Simultaneous Wire Sizing and Decoupling Capacitance Budgeting for Robust On-Chip Power Delivery*

Jingjing Fu[1], Zuying Luo[1], Xianlong Hong[1], Yici Cai[1],
Sheldon X.-D. Tan[2], and Zhu Pan[1]

[1] Department of Computer Science and Technology, Tsinghua University,
Beijing, 100084, P.R.China
{fujingjing00,pz01}@mails.tsinghua.edu.cn,
{luozy,hxl-dcs}@mail.tsinghua.edu.cn
[2] Department of Electrical Engineering, University of California at Riverside,
CA 92521, USA
stan@ee.ucr.edu

Abstract. In this paper, we present an efficient method to simultaneously size wire widths and decoupling capacitance (decaps) areas for optimizing power/ground (P/G) networks modeled as RLC linear networks subject to reliability constraints. We formulate the problem as a nonlinear optimization problem and propose an efficient gradient-based non-linear programming method for searching the solution. We apply a time-domain merged adjoint network to efficiently compute the gradients and a novel equivalent circuit modeling technique to speed up the optimization process. The resulting algorithm is very efficient and experimental results show that the new algorithm is capable of optimizing P/G networks modeled as RLC networks with million nodes within 10 hours in modern workstations.

1 Introduction

Signal integrity in VLSI is emerging as a limiting factor in the nano-regime VLSI chip designs as technology scales. This is especially true on global networks like P/G networks where noise margins have been reduced greatly in today's advanced VLSI designs due to decreasing supply voltages and the presence of excessive voltage drops coming from resistive and inductive effects. In order to reduce IR drops, wire sizing is typically employed [4][5][6][7]. As for dynamic voltage fluctuations on a P/G network, adding decap is the most efficient way to reduce such noises [8][9][18].

Because wires consume the routing resource while decaps cost spare die area, algorithm that can simultaneously optimize both of them is needed to reach the best trade-off between routing resource and die area as demonstrated in [10], which used a geometric multi-grid based method. But the geometric multi-grid approach was only applied to mesh-structured P/G grids modeled as resistor-only networks with decaps where capacitive and inductive parasitics on wire segments are ignored. Also the

* This work was supported by National Natural Science Foundation of China (NSFC) 90307017, 60176016 and 60121120706 and National Natural Science Foundation of USA (NSF) CCR-0096383.

E. Macii et al. (Eds.): PATMOS 2004, LNCS 3254, pp. 433–441, 2004.

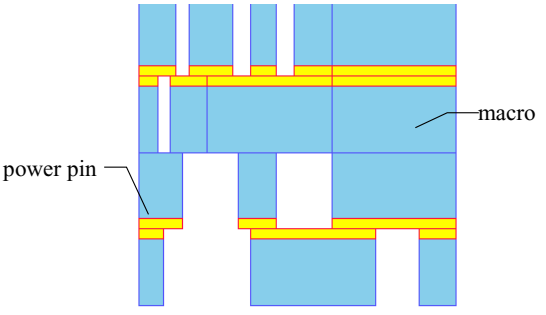

Fig. 1. An instance of standard cell layout

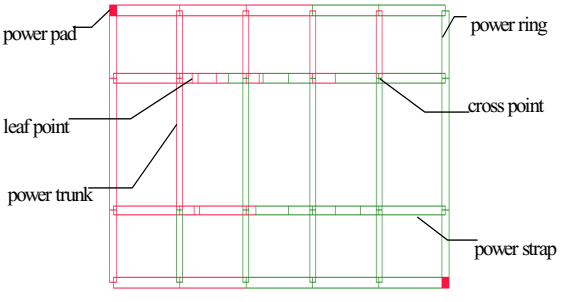

Fig. 2. Structure of power network

geometric multi-grid simulation method [10] is less flexible to deal with non-mesh-structured circuits [17], which are very common in various ASIC designs. Another problem with the multi-grid method is that it can easily lead to optimistic solutions [14][17], which makes this method less reliable.

In this paper, we focus on standard-cell like ASIC layout structures as shown in Fig.1 and try to optimize the areas of P/G networks as the objective [6][7][[9]. In the figure, the white space is the spare space for adding decaps to reduce dynamic noise. Fig.2 shows a power network of the standard-cell ASIC layout shown in Fig.1. Power trunks in the power network will be sized for reducing IR drop.

The P/G networks are modeled as lumped RLC circuits. To clearly describe our work, we make the following assumptions:

1) The package is predominantly inductive, and is modeled by serially connected inductors through power supply sources to their contact points on the power grid.
2) The power grid is modeled as a RLC network where R and L are the parasitical resistance and inductance of wires and C is the capacitance consisting of parasitical capacitance and decoupling capacitance (both built-in and added-on).
3) Connections between top and low levels are fine enough that we ignore the resistor of the connection.
4) The nonlinear devices or modules are modeled as time-varying current sources connected between each node on power grid and ground.

The problem we are concerned in this paper is how to efficiently optimize P/G networks by simultaneously sizing wires and decaps subject to the reliability and design rules for P/G grids, which consist of constraints due to dynamic voltage fluc-

tuations, electro-migration and other design rules. The advantages of our approach are:

1) The approach can optimize P/G networks modeled as RLC networks.
2) It is a very general approach and can be applied to any P/G network structures.
3) The new method is based on a very efficient and general nonlinear programming framework, which can deal with both convex and concave cases [7] and can be scaled to handle circuits with millions of nodes.
4) Gradients are efficiently computed using time-domain merged adjoint network combined with equivalent circuit modeling techniques.

This paper is organized as follows. Section 2 formulates the wire sizing and decap budgeting problem. The efficient non-linear programming technique and the equivalent circuit modeling method are described in Section3. Experimental results are presented in Section 4. Section 5 concludes the paper.

2 Problem Formulation

For the similarity of power and ground networks, we will only describe the algorithm for power networks in this paper. Let $G = \{N,B,M\}$ be a power network with n nodes $N = \{1, ..., n\}$ and b RLC branches of power trunks $B = \{1, ..., b\}$and m nodes that are permitted to connect tunable decaps $M = \{1, ...,m\}$. Each branch (p, q) in B connects two nodes p and q with current flowing from p to q. For a node p, $v_p(t)$ is the node voltage.

2.1 Objective Function

The objective is to minimize the combined area used by power grid wires and decaps subject to design rules and power network integrity related constraints:

$$\min A = \sum_{i \in B} \left(l_i wr_i \right) + \eta \sum_{j \in M} \left(wc_j \times H \right) \tag{1}$$

where l_i and wr_i are the length and width of branch i, H is the fixed height of standard cells and wc_j is the width of the decap connected to node j. And η is the weighted factor for total decap area. Designers can properly set the value of η to set up the priorities of the resources of the routing and die area.

2.2 The Constraints

1) The voltage drop constraints. To ensure the correct and reliable logic operation, the voltage drops from the power pads to the leaf nodes should be restricted. To efficiently measure the dynamic voltage drop, we follow the voltage drop noise metric definition in [9] and reformulate the voltage drop constraints as follows:

$$s_i = \int_0^T \max(V_{\min} - v_i(t), 0)dt = \int_{t_1}^{t_2} (V_{\min} - v_i(t))dt \,, \tag{2}$$

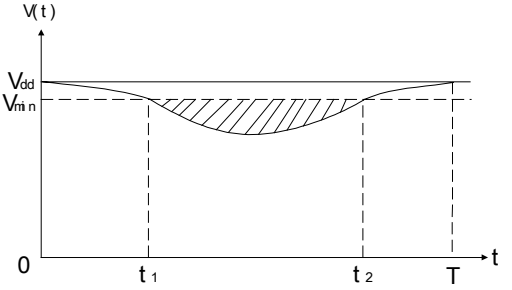

Fig. 3. Illustration of dynamic voltage drops

where $[t_1,t_2]$ is the time interval in which the constraint is violated as shown in Fig.3. There may exist several such intervals in one clock cycle.

2) The electro-migration constraints. Electro-migration effects on a power grid wire segment set an upper bound on the current density of the wire segment as the routing layer has fixed thickness. This constraint for branch (p,q) is expressed as $|i_{p,q}(t)| \le w_{p,q} \times \sigma$, and can be re-written as $|v_p(t)-v_q(t)| \le \rho l_{p,q} \sigma$, where σ is the maximal current density, ρ is the sheet resistance, and l_{pq} is the length of branch (p, q). Due to dynamic nature of $v_{p,q}(t)$, we reformulate the constraint as below:

$$\begin{aligned} t_{p,q} &= \int_0^T \max\left[\left(\left|v_p(t)-v_q(t)\right|-\rho l_{p,q}\sigma\right),0\right]dt \\ &= \int_{t_1}^{t_2}\left(\left|v_p(t)-v_q(t)\right|-\rho l_{p,q}\sigma\right)dt \end{aligned} \tag{3}$$

where $[t_1,t_2]$ is the time interval in which the constraint is violated. Similarly, there may exist several such intervals in one clock cycle.

3) Minimal wire width constraints. The wire width cannot be smaller than the minimal width of metal line w_{min}.

4) Decap area constraints. Because we essentially re-distribute the spare space around cells, the total width of all decaps added in a row should be limited by the width of total spare area in this row:

$$dw_{ir} = \max\Big(\sum_{jc \in ND(ir)} w_{ir,jc} - rw_{ir}, 0\Big), ir \in NR, \tag{4}$$

where NR is the row set defined as $\{1,2,\ldots,N_{row}\}$ and N_{row} is the row number of the cells in the placement; $ND(ir)$ is the decap position set of ir^{th} row defined as $\{1,2,\ldots,nr_{ir}\}$ and nr_{ir} is the number of nodes where decaps can be attached in the ir^{th} row; $w_{ir,jc}$ is the width of the decap at row ir and position jc; and rw_{ir} is the width of total spare area in row ir.

5) Decap maximum width constraints. The width of a decap in row ir should be bounded by the total width of spare area provided by this row.

$$ew_{ir,jc} = \max(w_{ir,jc} - rw_{ir}, 0), ir \in NR,\ jc \in ND(ir), \tag{5}$$

3 Solution Based on Non-linear Programming

The resulting problem is a nonlinear minimization problem as voltages and currents are nonlinear functions of the width vector of wires and decap areas. As a result, we propose to use a gradient-based non-linear programming method to solve the optimization problem as gradients can be efficiently computed in time-domain.

Specifically, at the beginning, all the decap widths are set to be zero while all the wire widths are assigned as the minimum wire width w_{min}. We then analyze the network to get the node voltage waveforms and the branch currents, and identify the constraint violations. In each optimization iteration, we use conjugate gradient method to update wire and decap widths. This process stops when all the constraints are satisfied or no improvement can be made.

3.1 Formulation of Penalty Function

The first step is to transform the original constrained problem into a sequence of unconstrained problems. The transformation is accomplished by adding to the objective function a penalty term that gives a high cost to the constraint violations. We adopt a penalty function as below:

$$f = A + \alpha \cdot \left(\sum_{i \in N} s_i^2 + \sum_{(p,q) \in B} t_{pq}^2 + \sum_{ir \in NR} dw_{ir}^2 + \sum_{jc \in ND(ir)} ew_{ir,jc}^2 \right), \tag{6}$$

where α is the penalty parameter. A, s_i, $t_{p,q}$, dw_{ir} and $ew_{ir,jc}$ are defined in equations (1) , (2) , (3) , (4) and (5) respectively. Let us set the penalty term p_t as

$$p_t = \alpha \cdot \left(\sum_{i \in N} s_i^2 + \sum_{(p,q) \in B} t_{pq}^2 + \sum_{ir \in NR} dw_{ir}^2 + \sum_{jc \in ND(ir)} ew_{ir,jc}^2 \right), \tag{7}$$

We then rewrite the penalty function as $f=A+P_t$. Meanwhile, we transform the original constrained problem into a problem of minimizing the penalty function (8),

$$\begin{aligned} \min f &= A + p_t \\ &= A + \alpha \cdot \left(\sum_{i \in N} s_i^2 + \sum_{(p,q) \in B} t_{pq}^2 + \sum_{ir \in NR} dw_{ir}^2 + \sum_{jc \in ND(ir)} ew_{ir,jc}^2 \right), \end{aligned} \tag{8}$$

In this optimization problem, the variable is the width vector $W=[wr_1, wr_2, \ldots wr_b, wc_1, wc_2, \ldots wc_m]^T$ that consists of two parts: the wire width vector $Wr=[wr_1, wr_2, \ldots wr_b]^T$ and the decap width vector $Wc=[wc_1, wc_2, \ldots wc_m]^T$.

3.2 Optimization Scheme

Given an initial penalty parameter α, we minimize the penalty function. We then increase the value of the penalty parameter α for the next minimization iteration such that the contribution of the penalty with respect to the actual cost is maintained at a

constant level. Such dynamic penalty factor adjustment can help speedup of the convergence of the optimization process. The process continues until all the constraints are satisfied or no improvement can be found. The solution procedure can be described briefly as below:

1. Set an initial value of penalty parameter α; initial wire and decap width vector $W^{(0)}=[wr_1, wr_2, \ldots wr_b, wc_1, wc_2, \ldots wc_m]^T=[w_{min}, w_{min}, \ldots w_{min}, 0, 0, \ldots 0]^T$; and error bound, $\varepsilon_b > 0$.
2. Solve unconstrained minimization problem (8), obtain current width vector $W^{(k)}$.
3. If $p_t^{(k)} < \varepsilon_b$, then stop, else update penalty parameter α, set $k = k+ 1$ and turn to step 2.

3.3 Time-Domain Merged Adjoint Network Method

In order to efficiently compute the gradients used in conjugate gradient method, we directly compute the required gradient vector used in the objective function instead of individual gradients of each wire and decap as traditional method did [9]. By using the convolution method for node voltage sensitivity calculation presented in [9], the final resulting method is a time-domain merged adjoint network method, which was described in detail in [7][18]. The gradient of penalty function f with respect to the width vector W can be expressed as following:

$$\nabla f(W) = \left[\frac{\partial f}{\partial wr_1}, \cdots, \frac{\partial f}{\partial wr_i}, \cdots, \frac{\partial f}{\partial wr_b}, \frac{\partial f}{\partial wc_1}, \cdots, \frac{\partial f}{\partial wc_j}, \cdots, \frac{\partial f}{\partial wc_m} \right]^T, \quad (9)$$

$$i \in B, j \in M$$

Using time-domain merged adjoint network method, we only need to simulate the power network twice. The first simulation is on time period from *0* to *T* with original time-varying current sources in original power network. The second simulation is on time period from *T* to *0* with merged current sources in the adjoint network. Then, voltages got from these two simulations are convolved to obtain the gradients. Such two-simulation approach can significantly speed up the optimization process as the computing cost is independent of violation nodes.

3.4 Equivalent Circuit Modeling for Transient Simulation

Although many efficient simulation methods have been proposed in the past [1]-[3][11]-[16], the regular structures of RLC P/G networks with standard-cell layouts are not explicitly exploited. It was shown that there exist many regular structures in the P/G networks of standard-cell layouts as shown in Fig.4. For such a RLC chain circuit, we can reduce it into a simple resistor-only equivalent circuit at each time step and thus speed up the transient simulation of the P/G networks.

In our method, we combine equivalent circuit modeling with preconditioned conjugate gradient method (PCG)[7][16] to perform the transient analysis. Specifically, at each time step, the numerical integration by using companion models in Norton's form is performed and the original RLC circuit will become a resistor-only circuit.

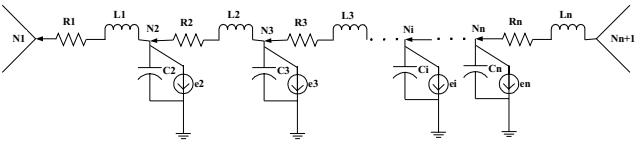

Fig. 4. A series RLC chain in a P/G network.

We then repeatedly apply the Y-Δ network transformation to reduce one node at a time until only the terminal nodes are left. After this, PCG algorithm solves the reduced network. Once the voltage at terminal nodes are solved, all the voltages at intermediate nodes of original circuits can be back solved using the superposition principle.

The new equivalent modeling technique can be viewed as a special pre-ordered Gaussian elimination process where no-fills are generated. By combining it with PCG, we take advantage of both direct (via Gaussian elimination) and iterative approaches.

4 Experimental Results

The proposed optimization algorithm has been implemented in *C* and *C++* programming languages. All the experimental results are obtained on *SUN UltraSparc* workstation V880 with 750MHz CPU and 2GB memory.

We tested our program with some real industry standard-cell circuits with preplacement information in LEF/DEF format. Those circuits have complexities ranging from 744 nodes to 1.6 million nodes. To demonstrate the efficiency of our algorithm we compare the experimental results with the results from method named two-step optimization method that first size the wire widths and then tune decap size. Table 1 shows the parameters of each circuit and the experimental results of the two-step method. In table 1, the violation nodes are nodes on the power grid that violate the voltage drop constraints or the electro-migration constraints. Column 4 and 5 in table 1 show the routing resources and space chip areas used by the two-step optimization method respectively. The last column shows the running time.

Table 1. Experimental results of two-step mothod

Name of circuits	#nodes	#violation nodes	Area of power grid (μm^2)	Area of decaps (mm^2)	Time(s)
Test1	744	91	6006.00	0.0057	110.23
Test2	3741	665	16828.50	0.0350	341.78
Test3	32112	3683	66452.40	0.1745	1315.63
Test4	112392	10755	125239.68	2.6465	2549.37
Test5	321120	11496	87264.00	3.2290	9485.69
Test6	1618026	612132	368288.00	3.0256	20499.89

Table 2. Comparison of new algorithm with the two-step method

Name of circuits	η (wire/decap)	Area of power grid (μm^2)	Area of power grid ratio	Area of decaps (mm^2)	Area of decaps ratio	Time(s)
Test1	0.70/0.30	4938.51	82.23%	0.0051	90.52%	28.79
Test2	0.62/0.38	13503.65	80.24%	0.0321	91.63%	736.18
Test3	0.43/0.57	52500.77	79.01%	0.1861	106.61%	1382.81
Test4	0.60/0.40	44202.21	35.29%	0.9596	36.26%	4297.58
Test5	0.76/0.24	87264.47	100.00%	2.7587	85.44%	17058.32
Test6	0.10/0.90	194976.00	52.94%	2.1878	72.31%	22252.80

Table 2 shows the experimental results of the proposed optimization algorithm and the comparison between these two algorithms. In column 2, η is the weighted factor for routing resource and total decap area. Designers can properly set the value of η to set up the priorities of the resources of routing and die area. In Table 2, for each circuit, the power grid area and decap area obtained from two-step method are normalized to 1 and are compared with the new method in terms of normalized ratios in column 4 and 6. From colunm 4, 6 and 7 we can find that the new algorithm is more area efficient although it takes more time to achieve this. It shows that our algorithm has the capability of optimizing very large scale circuit instance. For instance it takes 6.2 hours to optimize the largest circuits instance (Test6) with 16 millions nodes. After the optimization, all the violations are gone. Although we note that if the optimization problem is over constrained, we still may end up with violations after optimizations.

5 Conclusion

In this paper, we have proposed an efficient algorithm to optimize P/G networks modeled as RLC circuits by simultaneously sizing wires and tuning decap sizes subject to the reliability and design rule constraints of VLSI on-chip P/G grids. Our algorithm utilizes gradient-based nonlinear programming algorithm to search for the best solution and is flexible enough to handle any P/G circuits with capacitive and inductive parasitics. By applying time-domain merged adjoint network method combined with a novel equivalent circuit modeling technique, gradients used in our non-linear programming framework can be efficiently computed, which is key to the overall efficiency of the new algorithm. Experimental results show that by simultaneously sizing the areas of decaps and areas of power grid wire segments, the new algorithm is more area efficient compared with a two-step algorithm where wire sizing and decap allocation are done sequentially. Comparing with the existing methods, the new method can simultaneously perform the wire sizing and decap budgeting on P/G networks modeled as RLC networks with millions of nodes in a reasonable time for the first time.

References

1. H. H. Chen, D. D. Ling. "Power supply noise analysis methodology for deep-submicron VLSI chip design." Proc. 34th ACM/IEEE Design Automation Conf., pp. 638-643,1997.
2. G. Bai, S. Bobba and I. N. Hajj, "Simulation and optimization of the power distribution network in VLSI circuits", Proc. IEEE/ACM International Conf. on Computer-Aided Design., pp. 481–486, 2000.
3. H.-H. Su, K. H. Gala, and S. S. Sapatnekar. "Fast analysis and optimization of power/ground networks." Proc. IEEE/ACM International Conf. on Computer-Aided Design., pp. 477–482, 2000.
4. S. Chowdhury and M. A. Breuer, "Minimal area design of power/ground nets having graph topologies," IEEE Trans. on Circuits and Systems, pp.1441~1451, December. 1987
5. T. Mitsuhashi and E.S. Kuh, "Power and Ground Network Topology Optimization for Cell Based VLSIs", Proc. 29th ACM/IEEE Design Automation Conference, pp. 524~529, 1992
6. X.-D. Tan and C.-J. Shi. "Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programming." Proc. 36th ACM/IEEE Design Automation Conf., pp. 78-83,1999.
7. X.-H. Wu, X.-L. Hong, Y.-C. Cai, and et al. "Area minimization of power distribution network using efficient nonlinear programming techniques." Proc. IEEE/ACM International Conf. on Computer-Aided Design., pp. 153–157, 2001.
8. S.-Y. Zhao, K.K. Roy, C.-K. Koh. "Decoupling capacitance allocation for power supply noise suppression." Proc. IEEE/ACM International Symp. on Physical Design, pp.66-71,2001.
9. 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. " Proc. IEEE/ACM International Symp. on Physical Design, pp.68-75,2002.
10. K. Wang and M. Marek-Sadowask, "On-chip power supply network optimization using multigrid-based technique", Proc. ACM/IEEE Design Automation Conf., pp. 113–118, June, 2003.
11. X.-D. Tan and C.-J. Shi. "Fast power-ground network optimization using equivalent circuit modeling," Proc. 38th ACM/IEEE Design Automation Conf., pp. 550–554, 2001.
12. A. Odabasioglu, M. Celik, and L. T. Pilleggi, "PRIME: Passive reduction-order interconnect macromodeling algorithm," IEEE Trans. On Computer-Aided Design, vol. 17, no. 8, pp.645-654. 1998
13. A Dharchoudhury, R. Panda, D. Blaauw, R. Vaidyanathan and et al "Design and analysis of power distribution networks in power PC microprocessors", Proc. ACM/IEEE Design Automation Conf., pp. 738–743, June, 1998.
14. J. N. Kozhaya, S. R. Nassif, and F.N. Najm. "A multigrid-like technique for power grid analysis. IEEE Trans. Computer-Aided Design, vol. 21, no.10, pp. 1148-1160, Oct. 2002.
15. M. Zhao, R. V. Panda, S. S. Sapatnekar and et al, "Hierarchical analysis of power distribution networks", Proc. ACM/IEEE Design Automation Conf., pp. 150–155, June, 2000
16. T. Chen and C. C. Chen, "Efficient large-scale power grid analysis based on preconditioned Krylov-subspace iterative method", Proc. ACM/IEEE Design Automation Conf., pp. 559–562, June, 2001
17. S.R.Nassif, and J.N.Kozhaya. "Fast power grid simulation", Proc. ACM/IEEE Design Automation Conf. , pp.156-161, June, 2000.
18. J. Fu and Z. Luo and X. Hong and Y. Cai and S. X.-D. Tan and Z. Pan "A fast decoupling capacitor budgeting algorithm for robust on-chip power delivery", Proc. Asia South Pacific Design Automation Conf. (ASPDAC), pp. 505-510, Jan. 2004.