Power Delivery Exploration Methodology Based on Constrained Optimization

Rassul Bairamkulov, *Student Member, IEEE*, Kan Xu, *Student Member, IEEE*, Mikhail Popovich, *Member, IEEE*, Juan S. Ochoa, *Member, IEEE*, Vaishnav Srinivas, and Eby G. Friedman, *Fellow, IEEE*

***Abstract*—The conventional power network design process requires iterative modifications to the existing power network to eliminate hot spots and to converge to target impedance parameters. At later stages in the IC design process, this procedure may require significant time and human resources due to the limited flexibility to accommodate necessary changes. Power delivery exploration during early stages of the design process may bring considerable savings to the system development effort. The number of iterations may be greatly reduced by choosing the initial parameters sufficiently close to the optimum. This paper presents a power delivery exploration framework based on constrained global optimization. The power network parameters are estimated at early stages of the development process, while considering both electrical and nonelectrical factors, such as area and cost. A Laplace transform-based circuit simulator is described that is well suited for optimization purposes due to the high computational efficiency when a large number of iterations is required. The proposed framework has been applied to the distribution of voltage domains in a large scale complex integrated system, while minimizing the cost of the decoupling capacitor placement. The optimal number of voltage rails are determined, demonstrating an approximately 40% lower on-chip area than alternative solutions.**

***Index Terms*—Design methodology, design optimization, power quality, power system modeling, system-on-chip, time to market.**

I. Introduction

POWER delivery is pivotal to the performance of modern integrated systems [1]. Violating limitations in power delivery, such as load voltage droop, thermal characteristics, and power dissipation, may cause a variety of issues, such as circuit malfunction or overheating. Due to the high level of complexity in modern systems, it is difficult to monitor power delivery characteristics throughout the system development process. This approach adds risks to the entire development flow. Unsatisfied power quality constraints at later stages of the design process may require unacceptable time and resources.

One strategy for reducing the burden of modifying the power network is overdesign, such as using additional interconnections and pins for power or larger and more numerous decoupling capacitors. This strategy increases cost and allocates less metal and pin resources for signaling, and less area for the functional circuitry [2]. In addition, external factors, such as cooling power or cost, shift the resulting system even farther from the optimal objective.

Numerous works on power delivery optimization at various levels of abstraction exist. At the circuit level, several power regulator models have been proposed. On-chip voltage regulation has increased with minituarization of switching dc–dc converters [3] and on-chip integration of switched capacitor converters [4]. A notable improvement of power regulation is presented in [5], where power delivery in smart phones is improved using a combination of static and dynamic techniques. Integration of on-chip dc–dc regulators has been proposed in [6] and [7] to achieve higher regulation efficiency with smaller area.

Power management has been deeply investigated from an architectural perspective. The work of [8] presents a framework for system-wide dynamic voltage scaling with thermal considerations that improves overconstrained circuits based on worst case scenarios. In [9], the GradualSleep strategy has been proposed to minimize on-chip static energy dissipation. More recent works describe paradigms suitable for modern circuit-level power management solutions. A system-level theoretic framework for optimizing decoupling capacitor and parasitic inductance is proposed in [10]. In [11], a framework for combining switching and linear regulators within a single system is presented that provides high efficiency linear regulators and superior regulation characteristics in switching converters. A system-level power management system is described in [12], where the electrical and thermal characteristics are monitored to make appropriate adaptations, such as dynamic voltage and frequency scaling (DVFS) based on system temperature and workload.

Manuscript received May 17, 2018; revised November 7, 2018 and March 26, 2019; accepted May 18, 2019. Date of publication June 27, 2019; date of current version August 20, 2020. This work was supported in part by the National Science Foundation under Grant CCF-1329374, Grant CCF-1526466, and Grant CCF-1716091, in part by Intelligence Advanced Research Projects Activity under Grant W911NF-14-C-0089, in part by American Institute for Manufacturing Integrated Photonics under Award 059447-007, in part by the Intel Collaborative Research Institute for Computational Intelligence, in part by Cisco Systems, and in part by Qualcomm. This paper was recommended by Associate Editor H. Li. *(Corresponding author: Rassul Bairamkulov.)*

R. Bairamkulov, K. Xu, and E. G. Friedman are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627 USA (e-mail: rbairamk@ur.rochester.edu; kxu8@ur.rochester.edu; friedman@ece.rochester.edu).

M. Popovich is with the Platforms Department, Google Inc., Mountain View, CA 94043 USA (e-mail: mpopovich@google.com).

J. S. Ochoa and V. Srinivas are with QCT, Qualcomm Corporation, San Diego, CA 92121 USA (e-mail: jochoa@qti.qualcomm.com; vaishna@qti.qualcomm.com).

Digital Object Identifier 10.1109/TCAD.2019.2925397

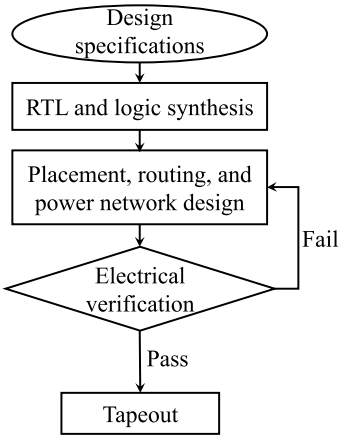


Fig. 1. Conventional IC development process [15].

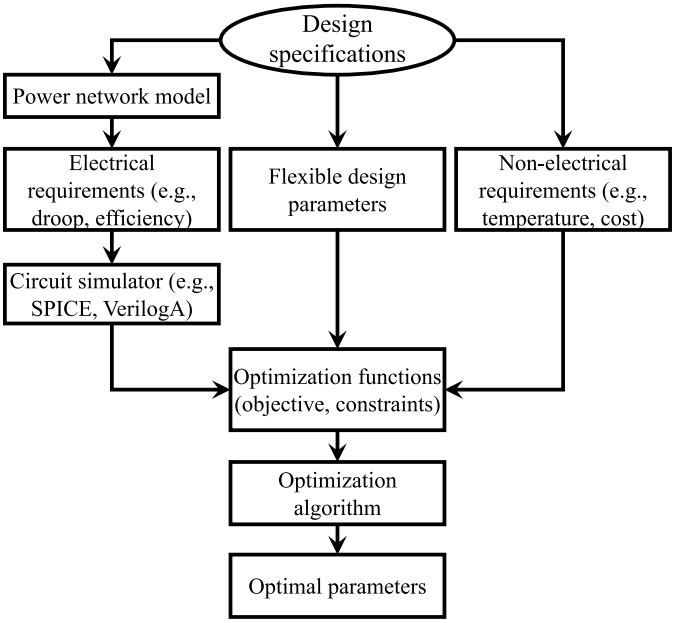


Fig. 2. Proposed power network optimization process.

Despite the maturity of the field, power delivery in VLSI systems is rarely approached from a constrained optimization perspective. In [13], quadratic programming methods are exploited to reduce the impedance profile of the power delivery network at frequencies of interest by sacrificing the impedance at less relevant frequencies. More recent work [14] utilizes differential evolutionary optimization to suggest the impedance profile of a physical structure. Constrained global optimization provides a natural framework for design exploration of power delivery systems. The primary strength of the technique is flexibility, allowing different design objectives and constraints to be considered, including thermal and cost parameters. The subsequent sections provide a deeper insight into this proposed methodology. In Section II, the necessary components of the proposed framework are described. Several case studies are provided in Section III to demonstrate the validity of the proposed approach. The strengths and limitations of the proposed methodology are discussed in Section IV, concluded by a summary in Section V.

II. Proposed Optimization Framework

The standard design process in the absence of power network design exploration is shown in Fig. 1 [15]. Due to the lack of preliminary information, power delivery network analysis is performed during the placement and routing stage [15]. If the circuit does not comply with power quality and voltage droop objectives, the power network is changed or resynthesized. The verification and redesign processes repeat until the resulting power network satisfies the required specifications. Due to the significant time required to evaluate and refine the power delivery network at the system level, multiple design iterations at later stages of the development process are highly undesirable, as these changes may cause delays on the order of days.

To mitigate potential losses, the number of power network redesigns needs to be minimized, preferably to zero. Power delivery exploration can provide valuable guidelines for power network synthesis, bringing the resulting system close to the optimal state. Two important characteristics of the early design stages are worth noting. First, the lack of accurate electrical data creates a high degree of uncertainty in the power network development process. The assumptions made at this stage are crucial. Second, before the primary design parameters are fixed, a high degree of flexibility exists. For example, the number of voltage domains may significantly affect the efficiency of the system at the expense of additional metal resources or increased power noise. Exploiting these tradeoffs is crucial to unlocking the full potential of the overall power delivery system.

The proposed power delivery exploration process is illustrated in Fig. 2. The process is general and varies greatly with different inputs. The process starts with the analysis of the design specifications. A model of the power network is used to estimate the electrical metrics. Nonelectrical metrics of interest are also identified and certain design flexibilities are identified. After the required components are characterized, the functions are passed to optimization algorithms. The result of the optimization process is a set of design guidelines that ensure proper operation without excessive overdesign. A more detailed explanation of the proposed exploration process is provided in the following sections.

A. Specification of the Electrical Design Requirements

A model of the power delivery network consists of four components: 1) topology; 2) voltage sources; 3) load currents; and 4) impedances. The topology reflects the relative placement of the elements within the netlist, supporting a comprehensive circuit analysis process. Technology information, such as the number of metal layers or interconnect conductivity, and design specifications, such as the interconnect dimensions, determine the parameters of the power network model. One of the simplest and most widespread power network models is the hierarchical model shown in Fig. 3 [1], composed of cascaded lumped sections consisting of series RL segments, representing the interconnects and solder bumps, interleaved with parallel RLC segments, representing the decoupling capacitors, with an equivalent series resistance and inductance.

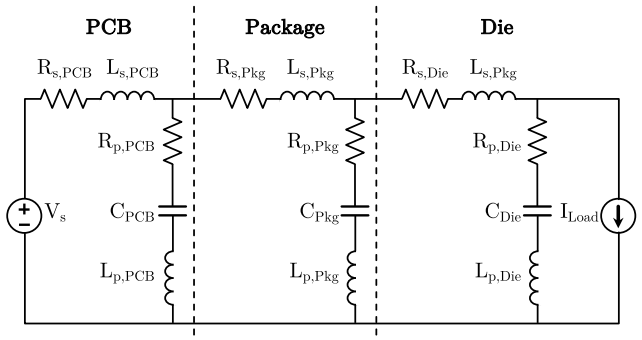


Fig. 3. Simplified model of power delivery network for optimization purposes.

More advanced topologies are necessary to evaluate the information from lower abstraction levels, such as the on-chip mesh. However, due to the lack of topology information during the early design phase, the development of a more accurate circuit model of a power network is a complex task.

The voltage source represents an idealized on-board regulator. For simplicity, a constant voltage supply is assumed. The main source of power consumption is modeled as a current source, representing the current delivered to the functional blocks, on-chip regulators, and leakage current. A current profile is necessary to evaluate the reliability of the network. Functional block information is used to model the profile of the load current [16]. Alternatively, the current profile may be modeled as a constant average current with a worst case current pulse [15].

Once the power network model is determined, it is necessary to convert the design goals and technology limitations into a functional form. For example, any limitations on voltage droop can be represented as

$$\text{Droop} = \frac{\min(V_{\text{Load}}(t))}{V_s} \tag{1}$$

where $V_{\text{Load}}(t)$ is the load voltage and V_s is the supply voltage. The power distribution efficiency, in turn, is

$$\text{Eff} = \frac{P_{\text{Load}}}{P_{\text{in}}} \tag{2}$$

where P_{Load} and P_{in} are, respectively, the power dissipated by the current source and the total dissipated power. These specifications are necessary to convert the metrics of interest into the optimization functions.

B. Specification of Nonelectrical Design Requirements

In this paper, the nonelectrical parameters are described as the system characteristics that are not directly inferred from the circuit model of the power network. These nonelectrical parameters include the on-chip temperature, manufacturing cost of the components, and area of the circuit elements. An externally supplied model is required to link the nonelectrical metrics and electric performance of the system. For example, if the mean time to failure (MTTF) is of concern, optimizing MTTF would place an upper limit on the current density and temperature, as shown in [17]

$$\text{MTTF} = \frac{K}{j^n}\exp\left(\frac{E_a}{kT}\right) \tag{3}$$

where K and n are material and process constants, E_a is the activation energy, k is the Boltzmann constant, T is the temperature, and j is the current density. Based on the analysis process, such as the individual currents, combined with external data, such as the wire dimensions, the current density in all of the elements is estimated to minimize this metric given the constraints.

C. Combination of Electrical and Nonelectrical Metrics

The final form of the optimization function is

$$\boldsymbol{x}_{\textbf{opt}} = \min(f(\boldsymbol{x})), \quad \text{subject to} \quad c(\boldsymbol{x}) \leq 0 \tag{4}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}_{\textbf{opt}}$ are the variables and correspond to the optimal parameter vectors, $f(\boldsymbol{x})$ is the function being optimized, and $c(x)$ is a set of constraint functions. The power delivery exploration process is formulated as in (4) to allow the application of constrained optimization algorithms.

The electrical analysis process needs to provide sufficient information to allow the nonelectrical metrics to be evaluated. The comprehensive optimization function requires an expression of the external metrics in terms of the variable parameters, electrical metrics, or both. For example, with adaption of [18], the MTTF of the interconnect segment can be approximated in terms of the interconnect dimensions and current

$$\text{MTTF} = \frac{K_1 W^n H^n}{I_{\text{rms}}^n}\exp\left(\frac{K_2 W^2 H^2}{I_{\text{rms}}^2}\right) \tag{5}$$

where W and H are, respectively, the interconnect width and thickness, I_{rms} is the RMS current through the segment, and K_1, K_2, and n are process and material related constants. Electrical metrics, such as the RMS current through the segment, are evaluated from simulations of the power network. The variable parameters determine the characteristics of the power network model. For example, the dimensional parameters can be used to determine the impedance of the circuit elements. The formulated metrics are combined to create the objective function and set of constraints.

If multiple design objectives exist, a weighted sum of each objective is used to minimize each objective. The resulting formulation is shown in (6) to (9), where V_s is the supply voltage, $W_{s,\text{Die}}$ and $H_{s,\text{Die}}$ are, respectively, the top level interconnect width and thickness, w_1 and w_2 are weight parameters, $A_{\text{int}}(\boldsymbol{x})$ is the total area of the metal expended for the interconnect, and $\text{Droop}_{\text{max}}$ and Eff_{min} are design constraints on, respectively, the voltage droop and efficiency. The objective function is the weighted sum of the MTTF and cost, minimizing both metrics. To be satisfied, both $c_1(\boldsymbol{x})$ and $c_2(\boldsymbol{x})$ need to be greater than or equal to 0, ensuring that the droop is not larger than $\text{Droop}_{\text{max}}$ and the efficiency is not less than Eff_{min}

$$\boldsymbol{x} = \left[V_s, W_{s,\text{Die}}, H_{s,\text{Die}}\right] \tag{6}$$

$$f(\boldsymbol{x}) = \frac{w_1}{\text{MTTF}(\boldsymbol{x})} + w_2 A_{\text{int}}(\boldsymbol{x}) \tag{7}$$

$$c_1(\boldsymbol{x}) = \text{Droop}(\boldsymbol{x}) - \text{Droop}_{\max} \tag{8}$$

$$c_2(\boldsymbol{x}) = \text{Eff}_{\min} - \text{Eff}(\boldsymbol{x}). \tag{9}$$

D. Circuit Simulation Procedure

During the optimization process, the circuit parameters are varied and the corresponding electrical parameters are evaluated. An efficient circuit simulator is the cornerstone of this procedure as the quality and timeliness depend upon the speed and accuracy of the simulator. Two simulation methods are utilized. The first method is commercial HSPICE [19] which requires a special interface with the programming language. The primary advantage of this approach is the versatility of the simulator. With the variety of available models, a wide range of circuits can be simulated and, therefore, optimized. The disadvantage of this approach is the communication overhead between the programming language and HSPICE which dramatically increases the simulation time.

Another approach is a custom Laplace transform-based simulator, requiring no interface with the programming language. The Laplace transform is widely used for simulation and optimization of linear circuits and systems [20], [21]. The primary advantage of this approach is the higher speed of the simulation due to the lack of communication with an external language and application-specific code optimization. A significant limitation is the narrow applicability of the method - only linear systems can be simulated using this approach due to the Laplace transform. A variety of methods exist, however, to extend the Laplace transform to nonlinear circuits. In [20], the switching transistors are replaced with lumped RC elements. A piecewise-linear model is another common approach for applying Laplace transforms to nonlinear systems. This method is particularly compatible with sequential switching [22], [23]. A modification of the Laplace transform applicable to a certain class of nonlinear systems is introduced in [24]. Incorporating this method into the proposed framework may significantly extend the applicability of the proposed tool.

The proposed optimizer is applied to a model of a power network, which typically consists of passive RL-RLC branches [1]. The active devices, such as a voltage regulator or load transistors, are replaced with equivalent linear models to offset the error due to the assumption of linearity, which enables the use of a Laplace transform-based optimizer. In cases where the power network model is nonlinear (e.g., a power gated network), typically slower, numerical simulation tools can be utilized, such as HSPICE [19] or Verilog-AMS [25]. The choice between an active and passive power network model, therefore, becomes a tradeoff between accuracy and computational speed.

The Laplace transform-based process is shown in Fig. 4. The circuit elements are represented in the s domain. The fixed parameters are expressed numerically, while the variables are represented as symbolic variables. For instance, the impedance of a capacitor with a variable capacitance, fixed equivalent series resistance of 1 mΩ, and fixed equivalent series inductance of 10 pH can be presented as

$$Z_c = 1\ \text{m}\Omega + 10\ \text{pH} \times s + \frac{1}{Cs} \tag{10}$$

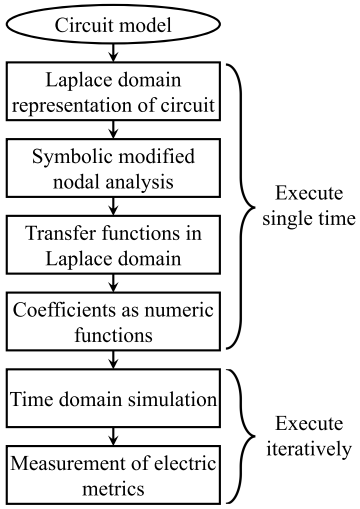


Fig. 4. Proposed Laplace transform-based optimization process.

where the capacitance C is shown as a symbolic variable, Z_c is the equivalent impedance of the capacitor, and s is the Laplace domain parameter.

After the circuit elements are expressed in the Laplace domain, a modified nodal analysis is applied. The circuit is modeled in terms of six input matrices, representing connections and parameter values, as shown in [26]

$$\begin{bmatrix} \boldsymbol{Y} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{V} \\ \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{F} \end{bmatrix} \tag{11}$$

where $\boldsymbol{V}$ and $\boldsymbol{I}$ are, respectively, the node voltages and currents through the voltage sources, $\boldsymbol{Y}$ is the matrix of nodal admittances, while $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{J}$, and $\boldsymbol{F}$ encode current and voltage sources, including controlled sources. The constructed matrix equation is solved for $[\boldsymbol{V}, \boldsymbol{I}]^T$ using left matrix division.

The resulting vector represents the node voltages and source currents in terms of symbolic parameters in the Laplace domain. Dividing the resulting vectors by the source produces the transfer function, as shown in

$$H(s) = \frac{b_n s^n + \cdots + b_0}{a_m s^m + \cdots + a_0}. \tag{12}$$

The coefficients of the transfer function are expressed as a function of the variable parameters

$$b_i = f_{i,\text{num}}(\boldsymbol{x}) \tag{13}$$

$$a_i = f_{i,\text{den}}(\boldsymbol{x}). \tag{14}$$

While the aforementioned procedure is computationally expensive, requiring a solution of the symbolic matrix system, the process only needs to be performed once for a particular circuit topology. Modifications of the variable parameters only change the value of the coefficients, $b_n \ldots b_0\ a_n \ldots a_0$, while the symbolic representation remains intact. The speedup due to the proposed simulator is, therefore, largely dependent upon the number of iterations N during the optimization process. The speedup is estimated as

$$\text{Speedup} = \frac{t_n}{\frac{t_{\text{setup}}}{N} + t_{\mathcal{L}}} \tag{15}$$

TABLE I
VOLTAGE DOMAIN SPECIFICATIONS OF POWER DELIVERY NETWORK ADAPTED FROM [29]

	Rail #	Voltage max, V	Voltage min, V	Current max, mA	Current min, mA	Peak slew rate, A/μs	Function
A	A1 to A4	1.42	0.97	5,830	416	1,000	CPU core
	A5	1.20	0.99	3,150	225	500	GPU
	A6	1.33	1.00	10	1	500	USB
	A7	1.93	1.67	10	1	500	GPS
	A8	1.93	1.72	30	1	500	DSP
	A9	1.93	1.67	10	1	500	Camera
	A10	1.93	1.67	10	1	500	Audio
	A11	1.93	1.67	1,500	58	500	LTE+WiFi
	A12	1.55	1.00	3,150	225	500	Memory
B	B1 to B4	1.42	0.97	5,830	416	1,000	CPU core
	B5	1.20	1.00	3,160	226	*	GPU+USB
	B6	1.93	1.67	1,500	58	500	LTE+WiFi
	B7	1.93	1.72	60	4	*	GPS+DSP+Camera+Audio
	B8	1.55	1.00	3,150	225	500	Memory
C	C1	1.42	1.00	26,470	1,889	*	CPU+Memory
	C2	1.20	1.00	3,160	226	*	GPU+USB
	C3	1.93	1.72	1,560	62	*	GPS+DSP+Camera+Audio+LTE+WiFi

where t_n and $t_{\mathcal{L}}$ are the time per iteration using, respectively, numerical analysis and the Laplace transform-based simulator, and t_{setup} is the time required to determine the transfer function (12). Note that typically $t_{\text{setup}} > t_n > t_{\mathcal{L}}$, thus the speedup converges to a positive value with large N, while approaching zero with small N. Since most optimization procedures require a large number of iterations to determine the global minimum, the creation of a symbolic transfer function represents a negligible fraction of the total computational time.

To simulate the transfer functions and extract the numeric data, the coefficients of the transfer functions of interest are calculated and converted into a state space model. A variety of efficient state space model simulation packages are available, such as LAPACK [27] and LTITR [28]. The input waveform and state space model are passed to the simulators to calculate the output waveform. This approach achieves significant speedup as compared to conventional, purely numerical algorithms. Applying a state-space model, the output waveform can be determined without solving the matrix equation during each time step. Conversion of a circuit into a matrix form is performed only once, greatly reducing the computational overhead. With the large number of circuit simulations during the optimization process, significant optimization speedup is achieved, as described in Section III.

III. CASE STUDY

The problem of choosing the optimal number of rails is an important power delivery exploration issue. Utilizing several voltage domains may bring considerable savings in terms of power, while achieving performance goals [30]. At early stages of the design process, planning the circuit topology is problematic since the resulting power delivery characteristics are difficult to estimate in advance. In particular, it is unclear whether the power network is sufficiently conductive to satisfy voltage droop requirements. Separation of the low voltage circuitry from the rest of the IC is an attractive option to reduce power consumption due to the quadratic relationship between power consumption and operating voltage. The scaled voltage is, however, less robust to sudden load current fluctuations, possibly violating droop requirements, allowing the device to malfunction. Moreover, utilizing separate power networks requires less metal resources for each rail, resulting in a power delivery network exhibiting higher impedance.

TABLE II
PARAMETERS OF DECOUPLING CAPACITOR COST [31]

	Die	Package	PCB
Cost per m^2 [× \$1,000]	200	8	0.5
Insulator thickness [μm]	0.7	10	50
Relative permittivity	3.9	4.6	4.5
Cost per capacitance [\$/nF]	4.06	1.97	0.63

To investigate this problem, three power networks are considered, 12 rail (A), 8 rail (B), and 3 rail (C) systems. The impedance characteristics of these networks are based on [29] and assume the power network topology shown in Fig. 3. The rail specifications are listed in Table I. The maximum and minimum voltages represent the range of allowed values of the voltage. The model of the load current is a worst case triangular current waveform [10].

System B merges the rails with the closest voltage levels to minimize energy losses due to the voltage conversion process. Rail A5 is merged with rail A6 to produce rail B5, and rails A7 through A10 are merged into rail B7, resulting in the eight rail system B. Further, rails B1 to B4 and B8 are merged, while rail B6 is merged with rail B7 to produce the three rail system C. The variables are the voltage supply of each rail, as well as the decoupling capacitance at each level of each rail. For simplicity, the power rails are assumed to be mutually isolated, allowing each rail to be evaluated separately.

The objective of the design exploration process is to determine the set of rails that delivers the lowest possible cost of decoupling capacitance area. In this case study, the proxy metric for the decoupling capacitance cost is the weighted sum of the decoupling capacitor area, as described by

$$\text{Cost} = \sum_{i \in S} w_i A_i \tag{16}$$

where S is the set of layers of the power network [printed circuit board (PCB), package, or die], and w_i and A_i are, respectively, the cost per unit area and the area of the decoupling capacitors on level i. In this objective function, both w_i and A_i are external parameters related to the electrical metrics. The area of a decoupling capacitor can be expressed as

$$A = \frac{Cd}{\varepsilon} \tag{17}$$

where C is the capacitance, and d and ε are, respectively, the thickness and electrical permittivity of the insulating material. Inserting (17) into (16) leads to the final form of the objective function

$$\text{Cost} = \sum_{i \in S} \frac{C_i d_i w_i}{\varepsilon_i}. \tag{18}$$

To estimate the relative cost of the unit area of each level, certain trends in Table II can be noted. The cost per unit area increases when approaching the die level. The opposite trend is noted as the insulator thickness becomes thinner as the die is approached. Overall, the cost per unit capacitance follows the unit area cost pattern, making on-die decoupling capacitor placement more than six times costlier than on the PCB. During the optimization process, this distribution of capacitance costs discourages placing a large capacitance on the die and tolerates placing more capacitance farther from the IC.

Moving the decoupling capacitance farther from the load makes the system more vulnerable to inductive noise [32], limiting the cost benefits of a small on-chip capacitance. The greater fluctuations in the load voltage result in a need for a higher voltage supply to offset the potential voltage droops, resulting in higher power consumption. In addition, the inductive system response may result in significant overshoots that may damage the transistors. In the proposed framework, the aforementioned tradeoffs are expressed as constraint functions, as shown in

$$c_1\left(V_s, C_{PCB}, C_{Pkg}, C_{\text{Die}}\right) = V_{\text{load,min}} - \min(V_{\text{load}}(t)) \tag{19}$$

$$c_2\left(V_s, C_{PCB}, C_{Pkg}, C_{\text{Die}}\right) = \max(V_{\text{load}}(t)) - V_{\text{load,max}} \tag{20}$$

$$c_3\left(V_s, C_{PCB}, C_{Pkg}, C_{\text{Die}}\right) = \text{Power}_{\text{total}} - \text{Power}_{\text{max}} \tag{21}$$

where $V_{\text{load}}(t)$ is the waveform of the load voltage, $V_{\text{load,min}}$ and $V_{\text{load,max}}$ are, respectively, the minimum and maximum bounds on the load voltage, and $\text{Power}_{\text{total}}$ and $\text{Power}_{\text{max}}$ are, respectively, the total power consumption and upper limit on the consumed power. The constraint functions place strict requirements on the quality of the power rails. If the voltage waveform violates the constraint functions, the objective function (or cost) is severely penalized, invalidating the result.

The power network model used in this case study does not include any nonlinear elements. A Laplace transform-based simulator has therefore been chosen. Particle swarm optimization is chosen as the optimization algorithm due to the robustness and efficiency characteristics of this algorithm. The optimization procedure is run on an eight core 3.40-GHz Intel Core i7-6700 machine. The results for 23 separate rail configurations are obtained in 26 minutes, with an average time of 67 s per rail. The results of the optimization are shown in Fig. 5. Note that the lowest value of the objective function is achieved with eight rails. In the 8 rail and 12 rail scenarios, certain rails (e.g., rails 7 to 11 in the 12 rail scenario) do not require decoupling capacitors due to the low load currents and high tolerance to variations.

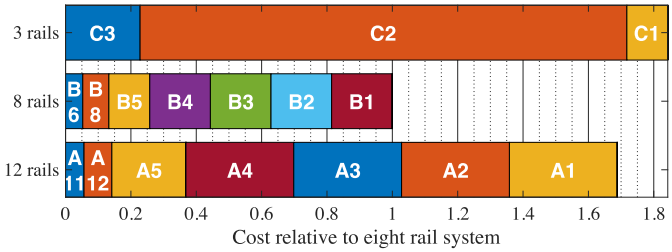


Fig. 5. Decoupling capacitor placement for three power delivery networks.

To evaluate the benefits of the Laplace transform optimization process, a similar optimization is performed using HSPICE [19]. The optimization results are identical to those results obtained from the Laplace transform optimization process due to the absence of nonlinear elements in the model. The total computational time, however, is 265 minutes, ten times greater than the Laplace simulator.

IV. Discussion

Distribution of the decoupling capacitor costs across the voltage domains normalized to the least expensive system is shown in Fig. 5. Certain patterns can be inferred. Comparing the 8 and 12 rail systems, allocation of metal resources for separate power rails is unjustified from a cost perspective. The higher contribution of the CPU cores (A1 to A4) in the 12 rail network indicates that voltage fluctuations in this network are greater due to less metal resources allocated to each CPU rail, as compared to the eight rail system. The combination of rails A5 and A6 allocates more metal resources for both networks, resulting in reduced decoupling capacitor cost in combined rail B5.

As compared to the three rail system, where rails B1 to B4 and B8 (CPU cores and memory) are merged into a single voltage domain, the three rail system requires a large decoupling capacitance for the combined rail C2. The reason for the increased decoupling capacitance is the poor compatibility between voltage ranges. While rails B1 to B4 require a range of 0.97 to 1.42 V, rail B8 has a range of 1.00 to 1.55 V. The combined rail, therefore, needs to satisfy both ranges and is effectively shrunk to 1.00 to 1.42 V, placing greater limitations on the voltage fluctuations. The narrow voltage range is compensated by placing a larger on-chip decoupling capacitance, increasing the overall cost of the power network.

A conventional power network design process may require a series of late design backtracking iterations to satisfy target noise performance requirements [33], [34]. Assuming that the post-floorplan power network model requires time t_{sim} for simulation and t_{correct} for hotspot correction, and N iterations are required to reach the acceptable characteristics, the total time for the power integrity analysis process without early exploration is

$$t_{noEE} = (N-1)t_{\text{sim}} + Nt_{\text{correct}} \tag{22}$$

where, typically, t_{sim} and t_{correct} are on the order of hours and days, and N typically ranges between two and ten iterations. Alternatively, early power delivery exploration requires time t_{exp}, which may require several hours to complete. An expected result of the power delivery exploration process is a significant reduction in the number of iterations. Assuming the updated number of iterations is N_{new}, the total time for the power integrity analysis process is

$$t_{EE} = t_{\text{exp}} + (N_{\text{new}} - 1)t_{\text{sim}} + N_{\text{new}}t_{\text{correct}}. \tag{23}$$

The savings in time due to the early power integrity analysis process is

$$t_{noEE} - t_{EE} = (N - N_{\text{new}})(t_{\text{sim}} + t_{\text{correct}}) - t_{\text{exp}} \tag{24}$$

therefore, to ensure that the power delivery exploration is justified from the perspective of computational time, the following condition must be satisfied:

$$(N - N_{\text{new}})(t_{\text{sim}} + t_{\text{correct}}) > t_{\text{exp}}. \tag{25}$$

Note that typically $t_{\text{sim}} + t_{\text{correct}} > t_{\text{exp}}$, therefore, to justify early design exploration, it is sufficient to reduce the number of post-floorplan backtracking iterations, i.e., $N_{\text{new}} < N$.

The proposed early power delivery exploration framework may reduce the number of costly iterations by providing an estimate of the optimal parameters at an earlier phase of the development process, shrinking both time and labor. The nonelectrical parameters, such as area and cost, are combined with the electrical parameters to produce a system with minimum cost while satisfying target performance metrics. This approach provides useful information for early system exploration, allowing more effective design decisions to be made.

Several limitations of the proposed framework exist. First, the computational time largely depends upon the circuit simulator. Therefore, optimization of more complex circuits with a larger number of nodes may require significant computational time. A Laplace transform-based simulator is proposed for optimization of linear circuits. The speedup due to the Laplace transform-based simulator, however, largely depends upon the number of iterations during the optimization process. Second, a function for the metrics of interest needs to be determined to conduct the power delivery exploration process. Practical assumptions, therefore, need to be made to achieve useful results. An issue of premature convergence exists, resulting in the optimization converging to a local minimum rather than a global minimum [35]. It is, therefore, necessary to ensure that the design space is thoroughly explored, for example, by increasing population sizes (evolutionary algorithms), mutation and migration rates (genetic algorithm), swarm velocities and inertia (particle swarm), and the initial temperature and frequency of reheating (simulated annealing).

V. Conclusion

A versatile methodology for power delivery design exploration is described in this paper. The primary strength of the framework is applicability to a wide range of objectives and constraints, including external, nonelectrical parameters. The procedure supports the application of robust, general purpose algorithms to solve power delivery problems. A fast, optimization oriented Laplace transform-based simulator is described. The limitations of the proposed framework include the dependence on the computational time of the circuit simulator, the need for optimization functions during the preliminary design stages, and careful tuning of the optimization algorithms. The effectiveness of the framework is demonstrated by a case study, where the appropriate power delivery network is chosen among existing options.

References

[1] I. P. Vaisband, R. Jakushokas, M. Popovich, A. V. Mezhiba, S. Köse, and E. G. Friedman, *On-Chip Power Delivery and Management*. Cham, Switzerland: Springer, 2016.

[2] K. Xu, R. Patel, P. Raghavan, and E. G. Friedman, "Exploratory design of on-chip power delivery for 14, 10, and 7 nm and beyond FinFET ICs," *Integration*, vol. 61, pp. 11–19, Mar. 2018.

[3] C. F. Lee and P. K. T. Mok, "A monolithic current-mode CMOS DC–DC converter with on-chip current-sensing technique," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 3–14, Jan. 2004.

[4] G. Patounakis, Y. W. Li, and K. L. Shepard, "A fully integrated on-chip DC–DC conversion and power management system," *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 443–451, Mar. 2004.

[5] W. Lee, Y. Wang, D. Shin, N. Chang, and M. Pedram, "Optimizing the power delivery network in a smartphone platform," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 1, pp. 36–49, Jan. 2014.

[6] V. Kursun, S. G. Narendra, V. K. De, and E. G. Friedman, "Low-voltage-swing monolithic DC–DC conversion," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 51, no. 5, pp. 241–248, May 2004.

[7] W. Kim, D. M. Brooks, and G.-Y. Wei, "A fully integrated 3-level DC/DC converter for nanosecond-scale DVS with fast shunt regulation," in *Proc. IEEE Int. Solid-State Circuits Conf.*, San Francisco, CA, USA, Feb. 2011, pp. 268–270.

[8] W. Liao, L. He, and K. M. Lepak, "Temperature and supply voltage-aware performance and power modeling at microarchitecture level," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 7, pp. 1042–1053, Jul. 2005.

[9] S. Dropsho, V. Kursun, D. H. Albonesi, S. Dwarkadas, and E. G. Friedman, "Managing static leakage energy in microprocessor functional units," in *Proc. IEEE/ACM Int. Symp. Microarchit.*, Feb. 2002, pp. 321–332.

[10] E. Salman, E. G. Friedman, R. M. Secareanu, and O. L. Hartin, "Worst case power/ground noise estimation using an equivalent transition time for resonance," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 5, pp. 997–1004, May 2009.

[11] I. Vaisband and E. G. Friedman, "Heterogeneous methodology for energy efficient distribution of on-chip power supplies," *IEEE Trans. Power Electron.*, vol. 28, no. 9, pp. 4267–4280, Sep. 2013.

[12] E. Rotem, A. Naveh, A. Ananthakrishnan, E. Weissmann, and D. Rajwan, "Power-management architecture of the Intel microarchitecture code-named sandy bridge," *IEEE Micro*, vol. 32, no. 2, pp. 20–27, Mar./Apr. 2012.

[13] A. E. Engin, "Efficient sensitivity calculations for optimization of power delivery network impedance," *IEEE Trans. Electromagn. Compat.*, vol. 52, no. 2, pp. 332–339, May 2010.

[14] A. Orlandi, "Differential evolutionary multiple-objective sequential optimization of a power delivery network," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 3, pp. 754–760, Jun. 2018.

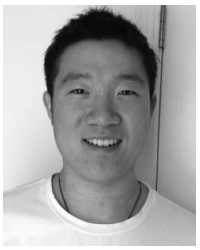
[15] B. Ko, J. Kim, J. Ryoo, C. Hwang, J. Song, and S. W. Kim, "Simplified chip power modeling methodology without netlist information in early stage of SoC design process," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 6, no. 10, pp. 1513–1521, Oct. 2016.

[16] S. Köse, E. G. Friedman, R. M. Secareanu, and O. Hartin, "Current profile of a microcontroller to determine electromagnetic emissions," in *Proc. IEEE Int. Symp. Circuits Syst.*, Beijing, China, May 2013, pp. 2650–2653.

[17] X. Huang, T. Yu, V. Sukharev, and S. X.-D. Tan, "Physics-based electromigration assessment for power grid networks," in *Proc. ACM/IEEE Design Autom. Conf.*, San Francisco, CA, USA, pp. 1–6, Jun. 2014.
[18] P. Salome, C. Leroux, P. Crevel, and J. P. Chante, "Investigations on the thermal behavior of interconnects under ESD transients using a simplified thermal RC network," in *Proc. Elect. Overstress Electrostatic Discharge Symp.*, Oct. 1998, pp. 187–198.
[19] *HSPICE Quick Reference*, Synopsys, Mountain View, CA, USA, Mar. 2017.
[20] H. Su, K. H. Gala, and S. S. Sapatnekar, "Analysis and optimization of structured power/ground networks," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 22, no. 11, pp. 1533–1544, Nov. 2003.
[21] A. J. Fleming, S. Behrens, and S. O. R. Moheimani, "Optimization and implementation of multimode piezoelectric shunt damping systems," *IEEE/ASME Trans. Mechatronics*, vol. 7, no. 1, pp. 87–94, Mar. 2002.
[22] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 22, no. 2, pp. 155–170, Feb. 2003.
[23] P. Moreno and A. Ramirez, "Implementation of the numerical laplace transform: A review," *IEEE Trans. Power Del.*, vol. 23, no. 4, pp. 2599–2609, Oct. 2008.
[24] E. I. Verriest, "Linear systems over the perspective field as a class of nonlinear systems for which a 'Laplace' transform can be defined," in *Proc. IEEE Int. Conf. Control Autom.*, Jun. 2016, pp. 271–276.
[25] *Verilog-AMS Language Reference Manual Analog and Mixed-Signal Extensions to Verilog-HDL*, Accellera, Napa, CA, USA, Jun. 2009.
[26] C.-W. Ho, A. Ruehli, and P. Brennan, "The modified nodal approach to network analysis," *IEEE Trans. Circuits Syst.*, vol. CSI-22, no. 6, pp. 504–509, Jun. 1975.
[27] E. Anderson *et al.*, *LAPACK Users' Guide*. Philadelphia, PA, USA: SIAM, 1999.
[28] *MATLAB Control Systems Toolbox 10.3*, MathWorks, Inc., Natick, MA, USA, Mar. 2018.
[29] *Qualcomm Snapdragon 600E Processor APQ8064E*, Qualcomm Technol., Inc., San Diego, CA, USA, Oct. 2017.
[30] V. Kursun and E. G. Friedman, *Multi-Voltage CMOS Circuit Design*. Chichester, U.K.: Wiley, 2006.
[31] R. Bairamkulov, K. Xu, E. G. Friedman, M. Popovich, J. Ochoa, and V. Srinivas, "Versatile framework for power delivery exploration," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2018, pp. 1–5.
[32] M. Popovich, M. Sotman, A. Kolodny, and E. G. Friedman, "Effective radii of on-chip decoupling capacitors," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 7, pp. 894–907, Jul. 2008.
[33] M. S. Tanaka, M. Toyama, R. Mori, H. Nakashima, M. Haida, and I. Ooshima, "Early stage chip/package/board co-design techniques for system-on-chip," in *Proc. IEEE Conf. Elect. Perform. Elect. Packag. Syst.*, San Jose, CA, USA, Oct. 2011, pp. 21–24.
[34] T.-W. Tseng, C.-T. Lin, C.-H. Lee, Y.-F. Chou, and D.-M. Kwai, "A power delivery network (PDN) engineering change order (ECO) approach for repairing IR-drop failures after the routing stage," in *Proc. Int. Symp. VLSI Design Autom. Test*, Apr. 2014, pp. 1–4.
[35] D. Pham and D. Karaboga, *Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*. London, U.K.: Springer, 2012.

Rassul Bairamkulov (S'17) received the B.Eng. degree in electrical and electronic engineering from Nazarbayev University, Astana, Kazakhstan, in 2016, and the M.S. degree in electrical engineering from the University of Rochester, Rochester, NY, USA, in 2018, where he is currently pursuing the Ph.D. degree under the supervision of Prof. E. G. Friedman.

He was an Intern with Qualcomm Inc., San Diego, CA, USA, in 2018. His current research interests include power delivery network design, electronic design automation, and optimization algorithms in very large scale integration.



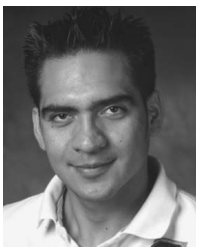
Kan Xu (S'15) received the B.S. degree in electrical engineering from the North China University of Water Resources and Electric Power, Zhengzhou, China, in 2012, and the M.S. degree in electrical and computer engineering from the University of Rochester, Rochester, NY, USA, in 2014, where he is currently pursuing the Ph.D. degree in electrical engineering under the supervision of Prof. E. G. Friedman.

His current research interests include on-chip and package level power delivery network, high current HPC systems, 3-D integration, and optical waveguides.

Mikhail Popovich (M'08) received the B.S. degree in electrical engineering from Izhevsk State Technical University, Izhevsk, Russia, in 1998, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Rochester, Rochester, NY, USA, in 2002 and 2007, respectively.

He was an Intern with Freescale Semiconductor, Inc., Tempe, AZ, USA, in 2005, where he researched on signal integrity in RF and mixed-signal ICs and developed design techniques for placing distributed on-chip decoupling capacitors. In 2007, he joined Qualcomm Corporation, San Diego, CA, USA, as a Senior Engineer. In more than a decade with Qualcomm Inc., he rose to Principal Engineer, managing low power implementation team with focus on power integrity (from on-die power distribution grids to system level PDN), low power design techniques, and power/performance/cost tradeoffs in Snapdragon SOCs. In 2018, he joined Google Inc., Mountain View, CA, USA, as a Power Integrity Engineer researching on system level PDN design. He holds over a dozen U.S. patents. His current research interests include noise, signal integrity, interconnect design, including on-chip inductive effects, optimization of power distribution networks, novel circuits and techniques for low power, and power/cost/performance tradeoffs.

Dr. Popovich was a recipient of the Best Student Paper Award at the ACM Great Lake Symposium on VLSI in 2005 and the GRC Inventor Recognition Award from the Semiconductor Research Corporation in 2007.



Juan S. Ochoa (S'13–M'13) received the B.S. degrees (*summa cum laude*) in physics and in electrical engineering from the Universidad San Francisco de Quito, Quito, Ecuador, in 2008, and the M.S. degree in electrical engineering and the Ph.D. degree in electrical engineering (in the field of computational electromagnetics) from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 2010 and 2013, respectively.

He is currently with Qualcomm Inc., San Diego, CA, USA, as a Senior Power Integrity Engineer. His current research interests include computational electromagnetics, power integrity, statistical techniques, stochastic model-order reduction methods, sparse grid interpolation, electromagnetic scattering, and disordered periodic structures.

Dr. Ochoa was a recipient of the Prof. Kung Chie Yeh Fellowship of the University of Illinois at Urbana–Champaign in 2012–2013 for outstanding scholastic record and contributions in wave propagation.

Vaishnav Srinivas received the B.Tech. degree from the Indian Institute of Technology Madras, Chennai, India, in 2000, and the M.S. degree in electrical engineering from the University of California at Los Angeles, Los Angeles, CA, USA, in 2002.

He is a Senior Director of engineering with Qualcomm, San Diego, CA, USA, researching on low-power high-speed interface architecture, power and signal integrity, and circuit-system co-design. His current research interests include roadmapping interface and PDN technology and design, advanced PPA modeling techniques, die/PKG/PCB co-design and hardware/firmware co-design and validation.

Eby G. Friedman (F'00) received the B.S. degree in electrical engineering from Lafayette College, Easton, PA, USA, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Irvine, Irvine, CA, USA, in 1981 and 1989, respectively.

He was with Hughes Aircraft Company, Los Angeles, CA, USA, from 1979 to 1991, rising to Manager of the Signal Processing Design and Test Department, where he was responsible for the design and test of high performance digital and analog ICs. He has been with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA, since 1991, where he is a Distinguished Professor and the Director of the High Performance VLSI/IC Design and Analysis Laboratory. He is also a Visiting Professor with the Technion—Israel Institute of Technology, Haifa, Israel. He has authored over 500 papers and book chapters, 19 patents, and the authored or edited 18 books in the fields of high-speed and low-power CMOS design techniques, 3-D design methodologies, high speed interconnect, and the theory and application of synchronous clock and power distribution networks. His current research and teaching interests include high performance synchronous digital and mixed-signal microelectronic design and analysis with application to high speed portable processors, low power wireless communications, and server farms.

Dr. Friedman was a recipient of the IEEE Circuits and Systems Mac Van Valkenburg Award, the IEEE Circuits and Systems Charles A. Desoer Technical Achievement Award, the University of Rochester Graduate Teaching Award, and the College of Engineering Teaching Excellence Award. He is the Editor-in-Chief of the *Microelectronics Journal*, an Editorial Board Member of the *Journal of Low Power Electronics* and the *Journal of Low Power Electronics and Applications*, and a Technical program committee member of numerous conferences. He was the Editor-in-Chief and the Chair of the steering committee of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, the Regional Editor of the *Journal of Circuits, Systems and Computers*, an Editorial Board Member of the PROCEEDINGS OF THE IEEE, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II: ANALOG AND DIGITAL SIGNAL PROCESSING, *Analog Integrated Circuits and Signal Processing*, the IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS, and the *Journal of Signal Processing Systems*, a member of the Circuits and Systems Society Board of Governors, program and technical chair of several IEEE conferences. He is a Senior Fulbright Fellow.