering input voltage, as shown in (b) of Fig. 6. According to (1) less \( V_{in} \) makes \( \frac{1}{\sqrt{Q}} \) larger, thus raises up the efficiency. With an extreme low input voltage, LDOs are pushed into dropout region and work merely as resistors with small resistance values. Therefore, as shown in (b) of Figure 6, the efficiency for this extreme case is very high. In conclusion, lowering input voltage would effectively improve power efficiency.

### 4.3 Design Tradeoffs of On-Chip Voltage Regulation

The design tradeoffs among maximum voltage droop, power efficiency, and overhead (number of LDOs) are summarized in Fig. 7. Four sets of simulations results are shown in this figure. In each set, the overhead (\( N \), number of LDOs) is fixed and the input voltages vary from 1.4V to 1.65V (from left to right in the figure). We have following important observations:

- In order to improve power efficiency, the most effective way is to reduce input voltage. As long as the input voltage is high enough to ensure that LDOs are working in regulation region, the power efficiency can get improved without jeopardizing voltage droop. However, if designers are too aggressive on raising up efficiency and using a very low input voltage, they may end up pushing LDOs to dropout region and make the voltage droop unacceptable. Reducing the number of LDOs can also improve the overall efficiency, however, the amount of improvement is small.

- In terms of reducing voltage droop, there are two ways: increasing the number of LDOs and increasing the input voltage. Increasing the LDO number seems to be a more effective way. However, it is noticed that when the number of LDOs is large enough, adding more LDOs to the PDN can only have a small amount of reduction on voltage droop, but it costs a lot for overhead. When LDOs are working in dropout region, increasing input voltage has a significant impact on improving voltage droop. However, once LDOs are in the regulation region, voltage droop is almost insensitive to input voltage changes.
5. OPTIMIZATION FOR ON-CHIP VOLTAGE REGULATION

As discussed in Section 4, for a given power domain, increasing its number of on-chip LDOs would decrease the maximum voltage droop, but leads to large overhead. At the same time, although, having a large input voltage would ensure that all the LDOs are working in the regulation region, power efficiency would be significantly low. Therefore, to reach the highest power efficiency objective while meeting the voltage ripple requirement, the number of LDOs for each power domain and the on-board input voltage need to be found out. Such a task leads to an optimization problem.

5.1 Optimization Formulation

Assume there are \( N \) power domains \( \{D_1, \ldots, D_N\} \), \( D_k \) has \( X_k \) LDOs \( \{Z_{k1}, \ldots, Z_{kX_k}\} \) uniformly distributed \( \{X_k\} \) is the number of LDOs in horizontal and \( Y_k \) is the number of LDOs in vertical. For a time period \( t_1 \) to \( t_2 \), the input powers of LDOs in \( D_k \) are \( \{P_{k1}^d, \ldots, P_{kX_k}^d\} \), while the output powers are \( \{P_{k1}^e, \ldots, P_{kX_k}^e\} \) respectively. The maximum voltage droops for power domain \( D_k \) during this period is \( V_k^d \). The input voltage supply from the on-board ideal voltage source is \( V_{in} \). Then the objective function for overall power efficiency can be expressed as

\[
\varepsilon = \sum_{k=1}^{N} \frac{X_k Y_k}{\sum_{j=1}^{N} X_k Y_k} \sum_{i=1}^{N} \frac{P_{ij}}{P_{ij}} \sum_{j=1}^{N} P_{ij} \tag{3}
\]

It is subject to:

1. Voltage droop constraint:

\[
V_k^m \leq V_k^d, \quad k = 1, \ldots, N \tag{4}
\]

where \( V_k^m \) is the voltage droop constraint for \( D_k \).

2. Overhead constraint:

\[
\sum_{k=1}^{N} X_k Y_k \leq M \tag{5}
\]

where \( M \) is the maximum number of LDOs that can be integrated on chip.

In the this optimization formulation, \( \{V_{in}, X_1, \ldots, X_N\} \) and \( \{Y_1, \ldots, Y_N\} \) are variables. \( \varepsilon \) and \( V_k^d \) \((k = 1, \ldots, N)\) are implicit functions of these variables and can only be obtained by analyzing the PDN for time period \( t_1 \) to \( t_2 \).

5.2 Optimization Using APPS

We propose to use the asynchronous parallel pattern search (APPS) [4] method to solve the above optimization problem. APPS solves unconstrained and bound-constrained optimization problems given by [4]

\[
\min f(x) \tag{6a}
\]

\[
s.t. \quad 1 \leq x \leq u \tag{6b}
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) and \( x, u \in \mathbb{R}^n \).

APPS is targeted for simulation-based optimization with a moderate number of variables and an objective function requiring complex simulation. Hence, it is suitable for solving the optimization problem in (3) (4) (5). Moreover, since APPS is a search-based optimization approach requiring no derivative information, by controlling the minimum search step length \( \Delta_m \) associated with variables \( \{X_1, \ldots, X_N\} \) and \( \{Y_1, \ldots, Y_N\} \) can be kept as integers by setting their \( \Delta_m = 1 \) while \( V_{in} \) can be treated as continuous number with a much less minimum step length (e.g. \( \Delta_m = 0.05 \)).

In order to utilize APPS, the optimization formulation are revised.

\[
\min f(x) = 1 - \varepsilon + P(V_d) + P(XY) \tag{7a}
\]

\[
s.t. \quad 1 \leq \{X_1, \ldots, X_N\}, 1 \leq \{Y_1, \ldots, Y_N\}, 0 \leq V_{in} \tag{7b}
\]

where \( P(V_d) \) and \( P(XY) \) are the penalty functions for voltage droop constraint (4) and overhead constraint (5).

\[
P(V_d) = \begin{cases} \alpha(V_k^d - V_k^m)^2, & \exists i : V_k^d > V_k^m \\ 0, & \text{others} \end{cases} \tag{8}
\]

\[
P(XY) = \begin{cases} \beta \left( \sum_{k=1}^{N} X_k Y_k - M \right)^2, & \exists i : X_k Y_k > M \\ 0, & \sum_{k=1}^{N} X_k Y_k \leq M \end{cases} \tag{9}
\]

where \( \alpha \) and \( \beta \) are two large coefficients.

6. EXPERIMENTAL RESULTS

The PDN simulator GSim has been implemented in CUDA [14] and C++, respectively. The optimization is carried out using APPSPACK 5.0.1. The GPU program is executed on a single GPU of the NVIDIA GeForce 9800 GX2 card (including two GPUs), with a total on board memory of 512Mb. All the C++ programs are executed on a workstation with Intel Xeon CPU@2.33GHz and 4G RAM running 64-bit Linux OS.

6.1 Gsim for PDN Simulation

Table 1: Transient simulation runtimes, and numbers of iterations of Gsim for PDNs with on-chip LDOs. 1200 time steps are simulated. CPU: the percentage of runtime spent on CPU. The runtime is in seconds.

<table>
<thead>
<tr>
<th>Num. Nodes</th>
<th>Num. LDOs</th>
<th>Runtime</th>
<th>Num. Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.25M</td>
<td>46</td>
<td>21</td>
<td>224</td>
</tr>
<tr>
<td>2.25M</td>
<td>144</td>
<td>1.5</td>
<td>224</td>
</tr>
<tr>
<td>2.25M</td>
<td>96</td>
<td>1.5</td>
<td>224</td>
</tr>
<tr>
<td>9M</td>
<td>250</td>
<td>2.5</td>
<td>1900</td>
</tr>
<tr>
<td>9M</td>
<td>550</td>
<td>2.5</td>
<td>1900</td>
</tr>
</tbody>
</table>

The transient simulation runtime is examined by analyzing PDNs with 2.25M or 9M on-chip nodes. Each PDN has a different number of LDOs. The results are shown in Table 1. Here, in an iteration, all the circuit blocks are individually solved once. For PDNs with more LDOs, since the voltage changes are smoother than the PDNs with less LDOs, they require less iterations to converge. But all 4 cases can converge in less than an average of 3 iterations per time step. As can be seen, the cost of analyzing on-chip power grids is dominant, and the overhead introduced by simulating LDOs is not significant because of the proposed partitioning scheme. Therefore, putting on-chip power grid simulation to fast GPU engine is extremely effective and speeds up the entire simulation process visibly. In summary, Gsim is an accurate and efficient solver to tackle multi-million-node PDNs with on-chip LDO regulators.
6.2 Optimization for On-Chip Voltage Regulation with Multiple Power Domains

The optimization scheme presented in Section 5 is used to design on-chip voltage regulations for two test circuits with multiple power domains. These two circuits have the same total area and same power domains, A, B, and C. But each circuit has a different area for each domain. Domain A has large current density and $V_{dd}$ level representing high-voltage high-current domain. Domain B has nominal current density and $V_{dd}$ level. Domain C has low current density as well as low $V_{dd}$ level. Here, current density mimics the corresponding clock speed. Circuit 1 is a low-voltage dominant circuit, in which most of the area is taken by domain C, whereas circuit 2 is a high-voltage dominant circuit since domain A takes most of its area. Initial numbers of LDOs and input voltages are picked such that the voltage droop and overhead constraints are both satisfied for these two circuits. The experimental parameter setups are shown in Table 2.

**Table 2: Experimental Parameters. Area is in unit area.**

<table>
<thead>
<tr>
<th>Domain</th>
<th>$J$ (in A/unit area)</th>
<th>$V_{dd}$</th>
<th>$V_{m}$</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2/3</td>
<td>1.4</td>
<td>70</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.1</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>1.0</td>
<td>50</td>
<td>6</td>
</tr>
</tbody>
</table>

**Table 3: Optimal on-chip voltage regulation designs for two test circuits.**

<table>
<thead>
<tr>
<th>TC</th>
<th>M</th>
<th>Num. LDOs</th>
<th>$V_{dd}$</th>
<th>$V_{m}$</th>
<th>$V_{in}$</th>
<th>$\epsilon$</th>
<th>Rt</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>50</td>
<td>3</td>
<td>61.8</td>
<td>50.4</td>
<td>45.2</td>
<td>1.45</td>
<td>7/3</td>
</tr>
<tr>
<td>C2</td>
<td>65</td>
<td>5</td>
<td>69.6</td>
<td>59.0</td>
<td>49.6</td>
<td>1.45</td>
<td>7/3</td>
</tr>
</tbody>
</table>

The optimization results are shown in Table 3, and we have following observations:

- In circuit 1, most of the LDOs are placed in domain C. There are two reasons. Firstly, domain C has more current and bigger area than domains A and B, thus requires more LDOs to suppress voltage droop caused by its large load currents. Secondly, since domain C has a tighter voltage droop constraint (50mV), a large number of LDOs are put here to reduce the maximum voltage droop so that the stringent requirement (4) can be satisfied. Therefore, the input voltage and overall power efficiency of the circuit is mostly determined by the LDOs in domain C. Since the $V_{dd}$ level in C is low and the LDOs in this domain have large input-to-output voltage difference, the power efficiency driven optimization tries to lower $V_{in}$ in order to raise up the efficiency. As we can see, the optimized input voltage is 1.45V, which is at the edge of dropout region for 1.4V-output LDOs in domain A.

- For circuit 2, most of the LDOs are placed in domain A since it consumes most of the current. As we can expect, the input voltage and overall efficiency of this circuit is mostly determined by LDOs in domain A. Although lowering input voltage would increase the overall efficiency, when the optimizer pushes the input voltage towards the dropout region of LDOs in A, a significant amount of extra LDOs are required to reduce the maximum voltage droop. This move is prohibited by the overhead constraint. Therefore, the input voltage level is at 1.5V (in the regulation region). The reason that the number of LDOs used is less than the overhead constraint is that having more LDOs in domain A would decrease the power efficiency as shown in Section 4.2.2.

In summary, our optimization scheme is very effective and optimizes the power efficiency for different circuits according to their characteristics.

7. CONCLUSION

By utilizing a fast combined GPU-CPU PDN analysis engine, the significant performance improvement brought by on-chip LDOs in terms of suppressing high-frequency local voltage droops and avoiding mid-frequency resonance are demonstrated. Comprehensive analysis on the trade-offs among overhead of on-chip LDOs, maximum voltage droop and overall power efficiency is performed. We conduct systematic design optimization by developing a simulation-based nonlinear optimization strategy that determines the optimal number of on-chip LDOs and on-board input voltage, to maximize overall power efficiency for PDNs with multiple power domains, while meeting voltage droop and overhead requirements. The results suggest that our simulation engine is very efficient and accurate, and our optimization scheme is effective to meet different design needs.

8. REFERENCES


