

# **IBM Research Report**

## **Improving the Accuracy of Power Grid Simulation**

**Sani R. Nassif, Haihua Su**  
IBM Austin Research Laboratory  
11501 Burnet Rd.  
Austin, TX 78758



**Research Division**  
**Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich**

# Improving the Accuracy of Power Grid Simulation \*

Sani R. Nassif  
IBM Austin Research Laboratory  
11501 Burnet Rd.  
Austin, TX 78758  
nassif@us.ibm.com

Haihua Su  
IBM Austin Research Laboratory  
11501 Burnet Rd.  
Austin, TX 78758  
haihua@us.ibm.com

## ABSTRACT

As technology scales the power supply  $V_{dd}$  is being lowered in order to lower operating power and to meet device reliability requirements. Because of the increase in noise and leakage, however, the threshold voltage  $V_T$  is not being lowered at the same rate as  $V_{dd}$ . This results in an increase in the sensitivity of circuit delay to power supply noise, and underscores the need to perform detailed analysis of the on-chip power distribution for noise, robustness and reliability. A common technique applied in the analysis of on-chip power grids is to separate the linear and non-linear components of the problem and treat them separately (see for example [2], [4] and [7]). With the increasing sensitivity of delay to power grid noise, this artificial separation can lead to subtle errors, especially for otherwise marginal designs.

This paper describes a technique to improve the accuracy of power grid analysis by improving the coupling between the linear and non-linear parts of the power analysis problem while maintaining the efficiency and scalability required for full chip analysis.

## Categories and Subject Descriptors

B.7.2 [Hardware]: Integrated Circuits—*Design Aids*

## General Terms

Power Grid Analysis

## 1. INTRODUCTION

With the advent of deep sub-micron technologies, we are observing a rapid increase in operating frequency and power dissipation (see for example [1]). Much of this progress is owed to aggressive technology lithography scaling whereby device and interconnect dimensions are reduced [8]. With

\*(Produces the permission block, copyright information and page numbering). For use with ACM PROC. ARTICLE-SPCLS V2.0. Supported by ACM.

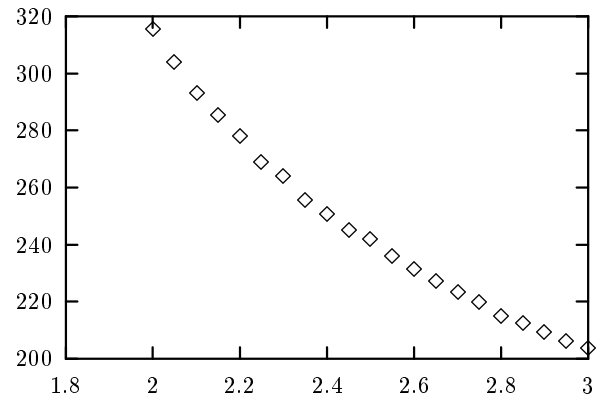


Figure 1: Delay in ps vs. power supply voltage.

MOSFET channel length reduction, however, the prevailing constant-field scaling theory [3] requires a reduction in the power supply voltage ( $V_{dd}$ ) in order to meet device reliability requirements.

Because of concerns about excessive background leakage currents and cross-coupling induced noise, the threshold voltage  $V_T$  is not being reduced at the same rate as the power supply voltage  $V_{dd}$ . This means that the maximum over-drive  $V_{dd} - V_T$  is decreasing as a percentage of  $V_{dd}$ . This causes an increase in the sensitivity of circuit delay to power supply fluctuations. Figure 1 shows the delay of a buffer with a fanout of 3 (i.e. loaded by three copies of itself) as a function of power supply. It is clear that the sensitivity of delay to  $V_{dd}$  increases as the over-drive is lowered.

Consider the circuit shown in figure 2 which can be thought of as a canonical model of a power grid and loading circuit. In the figure,  $L$  models the package inductance,  $R_g$  models the grid resistance,  $R_d$  and  $C_d$  model the local decoupling capacitance, and  $I_{load}$  models the time dependent current waveform of the load. We model  $I_{load}$  as the following periodic triangular waveform (where  $T$  denotes the period):

$$I_{load} = \begin{cases} 0 & : t < 0 \\ \mu t & : t < t_p \\ \mu(2t_p - t) & : t < 2t_p \\ 0 & : 2t_p < t < T \end{cases} \quad (1)$$

We use data from the International Technology Roadmap for Semiconductors [8], summarized in table 1 to predict the dependence of the maximum voltage drop  $V_{drop}$  on the var-

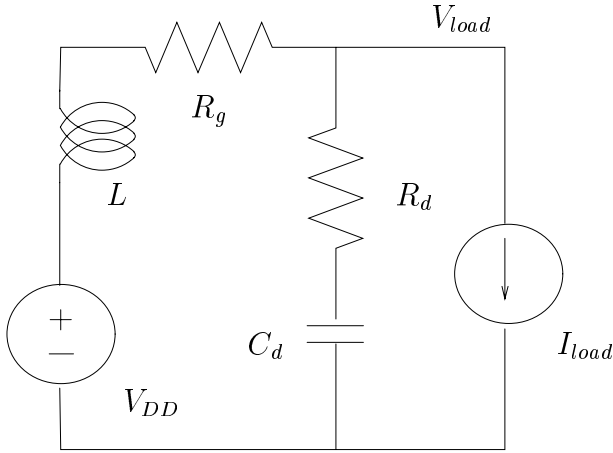


Figure 2: Canonical power circuit.

Year	$L_{eff}$ nm	$f_{max}$ MHz	$V_{dd}$ V	Size $mm^2$	Power W	Density $W/mm^2$
1999	140	1200	1.8	450	90	0.2
2000	120	1321	1.8	450	100	0.22
2001	100	1454	1.5	450	115	0.26
2002	85	1600	1.5	509	130	0.26
2003	80	1724	1.5	567	140	0.25
2004	70	1857	1.2	595	150	0.25
2005	65	2000	1.2	622	160	0.26

Table 1: Technology Roadmap Parameters.

ious circuit parameters in order to predict trends in power-grid-induced noise with technology scaling. It was recently shown [6] that the maximum noise for this circuit can be well approximated by

$$V_{max} = \mu L + \mu R_g t_p - \mu C_d R_g^2 (1 - e^{-t_p/\tau}) \quad (2)$$

where:

$$\tau = (R_g + R_d)C_d \quad (3)$$

We note that  $t_p \propto freq^{-1}$ , and that power density  $P_{\square} \propto V_{dd}\mu t_p$  thus  $\mu \propto P_{\square} freq/V_{dd}$ . Based on the trends in table 1  $t_p$  will decrease by 0.6X and  $\mu$  will increase by 3.21X. In order to keep  $V_{max}$  constant (i.e. keep the same amount of noise as a percentage of  $V_{dd}$ ) we need to dramatically increase the decoupling capacitance term in Eq. (2):  $C_d R_g^2 (1 - e^{-t_p/\tau})$ . This can only be accomplished by careful analysis and design of the power grid and by judicious placement and sizing of decoupling capacitance [10]. This also means that the *accuracy* of power estimation and power grid analysis becomes more critical with technology scaling.

## 2. FULL-CHIP POWER GRID ANALYSIS

Powergrid analysis requires modeling the grids and the power *sources* and *drains*. Power distribution within an integrated circuit starts from the top-level metal layer which is connected to the package, connects to lower level of metal through inter-layer *vias*, and terminates at contacts to the active devices. The metal wires and vias are modeled as a linear, passive time invariant network consisting of resistive

capacitive and inductive elements. Power *sources* are modeled as ideal voltage sources connected to the power grid either directly or through a linear network of parasitic elements.

Power *drains* are more difficult to model because they need to account for (a) the complex interaction between the power grid, the underlying non-linear circuit, and (b) the time-varying signals propagating across the integrated circuit. Current designs, however, have power grids with millions of nodes and tens of millions of elements and we will see below that such large networks can be solved much more efficiently when they are linear; hence the prevailing practice of modeling the power drains as time-varying current sources. The complete power grid model is composed of a linear network of RLC elements excited by constant voltage sources and time varying current sources. The behavior of such a system is described using an MNA formulation as the following ordinary differential equation:

$$Gx + C\dot{x} = u(t) \quad (4)$$

Where  $x$  is a vector of node voltages, and source and inductor currents;  $G$  is the conductance matrix;  $C$  includes the capacitance and inductance terms, and  $u(t)$  denotes the time varying sources modeling the sources and drains.

Due to the large size of typical power grids, general circuit simulators such as Spice [5] are not adequate for power grid analysis because they use general purpose solution methods meant to be robust in the face of stiff systems of equations. By contrast, power grids are well behaved spatially (nearly regular) and temporally (damped). This motivates special-purpose simulation tools for power grids which can make use of these properties [4, 9].

If we apply the Backward Euler integration formula to Eq. (4) we get a system of linear equations:

$$(G + C/h)x(t+h) = u(t+h) + C/hx(t) \quad (5)$$

which can be readily simplified to  $Ax(t+h) = b$  with  $A = G + C/h$  and  $b = u(t+h) + C/hx(t)$ .

The solution of Eq. (5) requires the factorization of the matrix  $G + C/h$  which is independent of  $x$ , time-invariant, large and sparse. If we hold the time step  $h$  constant then only *one* initial factorization is required, with a forward/backward solve at each time step. This results in *dramatic* computational effort savings that make full-chip power grid analysis possible.

Our purpose in this paper is to improve the accuracy of this analysis technique by modeling the nonlinear dependence of the drain currents ( $u(t)$  in Eq. (4)) on the local values of supply voltage  $x$ , while at the same time preserving the extraordinary efficiency of the single-factorization scheme outlined above.

## 3. DRAIN MODELING

Our objectives to come up with a first order model for the dependence of drain current (the components of  $u(t)$  in Eq. (5)) on the corresponding value of power grid voltage. For a drain connected to nodes  $a$  and  $b$ , the local power

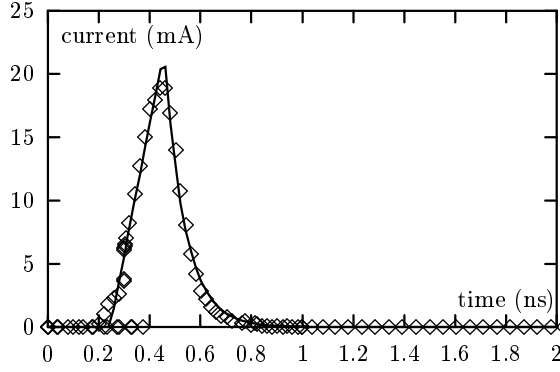


Figure 3:  $I_{dd}$  current waveform for a CMOS buffer.

grid voltage is simply  $x_a - x_b$ . We define the usual drain incidence matrix  $\mathcal{A}$ .

Consider a simple CMOS non-inverting buffer formed from the cascading of two CMOS inverters. Based on observing the waveform of the power supply current  $I_{dd}$  under a variety of conditions we choose to model the waveform as:

$$I_{dd} = \begin{cases} 0 & : t < 0 \\ I_p t / T_r & : t < T_r \\ I_p \eta e^{T_r - t} & : t > T_r \end{cases} \quad (6)$$

An example of a simulated and fitted waveform is illustrated in figure 3. For this example we used transistor model parameters from a  $0.25\mu$  CMOS process running at a voltage supply of 2.5 volts and the power supply current waveform model parameters were  $T_r = 0.212$  ns,  $I_p = 21.905$  mA and  $\eta = 12.351$  ns<sup>-1</sup>.

Our goal is to examine the dependence of  $I_{dd}$  on  $V_{dd}$  in order to find a suitable model to incorporate into the linear system Eq. (5). To explore this dependence we simulate the buffer over a supply voltage range of  $\pm 20\%$  and model the dependence of the  $I_{dd}$  model parameters ( $T_r$ ,  $I_p$  and  $\eta$ ) on  $V_{dd}$ . We find that all three parameters are well predicted by simple linear functions of  $V_{dd}$  with correlation coefficients in excess of 0.96.

We collect the individual load currents ( $I_{dd}$  components denoted by  $\mathcal{I}$ ) into the right-hand-vector of Eq. (5) via the standard nodal incidence matrix  $\mathcal{Z}$ , thus  $u(t) = \mathcal{Z}^T \mathcal{I}(t)$ . We then simply rewrite 5 as:

$$(G + C/h)x(t+h) = \mathcal{Z}^T \mathcal{I}(t+h) + C/hx(t) \quad (7)$$

Unfortunately, the dependence of  $\mathcal{I}(t+h)$  on the values of the current system variables  $x(t+h)$  invalidates the constant system matrix property. In order to regain it, we choose to delay the dependence of  $\mathcal{I}$  on  $x$  by rewriting the system as:

$$(G + C/h)x(t+h) = \mathcal{Z}^T \mathcal{I}(t) + C/hx(t) \quad (8)$$

thus regaining the constant  $G + C/h$  system matrix. This can also be thought of as a mixing of implicit and explicit integration methods. We have determined empirically that

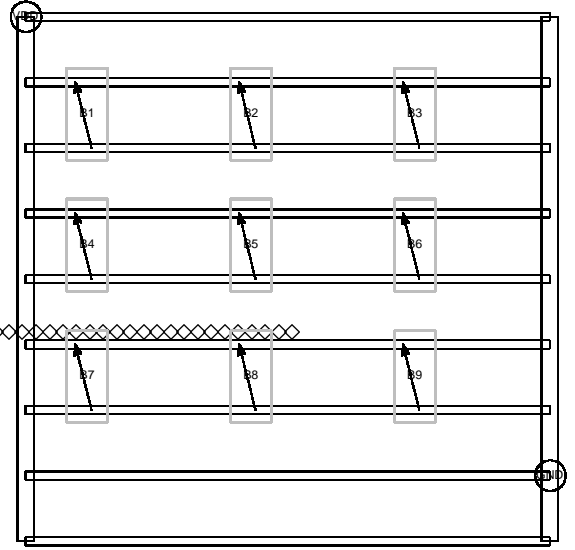


Figure 4: Example power grid.

for sufficiently small time step  $h$  (in the range of 1% of the clock period), the relatively weak dependence of  $\mathcal{I}$  on  $x$  does not play a part in the stability of the overall solution method.

#### 4. EXAMPLE 1

We construct a simple power grid composed of two vertical and 9 horizontal wire segments with a single connection to power supply and ground. We place nine buffers connected to the grid as illustrated in figure 4.

We perform three simulations of the system illustrated in figure 4:

1. We use a full Spice [5] simulation to solve the linear and non-linear parts together. The waveforms for this case (which we denote A) are shown in figure 5.
2. We use Eq. (8) to simulate the power grid with load current dependence on  $V_{dd}$ . The  $I_{dd}$  waveform model parameters were modeled as linear functions of the local supply voltage. The equations were:  $T_r = 0.5 - 0.11V_{dd}$  ns,  $I_p = 18.9V_{dd} - 25.3$  mA and  $\eta = 4.28V_{dd} + 1.72$  ns<sup>-1</sup>. The waveforms for this case (which we denote B) are shown in figure 6.
3. We assume that the buffer current does not depend on  $V_{dd}$  and that the  $I_{dd}$  waveform model parameters were  $T_r = 0.218$  ns,  $I_p = 21.589$  mA and  $\eta = 12.351$  ns<sup>-1</sup>. The waveforms for this case (which we denote C) are shown in figure 7.

Table 2 shows the minimum  $V_{dd}$  value at each of the buffers, the maximum voltage drop at each of the buffers, and the error (from case A) in the predicted drop for cases B and C.

We find that the average error in predicted maximum voltage drop for case C is 27.80% while the average error for

V(A)	V(B)	V(C)	$\Delta A$	$\Delta B$	$\Delta C$	err B	err C
2.314	2.295	2.259	0.186	0.205	0.241	10.2	29.6
2.329	2.312	2.280	0.171	0.188	0.220	9.94	28.7
2.331	2.313	2.283	0.169	0.187	0.217	10.7	28.4
2.347	2.331	2.304	0.153	0.169	0.196	10.5	28.1
2.361	2.347	2.321	0.139	0.153	0.179	10.1	28.8
2.370	2.355	2.336	0.130	0.145	0.164	11.5	26.2
2.380	2.368	2.347	0.120	0.132	0.153	10.0	27.5
2.386	2.374	2.357	0.114	0.126	0.143	10.5	25.4
2.420	2.411	2.398	0.080	0.089	0.102	11.3	27.5

Table 2: Analysis of waveforms for Example 1.

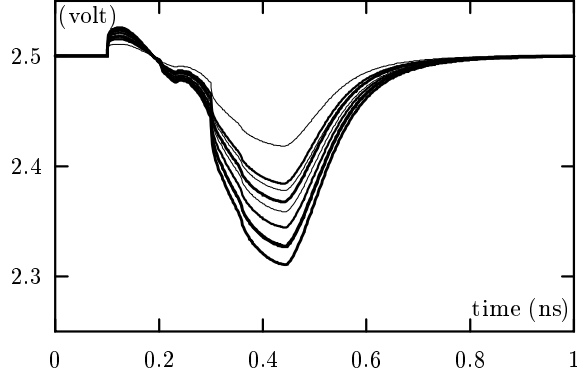


Figure 5: Waveforms from Spice.

case B is only 10.53%, a reduction in error of a factor of 2.6. The only computational difference between cases B and C was the evaluation of the linear equation relating the supply current waveform model parameters to  $V_{dd}$  the impact of which was too small to measure in comparison to the computational effort required to perform the matrix operations required at each time step.

## 5. EXAMPLE 2

For this example we use a real power grid consisting of five levels of metallization and encompassing an area of approximately 8mm by 8mm. At the lowest level, the grid has 5403 horizontal tal wires. At the topmost level the grid is connected

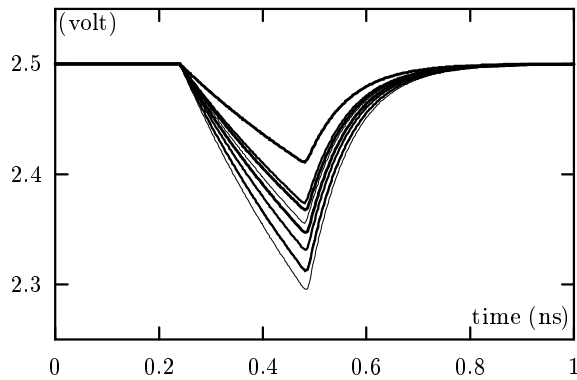


Figure 6: Waveforms with delayed  $I_{dd}$  dependence.

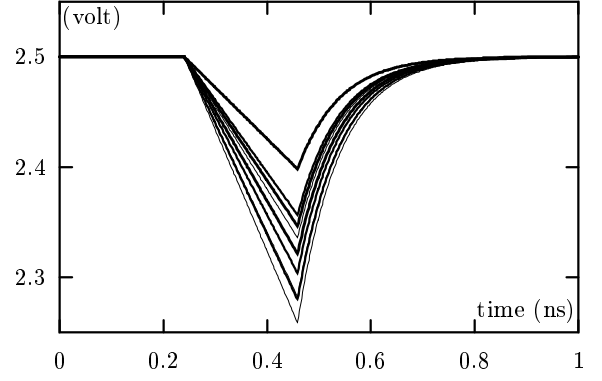


Figure 7: Waveforms with no  $I_{dd}$  dependence on  $V_{dd}$ .

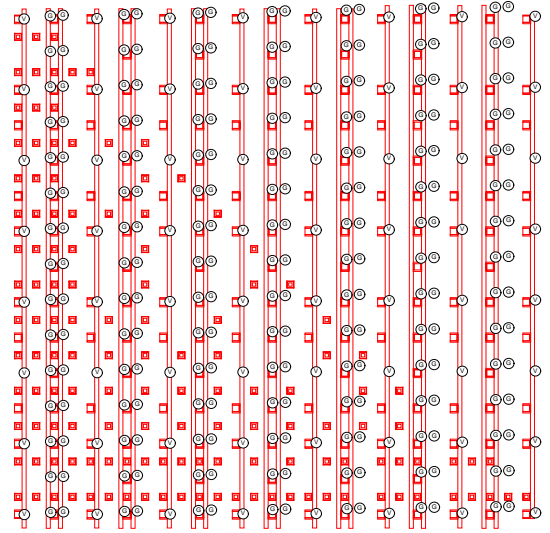


Figure 8: Example power grid (top level metal).

to the power supply via 210 ground C4s and 64  $V_{dd}$  C4s. The top level metal of the grid is illustrated on figure 8. The simulation model for the grid contains approximately 66000 nodes, making it impractical for simulation in Spice. Typical simulation times with a specialized internal IBM tool were about 60 seconds on a modest personal computer running Linux.

We populate the grid with 5000 latches, each of which is modeled in a manner similar to the model for the buffer in the previous example. We also included a uniform background decoupling capacitance to model the non-switching circuitry. We performed two simulations:

1. We use Eq. (8) to simulate the power grid with load current dependence on  $V_{dd}$ . The  $I_{dd}$  waveform model parameters were expressed as linear functions of the local supply voltage:  $T_r = 1.21 - 0.268V_{dd} \text{ ns}$ ,  $I_p = 7.42V_{dd} - 10.46 \text{ mA}$  and  $\eta = 3.28V_{dd} - 2.36 \text{ ns}^{-1}$ . We

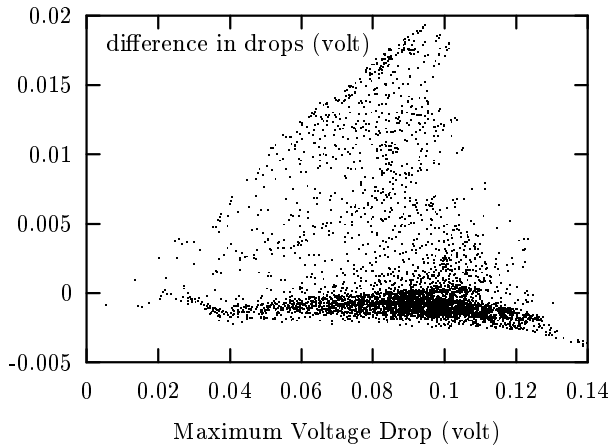


Figure 9: Error between cases D and E vs. results for D.

denote this simulation by case D.

2. We assume that the buffer current does not depend on  $V_{dd}$  and that the  $I_{dd}$  waveform model parameters were  $T_r = 0.54 \text{ ns}$ ,  $I_p = 8.09 \text{ mA}$  and  $\eta = 5.84 \text{ ns}^{-1}$ . We denote this simulation by case E.

For each of the simulations, we determined the maximum voltage drop at each of the latches. Figure 9 shows the difference between the maximums for cases D and E vs. the value of the maximum for case D. We note that the difference can be up to 20mV in a total drop of 100mV, an error of 20%.

For case D, in order to further understand the impact of time step on this mixed explicit/implicit integration method, we performed the same time domain simulation with a variety of time steps from 0.002 ns to 0.02 ns. We measured the percentage error in the integral of the voltage droop at each latch node in the circuit and compared it to the result with the smallest time step. The Y-axis of Figure 10 plots the average percentage error of those nodes with respect to the time step chosen. The errors show a linear relationship between average error and time step, with error values in the range of 1%.

## 6. CONCLUSION AND FUTURE WORK

In this paper we have represented an algorithm to improve the accuracy of power grid analysis by accounting for the coupling between power grid voltage and the loading on the power grid. The algorithm is very computationally efficient, but requires that compact analytical models of the dependence of the loading on the power supply voltage be available. We show examples of how such models can be generated and used with a significant improvement in accuracy over the case where such dependence is ignored.

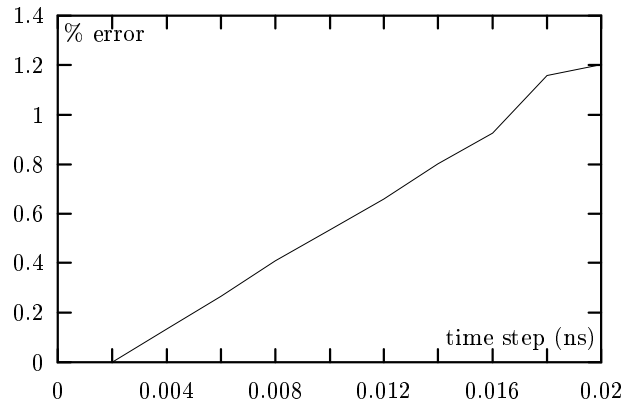


Figure 10: Error versus time step.

The next challenge is to include the impact of power supply noise on the overall timing of a complete chip. This will require modelling the power supply current and the delay as a function of local power supply voltage.

## 7. REFERENCES

- [1] Power4 focuses on memory bandwidth. *Microprocessor Report* Oct 1999.
- [2] H. Chen and D. Ling. Power supply noise analysis methodology for deep-submicron vlsi chip design. In *Proceedings of DAC* 1997.
- [3] R. Dennard et al. Design of ion-implanted mosfets with very small dimensions. *IEEE Journal Solid State Circuits*, Oct 1974.
- [4] A. Dharchoudhury, R. Panda, D. Blaauw, R. Vaidyanathan, B. Tutuianu, and D. Bearden. Design and analysis of power distribution networks in powerpc microprocessors. In *Proceedings of DAC* 1998.
- [5] L. Nagel. *SPICE2: A Computer Program to Simulate Semiconductor Circuits*. PhD thesis, University of California, Berkeley, 1975.
- [6] S. Nassif and O. Fakhouri. Technology trends in power-grid-induced noise. In *Proceedings of SLIP*, 2002.
- [7] S. Nassif and J. Kozhaya. Fast power grid simulation. In *Proceedings of DAC* 2000.
- [8] Semiconductor Industry Association, [http://public.itrs.net/Files/1999\\_SIA\\_Roadmap](http://public.itrs.net/Files/1999_SIA_Roadmap). *The International Technology Roadmap for Semiconductors*, 1999.
- [9] G. Steele, D. Overhauser, S. Rochel, and S. Z. Hussain. Full-chip verification methods for dsm power distribution systems. In *Proceedings of DAC* 1998.
- [10] H. Su, S. Sapatnekar, and S. Nassif. An algorithm for optimal decoupling capacitor sizing and placement. In *Proceedings of ISPD* 2002.