PAPER *Special Section on VLSI Design and CAD Algorithms*

A Fast Decoupling Capacitor Budgeting Algorithm for Robust On-Chip Power Delivery*

Jingjing FU[†a)], ***Member***, **Zuying LUO**[†], **Xianlong HONG**[†], **Yici CAI**[†], **Sheldon X.-D. TAN**[††], ***and*** **Zhu PAN**[†], ***Nonmembers***

SUMMARY In this paper, we present an efficient method to budget on-chip decoupling capacitors (decaps) to optimize power delivery networks in an area efficient way. Our algorithm is based on an efficient gradient-based non-linear programming method for searching the solution. Our contributions are an efficient gradient computation method (time-domain merged adjoint network method) and a novel equivalent circuit modeling technique to speed up the optimization process. Experimental results demonstrate that the algorithm is capable of efficiently optimizing very large scale P/G networks.

key words: *Power Delivery, Optimization, Decoupling Capacitor, Non-linear Programming*

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 power/ground (P/G) networks where noise margins have been reduced greatly in VLSI designs due to decreasing supply voltages and the presence of excessive voltage drops coming from resistive and inductive effects. For dynamic voltage fluctuations on a P/G network, adding decap is regarded as the most efficient way to reduce such noises [7][8]. To this end, some spare chip area will be reserved for decaps in the floorplanning and placement phases. In this paper, we focus on standard-cell like ASIC layout structures as shown in Fig.1 where the white space are the spare space that can be used for adding decaps for noise reduction of P/G networks as well as for adding buffers for optimizing timing in global signal networks and clock networks. Given the same voltage fluctuation constraints, smaller decap area is more desirable as more spare space can be used for buffering in the later routing phases. The problem we try to solve in this paper is how to minimize the decap area subject to the important P/G network integrity constraints, which include voltage fluctuation, electromigration and other design rules.

Due to the importance of reducing noise on the P/G networks, many significant research efforts have been carried out in the past to reduce noises while avoiding using excessive areas [1]-[8]. Methods in [2]-[6] size wire segments of P/G grids for optimizing area of P/G networks. Among them, conjugate gradient based method in [5] and equivalent circuit modeling method in [6] can only be applied to resistor-only P/G networks. Methods proposed in [1][7][8] allocate and optimize decaps to reduce transient noises in P/G networks. Methods [1][7] focus on the full custom design styles and [8] targets the standard cell layout style where quadratic programming was used to allocate decaps without explicitly optimizing the area. Recently a multigrid-based decap allocation approach was proposed in [19]. But the method can only be applied to very regular mesh-structured P/G networks due to the geometric oriented grid complexity reduction process. Furthermore, the important parasitic capacitances and inductances on P/G wires were also ignored (which actually are difficult to be incorporated) in the method.

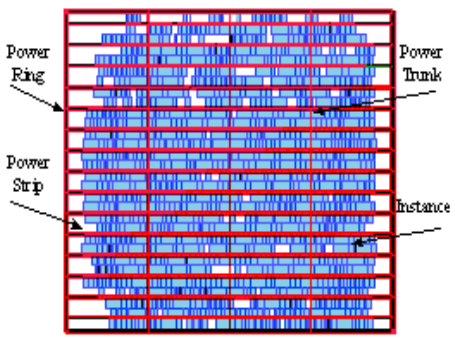


Fig. 1 A P/G network of standard cell based ASICs.

In this paper, we propose an efficient approach to reduce transient noises on P/G networks via decap allocation in an area efficient way based on standard-cell layout style. In our method, we model each segment of the power grid as a lumped RLC element, which is in contrast to [8] and [19] where P/G wire segments are modeled as RC/R components. Parasitics from power

Manuscript received March 22, 2004.

[†]The authors are with Department of Computer Science and Technology, Tsinghua University, Beijing, P.R.C., 100084

[††]The author is with Department of Electrical Engineering, University of California at Riverside, CA 92521, CA

a) E-mail: fujingjing00@mails.tsinghua.edu.cn

*This work was supported by National Natural Science Foundation of China 90307017 and 60176016.

Some preliminary results of this paper was presented at Asia South Pacific Design Automation Conference (ASPDAC), January, 2004. [20]

pads and packages are also modeled as lumped RLC elements similar to [1]. The nonlinear devices or modules are modeled as time-varying current sources, which can be obtained by off-line logic simulation. 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. We haven't considered the mutual inductance in this work. And it will be the goal of our future work.

Different from [8], we explicitly consider area reduction as the primary objective as spare areas in a chip are premium resources and should be used more economically for different physical optimization processes. For instance, the unused white space can be further allocated for buffering to improve timing in the later phases of physical design. The new approach also considers electron-migration constraints that existing approach [8] failed to consider. The new approach consists of an efficient gradient-based non-linear programming technique [5] to minimize the decap area. As the most time-consuming step in gradient-based approach is the gradient computation via transient simulation of P/G networks, an equivalent circuit modeling scheme for RLC chain circuit is proposed to reduce the complexities of P/G networks for speeding up transient simulations and a time-domain merged adjoint network method is used to speed up the computation of gradients. Experimental results show that our method uses smaller decap areas compared with a decap allocation scheme aiming at reducing voltage noise only and can optimize substantial large P/G networks compared with similar existing approaches.

This paper is organized as follows. Section 2 formulates the problem of the decap area minimization. The efficient non-linear programming technique and the equivalent circuit modeling method are described in Section 3. The time complexity of the algorithm is analyzed in Section 4. Experimental results are presented in Section 5. Section 6 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 $B = \{1, ..., b\}$ 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 which is function of time t.

2.1 Objective Function

The objective is to minimize the decap area of a power network subject to design rules and power grid integrity related constraints:

$$\min A(\mathbf{w}_r) = \min \sum_{j \in M} (w_j \times H), \tag{1}$$

where H is the fixed height of standard cells and w_j is the width of the decap connected to node j. $\mathbf{w}_r = [w_1, ..., w_j, ..., w_m], j \in M$ is the width vector of all the tunable decaps.

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 [8] and reformulate the voltage drop constraints in the following for power networks:

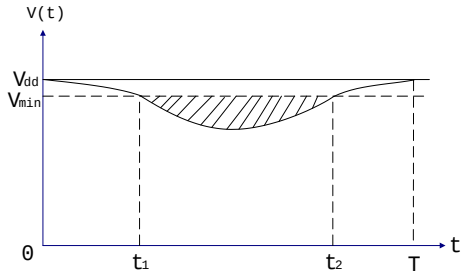


Fig. 2 Illustration of dynamic voltage drops.

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

Where $[t_1, t_2]$ is the time interval in which the constraint is violated as shown in Fig.2. V_{min} is the minimum threshold value of node voltage based on the technology and the routing layer. We require $s_i \leq 0$ to restrict the voltage drop at node i. 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 $|i_{p,q}(t)| \leq w_{p,q} \times \sigma$, and can be re-written as $|v_p(t) - v_q(t)| \leq \rho l_{p,q} \sigma$, σ is a constant for a particular routing layer with a fixed thickness. 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:

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

where $[t_1, t_2]$ is the time interval in which the constraint is violated. We require $t_{p,q} \leq 0$ to restrict the current density in branch (p, q). Similarly, there may exist several such intervals in one clock cycle.

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

$$d_r = \max(\sum_{l \in P(r)} w_{r,l} - w_{t,r}, 0), r \in R. \quad (4)$$

where R is the row set defined as $\{1, 2, \ldots, N_{row}\}$ and N_{row} is the row number of the cells in the placement; *P(r)* is the decap position set of row *r*. $w_{r,l}$ is the width of the decap at row r and position *l;* and $w_{t,r}$ is the width of total spare area in row *r*.

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

$$e_{r,l} = \max(w_{r,l} - w_{t,r}, 0), r \in R, l \in P(r). \quad (5)$$

3. Solution Based on Non-Linear Programming

The resulting problem is a nonlinear minimization problem as voltages and currents are nonlinear functions of decap sizes. As a result, we propose to use gradient-based conjugate gradient method to solve the optimization problem as gradients can be efficiently computed in time-domain by our new merged adjoint method to be discussed later.

Before the optimization, all the decap widths are set to be zero. We then analyze the network to obtain the node voltage waveforms and the branch currents, and identify the constraint violations. Conjugate gradient method proceeds in an iterative way. In each optimization iteration, a conjugate gradient is computed to guide the updating of the 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 in our approach 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:

$$f(\mathbf{w}_r) = A(\mathbf{w}_r) + \alpha \cdot \left(\sum_{i \in N} s_i^2 + \sum_{(p,q) \in B} t_{pq}^2 + \sum_{r \in R} d_r^2 + \sum_{l \in P(r)} e_{r,l}^2\right), \quad (6)$$

where α is the penalty parameter. $A(\mathbf{w}_r)$, s_i, $t_{p,q}$, d_r and $e_{r,l}$ are defined in equations (1), (2), (3), (4) and (5) respectively. Let us define the penalty term p_t as

$$p_t = \alpha \cdot \left(\sum_{i \in N} s_i^2 + \sum_{(p,q) \in B} t_{pq}^2 + \sum_{r \in R} d_r^2 + \sum_{l \in P(r)} e_{r,l}^2\right).$$

We then rewrite the penalty function as $f = A + p_t$. As a result, we transform the original constrained problem into the problem of minimizing the following penalized objective function,

$$f(\mathbf{w}_r) = A(\mathbf{w}_r) + p_t \quad (7)$$
$$= A(\mathbf{w}_r) + \alpha \cdot \left(\sum_{i \in N} s_i^2 + \sum_{(p,q) \in B} t_{pq}^2 + \sum_{r \in R} d_r^2 + \sum_{l \in P(r)} e_{r,l}^2\right).$$

3.2 Optimization Scheme (modified)

Given an initial penalty parameter α, we proceed to minimize the penalized objective function. Then we increase the value of the penalty parameter α for the next minimization iteration. The process continues until all the constraints are satisfied or no improvement can be made. The solution procedure can be described as below:

1. Set an initial value of penalty parameter α; initial decap width vector $\mathbf{w}_r^{(0)}$ and error bound $\varepsilon_b > 0$.
2. Solve unconstrained minimization problem (7), obtain current width vector $\mathbf{w}_r^{(k)}$.
3. If $p_t^{(k)} < \varepsilon_b$, stop, otherwise update penalty parameter α, set $k = k+ 1$ and turn to step 2.

In traditional penalty method, penalty parameter α is always updated by timing a fixed updating ratio, which may let experimental results rely on updating ratio. To solve this problem, we use a better way to update penalty parameter α. Let r_{po} be a constant ratio of penalty term to objective function. We then set penalty parameter α in last unconstrained optimization to be α_k. Then the new α_{k+1} can be calculated by the constant r_{po} as below,

$$\alpha_{k+1} = \alpha_k \times \frac{r_{po} \times A(w_r)}{p_t}$$

3.3 Time-Domain Merged Adjoint Network Method (modified)

As gradient is required in every iteration of conjugate gradient algorithm, it is crucial for efficiently computing the gradients. Gradients are typically computed by using adjoint method [15], which was extended into time domain [18]. The time-domain adjoint method can compute the sensitivities of a node voltage waveform for all the variables after two transient simulations. So the number of transient simulations will goes up linearly with number of node voltage waveforms we try to change, which is expensive if the number of violation nodes is large as shown in [7]. To mitigate the computation inefficiency problem of traditional adjoint method, merged adjoint method was proposed recently in [5] for gradient computation based on DC analysis and resistor-only P/G grid optimization.

In this paper, we extend the DC merged adjoint method to time domain for efficient computation of the gradients. The main idea of time-domain merged adjoint method is to directly compute all the derivative of the objective function (7) instead of each individual gradient of decap. We will show how this can be extended to the time domain. The resulting gradient computation will be significantly faster than the traditional adjoint method.

Specifically, the gradient of panelized objective function f with respect to the decap width vector $\mathbf{w}_r$ can be expressed as follows:

$$\nabla f(\mathbf{w}_r) = \left[\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}, \cdots, \frac{\partial f}{\partial w_j}, \cdots, \frac{\partial f}{\partial w_m}\right]^T, j \in M \quad (8)$$

With equations (1),(6), we can derive the following equation:

$$\begin{aligned}\frac{\partial f}{\partial w_j} &= \frac{\partial A}{\partial w_j} + \frac{\partial}{\partial w_j}\left[\alpha \cdot \left(\sum_{i\in N} s_i^2 + \sum_{(p,q)\in B} t_{pq}^2 \right.\right. \\ &\quad \left.\left. + \sum_{r\in R} d_r^2 + \sum_{l\in P(r)} e_{r,l}^2\right)\right] \\ &= H + 2\alpha\left(\sum_{i\in N} s_i \frac{\partial s_i}{\partial w_j} + \sum_{(p,q)\in B} t_{pq}\frac{\partial t_{pq}}{\partial w_j} \right. \\ &\quad \left. + \sum_{r\in R} d_r \frac{\partial d_r}{\partial w_j} + \sum_{l\in P(r)} e_{r,l}\frac{\partial e_{r,l}}{\partial w_j}\right).\end{aligned} \quad (9)$$

Assuming that the decap c_j is located at r^{th} row and l^{th} position, equation (9) can be transformed into

$$\begin{aligned}\frac{\partial f}{\partial w_j} &= H + 2\alpha\left(d_r + e_{r,l}\right) \\ &\quad + 2\alpha\left(\sum_{i\in N} s_i \frac{\partial s_i}{\partial w_j} + \sum_{(p,q)\in B} t_{pq}\frac{\partial t_{pq}}{\partial w_j}\right). \\ &\quad r \in R, l \in P(r)\end{aligned} \quad (10)$$

In the sequel, we focus on the derivative computations for the constraints s_i and t_{pq}.

Let us assume that the relation between decap and its width is

$$c = \frac{\varepsilon_{ox}}{T_{ox}} \times w \times H$$

where ε_{ox} and T_{ox} are the permittivity and thickness of the gate oxide respectively. Equation (10) can be transformed into

$$\begin{aligned}\frac{\partial f}{\partial w_j} &= H + 2\alpha\left(d_r + e_{r,l}\right) \\ &\quad + \frac{2\alpha\varepsilon_{ox}H}{T_{ox}}\left(\sum_{i\in N} s_i \frac{\partial s_i}{\partial c_j} + \sum_{(p,q)\in B} t_{pq}\frac{\partial t_{pq}}{\partial c_j}\right) \\ &\quad r \in R, l \in P(r).\end{aligned} \quad (11)$$

Let us assume decap c_j is connected to the l^{th} node on the power network. According to the convolution method for sensitivity calculation [18], The sensitivities $\frac{\partial s_i}{\partial c_j}$ and $\frac{\partial t_{pq}}{\partial c_j}$ can be computed as follows:

$$\frac{\partial s_i}{\partial c_j} = \int_0^T v'_{l,B_i}(t) \times \dot{v}_l(T-t)dt \quad (12)$$

$$\begin{aligned}\frac{\partial t_{pq}}{\partial c_j} &= \int_0^T v'_{l,t_{pq}}(t) \times \dot{v}_l(T-t)dt \\ &= \int_0^T \left[v'_{l,B_p}(t) - v'_{l,B_q}(t)\right] \times \dot{v}_l(T-t)dt,\end{aligned} \quad (13)$$

where $\dot{v}_l(T-t) = \frac{dv_l(T-t)}{dt}$ is the voltage derivative with respect to time t at the l^{th} node to which the decap c_j is connected in the original circuit, and $v'_{l,B_i}(t)$ is the node voltage obtained from the adjoint circuit where the current $i_s(t,t_1,t_2) = -u(t-t_1)+u(t-t_2)$ is applied at node i due to constraint s_i or t_{pq} and $u(t)$ is the unit step function. Specifically, we have

$$\begin{aligned}v'_{l,B_i}(t) &= Z(l)v'_{B_i} \\ &= Z(l)(A^T)^{-1}\left[B_i(t) + EC_i(t) + EL_i(t)\right] \\ B(i) &= [0,0,\cdots,0,i_s(t,t_1,t_2),0,\cdots,0]^T, \\ Z(l) &= [0,0,\cdots,0,1,0,\cdots,0],\end{aligned} \quad (14)$$

where A is the coefficient matrix of the nodal voltage equation $AV(t) = I(t)$, $B_i(t)$ is a current vector in which all currents are zero except for the current injected into node i which is $i_s(t,t_1,t_2)$, and $Z(l)$ is a

position vector in which all items are zero except for the position at the l^{th} node, which is 1. $EC_i(t)$ and $EL_i(t)$ are equivalent current vectors which are functions of the states of capacitor and inductance in the adjoint circuit according to node i.

In traditional adjoint method, we should build a series of adjoint circuits according to each violation node and do transient simulation for each adjoint circuit to get the node voltage waveforms to calculate the sensitivities $\frac{\partial s_i}{\partial c_j}$ and $\frac{\partial t_{pq}}{\partial c_j}$ separately. Then we can further compute the value of $\frac{\partial f}{\partial w_j}$. Thus it is very expensive if the number of violation nodes is large. To solve this problem, we propose merged adjoint method to efficiently calculate the value of $\frac{\partial f}{\partial w_j}$ by building the adjoint circuit only once.

The idea of merged adjoint network method is based on the fact that all adjoint networks (for different violations) have the same coefficient matrix as the original network, so we can merge all their inputs (therefore their response) to compute the gradient vector of the objective function (8) directly.

Specifically, by using superposition, we can obtain the merged stimulus vector $I_{new}(t)$ due to constraint violations as:

$$\begin{aligned} I_{new}(t) &= \sum_{i\in N}[s_i \times B_i(t)] \\ &\quad + \sum_{(p,q)\in B}\{t_{pq}\times[B_p(t)-B_q(t)]\}. \end{aligned} \tag{15}$$

With this merged stimulus, we can compute the responding voltage vector $V'(t)$ as follows:

$$\begin{aligned} V'(t) &= (A^T)^{-1}\left[I_{new}(t)+EC(t)+EL(t)\right] \\ &= (A^T)^{-1}I_{total}(t). \end{aligned} \tag{16}$$

where $I_{total}(t) = I_{new}(t) + EC(t) + EL(t)$. $EC(t)$ and $EL(t)$ are equivalent current vectors which are functions of the states of capacitors and inductances in the merged adjoint circuit with I_{new} as its current sources. As a result, we obtain node voltages of the merged adjoint network.

$$v'_{c_j} = Z(l)V'(t). \tag{17}$$

With equation (11) and (17), we can simply compute the gradient of penalized objective function as follows:

$$\begin{aligned} \frac{\partial f}{\partial w_j} &= H + 2\alpha\left(d_r + e_{r,l}\right) \\ +\frac{2\alpha\varepsilon_{ox}H}{T_{ox}} &\int_0^T \left\{\left[Z(l)\times V'(t)\right]\times \dot{v}_l(T-t)\right\}dt \\ &= H + 2\alpha\left(d_r + e_{r,l}\right) \\ +\frac{2\alpha\varepsilon_{ox}H}{T_{ox}} &\int_0^T \left\{\left[Z(l)\times (A^T)^{-1}\times I_{total}(t)\right]\right. \\ &\left.\times \dot{v}_l(T-t)\right\}dt \\ &r\in R, l\in P(r). \end{aligned} \tag{18}$$

As a result, we only need to simulate the power network twice. The first time period is from 0 to T with the original time-varying current sources. The second time period is from T to 0 with $I_{new}(t)$ as the current sources. The voltage waveforms at the l^{th} node obtained from the two simulations are convolved according to (18) to obtain the gradient with respect to w_j. Notice that we need to cache all the voltage waveforms at the violated nodes. But we can use the piece wise linear functions to represent the node voltage waveforms as done in [8] to make the algorithm more space efficient.

3.4 Transient Simulation Based on Equivalent Circuit Modeling

Although many efficient simulation methods have been proposed in the past [1]-[3], [8]-[17], the regular structures of RLC P/G networks on 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.3. For such a RLC chain circuit, we can reduce it into a simple resistor-only equivalent circuit and thus speed up the transient simulation of the P/G networks.

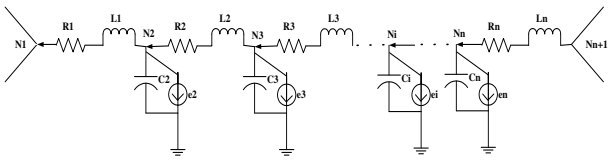


Fig. 3 A series RLC chain in a P/G network.

In our method, we combine equivalent circuit modeling with preconditioned conjugate gradient method (PCG) [5], [9] to solve the P/G network in time domain. Specifically, at each time step, the numerical integration by using companion models in Norton's forms is performed and the original RLC circuit will become a resistor-only circuit. We then repeatedly apply the Y-Δ network transformation to reduce one node at a time until only the cross (terminal) nodes are left. After this, PCG algorithm solves the reduced network. Once the voltages at cross nodes are solved, all the voltages at intermediate nodes of original circuits can be back solved using the superposition principle.

3.5 Decap Legalization

In typically standard cell designs, decaps are assigned

as dummy decap cells. Since the solutions we obtain are continuous in terms of decap widths, we need to *legalize* the decap widths in the solutions when actually decaps are assigned. In our approach, we use some bigger decaps for absorbing smaller ones around them, and then round off the areas of the clustered decaps to form a number of dummy decaps.

4. Analysis of Time Complexity (modified)

In this section, we briefly discuss the time complexity of the overall optimization algorithm. As mentioned above, in each area optimization iteration, we need to solve the original network and the adjoint network once, then compute new searching direction and do line searching along this direction. The overall time complexity can be expressed as

$$T = \sum_{i=1}^{N_{iter}} \left(T_{so}(i) + T_{sa}(i) + \sum_{j=1}^{N_{line}(i)} T_{so}(i,j) \right) \quad (19)$$

where, N_{iter} is the total number of area optimization iterations; $N_{line}(i)$ is the number of line searching in iteration i; $T_{so}(i)$ and $T_{sa}(i)$ are the time to solve original network and adjoint network with PCG algorithm in iteration i; $T_{so}(i, j)$ is also the time to solve original network with PCG during line searching.

The components in (19) which consume most of the running time are the time for solving the network and its adjoint network. In our method, as we apply equivalent circuit modeling method to reduce P/G network complexities first, computation costs have been greatly reduced. Specifically, if there are total n nodes in the P/G network, $n_{cross}$is the number of nodes left after reduction and n_{mid} is the number of nodes suppressed. $n = n_{cross} + n_{mid}$. Typically n_{cross} is far less than n in the P/G networks of standard-cell layouts as shown in Table 1. The circuits listed in the table are real industry standard-cell circuits. As a result, the reduced P/G network is significantly smaller than the original network.

Table 1 Circuit nodes and cross nodes number comparison.

Name of circuits	#node	#Cross node	Ratio
U_cnt100	744	96	12.9%
U_cnt500	3741	364	9.73%
U05614	32112	950	2.96%
U19649	112392	1408	1.25%
U56140	321120	2768	0.86%
Nec_600k	1618026	6104	0.38%

In Fig.4, we draw curve of transient simulation time versus number of nodes in the testing circuits. Six networks with different ratio of n_{cross} to n as shown in Table 1 are solved. It is shown that the CPU times used by the new algorithm grow almost linearly as the ratio of n_{cross} to n typically decreases significantly when the complexities of circuits increase.

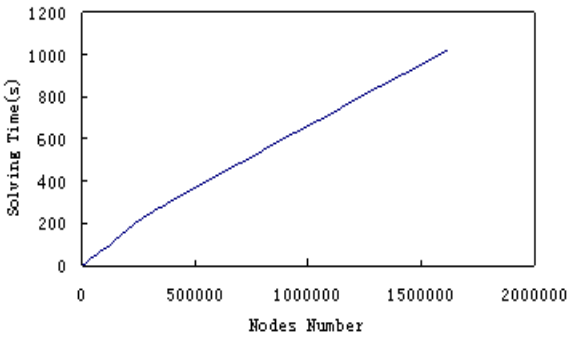


Fig. 4 Solving time versus nodes number.

5. Experimental Results

The proposed decap optimization algorithm has been implemented in C and C++. 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 pre-placement information in LEF/DEF format. Those circuits have complexities ranging from 744 nodes to 1.6 million nodes. Table 2 shows the parameters of three smaller circuits. The violation nodes are nodes on the power grid that violate the voltage drop constraints or the electro-migration constraints.

Table 2 Parameters of the P/G circuits.

Names of circuits	#nodes	#violation nodes
U_cnt100	744	91
U_cnt500	3741	665
U05614	32112	3683

Ideally, we would like to compare our work with some similar existing works such as [8]. But due to the proprietary nature of those algorithms and their test circuits, direct comparison is impossible. Instead, the proposed algorithm is compared with a simple sensitivity based decap allocation heuristic method that we developed, which is similar to sensitivity based method used in [2]. In the simple heuristic algorithm, we first calculate the sensitivities of all the decap candidate nodes using the time-domain merged adjoint method and sort them in a descending order. Then we choose a number of the nodes with largest sensitivities to add a small number of decaps around them and also adjust the cell positions in the row. If the violations still exist, the process is repeated until no violations are present or no improvement can be made. Table 3 summarizes the comparison results of two algorithms.

Table 3 clearly shows that area obtained by the algorithm proposed in this paper is significantly smaller than that obtained from the simple heuristic method

Table 3 Optimization efficiency comparison.

Name of circuits	Heuristic algorithm			Programming algorithm		
	Time (s)	Area of Decap (μm^2)	#decap	Time (s)	Area of Decap (μm^2)	#decap
U_cnt100	38.02	7951.429	128	109.62	5677.962	196
U_cnt500	223.59	40634.776	549	339.41	35007.882	797
U05614	954.09	245153.92	7736	1285.15	174591.12	9453

although the new method takes a little bit more CPU time.

Compared with a similar standard-cell based decap allocation algorithm [8], our algorithm is able to handle much larger circuits. For instance, the largest circuit with 1,618,026 nodes can be optimized in 7.38 hours, while the method in [8] can size circuit with only 828 nodes. We test our algorithm with some large circuits and the results are listed in Table 4. For all the test cases, the new algorithm successfully eliminates all the violation nodes.

Fig.5 shows the distribution profile of violation nodes for the circuit *c_cnt100* before optimization where violation nodes are marked by black color. Since the pads are located on the up-left and bottom-right corners of the chip, the violation nodes are mainly at the center of the circuit as expected.

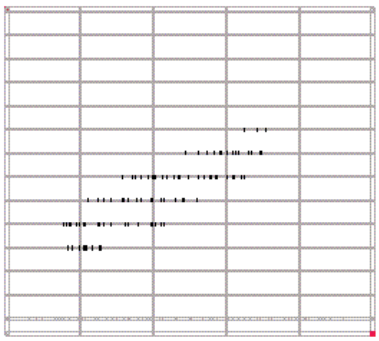


Fig. 5 Distribution of violation nodes.

Fig.6 and Fig.7 show the decap allocation results from the simple heuristic algorithm and our programming based algorithm for circuit *c_cnt100* respectively. The dark (red) blocks are added decaps and the light (white) ones are spare space. We observe that our new algorithm not only saves more spare area but also spreads decaps closer to the violation nodes while the simple heuristic approach places decaps in more narrow areas which are far from some violation nodes. As a result, our algorithm can save more spare area for later design phases.

6. Conclusion

In this paper, we have proposed an efficient non-linear programming based decoupling capacitor budgeting algorithm. Our method utilizes conjugate gradient algorithm to search for the best solution. The proposed

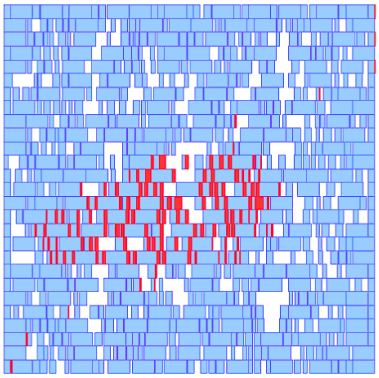


Fig. 6 Decap allocation result of the simple heuristic algorithm.

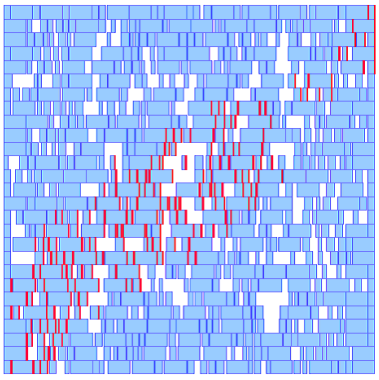


Fig. 7 Decap allocation result of the new algorithm.

time-domain merged adjoint network method combined with a novel equivalent circuit modeling technique leads to very efficient computation of gradients, which is key to the overall efficiency of the optimization algorithm. Experimental results have demonstrated that the proposed algorithm is capable of optimizing P/G networks with million nodes in 7.38 hours. Since the new algorithm explicitly minimizes the area of decaps, it uses significantly smaller decap area compared with a simple heuristic method where the main objective is to just reduce P/G noise. As our new algorithm scales well for large P/G networks and results in less budgeted decap area, it can easily be incorporated into other physical design steps to achieve fast design closure.

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-

Table 4 Statistic results of large P/G circuits.

Name of circuits	#nodes	#violations nodes	Running time (s)	#decaps	Area of decaps (mm^2)	Chip area (mm^2)
U19649	112392	10755	2455.03	10652	2.6465	9.4224
U56140	321120	11496	9216.20	14996	3.2290	36.7236
Nec_600k	1618026	61213	26579.31	5845	0.5223	183.3316

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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] A. Odabasioglu, M. Celik, and L. T. Pilleggi, "PRIMA: passive reduction-order interconnect macromodeling algorithm," *IEEE Trans. On Computer-Aided Design*, vol. 17, no. 8, pp.645-654. 1998.

[11] Y. Cao, Y. Lee, T. Chen and C. C. Chen, "HiPRIME: hierarchical and passivity reserved interconnect macromodeling engine for RLKC power delivery", *Proc. ACM/IEEE Design Automation Conf.*, pp. 379–384, June, 2002.

[12] D. Stark and M. Horowitz, "Techniques for calculating currents and voltages in VLSI power supply networks," *IEEE Trans. on Computer-Aided Design*, vol. 9, no. 2, pp. 126-132, February 1990.

[13] A. Dharchoudhury, R. Panda, D. Blaauw, R. Vaidyanathan, B. Tutuianu and D. Bearden, "Design and analysis of power distribution networks in power PC microprocessors", *Proc. ACM/IEEE Design Automation Conf.*, pp. 738–743, June, 1998.

[14] Y.-M. Lee, C.-P. Chen. "Power grid transient simulation in linear time based on transmission-line-modeling alternating-direction-implicit method." *Proc. IEEE/ACM International Conf. on Computer-Aided Design.*, pp. 75–80, 2001.

[15] R. A. Rohrer, I. Nagel, R. Meyer, and L. Weber,"Computationally efficient electronic circuit noise caculations", *IEEE Journal of Solid State Circuits*, vol. SC-6, pp. 204-213, August 1971.

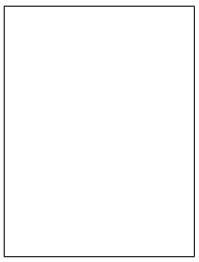
[16] J. N. Kozhaya, S. R. Nassif, and F.N. Najm. "A multigrid-like technique for power grid analysis. *IEEE Trans. Computer-Aided Design*, vol. 21, no.10, pp. 1148-1160, Oct. 2002.

[17] M. Zhao, R. V. Panda, S. S. Sapatnekar and D. Blaauw, "Hierarchical analysis of power distribution networks", *IEEE Trans. Computer-Aided Design*, vol. 9, no. 2, pp. 159–168, Apr. 1990.

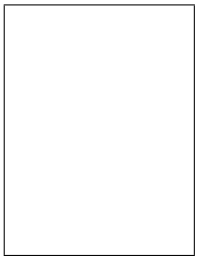
[18] L.O. Chua and P.M. Lin, *Computer-Aided Analysis of Electronic Circuits,* Englewood Cliffs: Prentice-Hall, Inc. 1975.

[19] 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.

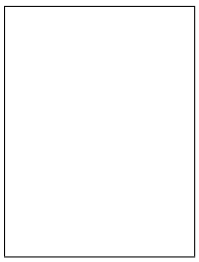
[20] J. Fu and Z. Luo and X. Hong and Y. Cai and S. X.-D. Tan and Z. Pan "A fast decoupling capacitor budgeting algorithm for robust on-chip power delivery", *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, pp. 505-510, Jan. 2004.



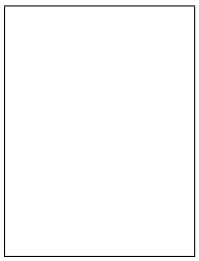
Jingjing Fu received B.S. degree in computer science and engineering department from Southeast University, Nanjing, China, in 2000. She is currently studying toward the Ph.D. degree in computer science & technology department at Tsinghua University, Beijing, China. Her research interests include Power/Ground Network simulation and optimization, Power delivery network noise reduction in the area of physical design.



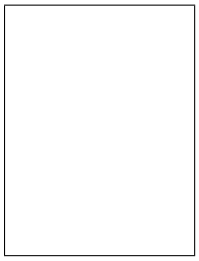
Zuying Luo received Ph.D. degree in department of automation from Tsinghua University in 2002. Now he is a postdoctor in Department of Computer Science & Technology, Tsinghua University, Beijing, China. His research interests include Low-power SOC design, and P/G grid design and optimization.



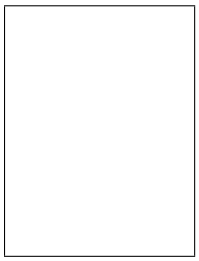
Xianlong Hong graduated from Tsinghua University, Beijing, China in 1964. He was a Visiting Scholar at the University of California, Berkeley, and worked in the research group of Prof. E.S. Kuh form April 1991 to October 1992 and from June to September 1993. Since 1988, he has been a professor in the Department of Computer Science & Technology, Tsinghua University. His research interests include VLSI layout algorithms and DA systems. He is IEEE Follow and the Senior Member of Chinese Institute of Electronics.



Yici Cai received B.S degree in Electronic Engineering from Tsinghua University in 1983 and received M.S. degree in Computer Science from Tsinghua University in 1986. She has been an associate professor in the Department of Computer Science & Technology, Tsinghua University, Beijing, China. Her research interests include VLSI layout algorithms and theory.



Sheldon X.-D. Tan received his B.S. and M.S. degrees in electrical engineering from Fudan University, Shanghai, China in 1992 and 1995, respectively and the Ph.D. degree in electrical and computer engineering from the University of Iowa, Iowa City, in 1999. He was a faculty member in the electrical engineering Department of Fudan University from 1995 to 1996. Now he is with the Department of Electrical Engineering, University California at Riverside as an Assistant Professor.
His research interests include several aspects of design automation for VLSI integrated circuits – signal integrity issues in VLSI physical design, high performance power/ground distribution network design and optimization, symbolic analysis and modeling of mixed-signal/RF/analog circuits.



Zhu Pan received his B.S. degree in Nanjing University of Postage and Communication, Nanjing, China, 2001. Now he is studying in the Dept. of Computer Science & Technology of Tsinghua University as a master candidate. His reasearch interest is the algorithm for transient simulation of Power/Ground network in VLSI.