Verification and Co-Design of the Package and Die Power Delivery System Using Wavelets

Imad A. Ferzli, Eli Chiprout, and Farid N. Najm

Abstract— **We introduce a wavelet-based framework to characterize circuit currents and compute worst-case supply/ground voltage fluctuations. This framework is apt at determining the impact of the various stages of the power delivery network, enabling their co-design.**

I. INTRODUCTION

The intense drive toward lower power designs has highlighted the need for robust design of a chip's power delivery network (PDN). The PDN, starting at the voltage regulation module (VRM), through the motherboard, package and finally the on-die power grid, must supply a reliable source of power that is fairly free from fluctuations over time. To determine the PDN response, designers typically model the PDN using either RLC elements or electromagnetic (EM) models. While modeling the PDN system is generally accurate, there remains a significant source of error in modeling the worst-case current draw of the die: The number of state transitions is astronomical and searching the current space for the worst case is a daunting task. Several attempts have been made to address this problem, including a more formal approach [1] that constrains the problem with known design bounds, rather than simulate or search through all possible scenarios.

In this paper we introduce the concept of time-frequency descriptions of die currents using wavelets. Wavelet analysis of early-design die current profiles has been proposed previously in [2] to compute current statistics. In this work we show that wavelets are a natural way to comprehensively characterize die behavior in the time-frequency plane and that this description may be used to extract new relevant bounds that will serve in determining the worst-case current draw. The wavelet framework will allow us to find the worst-case current draw for a given PDN on a systematic and general basis, opening up the possibility of obtaining realistic and non-obvious worst-case die current waveforms that mimic complex circuit behaviors. The wavelet framework combines both the time and frequency domains. Since finding worst-case voltage drop on the PDN has time and frequency dimensions, this framework will give us the best of both worlds. Purely time-domain techniques only deal with the simplest system descriptions, whereas purely frequency-domain methods lack accuracy, especially with state-of-the-art, multiresonant PDN systems.

II. WORST-CASE STIMULUS CHARACTERIZATION USING WAVELETS

Our overarching aim is to use wavelet analysis to *construct* a synthetic worst-case PDN stimulus that maximizes voltage drop at a node of interest on the PDN, and to compute the resulting maximum drop. The node could be physically located on the chip, package or anywhere else on the PDN. We call the stimulus "synthetic" because we do not observe it in actual traces or simulations, but rather, we construct it by optimization to yield the worst-case voltage drop.

Assume the PDN is a linear time-invariant RLC circuit, with v nodes and q current sources. We define a PDN stimulus as a collection of current waveforms $i_1(t),\ldots, i_q(t)$, that simultaneously load the PDN, such that $i_j(t)$ attaches to the jth current source. Wavelet analysis enables us to write an arbitrary waveform $i_j(t)$ as [3]:

$$i_j(t) = i_{dc,j} + \sum_{m=1}^{m_0} \sum_{n=0}^{n_{m,j}} T_{m,n,j} \psi_{m,n}(t) \tag{1}$$

where $i_{dc,j}$ is the DC component and $\psi_{m,n}(t)$ are dilated and time-shifted versions of the same function $\psi(t)$, such that $\psi_{0,0}(t) = \psi(t)$ and $\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n)$. The choice of $\psi(t)$, m_0, and $n_{m,j}$ is dictated by the frequency band of interest and the properties of the PDN system (e.g., decay in the step response). The coefficients $T_{m,n,j}$ and the DC component $i_{dc,j}$ are unknown and serve as variables in an optimization problem that maximizes voltage drop under a set of constraints, as discussed below.

$\psi_{m,n}(t)$ are wavelets, which are basically time-functions satisfying certain finiteness and localization properties in time and in frequency [3]. Intuitively, one could think of m, referred to as the *scale*, as an inverse frequency, and of n, known as the

I. A. Ferzli and F. N. Najm are with the Department of Electrical and Computer Engineering, the University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada (phone: 416-946-5175; fax: 416-946-8734; e-mail: {ferzli, najm}@ eecg.toronto.edu).

E. Chiprout is with the Strategic CAD Labs, Intel Corporation, 2501 NW 229th Avenue, Hillsboro, Oregon 97124, USA (e-mail: eli.chiprout@intel.com).

translation, as a simple time shift. If m is *increased*, the time span of $\psi_{m,n}(t)$ would *increase* while its center frequency *decrease* and its bandwidth *shrink*, in accordance with the Fourier transform time-scaling property: $\mathscr{F}\{\psi(t/a)\} = |a|\,\Psi(af)$, where $\Psi(f)$ is the Fourier transform of $\psi(t)$. Hence the name *time-frequency analysis*.

We illustrate our analysis with the simple Haar wavelet, which is a section of a square wave, defined by $\psi(t) = +1$, for $0 < t < u/2$, -1, for $u/2 < t < u$, and 0 otherwise, with u being a suitably chosen time unit. We note that the analysis to follow can be extended to most other wavelet systems (technically, those having a scaling function [3], as do most wavelets in practice). The choice of u stems from the highest frequency of interest, f_H, and can be shown to be $u = 1.165/(\pi f_H)$. The largest scale m_0 follows from f_H and from the lowest frequency of interest, denoted by f_L, and it can be further shown that $m_0 = \lceil \log_2(2 f_H/f_L) \rceil$. The choice of f_H and f_L in practice is guided by the PDN's impedance plots, commonly available in the early design stages.

Computing the PDN response to wavelets at every scale and time shift is key to choosing the correct duration for $i_j(t)$ and to setting up the voltage-drop-maximizing optimization problem. Let $h_{m,n,j,z}(t)$ be the voltage drop waveform on some node z when the PDN has a single stimulus set to $\psi_{m,n}(t)$ on current source j. We are interested in computing $h_{m,n,j,z}(t), \forall z = 1,\ldots,v, \forall j = 1, \ldots, q, \forall n = 0,\ldots, n_{m,j}$, and $\forall m = 1,\ldots, m_0$. A brute force approach is inefficient and would require simulation of the PDN with a single stimulus on different current loads consisting of wavelets at all scales and translations.

Instead, we leverage the linearity inherent in the wavelet system and the linearity and time-invariance of the PDN to require PDN simulation at only one scale and translation, and we compute all the needed $h_{m,n,j,z}(t)$ from the simulation results by efficient linear operations and time shifts. To achieve this, we introduce a companion function to the wavelet, known as the *scaling function* [3], denoted by $\varphi(t)$. In the Haar system the scaling function $\varphi(t) = \varphi_{0,0}(t)$ is a unit-amplitude, unit-width square pulse starting at $t = 0$ (with $\varphi_{0,n}(t) = \varphi_{0,0}(t-nu)$) which enables transitions from scale m to $(m+1)$ according to $\psi_{m+1,n}(t) = (1/\sqrt{2})(\varphi_{m,2n}(t) - \varphi_{m,2n+1}(t))$ and $\varphi_{m+1,n}(t) = (1/\sqrt{2})(\varphi_{m,2n}(t) + \varphi_{m,2n+1}(t))$ (wavelets other than Haar have analogous expressions). Denote by $s_{m,n,j,z}(t)$ the voltage drop on node z when a stimulus of $\varphi_{m,n}(t)$ is applied on $i_j(t)$. From the linearity of the PDN, it suffices to simulate for $s_{0,0,j,z}(t)$ (or for step responses, since $\varphi(t)$ is a square pulse in the Haar system) to deduce $h_{m,n,j,z}(t)$ as

follows:
$$s_{m,1,j,z}(t) = s_{m,0,j,z}(t - 2^m u) \quad s_{m,0,j,z}(t) = (1/\sqrt{2})(s_{m-1,0,j,z}(t) + s_{m-1,1,j,z}(t))$$

$$h_{m,0,j,z}(t) = (1/\sqrt{2})(s_{m-1,0,j,z}(t) - s_{m-1,1,j,z}(t)) \quad h_{m,n,j,z}(t) = h_{m,0,j,z}(t - n2^m)$$

We use the $h_{m,n,j,z}(t)$ in two ways. First, we choose $n_{m,j}$ based on the decay in $h_{m,n,j,z}(t)$, i.e., we set $n_{m,j}$ to be such that $h_{m,n,j,z}(n_{m,j}2^m)$ has died down to some negligible value. Second, the $h_{m,n,j,z}(t)$ will serve in the objective function of the voltage-drop-maximizing optimization problem, to be discussed next.

III. OPTIMIZATION PROBLEM FORMULATION

In this section we formulate a *linear program* (LP) that maximizes voltage drop at a node of interest on the PDN. Consider an *arbitrary* time point t_0. Our objective is to construct, for every node z, a worst-case stimulus which maximizes voltage drop on that node at $t = t_0$. It will be clear in our formulation that t_0 is indistinguishable from any other time point. Therefore, *maximizing voltage drop at $t = t_0$ is tantamount to maximizing voltage drop at any arbitrary time point during circuit operation*. This point is key: although our problem is that of maximizing voltage drop, a transient quantity during circuit operation, we approach it by a single-time-point optimization. We work backward from t_0 and construct *indirectly* the waveforms $i_1(t),\ldots, i_q(t)$ that make up the PDN stimulus, computing for each a set of coefficients $T_{m,n,j}$ and DC components, therefore fully specifying each waveform $i_j(t)$ as per (1).

Leveraging once more the linearity and time invariance of the PDN, denoting by $k_{j,z}$ the DC gain from the jth load to node z (i.e., the voltage drop on node z when the stimulus is $i_j(t) = 1$A DC), and using he PDN response to wavelet functions, $h_{m,n,j,z}(t)$, discussed in the previous section, we can write the voltage drop on node z at time t_0 as:

$$v(t_0) = \sum_{j=1}^{q} k_{j,z} i_{dc,j} + \sum_{j=1}^{q}\sum_{m=1}^{m_0}\sum_{n=0}^{n_{m,j}} h_{m,0,j,z}((n_{m,j}-n)2^m u)T_{m,n,j} \tag{2}$$

Equation (2) is our objective function, and is linear in terms of the optimization variables $T_{m,n,j}$ and $i_{dc,j}$. Maximizing (2) yields the maximum voltage drop on node z. We distinguish two types of constraints that can be imposed on the problem: power/current constraints and wavelet/frequency constraints.

A broad range of current/power bounds can be imposed on the PDN stimulus, given *specifications* known about the design at an early stage. Such bounds have been used in prior art [1], and can be easily embedded into our wavelet framework. The simplest bounds are on the minimum and peak-currents, such as: $0 < i_j(t) < i_{max,j}$, where $i_{max,j}$ is known by specification or simulation of the design block represented by $i_j(t)$. Another bound commonly available from design specification is a global power envelope P_{max} for a given chip, which can be expressed as $\Sigma\, i_j(t) < P_{max}/V_{dd}$. A third type of constraints that designers can infer is in the form of *max delta*, i.e., a bound on the change in current between successive time units. It can be written in the form $-\delta < i_j(t) - i_j(t-u) < +\delta$. Since the expression (1) of every $i_j(t)$ is linear in the optimization variables, expanding bounds such as peak envelope, power envelope, or max delta results in linear constraints on the problem. Evidently, these three types of constraints are not the only ones possible: any bounds which can be expressed as linear combinations of circuit currents will do.

Wavelet analysis yields further bounds in the time-frequency domain. From the perspective of PDN design, we would want to analyze commonly used high-level benchmark power simulation traces (architectural, RTL or logic) in order to generate time-frequency constraints that would complement current and power bounds. One such constraint is the *wavelet frequency envelope* that we derive by applying a wavelet transform algorithm (such as the Discrete Wavelet Transform [3]) on the available traces and taking the maximum observed values of the resulting transforms for every scale, resulting in new optimization constraints in the form: $-T_{m,max,j} < T_{m,n,j} < T_{m,max,j}$, $n = 0,\ldots, n_{m,j} - 1$. $m = 1,\ldots, m_0$, and $j = 1,\ldots, q$. The $T_{m,max,j}$ thus obtained form a wavelet frequency envelope for die currents and are a proxy for the maximum energy burst observed in these currents at a given frequency band (the wavelet's pass-band at each scale). A wavelet frequency envelope obtained after analyzing 50 power traces of an early-stage microprocessor design is shown in Fig. 1.

IV. Experimental Results

We tested the proposed approach on several models of a high-performance microprocessor design where the PDN was modeled from the VRM down to the die. Fig. 2 illustrates a worst-case current stimulus lasting $t_0 = 387$ nsec and covering a mid-to-high frequency band. Our approach covers arbitrary frequency bands by generating waveforms of various durations. As a general trend, worst-case stimuli were characterized by fairly regular, low-frequency variations at the beginning, with higher-frequency components (wavelets) progressively kicking in, creating local fine patterns that intensify near the end point, t_0.

Our approach may be construed as a generalization of the *reverse pulse technique* RPT [4]. While RPT works backward from t_0 to construct a worst-case stimulus simply as a sequence of full-swing step functions I_{max} to I_{min} and I_{min} to I_{max}, our approach yields finer patterns in the current waveform, as visible on Fig. 2. The reason is that it embeds sophisticated constraints in ways RPT can't, but when we stripped our optimization formulation from these constraints, the worst-case waveforms matched with RPT, but overestimated the maximum voltage drop versus their constrained counterparts by 16-25%. Therefore, our technique has the benefit of offering less pessimistic predictions, and broader user-characterization of currents than is possible with RPT.

Our approach can help designers predict voltage drop *trends* in an early design stage. By that, we are referring to the ability to systematically quantify the relative magnitudes of different voltage drop components. To illustrate this idea, we applied our analysis on a 4-core early-stage design, and measured the contribution of leakage, IR-drop, and di/dt switching activity for each core, on the worst-case voltage drop on Core 1. Fig. 3(a) shows sample results: IR-drop was found to contribute 18.2% of the total drop (with a separate 1.5% share for leakage). And while it is expected that the largest individual contributor to di/dt drop on Core 1 is switching activity on the same core (30.8%), the combined impact of the other cores exceeds that individual contribution (51.5%), thereby firmly establishing the need for an integrated cross-core analysis.

A key advantage of our framework is frequency-awareness. Modern PDNs have several resonant modes, each influenced by a set of electrical and design parameters. For example, the resonant mode at the highest frequency, commonly referred to as "first droop", is a strong function of package inductance and die capacitance. Lower frequency resonance, i.e., second and third droops, depend on other parameters including motherboard and bulk capacitors. Designers of different stages of the PDN will therefore be interested in gauging voltage drop across specific frequency ranges. While current methods for this end are tedious or measurement-heavy, ours naturally incorporates these considerations by specifying f_H and f_L and leveraging the wavelet time-frequency properties. Fig. 3(b) shows the breakdown of the three resonant modes in terms of their share of worst-case di/dt voltage drop. It is worth noting that each core required about 500 optimization variables to cover the frequency range of first, second, and third droop.

We also applied our approach to optimize the package inductance L_{pkg} for an early-design microprocessor model. The idea was to vary L_{pkg} around the design point and compute the expected worst-case voltage drop with die constraints unchanged. The alternative for designers is to discover the optimal L_{pkg} by simulating the PDN for given traces and candidate L_{pkg} values. These simulations, however are not guaranteed to uncover the worst-case corner, and simulating the PDN for several (we used 50) multi-million-cycle traces becomes impracticable. In this respect, our approach is both faster and more accurate. Fig. 4 is a plot of the predicted maximum voltage drop versus L_{pkg}, revealing that designers have approximately 30% headroom in their choice of L_{pkg} without incurring a voltage drop penalty. In fact, the figure predicts a slight improvement in the worst-case voltage drop if L_{pkg} were to be increased by about 25% from its planned value. This, somewhat surprising finding, stems from the fact that a change in L_{pkg} changes the PDN frequency response, and the interplay between the new frequency response and the wavelet envelope bounds (which are frequency constraints), results in a net decrease in the maximum possible voltage drop.

We also tested our approach on on-die power grids. We carried out an experiment where we fixed the dimension of a grid (5 mm x 5 mm), total power budget (1 W), and die capacitance (200 nF), then progressively added metallization layers and verified the grid with every new layer. The grid was modeled as an RC network and the power budget non-uniformly distributed over 182 current loads. Grid verification consisted of finding the worst-case voltage drop in the 50 MHz - 1 GHz range, over the 182 nodes with loads attached. Table 1 shows the results, obtained on a server with two dual-core, 2.6 GHz processors. We divided runtime into two components: 1) simulation of step-responses (see section II), carried out with HSpice, and 2) solution of the optimization problem (section III) done with PCx, a freely available LP package. It is clear that simulation is the overwhelming bottleneck. However, simulation is a pre-characterization step: users need to do it only once for a given PDN and

can re-run the verification with different constraints or over different frequency bands. More importantly, these results show that the cost of maximizing voltage drop, which essentially searches the feasibility space of currents in time, came down to the computation of step responses. The second highlight is the efficiency of the optimization itself, due to the relatively small number of variables per optimization problem (column 4 reports the average number over the 182 problems solved). This follows from the grid's spatial locality, which our framework picks up seamlessly: if $h_{m,n,j,z}(t_0)$ is negligible, then the corresponding variable $T_{m,n,j}$ is dropped from the optimization problem, leaving for every node only the set of relevant sources and scales (frequencies) as optimization variables.

V. CONCLUSION

We introduced the concept of time-frequency description of circuit currents using wavelets, and formulated an optimization framework that solves for the worst-case supply voltage drop. We applied this framework on an early-stage design process, for package-die co-design, and on power grids, showing how it naturally fills designers' needs for systematic predictions and characterizations of the power delivery network.

REFERENCES

[1]   D. Kouroussis and F. N. Najm. A static pattern-independent technique for power grid voltage integrity verification. In *ACM/IEEE Design Automation Conference*, Anaheim, California, pages 99-104, June 2-6, 2003.
[2]   R. Joseph, Z. Hu, and M. Martonosi. Wavelet analysis for microprocessor design: experiences with wavelets-based dI/dt characterization. In *International Symposium on High Performance Computer Architecture,* Madrid, Spain, pages 36-46, February 14-18, 2004.
[3]   P. S. Addison. *The Illustrated Wavelet Transform Handbook*. Taylor & Francis Group, New York, NY, 2002.
[4]   V. Drabkin, C. Houghton, I. Kantorovich, and M. Tsuk. Aperiodic resonant excitation of microprocessor power distribution systems and the reverse pulse technique. In *IEEE Topical Meeting on Electrical Performance of Electronic Packaging*, Monterey, California, pages 175-178, October 21-23 2002

Grid size (#nodes)	Simulation time	Optimization time	Average #variables	Max Vdrop
11,100	1.6 hrs.	11 sec.	675	59 mV
14,300	1.7 hrs.	8 sec.	604	35 mV
26,000	3.6 hrs.	5 sec.	543	20 mV
36,500	11.3 hrs.	4 sec.	522	13 mV

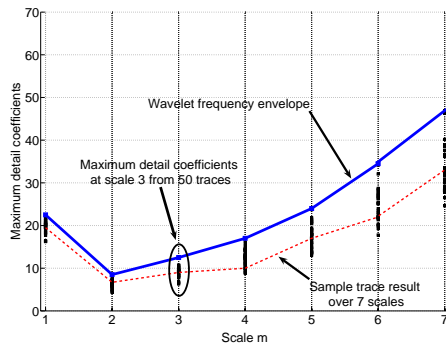Table 1. Application to power grid verification.

Fig.1. Wavelet frequency envelope in an early-design microprocessor. Scale $m = 1$ represents a wavelet lasting 10 clock cycles.
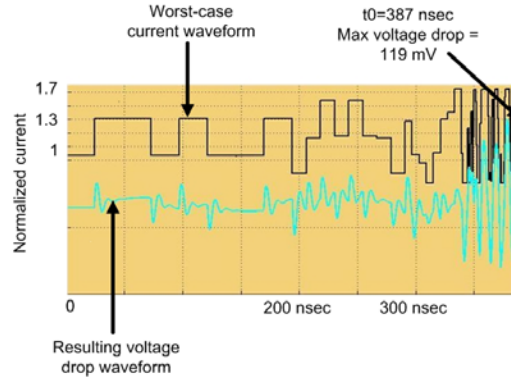

Fig.2. Sample worst-case current stimulus and resulting voltage drop waveform.
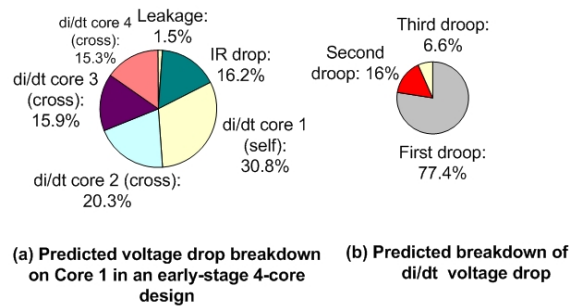

(a) Predicted voltage drop breakdown on Core 1 in an early-stage 4-core design

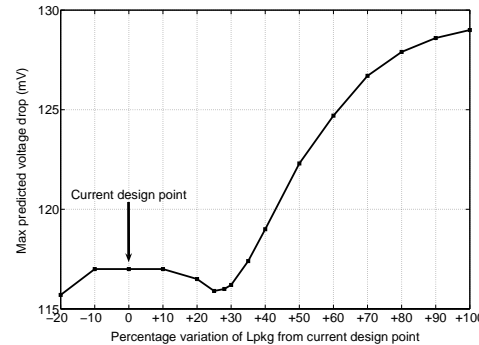(b) Predicted breakdown of di/dt voltage drop

Fig.3. Predicted trends in voltage drop

Fig.4. Application to L_{pkg} optimization