

Vectorless Analysis of Supply Noise Induced Delay Variation

Sanjay Pant*, David Blaauw*, Vladimir Zolotov**, Savithri Sundareswaran**, Rajendran Panda**
{spant,blaauw}@umich.edu, {vladimir.zolotov,savithri.sundareswaran,rajendran.panda}@motorola.com

*University of Michigan, Ann Arbor, MI

**Motorola, Inc., Austin, TX

Abstract

The impact of power supply integrity on a design has become a critical issue, not only for functional verification, but also for performance verification. Traditional analysis has typically applied a worst case voltage drop at all points along a circuit path which leads to a very conservative analysis. We also show that in certain cases, the traditional analysis can be optimistic, since it ignores the possibility of voltage shifts between driver and receiver gates. In this paper, we propose a new analysis approach for computing the maximum path delay under power supply fluctuations. Our analysis is based on the use of superposition, both spatially across different circuit blocks, and temporally in time. We first present an accurate model of path delay variations under supply drops, considering both the effect of local supply reduction at individual gates and voltage shifts between driver/receiver pairs. We then formulate the path delay maximization problem as a constrained linear optimization problem, considering the effect of both IR drop and Ldl/dt drops. We show how correlations between currents of different circuit blocks can be incorporated in this formulation using linear constraints. The proposed methods were implemented and tested on benchmark circuits, including an industrial power supply grid and we demonstrate a significant improvement in the worst-case path delay increase.

1 Introduction

Power supply networks are essential in providing the devices on a die with a reliable and constant operating voltage. Due to the interconnect resistance and inductance of the on-chip and package supply networks, the supply voltage delivered to various devices on a die is non-ideal and exhibits both spatial and temporal fluctuations. These fluctuations in the supplied voltage can result in a reduction in operating frequency and can compromise the functional stability. Power supply integrity is therefore a critical concern in high-performance designs.

The voltage drop that develops in a supply network can be broadly classified into *IR-drop*, which is the voltage drop due to the parasitic resistances of the interconnects and *Ldl/dt drop*, which is the voltage drop due to the inductance of I/O pads and the parasitic inductance of the supply interconnects. In today's high-end designs, it is not uncommon for the supply network to conduct as much as 50-100 Amperes of total current [1,6]. As semiconductor technology is scaled and the supply voltage is reduced, the total current that must be supplied by the power network is expected to increase even further, making it more difficult to meet stringent supply integrity constraints. In particular, the Ldl/dt voltage drop is expected to become more prominent as it worsens with both increasing current demand and clock frequency [2]. Furthermore, IR-drop and Ldl/dt drop interact in a non-trivial manner and total drop is not always the sum of the individual voltage drops.

The voltage fluctuations in a supply network can inject noise in a circuit, leading to functional failures in the design. Extensive work has therefore been focussed on modelling and efficient analysis of the worse-case voltage drop in a supply network [2-7]. However, with decreasing supply voltages, the gate delay is becoming increasingly sensitive to supply voltage variation as the headroom between V_{dd} and V_t is consistently reduced [12]. For instance, in 0.13 μ m technology, a 10% variation in the V_{dd} and Gnd voltages can result

in a 30% delay variation for typical gates. With ever diminishing clock cycle times, accurate analysis of the supply voltage impact on circuit performance has therefore become a critical issue.

In this paper, we present a new approach for the analysis of supply voltage induced delay variations. Power supply analysis has been complicated by the enormous size of the supply network. For modern processors, it is not uncommon for the supply network to be represented by an RLC circuit requiring more than 60 million elements. Simulation of such a large circuit is extremely challenging and significant progress has been reported in developing efficient simulation approaches [3,5,7]. However, even with effective acceleration methods, it is typically not possible to simulate a supply network for more than a handful of clock cycles in reasonable time. Selecting the simulation vectors that exhibit the worst-case supply voltage drops is therefore a key issue in supply network verification. The supply voltage fluctuation is strongly dependent on the simulation vectors that determine the currents drawn by the devices from the supply network. Hence, critical supply integrity problems can go undetected if worst-case simulation vectors are not applied, regardless of the simulation accuracy.

A number of methods have therefore been proposed that use Genetic Algorithms or other search methods to automatically find vectors that maximize the total current drawn from the supply network [8,10]. These approaches typically are computationally intensive and are limited to circuit blocks, rather than full chip analysis. In addition, a number of vectorless approaches for constructing worst-case currents have been proposed using either propagation of timing windows [8] or constraint graph formulations [11]. Vectorless approaches have the advantage that they are conservative, meaning that the supply drop will be overestimated, rather than underestimated. However, these methods address only static IR-drop analysis, and not Ldl/dt drop, which has become a key concern in supply integrity analysis. Also, they do not consider the impact of supply fluctuations on delay. Recently, a statistical approach for analyzing the impact of supply noise on delay was also presented [14].

Power supply variation can impact the circuit delay in two ways: First, a reduced supply voltage lessens the gate drive strength, thereby increasing the gate delay. Second, a difference in the supply voltage between a driver and receiver pair creates an offset in the voltage with which the driver/receiver gates reference the signal transition. This has the effect of creating either a positive or negative time shift in perceived signal transition at the receiver gate, as illustrated in Figure 1. This dual nature of the supply voltage impact on circuit delay was observed in [13], and complicates the generation of simulation vectors that maximize the delay along a particular circuit path. Increasing the voltage drop at a particular location may worsen the delay of one gate while improving the delay of another. Therefore, a vector must be determined that results in an optimal combination of the often conflicting goals to introduce both reduced drive strengths and supply voltage shifts such that the total delay along a path is maximized.

Traditionally, the impact of supply noise on delay has been accounted for by reducing the operating voltage of all library cell by the expected supply voltage drop during library characterization. This assumes that the worst-case expected voltage drop occurs in all places of the design. This yields a very conservative analysis since, in practice, the worst-drop can occur in only a small region at any

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICCAD'03, November 11-13, 2003, San Jose, California, USA.

Copyright 2003 ACM 1-58113-762-1/03/0011 ...\$5.00.

one point in time. On the other hand, this approach ignores the impact of voltage shifts between driver/receiver pairs, thereby possibly underestimating the worst-case delay in certain situations. Also, it only accounts for static IR-drop.

In this paper, we therefore present a new approach for the analysis of power supply drops on circuit delay. The proposed approach is vector-less, allowing for efficient analysis, and addresses for both IR-drop and Ldi/dt drop effects. We develop a linear model that accounts for both the impact of driver strength reduction and voltage shifts between driver/receiver pairs. Based on this model, we formulate the task of determining the worst-case impact of supply noise on a path delay using a constrained linear optimization model where the currents of the different blocks are the optimization variables. We use both spatial and temporal super-position of the voltage drops resulting from currents of individual circuit blocks. Linear constraints are then formulated both for the total power consumption of a chip, as well as for individual block currents. Constraints between currents of different blocks or a single block in consecutive clock cycles can be formulated expressing both spatial and temporal correlations that exist between circuit blocks. The proposed approach has the advantage that accurate constraints can be extracted from extensive gate level simulation data that is readily available during the design process, thereby significantly improving the accuracy of the analysis while avoiding the need for lengthy and time consuming power grid simulation. We implemented the proposed methods and tested them on benchmark circuits, including a power grid from an industrial processor design. We show that the traditional analysis may overestimate the change in delay of a path by more than 50% and demonstrate the effectiveness of our analysis.

The remainder of this paper is arranged as follows. Section 2 describes our model for delay variations with respect to supply voltage fluctuations. Section 3 presents the problem formulations and optimization method for maximizing the impact of power grid fluctuations on delay. Section 4 presents the results obtained for different power grids. In Section 5, we draw our conclusions.

2 Delay Model for Supply Fluctuations

In this Section, we present our approach for modeling the impact of voltage variations on the delay of a circuit path. Since the voltage variations in a power grid are typically very slow compared to the transition time of a switching gate [15], we can make the simplifying assumption that the supply voltages are constant during the switching transitions. From the perspective of the path delay, we are therefore concerned with the impact of fixed voltage offsets from the nominal V_{dd} and V_{ss} voltages on the delay of a circuit path. Note however that *dynamic* IR-drop and Ldi/dt drop effects will be the cause of these voltage offsets.

A voltage drop at a power supply point can impact the delay of a gate through one of the following two mechanisms:

1. A decrease in the V_{dd} voltage or an increase in the V_{ss} voltage at the gate under consideration decreases the *locally* observed supply voltage of the gate and will reduce its drive strength and hence increase its delay. The worst case voltage drop is typically localized to a small region in the chip, as it requires all currents to be concentrated in that region. Hence, only a few gates in a path will typically be operated with a worst-case drive strength. Gates with higher local supply voltage therefore compensate for the increased delay of gates with reduced local supply voltage in the path and a global analysis of the impact of supply voltage on the path delay is therefore required.
2. A relative shift in the V_{dd} or V_{ss} voltages between the driver and receiver gates of a signal net can introduce a voltage offset that will impact the delay of a gate. This is illustrated in Figure 1

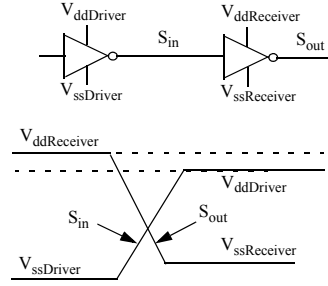


Figure 1. A driver-receiver pair in a non-ideal supply network

where the V_{ss} voltage of the receiver gate is increased relative to the V_{ss} voltage of the driver gate. Since the input signal has a rising transition, the NMOS transistor of the receiver gate senses the input voltage relative to the local V_{ss} voltage level. The shown voltage shift therefore results in an effective (negative) noise voltage at the receiver gate input that increases the delay of the receiver gate. Note that a shift in the supply voltage impacts the rising and falling transitions of a gate in opposite ways, meaning that an increase in the V_{ss} voltage from driver to receiver results in an increased delay for a rising input transition while an increase in the V_{dd} voltage improves the delay for a falling input transition.

The relative shift between the driver and receiver gates is likely to be larger if the gates are separated farther apart as compared to the case when they are closer together. Therefore, nets that transmit signals across the chip will have a higher likelihood of shifts in supply voltage between their driver and receiver pair and hence are more susceptible to power grid noise.

The relative magnitude of the above two mechanisms depends on the input slope and output loading of a gate. The sensitivity of gate delay to driver strength reduction will increase with output loading, while the sensitivity to voltage shifts will increase with slower input signal transition times.

In order to maximize the delay of a path, it is necessary to induce voltage drops in the supply network such that the delay of each gate is increased through both mechanisms: reduction of driver strength and voltage shifts between successive gates in the path. A possible voltage assignment that maximizes the voltage shift between consecutive gates in a circuit path is shown in Figure 2. However, this

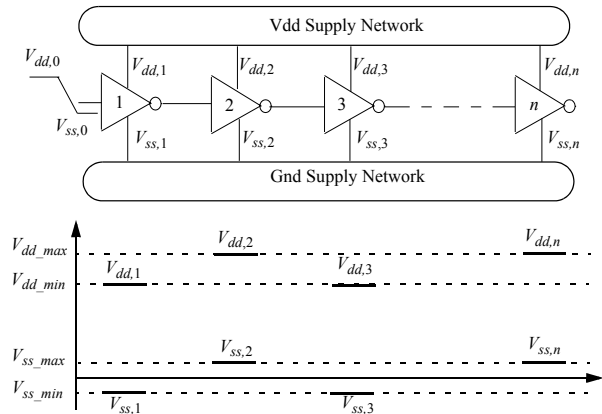


Figure 2. A path in a power supply network with worst-case voltage shifts

assignment does not reduce the driver strength of each gate by the maximum possible amount. Maximizing the delay through reduced drive strength and through voltage shifts therefore, requires conflicting voltage assignments that cannot be realized simultaneously. A worst-case *realizable* voltage assignment that maximizes the overall path delay is therefore not intuitively obvious and will depend on the specific conditions of the gates and their sensitivities to the different

voltage drop phenomena.

We now present our model for the dependence of the delay of a single gate on the voltage drops at that gate and at its preceding gate. We then extend this model to the delay of a circuit path.

Individual gate delay model

We consider the delay of a gate G , shown in Figure 3(a), with

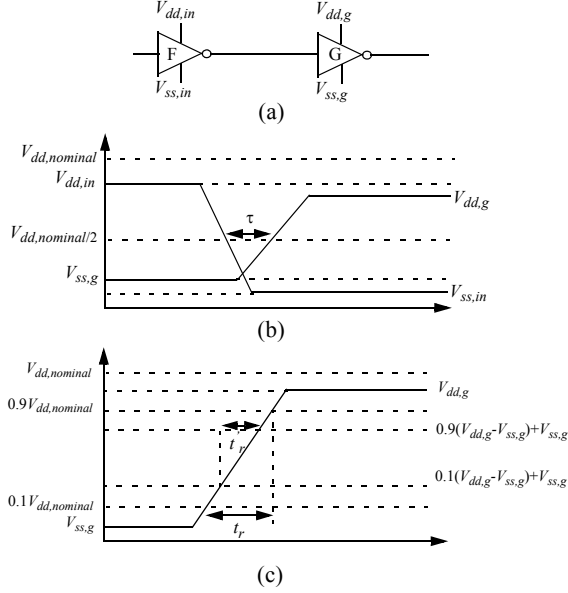


Figure 3. A Driver-receiver pair in a non-ideal supply network

local supply voltages $V_{dd,g}$ and $V_{ss,g}$ and supply voltages $V_{dd,in}$, $V_{ss,in}$ at the preceding driver gate. As shown in Figure 3(b), the propagation delay τ between the input and output transitions of a gate is measured at $1/2$ the *nominal* supply voltage point to ensure a common reference between successive gates. The delay of the receiver gate depends on the $V_{dd,g}$ and $V_{ss,g}$ voltages at the receiver gate itself, the voltages $V_{dd,in}$, $V_{ss,in}$ at the preceding driver gate, the input transition time and the output load. For the purpose of our discussion, we consider a fixed output load, although in our actual implementation gates are characterized over a range of output loads.

The input transition time at gate G is a function of the delay of the preceding gate F , which, in turn, is a function of the supply voltages. It is therefore necessary to include the impact of the supply voltage fluctuations on the signal transition times in the delay model. To provide a common reference for transition time, we again define the transition time t_r of a signal as the time between the 10% to 90% crossing of *nominal* Vdd for an equivalent full swing transition, as shown in Figure 3(c). Given the signal transition at the output of gate G , and given the *local* transition time t_r' , measured between the 10% to 90% crossing of the local supply voltage $V_{ss,g}$ to $V_{dd,g}$, the equivalent full-swing transition time t_r is computed as follows:

$$t_r = t_r' \cdot \frac{V_{dd,nominal}}{V_{dd,g} - V_{ss,g}} \quad (\text{EQ 1})$$

We now express the delay and transition time at the output of gate G as follows:

$$\tau = f(V_{dd,g}, V_{ss,g}, V_{dd,in}, V_{ss,in}, t_{rin}) \quad (\text{EQ 2})$$

$$t_{r,out} = g(V_{dd,g}, V_{ss,g}, V_{dd,in}, V_{ss,in}, t_{rin}) \quad (\text{EQ 3})$$

In general, f and g are nonlinear functions of their variables. However, the voltage drop in a power grid network is restricted and is typically within the range of $\pm 10\%$ of $V_{dd,nominal}$. We found that within this range, the delay of a gate is close to linear. Figure 4

shows the rise and fall delays of a typical gate as $V_{dd,g}$, $V_{ss,g}$, $V_{dd,in}$

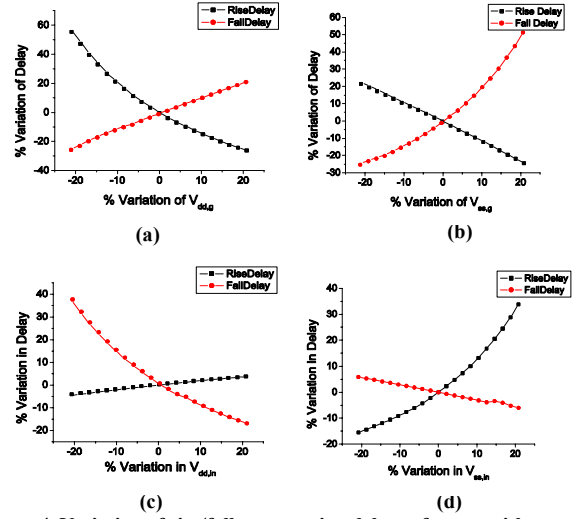


Figure 4. Variation of rise/fall propagation delays of a gate with respect to (a) $V_{dd,g}$, (b) $V_{ss,g}$, (c) $V_{dd,in}$ and (d) $V_{ss,in}$

and $V_{ss,in}$ are varied by $\pm 20\%$. The delay curves in Figure 4 show that f and g can be accurately modeled as linear functions for reasonable supply voltage variations. We therefore express the change in delay, $\Delta\tau$ of a gate with respect to its delay at nominal supply voltages as follows:

$$\Delta\tau = a_1\Delta V_{dd,g} + a_2\Delta V_{ss,g} + a_3\Delta V_{dd,in} + a_4\Delta V_{ss,in} + a_5\Delta t_{r,in} \quad (\text{EQ 4})$$

where $\Delta V_{dd,g}$, $\Delta V_{ss,g}$, $\Delta V_{dd,in}$ and $\Delta V_{ss,in}$ are the deviation of the four supply voltages from their nominal values and $\Delta t_{r,in}$ is the change in the input transition time from its nominal value. Similarly, we express the change in the transition time $\Delta t_{r,out}$ at the output of a gate with respect to its transition time at nominal supply voltages as follows:

$$\Delta t_{r,out} = b_1\Delta V_{dd,g} + b_2\Delta V_{ss,g} + b_3\Delta V_{dd,in} + b_4\Delta V_{ss,in} + b_5\Delta t_{r,in} \quad (\text{EQ 5})$$

The constants a_1 - a_5 and b_1 - b_5 are determined using multiple regression analysis where each gate is simulated over a range of supply voltage variations and rise/fall transition changes. Table 1 compares the delay values determined using our linear model with delay values obtained through SPICE simulation for a low to high propagation delay of an inverter in 0.13 micron technology with a nominal power supply of 1.2V. Different combinations of maximum supply voltage variations are shown. We also compared the accuracy of the

Table 1. Low-to-High Propagation delay Regression Results

$\Delta V_{dd,g}$	$\Delta V_{ss,g}$	$\Delta V_{dd,in}$	$\Delta V_{ss,in}$	$t_{r,in}$ Rise Time (ps)	Low-to-High Delay		% Error
					Reg.	spice	
0.10V	0.10V	0.10V	0.10V	50	11.19ps	11.72ps	4.5%
0.05V	0.00V	-0.10V	0.05V	50	16.71ps	16.34ps	2.3%
0.00V	0.05V	-0.10V	0.00V	75	19.12ps	19.48ps	1.85%
0.00V	-0.10V	-0.05V	0.10V	75	28.17ps	26.97ps	4.45%
-0.05V	-0.10V	0.10V	-0.10V	100	33.43ps	33.54ps	0.33%
-0.05V	0.00V	0.05V	-0.05V	100	27.31ps	27.42ps	0.40%
-0.10V	0.10V	-0.05V	-0.10V	125	26.89ps	26.17ps	2.75%
-0.20V	-0.05V	-0.10V	0.20V	125	48.04ps	44.98ps	6.80%

compares the delay values determined using our linear model with delay values obtained through SPICE simulation for a low to high propagation delay of an inverter in 0.13 micron technology with a nominal power supply of 1.2V. Different combinations of maximum supply voltage variations are shown. We also compared the accuracy of the

proposed delay model for more than 3000 randomly generated voltage and transition time variations of $\pm 10\%$, which resulted in a average error of 0.74% and maximum error of 8.1%.

It should be noted that while we linearly model the *change* in delay due to supply voltage variations, the *nominal* delay itself is not a linear function of output load and nominal input transition time. We therefore used a non-linear, table based model, similar to that used in Synopsys Design Compiler, to model the dependence of nominal delay and output transition time on output load and nominal transition input time. For each possible load and input transition time condition, we also determined different linear fitting constants a_1 - a_5 and b_1 - b_5 , which are stored in a table along with the nominal delay and output transition time values.

Circuit path delay model

We now consider the variation of the delay, $\Delta\tau_{path}$ of a circuit path due to supply voltage variations at different supply connections along a path as shown in Figure 2(a). In general, the change in the delay of the n th gate is given by:

$$\begin{aligned} \Delta\tau_n = & a_{1,n}\Delta V_{dd,n} + a_{2,n}\Delta V_{ss,n} + a_{3,n}\Delta V_{dd,n-1} \\ & + a_{4,n}\Delta V_{ss,n-1} + a_{5,n}\Delta t_{r,n-1} \end{aligned} \quad (\text{EQ 6})$$

and the change in its output transition time is given by:

$$\begin{aligned} \Delta t_{r,n} = & b_{1,n}\Delta V_{dd,n} + b_{2,n}\Delta V_{ss,n} + b_{3,n}\Delta V_{dd,n-1} \\ & + b_{4,n}\Delta V_{ss,n-1} + b_{5,n}\Delta t_{r,n-1} \end{aligned} \quad (\text{EQ 7})$$

where $a_{i,n}$, $b_{i,n}$ are the regression coefficients for gate n ; $\Delta V_{dd,n}$, $\Delta V_{ss,n}$ are the supply voltage drops at gate n ; $\Delta V_{dd,n-1}$, $\Delta V_{ss,n-1}$ are the supply voltage drops at its driver gate, $n-1$.

The delay of gate n is therefore defined in terms of the change of the output transition time of gate $n-1$, leading to a recursive definition of the overall path delay. The total delay change of a circuit path, $\Delta\tau_{path}$, is the sum of the changes of the gate delays along the path and is expressed as follows:

$$\begin{aligned} \Delta\tau_{path} = & \sum_{i=1}^n a_{1,i}\Delta V_{dd,i} + a_{2,i}\Delta V_{ss,i} + a_{3,i}\Delta V_{dd,i-1} \\ & + a_{4,i}\Delta V_{ss,i-1} + a_{5,i}\Delta t_{r,i-1} \end{aligned} \quad (\text{EQ 8})$$

where,

$$\begin{aligned} \Delta t_{r,i} = & b_{1,i}\Delta V_{dd,i} + b_{2,i}\Delta V_{ss,i} + b_{3,i}\Delta V_{dd,i-1} \\ & + b_{4,i}\Delta V_{ss,i-1} + b_{5,i}\Delta t_{r,i-1}, \end{aligned} \quad (\text{EQ 9})$$

and where n is the number of gates in the circuit path.

For simplicity of our discussion, we assume an ideal transition between 0V and nominal V_{dd} at the input of the path, and hence,

$$\Delta V_{dd,0} = \Delta V_{ss,0} = \Delta t_{r,0} = 0 \quad (\text{EQ 10})$$

However, the analysis can be easily extended to account for non-ideal input signal transitions.

Equations 8 and 9 model the change in the delay of a path as a linear function of supply voltages at the individual gate connections. In the next section, we propose a method to express these supply voltages as a linearly function of block currents and formulate the problem of maximizing delay as a linear optimization problem.

3 Maximum Delay Variation Formulation

We now discuss how the supply voltages can be expressed as a linear function of the current sources using both spatial and temporal superposition and accounting for both IR-drop and Ldl/dt drop. We

then show how the problem of maximizing delay change for a circuit path can be formulated as a linear optimization problem with linear constraints.

We consider a power supply network composed of RLC elements, current sources and voltage sources. We first consider an independent current source $i_m(t)$, applied at node m , and denote the voltage response generated at any node n due to the current $i_m(t)$ as $V_{m,n}(t)$. Given a set of current sources $i_m(t)$, the response at any node n in the circuit due to this set of current sources acting together is the summation of all the responses at node n caused by the individual current sources:

$$V_n(t) = \sum_m V_{m,n}(t) \quad \text{for all } m \quad (\text{EQ 11})$$

This is the well known principle of superposition, applied spatially across the different current sources of a supply network.

However, $V_n(t)$ in EQ11 depends on the entire current waveform $i_m(t)$, and requires that the entire current waveform is simulated for each current source. This complicates the formulation of the delay maximization problem since the number of possible current waveforms $i_m(t)$ can be very large and enumerating all possibilities would be impossible. We therefore approximate an arbitrary current waveform $i_m(t)$ using a piece-wise constant waveform with a discretization of time into time steps T_s , as shown in Figure 5(a). Given the

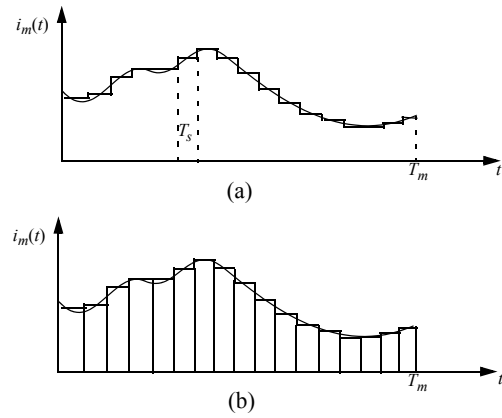


Figure 5. Temporal discretization and superposition approach.

total duration T_m of waveform $i_m(t)$ and the time step size T_s , the number of discretizations S is given by: $T_m = T_s * S$. If the discretization time step T_s is chosen sufficiently small, the piece-wise constant approximation of the continuous waveform has negligible error. We now represent the piece-wise constant current waveform as the sum of a series of current pulses of duration T_s , each shifted in time by one time step, as shown in Figure 5(b) and expressed as follows:

$$i_m(t) = \sum_{i=0}^{S-1} I_{m,i} p(t - iT_s) \quad (\text{EQ 12})$$

$$\begin{aligned} \text{where, } p(t) = & 1, \text{ if } 0 < t < T_s \\ & 0, \text{ otherwise.} \end{aligned}$$

and $I_{m,i}$ is the magnitude of the piece-wise constant approximation of current pulse $i_m(t)$ in the interval iT_s to $(i+1)T_s$.

Conceptually, we can therefore replace each current source $i_m(t)$ at node m with a set of S current pulse sources $i_{m,i}(t)$ connected to the same node in the grid. Note that each current pulse $i_{m,i}(t)$ is a scaled and shifted version of the unit current pulse $i_u(t)$ with a unit pulse height and a pulse width of T_s :

$$\begin{aligned} i_u(t) = & 1, \text{ if } 0 < t < T_s \\ & 0, \text{ otherwise} \end{aligned} \quad (\text{EQ 13})$$

Due to the nature of a power supply network, the voltage response $V_n^u(t)$ at node n due a single unit current pulse $i_n(t)$ will reach steady-state and approach the nominal supply voltage given sufficient time. The difference of the voltage $\Delta V_n^u(t)$ at node n from the nominal supply voltage $V_{dd,nominal}$ therefore approaches zero given sufficient time. We assume that this voltage difference has diminished below a specified error threshold at time $T_k = K * T_s$.

Since any finite length current waveform $i_m(t)$ can be represented by a finite set of current pulse sources, we can compute the voltage response $V_n(t)$ at node n by summing the response from each of the individual current pulse sources, using linear superposition. However, since the power supply network is linear, the response resulting from each current pulse is simply a shifted and scaled version of the response $V_n^u(t)$ resulting from a unit current pulse. We can therefore express the change in the voltage response $\Delta V_{m,n}(t)$ from the nominal supply voltage due to the current source $i_m(t)$ as follows:

$$\Delta V_{m,n}(t) = \sum_{i=0}^{K-1} \Delta V_n^u(t - iT_s) I_{m,i} \quad (\text{EQ 14})$$

where $I_{m,i}$ is the magnitude of the piece-wise constant current waveform approximated in interval iT_s to $(i+1)T_s$.

Using superposition in this temporal manner, we can therefore compute the response of any node in the network due to an arbitrary current source $i_m(t)$ using a single simulation of a unit current pulse and combining scaled and shifted versions of this response, using EQ14. The only approximation in this approach arise from the piece-wise constant approximation of the current waveform and the finite simulation length of the unit current pulse response. Given a sufficiently fine grain discretization and sufficient simulation length of the unite current pulse response, arbitrary accuracy can be obtained. Also, the computational complexity grows linearly with respect to the unit pulse response simulation length T_k and the number of discretizations S of the current waveform $i_m(t)$. Typically, the length T_m of waveforms $i_m(t)$ will be much greater than the unit pulse response time T_k . Since the simulation time of the supply network will by far dominate the run time effort, the proposed approach will provide a speedup of approximately T_m/T_k compared to simulating the entire current waveform $i_m(t)$. It should also be noted that the current waveform $i_m(t)$ can be approximated not only by a sequence of square current pulses, but also by other current pulse shapes, using a similar analysis.

Finally, we combine the temporal superposition with spatial superposition to obtain the voltage fluctuation $\Delta V_n(t)$ at a node n due to a set of arbitrary current sources $i_m(t)$ at nodes m as follows:

$$\Delta V_n(t) = \sum_{m=0}^{M-1} \sum_{i=0}^{K-1} \Delta V_{m,n}^u(t - iT_s) I_{m,i} \quad (\text{EQ 15})$$

where M is the number of current sources and K is the number of discretizations of the unit current voltage response. $\Delta V_{m,n}^u(t)$ is the difference in the voltage response at node n from nominal supply resulting from a unit current source at node m . $I_{m,i}$ is the magnitude of the current source m at time t . The proposed formulation requires that each current source is simulated, in turn with a unit current pulse for a simulation period of T_n and the voltage responses $V_{m,n}^u(t)$ are recorded at all nodes of interest. The formulation of EQ15 has the advantage that it is linear in terms of the current values $i_m(t)$ and hence allows the delay maximization problem to be cast as a constraint linear optimization problem as explained in the following Section.

Delay maximization formulation

We apply the above formulation to the problem of delay maximization, using a linear optimization formulation with the current values as optimization variables. We first divide the chip into circuit blocks and simulate the minimum and maximum currents of each circuit block using Powermill or Verilog simulations or estimate them on the basis of a previously fabricated part. In a microprocessor design, these circuit blocks could be, for example, the instruction fetch stage, instruction decode stage, execute stage, caches and the main memory control units. We make the simplifying assumption that the total current in a circuit block is evenly divided among its power supply points. This has the advantage that the voltage sensitivities, $\Delta V_n^u(t)$ can be computed with respect to the total current of a circuit block, instead of with respect to each individual current source point in a circuit block. This therefore greatly reduces the number of optimization variables in our formulation and improves its efficiency.

When selecting circuit blocks, it is therefore important that each block is sufficiently small to ensure that the spatial distribution of the currents within a circuit block do not significantly impact the voltage response. For high-performance processors, with tight and uniform supply grids over multiple layers of metal, the spatial distribution of the total block current is typically not significant for moderate size blocks [17]. If however, necessary, the proposed approach can be extended for non-uniform current distributions. It is also desirable that circuit blocks are selected such that their currents are independent, reducing the need to incorporate constraints between the currents of different blocks in the delay maximization formulation.

The current waveform for a circuit blocks typically has an approximately triangular shape within an clock cycle, as shown in Figure 6, reflecting a higher switching activity at the start of the clock cycle then at the end of the clock cycle [16]. We currently approximate the current waveform for a circuit block in a single clock cycle with a trapezoidal waveform, as shown in Figure 6. We

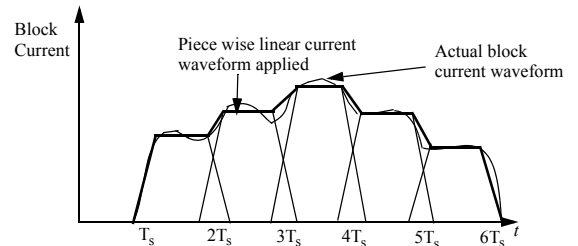


Figure 6. Current modeling for circuit blocks

then set the step size T_s in the superposition formulation equal to one clock period and approximate the total block current as the sum of shifted and scaled trapezoidal current pulses. This results in a piece-wise linear approximation of the total block current, as shown with darkened lines in Figure 6. However, our approach is not restricted to a specific current profile and different current profile approximations could be used as well.

The block current within a clock cycle may vary not only in magnitude but also in shape with different input data. Some input vectors will result in more switching activity at the start of the cycle, while other input vectors may result in more switching activity at the end of the cycle. However, with the scaling of process technology, the clock frequency has increased significantly while the resonance frequency of the supply network has steadily decreased. For a 1-2Ghz processor, typical resonance frequencies of the power supply network are in the range of 30-80Mhz [15]. Any change in the shape of the current waveform within a *single* clock cycle therefore impacts

frequencies that are well above the resonance frequency of the power distribution network and have little impact on the voltage waveforms. This is illustrated in Figure 7, where the voltage

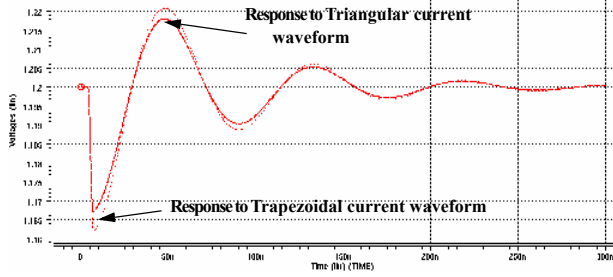


Figure 7. Variation of voltage at a node in the power grid with different clock cycle waveform shapes.

response of a node in the grid resulting from two different block current waveform shapes with equal total charge, is shown. One waveform uses a triangular current waveform shape and the other waveform uses the trapezoidal approximation, as shown in Figure 6. The simulations show that the response of the voltage is nearly indistinguishable. Note that, if necessary, the proposed approach can be extended such that each clock cycle is divided into multiple time-steps and is represented with a series of consecutive current pulses, allowing for different waveforms within a clock cycle.

Based on Figure 6, we also observe that the voltage response $V_{m,n}^u(t)$ within a clock cycle is nearly constant and can be approximated with a fixed voltage value $V_{m,n,i}^u$. Based on EQ15 we now express the voltage variation of a Vdd node n as a function of the current $i_m(t)$ of circuit block m as follows:

$$\Delta V_{dd,n} = \sum_{m=0}^{M-1} \sum_{i=0}^{S-1} \Delta V_{m,n,i}^{u,V_{dd}} I_{m,S-i} \quad (\text{EQ } 16)$$

where $I_{m,i}$ is the average current of the circuit block m in clock cycle i and $\Delta V_{m,n,i}^{u,V_{dd}}$ is the sensitivity of the V_{dd} voltage node n with respect to the current of block m after i clock cycles of delays. Similarly, we express the voltage variation of a V_{ss} node as:

$$\Delta V_{ss,n} = \sum_{m=0}^{M-1} \sum_{i=0}^{S-1} \Delta V_{m,n,i}^{u,V_{ss}} I_{m,S-i} \quad (\text{EQ } 17)$$

where $\Delta V_{m,n,i}^{u,V_{ss}}$ is the sensitivity of the V_{ss} node n with respect to the current of block m after i clock cycles of delays. We now formulate the problem of maximizing delay as a linear optimization problem as follows:

Maximize:

$$\Delta \tau_{Path} = \sum_{i=1}^n a_{1i} \Delta V_{dd,i} + a_{2i} \Delta V_{ss,i} + a_{3i} \Delta V_{dd,i-1} + a_{4i} \Delta V_{ss,i-1} + a_{5i} \Delta t_{r,i-1} \quad (\text{EQ } 18)$$

such that:

$$\Delta V_{dd,n} = \sum_{m=0}^{M-1} \sum_{i=0}^{S-1} \Delta V_{m,n,i}^{u,V_{dd}} I_{m,S-i} \quad (\text{EQ } 19)$$

$$\Delta V_{ss,n} = \sum_{m=0}^{M-1} \sum_{i=0}^{S-1} \Delta V_{m,n,i}^{u,V_{ss}} I_{m,S-i} \quad (\text{EQ } 20)$$

$$\Delta t_{r,i} = b_{1i} \Delta V_{dd,i} + b_{2i} \Delta V_{ss,i} + b_{3i} \Delta V_{dd,i-1} + b_{4i} \Delta V_{ss,i-1} \quad (\text{EQ } 21)$$

$$I_{min,i} \leq I_{m,i} \leq I_{max,i} \quad (\text{EQ } 22)$$

$$\sum_{i=1}^N I_{m,i} \leq I_{peak} \quad (\text{EQ } 23)$$

The constraint in EQ22 expresses that the current of a block must have a value between its maximum and minimum possible value, as

determined from Powermill or Verilog simulation. The constraint in EQ23 forces an upper-bound on the total current of the chip. This expresses that, while individual blocks may vary dramatically from cycle to cycle, the total power of the chip typically has a well known maximum current consumption. This upper-bound on the total current can be computed using either chip-level Verilog simulation or by scaling the maximum power of a similar design in an older technology. Other constraints expressing dependencies between different circuit blocks or expressing dependencies between different clock cycles can be added as well using linear inequalities, as explained in the following Section.

To compute $\Delta V_{m,n,i}^{u,V_{dd}}$ and $\Delta V_{m,n,i}^{u,V_{ss}}$, a unit trapezoidal current source waveform is, in turn, applied at each circuit block and the voltage drop of all nodes is measured for S subsequent clock cycles, till the voltage drop becomes insignificant. This is a time consuming step but for typically processor design at most a few tens of circuit blocks are required and the simulation is performed only once for each circuit block, after which the results can be reused for the analysis of any number of circuit paths. The optimization in EQ18 through EQ23 is implemented using a CPLEX linear optimization package. For typical power grids, the number of variables is of the order of thousands of variables, which can be easily solved using standard linear solution methods. Finally, we note that the optimization solution not only provides the maximum expected increase in the circuit path delay, but also will provide the exact current waveforms for each circuit blocks that produce this delay variation. Such a worst-case ‘‘block current trace’’ can be simulated by the design to verify the predicted delay change and can give insight into the operation of the supply grid.

Generation of block current constraints

Equations 22 and 23 express simple constraints on the current of individual blocks or the total current of the processor as a whole. However, in most processor designs, correlations between the currents of different blocks, or between currents of a block in consecutive clock cycles will also arise. For instance, positive correlation between the current of two pipeline stages can arise when data is passed from one pipeline stage to the next, or negative correlation may exist between the currents of two circuit blocks that operate mutually exclusively.

We therefore incorporate linear constraints in the proposed formulation to express such correlations. It should be noted that the delay maximization formulation is conservative, meaning that it will over estimate the change in delay due to supply voltage fluctuations. This is the result of the optimization formulation, which automatically maximizes the delay change within the bounds of the provided constraints. Incorporating additional constraints in the analysis is therefore an effective method to reduce the conservatism of the analysis.

Any linear constraint can be represented in the proposed formulation and a number of different approaches of automatically generating such constraints can be used. In this paper, we propose the use of gate level power simulation, such a Verilog based simulator, to extract correlation constraints. By simulating a large set of chip level simulation vectors, the correlation between the currents of different blocks in one clock cycle or between currents of blocks in different clock cycles can be observed and can be represented using linear constraints.

In Figure 8, we show an example of the correlation between the currents of a Multiplier and an ALU block in an Alpha processor. The X-axis of the scatter plot corresponds to the current of the Multiplier block and the Y-axis corresponds to the current of the ALU. The entire processor design was simulated, and the current of the ALU and Multiplier blocks were computed using pre-characterized

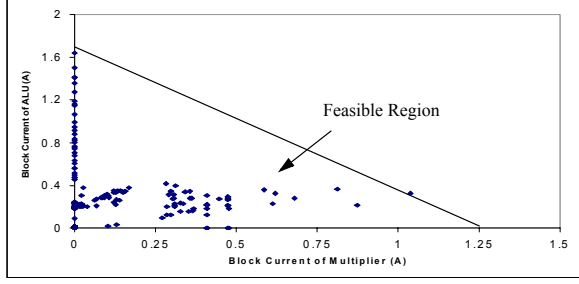


Figure 8. Correlation between Multiplier and ALU block currents.

power data in the cell library. Each point in the scatter plot represents a simulated clock cycle. In total, more than ten thousand clock cycles were simulated using a number of benchmark programs. Note that many of the scatter points coincide. Since the Alpha processor is a single issue machine and was designed with clock gating for reduced power consumption, the Multiplier and ALU blocks cannot be active in the same clock cycle. This negative correlation is evident from the L-shaped scatter points in Figure 8. To express this correlation in the delay maximization formulation, we generate the linear constraint as shown by the solid line in Figure 8 and expressed it with the following inequality:

$$I_{mult,t} + 1.36I_{ALU,t} \leq 1.7 \quad (\text{EQ 24})$$

It is clear that the constraint in EQ24 will reduce predicted delay increase of the analysis by preventing the Multiplier and the ALU from simultaneously exhibiting their maximum current values.

An example of a correlation between currents in different clock cycles is shown in Figure 9, where the current of the instruction

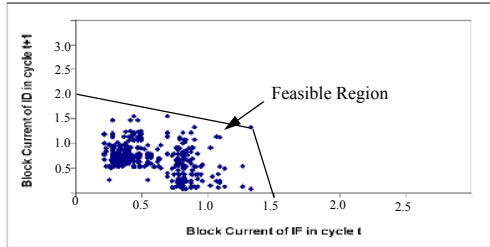


Figure 9. Correlation between IF stage in cycle t and ID stage in cycle $t+1$.

fetch stage in cycle t is plotted against the current of the instruction decode stage in cycle $t+1$. Since data is passed from the instruction fetch stage to the instruction decode stage, a correlation can arise, as clearly visible from the scatter plot in Figure 9. In this case, the correlation is captured using two constraints, as illustrated in Figure 9 and expressed as follows:

$$1.7I_{IF,t} + I_{ID,t+1} \leq 3.5 \quad (\text{EQ 25})$$

$$9.6I_{IF,t} + I_{ID,t+1} \leq 14.4 \quad (\text{EQ 26})$$

Although in this paper we manually extract constraints from the correlation data, it is clear that such constraints could be easily generated automatically by finding a polyhedron that encompasses all generated current points. The use of gate level power simulation has the advantage that very extensive suites of test vectors are readily available and block current data can be obtained from them with minimum overhead during the design process. Also, gate level simulation is typically performed for many millions of clock cycles. The proposed approach allows realistic constraints to be extracted, based on extensive simulation data while at the same time avoiding the need to evaluate long power grid vectors, that will lead to intractable simulation times.

3.1 Voltage drop formulation

We observe that the proposed method for delay maximization can be easily reformulated to compute the maximum voltage drop at a particular circuit node. In this case, we maximize the voltage drop, again subject to linear constraint and with the block currents as optimization variables, as follows:

Maximize

$$\Delta V_{dd,n} = \sum_{m=0}^{M-1} \sum_{i=0}^{S-1} \Delta V_{m,n}^{u,V_{dd}} I_{m,S-i} \quad (\text{EQ 27})$$

such that

$$\sum_{i=1}^N I_{m,i} \leq I_{peak} \quad (\text{EQ 28})$$

$$I_{min,i} \leq I_{m,i} \leq I_{max,i} \quad (\text{EQ 29})$$

for all $n = 1, 2, \dots, M$. Note that this formulation accounts for both IR-drop and Ldl/dt drop.

4 Results

The proposed approaches for determining the worst-case voltage drop and maximum increase in delay of a path were implemented and tested on a number of grids of different sizes for both flip-chip and wire bond package models. Grid-1 through Grid-8 are different size grids in 9 layers of metal, generated using pitches and widths of an industrial microprocessor design. Grid-9 is the grid of an industrial processor, extracted using a commercial extraction tool and consists of over 1 million elements. For each chip, design was partitioned into a number of blocks. The maximum and minimum current of each block and the total maximum power of the chip was then obtained through either Verilog simulation or chip area estimates.

Table 2 shows the results for worst-case voltage drop computation, using the approach described in Section 3.1. We compare the obtained results with two traditional approaches for voltage drop analysis. In the first approach (*Peak Curr*) shown in Table 2, all blocks are assigned their maximum switching current, so as to draw peak simultaneously. In the second approach (*Avg. Curr*), we assign an average current to each block. The last column shows the voltage drop obtained from the constrained maximization approach, where blocks with low sensitivity will be switching with lower currents while blocks with higher sensitivity will switch with higher currents. The current drawn by each block will change in every clock cycle so as to maximize the voltage drop at a given node due to both IR-drop and Ldl/dt drop. Table 2 shows that the peak current approach overestimates the worst-case voltage drop by a maximum of 64% and by 37% on average over all test cases. On the other hand, the average current approach underestimates the worst-case drop by as much as 61% and by 51% on average.

Table 2. Comparison of Worst-Case Voltage Drops Using Different Approaches

Grid	Grid Type	# of nodes	# of Blocks	Worst voltage drops		
				Peak Curr/Block (mV)	Average Curr/Block (mV)	Constr. Max (mV)
Grid-1	WB	1051	10	258.2	96.8	170.8
Grid-2	WB	1051	16	295.3	105.5	193.3
Grid-3	FC	1691	16	121.9	43.5	109.0
Grid-4	WB	1691	20	195.2	90.1	166.8
Grid-5	FC	2438	20	172.2	57.4	147.7
Grid-6	WB	2438	25	232.8	76.7	141.9
Grid-7	FC	3818	25	149.1	43.9	112.9
Grid-8	WB	3818	30	247.2	81.9	178.3
Grid-9	FC	1,57,180	30	190.3	69.2	134.7

Table 3, show the results of the proposed delay maximization approach. Table 3 shows the maximum expected delay increase of a critical path for each chip as determined by the proposed constrained optimization approach (*Constr. Max*). The results are compared with two traditional approaches. In traditional approach 1, the worst-case voltage drop of power supply network is applied at all voltage supply points of the gates constituting the critical path. This is equivalent to the common practice of lowering the operating voltage of all cells in the library by the worst-case expected voltage drop during timing characterization. Table 3 shows that this approach over-estimates the increase in delay compared to the constrained maximization approach by 135% on an average. It should be noted however, that the over-estimation depends on the placement of the gates in the path on the chip, giving a worse over-estimation of the delay increase for paths that are distributed over a significant area of the die. In traditional approach 2, the worst voltage drop at each gate location is first determined using the constrained voltage maximization formulation described in Section 3.1. Each local worst-case drop is then applied simultaneously at all gates in the path. This approach is therefore less conservative than traditional approach 1 since many nodes have a local worst-case drop that is less than the worst-case drop of the chip as a whole. Nevertheless, this approach is also conservative and Table 3 shows that this approach still over-estimates the delay by 44.7% on average compared with the constrained delay maximization approach.

Table 3. Comparison of Increase in Delay Using Different Approaches

Grid	# of Nodes	Increase in Delay of a path		
		Traditional Approach 1	Traditional Approach 2	Constr Max
Grid-1	1051	18.39%	14.42%	10.34%
Grid-2	1051	10.37%	8.07%	5.25%
Grid-3	1691	15.87%	7.95%	4.96%
Grid-4	1691	14.05%	4.83%	2.60%
Grid-5	2438	13.50%	9.64%	6.41%
Grid-6	2438	10.74%	6.95%	3.81%
Grid-7	3818	16.97%	10.40%	8.44%
Grid-8	3818	12.82%	8.94%	6.54%
Grid-9	1,57,180	16.25%	8.16%	6.50%

In Table 4, we demonstrate the effectiveness of incorporating additional constraints between block currents into the formulation. We repeated the analysis of Grid-1 of the Alpha processor, but added several linear constraints expressing correlations between currents of different blocks and between block currents in different clock cycles. The constraints were obtained using extensive Verilog simulation, as described in Section 3. Table 4 shows the increase in delay of 5 critical paths with and without these correlation constraints. Although only a few constraints were added to the analysis, the delay increase improved by as much as 21.7%, and by 16.5% on average, showing the effectiveness of this approach.

Table 4. Impact of Correlation Constraints on Increase in Delay

Critical Paths	Increase in Delay of Critical Path		% Improvement
	Without Correlation Constraints	With Correlation Constraints	
Path 1	8.92%	6.98%	21.7%
Path 2	8.40%	7.03%	16.3%
Path 3	10.91%	9.11%	16.49%
Path 4	10.88%	9.58%	11.95%
Path 5	9.36%	7.85%	16.13%

In Figure 9, the current waveforms generated by delay maximization approach for Grid-1 are shown. As can be seen, the currents generated by the analysis are time varying and exploit the time dependence of IR-drop and Ldi/dt drop. The run time for the linear

optimization was less than 1sec for all the grids since the linear optimizer can solve linear maximization problems very quickly. The initial step of computing sensitivities is computationally intensive in this approach but it can be considerably reduced using fast linear solvers.

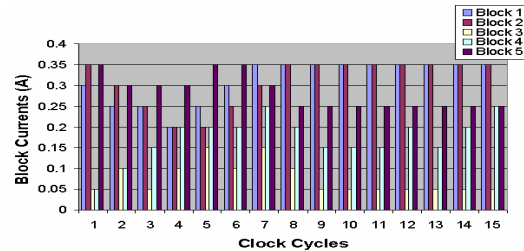


Figure 10. Variation of Block currents with time to maximize delay

5 Conclusion

In this paper, we have presented a new approach for computing the maximum delay increase of critical path due to power supply voltage fluctuations. The analysis is vectorless while considering both IR-drop and Ldi/dt drop. We presented an accurate model for the path delay as a function of the supply voltages and then formulated the delay maximization problem as a constrained linear optimization problem. We also discuss how linear constraints can be added to the formulation to represent correlations between block currents. The analysis was implemented and tested on a number of benchmark grids, including the power grid of an industrial processor and we demonstrate the effectiveness of the proposed approach.

6 Acknowledgement

This work was funded by research grants and contracts from SRC, NSF, Intel and IBM.

References

- [1] G. Steele, D. Overhauser, S. Rochel and Z. Hussain, "Full-chip verification methods for DSM power distribution systems," in *DAC*, 1998.
- [2] H. Chen and D. Ling, "Power supply noise analysis methodology for deep-submicron VLSI chip design," in *DAC*, pp. 638-643, 1997.
- [3] S. Zhao, K. Roy and C. K. Koh, "Frequency domain analysis of switching noise on power supply network," in *ICCAD*, pp. 487-492, 2000.
- [4] R. Panda, D. Blaauw, R. Chaudhry, V. Zolotov, B. Young and R. Ramaraju, "Model and analysis for combined package and on-chip power grid simulation," in *Proc. of the ISLPED*, pp. 179-184, 2000.
- [5] S. R. Nassif and J. N. Kozhaya, "Fast power grid simulation," in *Proc. Design Automation Conference*, pp. 156-161, 2000.
- [6] S. Taylor, "The challenge of designing global systems," in *Proc. IEEE Custom Integrated Circuits Conference*, pp. 429-435, 1999.
- [7] M. Zhao, R. V. Panda, S. S. Sapatnekar and D. Blaauw, "Hierarchical analysis of power distribution networks," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pp. 159-168, 2002.
- [8] H. Kriplani, F. Najm and I. Hajj, "Pattern independent minimum current estimation in power and ground buses of CMOS VLSI circuits," *IEEE Trans. on Computer-Aided Design*, pp. 998-1012, 1995.
- [9] A. Krstic and K. Cheng, "Vector generation for maximum instantaneous current through supply lines for CMOS circuits," in *Proc. Design Automation Conference*, pp. 383-388, 1997.
- [10] Y. M. Jiang, T. Young and K. Cheng, "VIP - an input pattern generator for identifying critical voltage drop for deep submicron designs," *Proc. ISLPED*, pp. 156-161, 1999.
- [11] S. Bobba and I. N. Hajj, "Maximum voltage variation in the power distribution network of VLSI circuits with RLC models," in *Proc. Intl. Symposium of Low Power Electronics and Design*, 2001.
- [12] D. Sylvester and K. Keutzer, "Getting to the bottom of deep submicron," *Proc. Computer-Aided Design*, pp. 203-211, 1998.
- [13] L. H. Chen, M. Sadowska and F. Brewer, "Coping with buffer delay change due to power and ground noise," *Proc. DAC*, 2002.
- [14] Y. M. Jiang and K.T. Cheng, "Analysis of performance impact caused by power supply noise in deep submicron devices," *Proc. Computer-Aided Design*, pp. 760-765, 1999.
- [15] A. Chandrakasan, W. J. Bowhill and F. Fox, *Design of high performance microprocessor circuits*. NY: IEEE Press, 2001.
- [16] R. Panda, tutorial, "On chip inductance extraction and modelling," *Intl. Symposium on Quality Electronics Design*, tutorial.
- [17] G. Bai, S. Bobba and I.N. Hajj "RC power bus maximum voltage drop in digital VLSI circuits," *Intl. Symposium on Quality Electronics Design*.