

Power/Ground Network Optimization Considering Decap Leakage Currents

Yici Cai, *Member, IEEE*, Jingjing Fu, Xianlong Hong, *Fellow, IEEE*, Sheldon X.-D. Tan, *Senior Member, IEEE*, and Zuying Luo

Abstract—In this brief, the authors take a first look at the leakage effects of decaps in power/ground (P/G) grid optimization. Through the use of an approximate leakage current model, it is revealed that simple usage of the leakage model in traditional optimization methods cannot help in reducing noises on P/G grids, and it even hurts power consumption due to overadded decaps. Therefore, it is necessary to develop an efficient method to budget decaps when leakage effect is considered. Here, a new two-stage approach to solve this problem is proposed. Experimental results demonstrate the effectiveness of our new method.

Index Terms—Decap, leakage, optimization, power/ground (P/G) network.

I. INTRODUCTION

AS TECHNOLOGY scales down to 90 nm and below, power consumption is becoming the limiting factor in high-performance VLSI chip design, and power reduction becomes a research-intensive area [1], [2]. At the same time, robust power delivery is also considered as one of the grand challenges. As predicted in the roadmap for integrated circuit (IC) development from the ITRS-2002 update [1], the chip working frequency and supply voltage will continue to scale aggressively, both of which will lead to significant Ldi/dt noises and IR drops on power/ground (P/G) networks [3]–[11], which will further affect the performances and reliabilities of VLSI chips. In the past, power budgeting and power delivering optimization were considered as separate issues. However, this is no longer the case, as leakage currents start to become dominant [12]–[15].

Usually, decaps in different levels (on-board, on-package, and on-die) are used together to reduce dynamic voltage noise of different frequencies [16]. Since on-die metal–insulator–metal or poly–insulator–poly decaps tend to consume large die area [3], MOS-transistor-based decaps are widely used in on-die decoupling, as they can easily be inserted close to the nodes that have

large Ldi/dt noises [6]–[8]. However, it was shown in [12] and [13] that when the grid oxide thickness T_{ox} is smaller than 2 Å, under 90-nm technology or smaller, the leakage current of MOS-based decaps will increase significantly ($10\times$) as T_{ox} decreases within 2 Å [13]. There are mainly two methods to reduce the decap leakage; one is to use less-leakage dielectric material, which may not be available until 2007 [13], and the other is to use thick-oxide MOS transistors as decaps in chip design, which may occupy more chip areas to achieve the same capacitance. Due to the complexity and costs of manufacture, leaky decaps are still used. As overadding leaky decaps will hurt power consumption, new P/G optimization methods, which can consider decap leakage effect and make good tradeoff among noise level, power consumption, decap area, and routing resources, are required for robust and low-power P/G network designs in the near future.

In this brief, we start with analyzing leakage currents for MOS-based decaps, and then based on a simple leakage model, we reveal that voltage violation caused by leakage currents is more sensitive to wire sizing than adding decaps. Therefore, a more sophisticated strategy is developed to optimize the power distribution. The two-stage optimization approach first reduces noise by inserting decaps without considering the leakage, then wire-sizing strategy is employed to reduce IR drops caused by leakage currents of decaps. This brief is organized as follows. Section II presents the leakage current model of decap. Section III briefly introduces how to model and optimize the P/G network and demonstrates the performance of traditional optimizers under new environment. Section IV reveals that optimization methods will harm power consumption and become less reliable when decap leakage currents are modeled in a simple way. In Section V, we present a new effective two-stage leakage-current-aware P/G grid optimization framework, which uses both decap allocation and wire-sizing strategies; in addition, experimental results are given to show effectiveness of our method. Finally, Section VI concludes this brief.

II. LEAKAGE MODEL OF MOS-BASED DECAPS

On-chip decaps can be made of nMOS or pMOS transistors. Because the substrate node of nMOS must be grounded while the substrate node of pMOS must be connected to V_{dd} , connection styles of nMOS-based and pMOS-based decaps are thus different, as shown in Fig. 1.

For a MOS transistor, as technology is scaling down to 90 nm, three types of leakage currents should be considered, including subthreshold leakage, gate leakage, and reverse-biased drain/

Manuscript received April 2, 2005; revised September 16, 2005 and December 30, 2005. This work was supported by the National Natural Science Foundation of China under Grant 60476014 and by the Hi-Tech Research & Development (863) Program of China under Grant 2005AA1Z1230. The work of S. X.-D. Tan was supported by the National Science Foundation (NSF) CAREER Award CCF-0448534, by the University of California MICRO Program Award 04-088, and NSF Grant OISE-0451688. This paper was presented in part at the Asia and South Pacific Design Automation Conference Shanghai, China, January 2005. This paper was recommended by Associate Editor R. W. Newcomb.

The authors are with the Electronic Design Automation Laboratory, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: shi-j03@mails.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TCSII.2006.882207

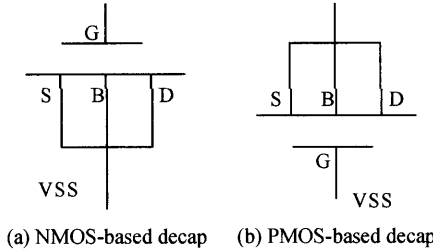


Fig. 1. On-chip decaps made of nMOS or pMOS.

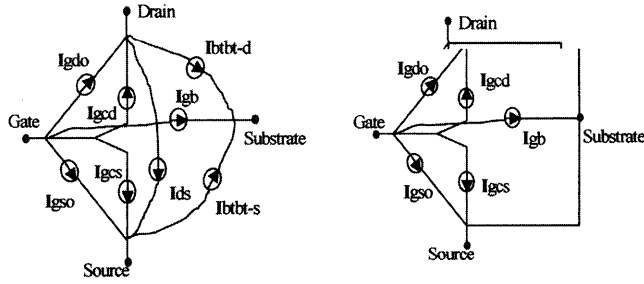


Fig. 2. Leakage models of MOS transistor and MOS-based decap.

substrate and source/substrate junction band-to-band tunneling (BTBT) leakage [12], as shown in Fig. 2(a).

In Fig. 2(a), I_{ds} is a current source to model the subthreshold leakage. It should be ignored when drain/source nodes are shorted. I_{btbt-d} and I_{btbt-s} are two sources to model the BTBT leakage, which should also be canceled when drain/source/substrate nodes are shorted. However, I_{gdo} , I_{gso} , I_{gcd} , I_{gcs} , and I_{gb} are another five sources to model the gate leakage current I_{gate} , which should be considered. That is why the resulting leakage current of the MOS-based decap mainly comes from gate leakage currents as shown in Fig. 2(b).

The leakage current of a MOS-based decap can be formulated as follows [17]–[19]:

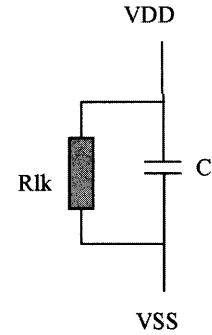
$$I_{gate} = k \times \left(\frac{V_{dd}}{T_{ox}} \right)^2 \times e^{-\alpha \cdot T_{ox}/V_{dd}} \times w_n \quad (1)$$

where α and k are parameters related to specific technology, w_n is the gate width of nMOS (w_p for pMOS), T_{ox} is the oxide thickness, and V_{dd} is the supply voltage. The capacitance of a MOS-based decap can be computed using the following formula:

$$C_d = \gamma \times \frac{\epsilon \times \epsilon_0}{T_{ox}} \times S. \quad (2)$$

where ϵ and ϵ_0 are relative and absolute dielectric constants, respectively, γ is a parameter related to the shape of the MOS transistor, and S is the area that is equal to width w_n times the length L_{eff} of a gate.

It is shown in [12] and [13] that I_{gate} increases by $10\times$ when T_{ox} decreases by 10% around the value of 20 Å. Based on the model presented above, we use BSIM [20] 70-nm technology to simulate an nMOS-based decap with $w_n = 0.2 \mu\text{m}$ and $L_{eff} =$

Fig. 3. $Rlk \parallel C$ model for decap leakage.

$0.038 \mu\text{m}$. The resulting decap is 0.26 fF with $\gamma = 1.5462$ when $\epsilon = 4$. Since different shapes cause different γ , we assign $\gamma = 1$ for simplicity as we do not know the exact shape of each decap.

To consider the leakage current of decaps, we propose a decap model $Rlk \parallel C$ that can capture the main leakage current as shown in Fig. 3. In this model, a resistor Rlk is connected in parallel with the capacitor C , which is used to model the gate leakage current. The value of Rlk can be determined by the supply voltage and the leakage current computed from (1). Unfortunately, this model is only an approximate model because practical leakage current also relate to node voltage. However, for guiding the optimization algorithm, this model is accurate enough to help us obtain very useful profile about how to do optimization efficiently. This simple leakage model used in this brief takes the same spirit of Elmore models that is used to approximate the delay of interconnects [21]. Later, more accurate (thus more complicated) leaky decap models can be added to our framework easily.

III. TRADITIONAL P/G NETWORK MODEL AND OPTIMIZATION METHODS

For the similarity of power and ground networks, we will only describe the optimization for power networks in this brief.

Early optimization approaches size widths of wire segment to reduce static IR drops [4], [5]. The optimization object is to minimize the routing area consumed by P/G grids. However, due to the increasing working frequency and decreasing supply voltage, recent P/G optimization approaches allocate on-chip decaps to reduce dynamic voltage noise [6]–[8]. The optimization object should now include minimizing the die area consumed by decaps under the same noise threshold. Although the approach in [9] simultaneously sizes wire segment widths and decaps for robust transient P/G networks, the leakage currents of decaps are not considered in that algorithm.

In this section, we first introduce our model to analyze and optimize P/G grids. Then, we give some experimental results to show the drawbacks of traditional optimization algorithm without considering the leakage current.

A. Power Delivery Network Modeling

For the row-based standard-cell design style, it is common to use a predefined mesh-like power delivery network. We model the network as follows.

Each wire segment in the power grid is modeled as lumped RLC elements. This is also the case for the parasitics of power

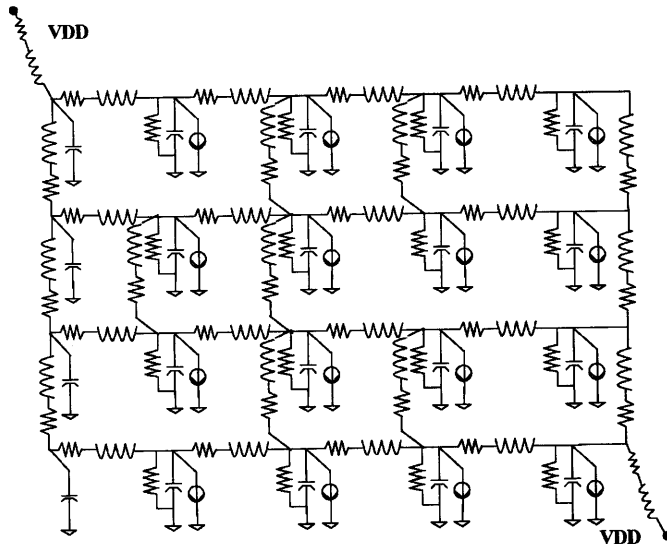


Fig. 4. Model of power network.

pads and packages. The nonlinear CMOS modules are modeled as time-varying current sources, which can be obtained by off-line logic simulations with the assumption that the supply voltage is ideally V_{dd} . We also consider both built-in on-die decaps (n-well capacitors and circuit capacitors) and add-on decaps (thin-oxide capacitors) connected between the power grid and the ground grid. Using the new decap model, which considers the leakage current described in Section II, the mesh-structured P/G grids as a pseudodistributed RLC network is illustrated in Fig. 4.

B. Noise Analysis of P/G Grids

The power delivery network modeled above can be formulated as a first-order matrix differential equation using the modified nodal analysis (MNA) method [22]. The solution of this matrix equation can give the dynamic voltage waveforms of any node in the P/G networks. Then, by giving a safety boundary (design rule), the performance of P/G networks can be verified. For an unreliable P/G design, there exist some nodes that violate the design rule. Fig. 5 shows a voltage waveform at a violation node (VN) on a power grid. This kind of waveform is caused by simultaneous switching events (SSEs). In this case, if decaps are not big enough to sustain the huge current surge, the node voltage would become an unsafe value, which may be lower than V_{min} even at the end of SSE period as shown in Fig. 5. Here, the worst point at time t_1 is probably triggered by the rising edge of a clock period in an inductive large fan-out system [23].

C. Drawbacks of Traditional Optimization Considering the Leakage Current of Decaps

As described in [4]–[8], the design of the P/G network has two stages. The first stage is to size wire widths to reduce static IR voltage drops based on empirical values of power consumption. The second stage is to add on-die decaps to reduce transient voltage drops given the time-varying input vectors. Almost all kinds of traditional optimization processes are based on the fact that more decaps will not hurt the system perfor-

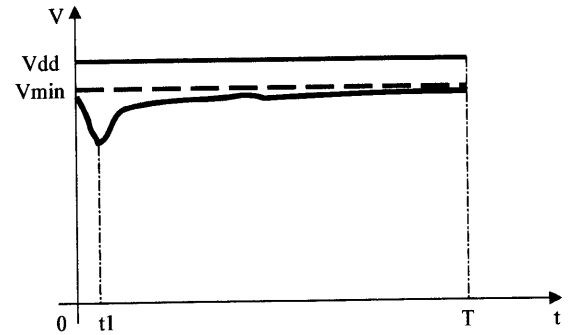


Fig. 5. Voltage waveform of a VN.

TABLE I
VN STATISTICS ABOUT NONLEAKAGE-CURRENT-AWARE OPTIMIZED P/G NETWORK WHEN LEAKAGE CURRENTS OF DECAPS ARE CONSIDERED

Circuit Name	Node Num	Eliminated VN Num	Added Decaps (nf)	Newly Crated VN	New VN Ratio
Test1	744	91	0.05403	11	12.08%
Test2	3741	665	0.30515	24	3.61%
Test3	32112	3683	1.28128	294	7.98%

mances while they can reduce the noise [6]–[9]. Therefore, previous methods will be terminated as long as all the violations are removed by adding enough on-die decaps [8]. However, when T_{ox} falls below 20 Å, more decaps will introduce more leakage current, therefore, the best choice to place decaps is not the best choice for system power delivery. Actually, traditional optimization methods tend to overestimate decaps because almost all optimization algorithms in the second optimization stage are more sensitive to the noises other than voltage level. This unawareness of leakage current in dynamic optimization will cause the optimal solutions obtained before to become unreliable due to the increase in power consumption of leaky decaps. A typical example to demonstrate the drawbacks is given in the following.

Here, we randomly select three P/G grids optimized without considering leakage currents of decaps using the optimization algorithm described in [8]. Then, transient analysis is performed to verify the actual voltage distribution when leakage current is considered. The transient analysis method we used is introduced in [24], and the analysis results are shown in Table I.

From the results, we can find that the traditional method can reduce the VN number to zero after adding enough decaps. However, when leakage current is considered, some new VNs appeared, the ratio to the eliminated VN number is about 7%.

Hence, one observation that we can have from the data is that it is necessary to consider the decap leakage currents during the decap allocation process to maintain a robust optimization.

IV. SIMPLE OPTIMIZATION METHOD CONSIDERING LEAKAGE

One simple way to consider the leakage effect is to introduce the leakage model directly to a traditional optimization method.

Table II shows the results after we add the leakage model into the old optimization process. Now, we can see that to compensate for the leakage effect, about 18% of extra die area is required on the average, which is an obvious area increase. Be-

TABLE II
OPTIMIZATION RESULTS WITH AND WITHOUT LEAKAGE CURRENT

Circuit Name	Node Num	Added Decap (nf) Without Leakage	Added Decap (nf) With Leakage	Ratio of Die Area Increase
Test1	744	0.05403	0.06303	16.66%
Test2	3741	0.30515	0.38015	24.58%
Test3	32112	1.28128	1.44128	12.49%

cause the basic area used by decaps is very large, it is too difficult to get an increase especially in the late design stage.

However, further analysis shows that the overestimated decaps are due to less sensitivity of IR drops to decaps in the presence of the leakage currents of decaps. Traditional optimization algorithms only try to insert decaps when the node voltage falls below the required boundary. It works fine when the voltage drop is caused by SSE. However, if the voltage drop is also caused by the leakage current, the optimization effect of adding decaps will become less effective as more decaps are needed to get the same optimization results. Obviously, the added decaps may bring more leakage currents, which, in turn, lead to more IR drops and thus more power consumption. One extreme situation would be that at a certain VN, the more decaps inserted, the larger IR drop there will be. That is why the method that directly uses leaky decaps to optimize P/G grids is less effective in terms of both noise reduction and power reduction, which, in turn, causes overestimated decaps. From this perspective, for the low-power design purpose, we should use decaps carefully and consider more sophisticated optimization options as shown in the following section.

V. MORE SOPHISTICATED OPTIMIZATION METHOD

Leakage current is typically dc current, and more decaps definitely increase the total power consumption. Insertion of leaky decaps cannot help in reducing the static IR drop at each node on the P/G network. That is why the previous algorithm will lose efficiency when a leakage effect is considered. However, wire sizing of the power grid can effectively reduce static IR drops. Therefore, if the use of a wire-sizing strategy is more efficient in both VN reduction and area optimization than iterative insertion of leaky decaps, we can solve the overestimated decaps problem in the dynamic optimization stage. We propose a more effective and practical two-stage P/G optimization method.

In the first stage, this method assumes that all decaps are leakage free, and then it optimizes the dynamic voltage noise using the algorithm described in [8]. That algorithm is an efficient nonlinear programming based decoupling capacitor budgeting algorithm, which utilizes the conjugate gradient algorithm to search for the best solution. Furthermore, the proposed time-domain-merged adjoint network method combined with a novel equivalent circuit modeling technique in [8] leads to a very efficient implement of the conjugate gradient algorithm.

Different from the traditional methods, in the second optimization stage, the new algorithm tries to detect the most sensitive nodes to perform a wire-sizing strategy using a branch-and-bound method to minimize the added wire area used to compensate the IR drops caused by leakage currents. The branch-and-

TABLE III
OPTIMIZATION RESULTS OF THE TWO-STAGE METHOD

Circuit Name	Node Num	Added Decap (nf) with leakage	Ratio of Routing Area Increase
Test1	744	0.05403	1.88%
Test2	3741	0.30515	5.83%
Test3	32112	1.28128	3.00%

bound method uses currents in wire segments as boundary variables; it only tries to size the wires on which current is lower than the boundary current. Furthermore, it slightly sizes wires segment by segment so it can provide a feasible solution with the routing area only slightly increased.

The optimization results of the two-stage method are shown in Table III. Comparing with the 18% area increase in the traditional method, which simply deals with the leakage effect, the new two-stage method only increases the routing area within an average of 4%. Because the second optimization stage is not very time consuming, it provides a good tradeoff between the computation complexity and area usage to handle the leakage problem.

VI. CONCLUSION

With technology scaling down to 90 nm and below, gate leakage current of decaps becomes so significant that we have to explicitly consider decap leakage currents for robust P/G grid design. This is especially the case when T_{ox} shrinks below 20 Å. In this brief, we take a first look at the impacts of gate leakage of practical MOS-based decaps on P/G grid designs. To clearly show the influence of gate leakage currents of decaps on the P/G grids, we propose a leakage current model for practical decaps. Then, we analyze the effect of leakage currents and show that practical (leaky) decaps may increase power consumption significantly and thus demand more routing resource or die area for achieving robust power delivery. On top of our analysis, we propose a more effective two-stage optimization approach to efficiently optimize P/G grids in the presence of decap leakage currents. Compared with the simple approach that directly adds more leaky decaps to reduce dynamic voltage noise, the two-stage approach is more area efficient as it consumes less routing resources (4% increase on the average) compared to the leaky-decap-only method, which uses much more die area (18% increase on the average).

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Huang from Intel Corporation for introducing the decap leakage problem.

REFERENCES

- [1] NAME/LOCATION OF COMPANY????, The International Technology Roadmap for Semiconductors (ITRS). (2002 update) [Online]. Available: <http://public.itrs.net/>
- [2] T. Karnik, S. Borkar, and V. De, "Sub-90 nm technologies—Challenges and opportunities for CAD," in *Proc. ICCAD*, San Jose, CA, 2002, vol. 11, pp. 203–207.
- [3] G. Bai, S. Bobba, and I. N. Hajj, "Simulation and optimization of the power distribution network in VLSI circuits," in *Proc. ICCAD*, San Jose, CA, 2000, vol. 11, pp. 481–486.
- [4] X. D. Tan and C. J. Shi, "Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programming," in *Proc. 36th ACM/IEEE DAC*, New Orleans, LA, 1999, pp. 78–83.
- [5] X. H. Wu, X. L. Hong, and Y. C. Cai *et al.*, "Area minimization of power distribution network using efficient nonlinear programming techniques," in *Proc. ICCAD*, San Jose, CA, 2001, pp. 153–157.
- [6] S. Y. Zhao, K. K. Roy, and C. K. Koh, "Decoupling capacitance allocation for power supply noise suppression," in *Proc. ISPD*, Sonoma, CA, 2001, pp. 66–71.
- [7] H. H. Su, K. K. Roy, S. S. Sapatnekar, and S. R. Nassif, "An algorithm for optimal decoupling capacitor sizing and placement for standard cell layouts," in *Proc. ISPD*, San Diego, CA, 2002, pp. 68–75.
- [8] J. J. Fu, Z. Y. Luo, X. L. Hong, Y. C. Cai, S. X. D. Tan, and Z. Pan, "A fast decoupling capacitor allocation algorithm for noise reduction of power delivery networks," in *Proc. ASP-DAC*, Yokohama, Japan, 2004, vol. 1, pp. 505–510.
- [9] K. Wang and M. M. Sadowska, "On-chip power supply network optimization using multigrid-based technique," in *Proc. 40th DAC*, Anaheim, CA, 2003, pp. 113–118.
- [10] H. H. Chen and D. D. Ling, "Power supply noise analysis methodology for deep-submicron VLSI chip design," in *Proc. 34th ACM/IEEE DAC*, Anaheim, CA, 1997, pp. 638–643.
- [11] Z. Y. Zhu, B. Yao, and C. K. Cheng, "Power network analysis using an adaptive algebraic multigrid approach," in *Proc. 40th DAC*, Anaheim, CA, 2003, pp. 105–108.
- [12] S. Mukhopadhyay, A. Raychowdhury, and K. Roy, "Accurate estimation of total leakage current in scaled CMOS logic circuits based on compact current modeling," in *Proc. 40th DAC*, Anaheim, CA, 2003, pp. 169–174.
- [13] D. W. Lee, W. Kwong, D. Blaauw, and D. Sylvester, "Analysis and minimization techniques for total leakage considering gate oxide leakage," in *Proc. 40th DAC*, Anaheim, CA, 2003, pp. 175–180.
- [14] S. Bobba, T. Thorp, K. Aingaran, and D. Liu, "IC power distribution challenges," in *Proc. ICCAD*, San Jose, CA, 2001, pp. 643–650.
- [15] M. D. Pant, P. Pant, and D. S. Wills, "On-chip decoupling capacitor optimization using architectural level prediction," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 10, no. 3, pp. 319–326, Jun. 2002.
- [16] D. J. Herrell and B. Beker, "Modeling of power distribution systems for high-performance microprocessors," *IEEE Trans. Adv. Packag.*, vol. 22, no. 3, pp. 240–248, Aug. 1999.
- [17] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, "Leakage current: Moore's law meets static power," *IEEE Trans. Comput.*, vol. 36, no. 12, pp. 68–75, Dec. 2003.
- [18] D. Helms, E. Schmidt, and W. Nebel, "Leakage in CMOS circuits—An introduction," in *LNCS (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2004, vol. 3254, pp. 17–35.
- [19] D. Lee, D. Blaauw, and D. Sylvester, "Gate oxide leakage current analysis and reduction for VLSI circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 2, pp. 155–166, Feb. 2004.
- [20] [Online]. Available: <http://www-device.eecs.berkeley.edu/~ptm/>
- [21] C. K. Cheng, J. Lillis, S. Lin, and N. Chang, *Interconnect Analysis and Synthesis*. Hoboken, NJ: Wiley, 2000, 0-471-29366-0.
- [22] T.-H. Chen and C. C.-P. Chen, "Efficient large-scale power grid analysis based on preconditioned Krylov-subspace iterative methods," in *Proc. 38th DAC*, Las Vegas, NV, 2001, pp. 559–562.
- [23] I. A. Ferzli and F. N. Najm, "Statistical verification of power grids considering process-induced leakage current variations," in *Proc. ICCAD*, San Jose, CA, 2003, pp. 770–777.
- [24] Z. Pan, Y. Cai, S. X.-D. Tan, Z. Luo, and X. Hong, "Transient analysis of on-chip power distribution networks using equivalent circuit modeling," in *Proc. ISQED*, San Jose, CA, 2004, pp. 63–68.