VLSI On-Chip Power/Ground Network Optimization Considering Decap Leakage Currents*

Jingjing Fu[1], Zuying Luo[1], Xianlong Hong[1], Yici Cai[1], Sheldon X.-D. Tan[2], Zhu Pan[1]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, P.R.China
[2]Department of Electrical Engineering, University of California at Riverside, USA
Contact Author: Jingjing Fu, fujingjing00@mails.tsinghua.edu.cn

Abstract — In today's power/ground(P/G) network design, on-chip decoupling capacitors(decaps) are usually made of MOS transistors with source and drain connected together. The gate leakage current becomes worse as the gate oxide layer thickness continues to shrink below 20Å. As a result, decaps will become leaky due to the gate leakage from CMOS devices. In this paper, we take a first look at the leaky decaps in P/G network optimization. We propose a leakage current model for practical decaps and also present a new two-stage leakage-current-aware approach to efficiently optimize P/G networks in a more area efficient way.

I Introduction

As technology scales into 90nm and below, power consumption is becoming the limiting factor in high-performance VLSI chip design and power reduction becomes an intensive research area [1-2]. On the other hand, robust power delivery is also considered as one of grant challenges [1]. As predicted in the roadmap for IC development from ITR-2002Update, the chip working frequency and supply voltage will continue to scale aggressively, which will lead to significant large Ldi/dt noise and current flows on P/G networks [7-13]. The excessive voltage flotation and current flows will adversely affect the performances and reliabilities of the VLSI chips. Power reduction and power delivery system optimization were considered as separate issues in the past. However, this is no longer the case as leakage currents start to become significant and leakage current due to the leaky decaps can no longer be ignored during the P/G grid optimization as technology scales into 90nm and below [3-4].

With huge currents delivered from P/G grids, on-board, on-package and on-chip decaps are used to reduce dynamic Ldi/dt noises. Since large on-chip metal-insulator-metal or poly-insulator-poly decaps tend to consume large die area [7], designers usually use decaps made of MOS transistors with connected source/drain/base nodes. The MOS-based decaps can effectively reduce Ldi/dt noise as they can easily be inserted very close to the nodes that have large Ldi/dt noise [10-12]. But it was shown in [3,4] that when the oxide thickness Tox is smaller than 20Å for the technology node below 100nm, the gate leakage of MOS-based decaps will become significant as gate leakage increases 10X when Tox decreases by 2Å [4]. To reduce the decap leakage, one way is to use less-leakage dielectric material, which may not be available until 2007[4]. The second way is to use thick oxide MOS transistors as decaps in chip designs, which may occupy more chip areas. And the more important thing is that it may increase the complexity of manufacture and make the costs go up greatly. So it is not an economical way. As a result, adding decaps will hurt power consumption, which in turn will make the added decaps less effective to reduce the voltage drops. Therefore explicit consideration of decap leakage powers and P/G optimization methods, which make the best tradeoff among power consumption, chip area and P/G routing sources in the presence of decap leakage powers, are required for robust and low-power on-chip P/G network design in the near future.

In this work, first, we analyze gate leakage current for MOS-based decap, and propose a decap leakage current model for optimizing P/G grids. Second, based on the optimal decap allocation solution that we obtain from the P/G grid optimization without considering gate leakage of decaps, we use an approximate leaky decap model, Rlk||C(leakage resistor) to analyze the optimal P/G grid. We find that leakage current of decaps will increase the voltage drop. We further show that given enough wire segment widths of P/G grids, it is impractical to simply add more decaps to reduce Ldi/dt noise because more decaps give rise to more IR drops that then lead to more decaps to be added. Instead, we propose an effective two-stage P/G grid optimization approach. The approach first assumes leakage-free decaps to reduce Ldi/dt noise and then sizes wire segment widths to take care of the IR drops caused by decap leakage currents. Experimental results demonstrate the effectiveness of the leakage-aware P/G optimization approach.

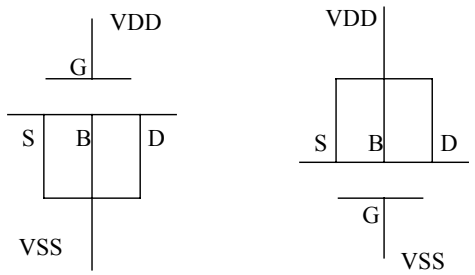
This paper is organized as follows. Section 2 presents the decap leakage current model. Section 3 shows that it is imperative for designers to consider decap leakage current in P/G network optimization in the sub-100nm VLSI design regime. Section 4 shows that adding decaps only can't effectively reduce voltage drops in the presence of decap leakage currents. In section 5, we present a new effective two-stage leakage-current-aware P/G grid optimization framework, which uses both decap allocation and wire sizing. Section 6 concludes the paper.

II. Models of MOS-Based Decaps

On-chip decaps can be made of NMOS or PMOS transistors whose source/drain/base nodes are simply connected together. Due to the fact that base node of NMOS must be connected to ground while the base node of PMOS to supply voltage, connection styles of NMOS-based and PMOS-based decaps are thus different as shown in Figure 1.

For MOS transistors, as technology scales into 100nm, three types of leakage currents should be considered. They are sub-threshold leakage, gate leakage and BTBT leakage current [3] as shown in Figure 2(a).

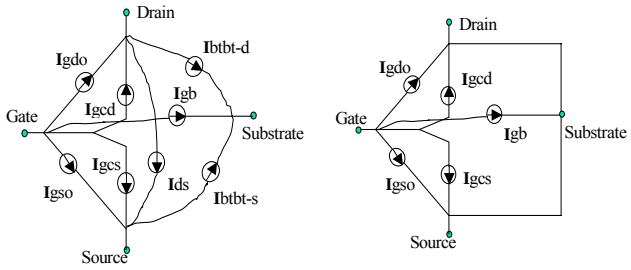
*This work is supported by National Natural Science Foundation of China (NSFC) 90307017, 60176016, the National Foundation Research (973) Program of China G1998030403, Hi-Tech Research & Development (863) Program of China 2002AA1Z1460.



(a) NMOS-based decap (b)PMOS-based decap

Figure.1 On-chip decaps made of NMOS or PMOS

In Figure 2(a), *Ids* is the sub-threshold leakage and will be ignored as drain/source nodes are short-circuited. *Ibtbt-d* and *Ibtbt-s* are two parts of BTBT leakage and will be also canceled as drain/source/base nodes are connected. *Igdo, Igso, Igcd, Igcs*, and *Igb* are five parts of gate leakage current *Igate* and should be considered. The resulting leakage current of MOS-based decap mainly comes from gate leakage current as shown in Figure 2(b).



(a) Leakage model of MOS transistor (b) Leakage model of MOS-based decap

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

It is shown in [3,4] that *Igate* increases 10X as Tox decreases 2Å after Tox scales below 20Å. The leakage current of MOS-based decap can be formulated as:

$$Igate = \alpha \times e^{-\beta Tox} \times w_n , \qquad (1)$$

where α ,β are parameters related to specific technology, w_n is the gate width of NMOS (w_p of PMOS). The capacitance of a decap can be formulated as following:

$$C_d = \gamma \times \frac{\varepsilon \times \varepsilon_0}{Tox} \times S , \qquad (2)$$

where ε, ε_0 are relative and absolute dielectric constants, γ, is a parameter related to the shape of MOS transistor, and *S* is the area that is equal to width w_n times length Leff of the gate. We use BSIM [5] 70nm technology to simulate a NMOS-based decap with w_n =0.2um and Leff=0.038um. The resulting decap is 0.26fF with γ=1.5462 when ε=4. Since different shapes cause different γ, we assign γ=1 for the simplicity.

To consider the gate leakage of decaps, we propose a decap model Rlk||C that can capture the main gate leakage as shown in Figure 3. In this resistor-based model, a resistor *Rlk* is connected in parallel to the capacitor C. The value of *Rlk* can be determined according to the supply voltage and the leakage current of C. This is obviously an approximate model because gate leakage is typically exponential to node voltage. However, since it is very difficult to calculate the exact value of the leakage current and to model it, we choose to use this simple and direct model to image the leakage problem of using decaps although it is not very accurate. In this paper, we just take a first look at the impacts of gate leakage of practical MOS-based decaps on P/G grid optimization. To build a better leaky decap model is our future work.

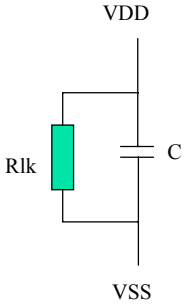


Figure.3 Rlk||C Model for decap leakage

III. P/G Network Optimization With/Without Decap Leakage Currents

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

Early optimization approaches size wire segment widths to reduce static IR drops [8,9]. The optimization object is to minimize the routing-area consumed by P/G grids. Due to increasing working frequency and decreasing supply voltage, recent P/G optimization approaches budget on-chip decaps to reduce Ldi/dt noise [10-12]. The optimization object is to minimize the chip-area consumed by decaps.

The approach in [13] simultaneously sizes wire segment widths and decaps for robust transient P/G networks, but no decap leakage currents are considered. In this paper, we focus on sizing wire segment widths or on-chip decaps separately.

In order to facilitate decap budgeting, we use standard cell layout structure which is similar to methods in [8,9,11,12], but the new method can be extended to other layout structures easily. To facilitate gradient-based optimization, the merged adjoint method [9,12] is used to determine the search directions in each optimization step.

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 segment of the power grid is modeled as lumped RLC element. Parasitics from power pads and packages are also modeled as lumped RLC elements. The nonlinear CMOS modules are modeled as time-varying current sources, which can be obtained by off-line logic simulation with the assumption that supply voltage is ideal V_{dd}. We also consider both built-in on-chip decoupling capacitors (n-well capacitors and circuit capacitors) and add-on decoupling capacitors (thin-oxide capacitors) connected between power and ground. Using the new decap model, which considers leakage current described in Section 2, the mesh-structured

P/G grids as a pseudo-distributed RLC network is illustrated in Figure 4.

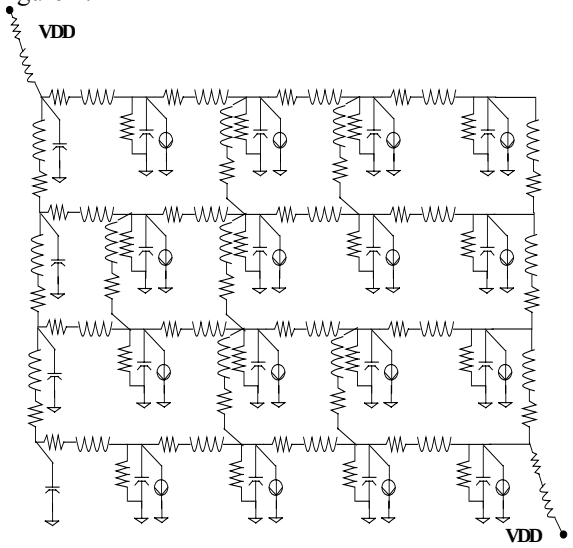


Figure. 4 Model of Power Network

B. Noise Analysis of P/G Grids

The behavior of power delivery network modeled above is analyzed with a first order differential equation formulated by modified nodal analysis (MNA) for transient analysis. For an unreliable P/G grid design, there are some constraint-breaking nodes or branches. Figure 5 shows a voltage waveform on a constraint-breaking node on power grid in the presence of decap leakage currents.

Such waveforms come from switching current, large leakage current caused by CMOS circuit modules and thin wire width of P/G grids. As a result, the voltage may be still smaller than Vmin when the chip ends its logic operation [6] as shown in Figure 5. The bottom point at time t1 is typically due to the sharp current increase caused by clock and the influence of inductance components of P/G grids.

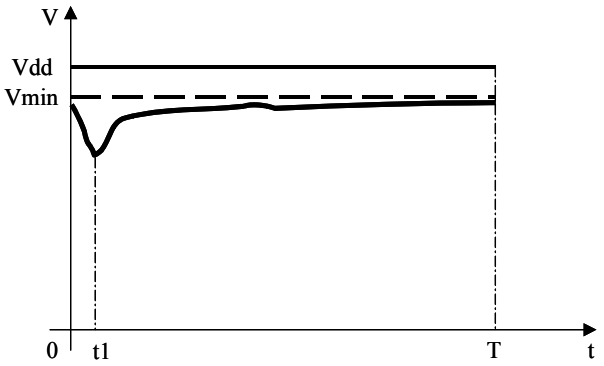


Figure. 5 Voltage waveform of a constraint-breaking node

C. Analysis of Optimization Results Considering Leakage Current of Decaps

Shown in [8-12], design of P/G network has two stages. The first stage is to size wire segment widths to reduce static IR voltage drops based on empirical values of absorption currents. The second stage is to add on-chip decaps to reduce transient voltage drops given the dynamic power consumption excited by the time-varying input vectors.

Because leakage current of on-chip decaps is not significant when Tox > 20Å, nearly all previous works on optimization of power delivery network did not consider leakage current of decaps [10-13]. As long as all the violations are removed by adding on-chip leakage-free decaps, the optimization is stopped and the resulting P/G grid is optimized and considered to be robust [12].

But when Tox scales below 20Å, leakage current of on-chip decaps becomes significant that the optimal solutions obtained before may become unreliable due to increasing power consumptions of the leaky decaps. To show this, we select three optimized P/G grids without considering decap leakage currents to do the transient analysis considering decap leakage currents. The results are shown in Table 1.

From the results, we find that the voltage drops on some nodes still exceed a safe bound. The 5th column of Table 1 shows the numbers of constraint-violating nodes that still exist after optimization with leakage-free decaps. Obviously, the power grid is not robust any more if we consider the leakage current of decaps.

Table 1: The non-leakage-current-aware optimized P/G network when leakage currents of decaps are considered.

Name of circuits	Num of nodes	#Violation nodes	Amount of decap (nf)	#Violation nodes
Test1	744	91	0.05403	11
Test2	3741	665	0.30515	24
Test3	32112	3683	1.28128	294

As a result, we need to consider the decap leakage currents during the decap allocation process. However, simply adding more decap to resolve this problem during the P/G optimization process will not be effective as decaps and power consumption will interplay with each other as shown below.

IV. Impractical Optimization Method

Existing optimization approaches, which assume leakage-free decaps, tend to use less decaps as shown in column 4 of Table 2. Because wire segments of robust DC P/G grids are wide enough, we can still obtain robust transient P/G grids by adding more decaps as shown in column 5 of Table 2. Shown in Table 2, practical decaps consume average 18% more die area than leakage-free decaps to get robust transient P/G grids. Thus, the method that directly uses practical decaps to optimize P/G grids is not very efficient.

Table. 2 Optimization Results with and without Leakage Current of Decaps

Name of circuits	Num of nodes	#Violati-on nodes	Amount of decap (nf) without leakage	Amount of decap (nf) with leakage	Ratio of increase
Test1	744	91	0.05403	0.06303	16.66%
Test2	3741	665	0.30515	0.38015	24.58%
Test3	32112	3683	1.28128	1.44128	12.49%

Since the optimization effect of decaps will become less effective when the decap leakage currents are further considered, more decaps are needed to get the same

optimization results. The added decaps may bring more leakage currents which will lead to more IR drops and thus more power consumption. Therefore, it is not wise to further add decaps to reduce voltage drops. From this perspective, for the low-power design purpose, we should use decaps carefully.

V. Practical Optimization Method

Leakage current of decaps increases the total power consumptions. The static IR drop of each node on the network becomes larger due to decap leakage current because decap leakage is DC current. Wire sizing of power grid can effectively reduce static IR drops. Therefore, if there are enough routing areas for power grids, we can use wire sizing to further reduce the voltage drops caused by decap leakage current instead of adding more decaps. The whole optimization problem essentially is about the best tradeoff between routing area consumption and decap area consumption.

Thus, to get better optimization results, we propose a more effective and practical two-stage P/G optimization method. This method first assumes leakage-free decaps to optimize the Ldi/dt noise. Then it sizes wire segment widths of power grids to deal with the additional IR drops due to decap leakage currents.

The results of two-stage method are shown in Table 3, which compare very favorably with the optimization results without considering decap leakage currents. Column 5 shows that the new method only increases average 4% routing area, which is in contrast to 18% area increasing by adding decaps only. We can see that by sizing a little more wire segment widths of power grids, the problem caused by leakage current of decaps is more easily to be resolved. Therefore, to efficiently reduce the impacts of leakage current of decaps in power grid optimization, we should use wire sizing instead of direct decap budgeting.

Table. 3 Optimization Results of the Two-Stage Method

Name of circuits	Num of nodes	#Violation nodes	Amount of decap (nf) without leakage	Increase ratio of wire area
Test1	744	91	0.05403	1.88%
Test2	3741	665	0.30515	5.83%
Test3	32112	3683	1.28128	3.00%

VI. Conclusion

With technology scaling into 100nm regime, gate leakage current of decaps become significant that we have to explicitly consider decap leakage currents for robust P/G grid design. This specially the case when Tox shrinks below 20Å. In this paper, we take a first look at the impacts of gate leakage of practical MOS-based decaps on P/G grid optimization. In order to clearly show influence of gate leakage 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 grid in the presence of decap leakage currents. Compared with the simple approach that directly adds more decaps to optimize P/G grids, the two-stage approach is more area efficient as it only consumes less routing resources (4% increase) while the simple approach consumes far more (18% increase) die area.

Acknowledgements

The authors would like to thank Dr. Jianhui Huang from Intel Corporation for providing us the decap leakage problem.

References

[1] The International Technology Roadmap for Semiconductors (ITRS), http://public.itrs.net/, 2002update.

[2] T. Karnik, S. Borkar, and V. De. Sub-90nm technologies-challenges and opportunities for CAD. In Proceeding of ICCAD2002, San Jose, 2002.11:203-207.

[3] S. Mukhopadhyay, A. Raychowdhury, and K. Roy. Accurate Estimation of Total Leakage Current in Scaled CMOS Logic Circuits Based on Compact Current Modeling. In Proceeding of 40th DAC, Anaheim, 2003.6: 169-174.

[4] D. W. Lee, W. Kwong, D. Blaauw, D. Sylvester. Analysis and Minimization Techniques for Total Leakage Considering Gate Oxide Leakage. In Proceeding of 40th DAC, Anaheim, 2003.6: 175-180.

[5] http://www-device.eecs.berkeley.edu/~ptm/

[6] I. A. Ferzli, F. N. Najm. Statistical Verification of Power Grids Considering Process-Induced Leakage Current Variations. In Proceeding of ICCAD2003, San Jose, 2003.11: 770-777.

[7] G. Bai, S. Bobba, and I. N. Hajj. Simulation and optimization of the power distribution network in VLSI circuits. In Proceeding of ICCAD2000, San Jose, 2000.11: 481-486.

[8] X. D. Tan and C. J. Shi. Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programming. In Proceeding of 36th ACM/IEEE DAC, New Orleans, 1999: 78-83.

[9] X. H. Wu, X. L. Hong, Y. C. Cai, and et al. Area minimization of power distribution network using efficient nonlinear programming techniques. In Proceeding of ICCAD2001, San Jose, 2001: 153–157.

[10] S. Y. Zhao, K. K. Roy, C. K. Koh. Decoupling capacitance allocation for power supply noise suppression. In Proceeding of ISPD2001, Sonoma, 2001: 66-71.

[11] H. H. Su, K. K. Roy, S. S. Sapatnekar, S. R. Nassif. An algorithm for optimal decoupling capacitor sizing and placement for standard cell layouts. In Proceeding of ISPD2002, San Diego, 2002: 68-75.

[12] J. J. Fu, Z. Y. Luo, X. L. Hong, Y. C. Cai, Sheldon. X. D. Tan, Z. Pan. A Fast Decoupling Capacitor Allocation Algorithm for Noise Reduction of Power Delivery Networks. In proceeding of ASP-DAC2004, Yokohama, 2004.1: 505-510.

[13] K. Wang and M. M. Sadowska. On-chip Power Supply Network Optimization using Multigrid-based Technique. In Proceeding of. 40th DAC, Anaheim, 2003.6: 113-118.