

# Congestion-Aware Topology Optimization of Structured Power/Ground Networks

Jaskirat Singh, *Student Member, IEEE*, and Sachin S. Sapatnekar, *Fellow, IEEE*

**Abstract**—This paper presents an efficient method for optimizing the design of power/ground (P/G) networks by using locally regular, globally irregular grids. The procedure divides the power grid chip area into rectangular subgrids or tiles. Treating the entire power grid to be composed of many tiles connected to each other enables the use of a hierarchical circuit analysis approach to identify the tiles containing the nodes having the greatest drops. Starting from an initial configuration with an equal number of wires in each of the rectangular tiles, wires are added in the tiles using an iterative sensitivity based optimizer. A novel and efficient table lookup scheme is employed to provide gradient information to the optimizer. Incorporating a congestion penalty term in the cost function ensures that regularity in the grid structure does not aggravate congestion. Experimental results on test circuits of practical chip sizes show that the proposed P/G network topology, after optimization, saves 12%–23% of the chip-wiring area over other commonly used topologies.

**Index Terms**—Congestion-aware, ground, nonuniform grid, power, topology.

## I. INTRODUCTION

TECHNOLOGY scaling in very large scale integration (VLSI) chips has led to an increase in chip densities and, hence, an increase in the currents drawn from the power/ground (P/G) networks. On the other hand, the resistances of the interconnects have also increased due to the decrease in wire widths. As a result, the IR drop in the supply grid wires has become significant and this can lead to logic failures. Specialized techniques for the design and optimization of P/G networks have thus become essential to tackle the problem of signal integrity for the current and future VLSI chips. The key constraints in the design of power supply network are those of IR drop, electromigration (EM) and ground bounce due to inductive effects. To meet these constraints, the typical techniques available to the designers of supply networks are by wire sizing [1]–[6], adding decoupling capacitors (decaps) [7], [8], and using appropriate topologies for the P/G network [9]–[11].

In [1], the authors propose a wire width sizing method based on transforming the constrained nonlinear programming problem into a sequence of linear programs. The problem is formulated to determine optimal wire widths by assuming the currents in segments to be fixed and branch voltages as

variables. An optimization scheme to calculate both the lengths and widths of P/G networks using a sequential network simplex method is proposed in [2]. In [3], the wire widths are optimized by building equivalent circuit models of many series resistors in original network, and then using a sequence of linear programs. The authors of [4] size the wire widths of P/G networks by building an optimization engine that first finds an initial feasible solution, and then iteratively looks for the optimal solution along a feasible direction using a sensitivity analysis method. In [5], another wire sizing algorithm is proposed based on the conjugate gradient method and circuit sensitivity analysis. In this problem formulation, only conductance is used as the variable and the adjoint method [12] is used to calculate the gradient. A heuristic based on minimizing total wire area is developed in [6]. This method determines the optimal wire widths and the network topology, i.e., a tree or a mesh structure by solving a nonlinear convex optimization problem.

A combined technique to appropriately size the wire widths and add decaps through a heuristic based on the transient adjoint sensitivity analysis is proposed in [7]. A decap optimization procedure involving an iterative process of circuit simulation and floor planning is proposed in [8].

The authors of [9] provide a method for topology optimization by removal of selected wires. In this work, they solve the problem by formulating it as a nonlinear combinatorial optimization problem and relaxing some of the constraints. The wires are represented as conductance links between the nodes, and a decision variable vector is used to determine the presence or absence of these conductance links. Some other works in topology optimization [10], [11] deal with the problem of optimal pad assignment to the power/ground grid structures.

The wire-sizing and decap placement methods of [1]–[8] all assume that the topologies of P/G networks are fixed, and only the widths of the wire segments and the positions of decaps need to be determined. These techniques of power grid design by wire sizing and decap placement have a significant cost of overutilization of the chip area. Furthermore, if the wire widths of the supply network vary throughout the chip, the routing of signal nets becomes much more difficult as a lot of bookkeeping must be done to keep track of the locations and widths of P/G wires. In the works on topology optimization, the emphasis is on optimal assignment of the pins to the pads and placement of pads on the power grid. The fact that the topology has a significant influence on the final layout area is recognized, but the quest for a good topology design technique remains an open problem.

In general, it is desirable to have as much regularity as possible in the power grid in order to permit the locations of power

Manuscript received June 14, 2004. This work was supported in part by the Semiconductor Research Corporation under Contract 2003-TJ-1084 and in part by the National Science Foundation under Award CCR-0205227. This paper was recommended by Guest Editor P. Groeneveld.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: jsingh@ece.umn.edu; sachin@ece.umn.edu).

Digital Object Identifier 10.1109/TCAD.2005.846369

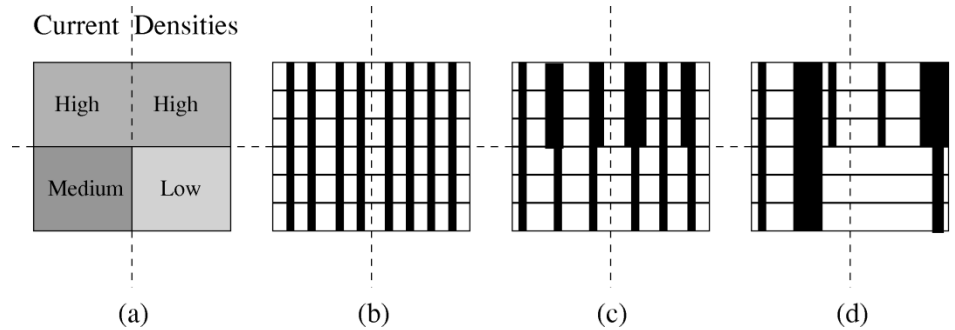


Fig. 1. (a) Current densities for the chip. Possible designs of P/G network. (b) Grid with four wires in each tile and regular wire sizing. (c) Grid with three wires in each tile and irregular wire sizing. (d) Nonuniform grid with variable pitches throughout and irregular sizing. (e) Piecewise uniform grid with four wires in the upper half tiles, two wires in the bottom-left tile and one wire in the bottom-right tile and uniform sizing.

grid wires to be easily accounted for during signal routing. Furthermore, a regular grid structure can easily be analyzed [13] as it results in simpler circuit models. A highly irregular grid (for example, one that has been sized irregularly, or one in which wires have been selectively removed) may well provide an excellent solution if the power grid design problem is viewed in isolation. However, if we consider the *entire* design flow, a high degree of irregularity can be an impediment to the design methodology as it may require a large amount of book-keeping to keep track of the precise locations of the power wires, and to determine which regions have excessive wiring congestion. Moreover, the number of optimizable parameters for such a problem can be very high, which may make the optimization highly computational. At the other extreme, a fully regular grid has few optimizable parameters and is ideal for the signal router. However, the constraint of full regularity can be overly limiting, since the requirement of regularity may cause the design to use excessive wiring resources. For instance, if all wire widths were to be required to be identical, the wires could be oversized in regions that have relatively low current densities.

Our work uses structured regularity in the power grid, with a topology that is intermediate to fully regular grids and highly irregular grids. These grids can be thought of as being a piecewise uniform grid that is *globally irregular* and *locally regular*. This structure combines the best of both worlds. It has the advantages for faster routing that is afforded by fully regular nets, while offering the flexibility in optimization and better resource utilization permitted by irregular topologies. It is possible, in some cases, that due to the regularity in the grid structure, the number of tracks available for signal net routes in high congestion regions are insufficient. However, this aggravation of congestion arising due to a regular grid design, can be checked and controlled if such a problem is anticipated and accounted for in the grid design procedure. In [14], a procedure is developed to simultaneously design the supply grid, under voltage droop constraints, and the signal net, under congestion considerations. Starting from an initial dense regular grid, the noncritical power grid wires in the high congestion regions are removed followed by a heuristic wire sizing step to overcome the effects of wire removal. The resulting grid is irregularly sized and thus, loses the advantages of structured regularity.

In this paper, we extend our work of [15] to incorporate congestion awareness in our supply net design heuristic without sacrificing the desirable property of local regularity in the grid

structure. We also add the EM constraints in the grid design procedure of our previous work. The main features of our method are as follows.

- 1) We present a topology optimization heuristic algorithm to design a supply grid, having a nonuniform mesh-like structure, that meets the IR drop and EM constraints. The optimization that determines the grid structure is carried out under dc conditions using a sensitivity-based heuristic.
- 2) A weighted sum of area and congestion is used as the objective function. Hence, having a penalty term for congestion controls the congestion aggravation, which may arise due to local regularity in the grid structure.
- 3) The macromodeling approach proposed in [16] is used to solve the circuit and determine the most critical node. We introduce reasonable approximations, specific to our scheme, that can be used to further speed up this method.
- 4) Starting from an initial sparse uniform grid, having the same number of wires in all the tiles in both horizontal and vertical directions, we iteratively add wires to the grid until the voltage drop constraints are met. These wires are added so as to maximize the alignments between wires in different grids. We choose to add wires to a sparse grid rather than remove wires from a dense grid because for the dc solution of the power grid circuit, unlike removal of wires, wire addition is guaranteed not to worsen the voltage drop situation.
- 5) We guide the optimization using a novel table look-up-based sensitivity calculation method to generate the gradient.

## II. PROBLEM DESCRIPTION

Our task is to optimize the design of the proposed P/G network given the worst case currents drawn by P/G nets in different regions of the chip, the position of  $V_{dd}$  and ground pads on the chip and the voltage drop specifications.

We use a toy example to illustrate the possible design choices. For simplicity, let us assume that for the given example problem we can divide the chip into four rectangular regions or tiles having different current densities as shown in Fig. 1(a).

The design in Fig. 1(b) corresponds to the case where the voltage drop constraints are met by constructing a regularly

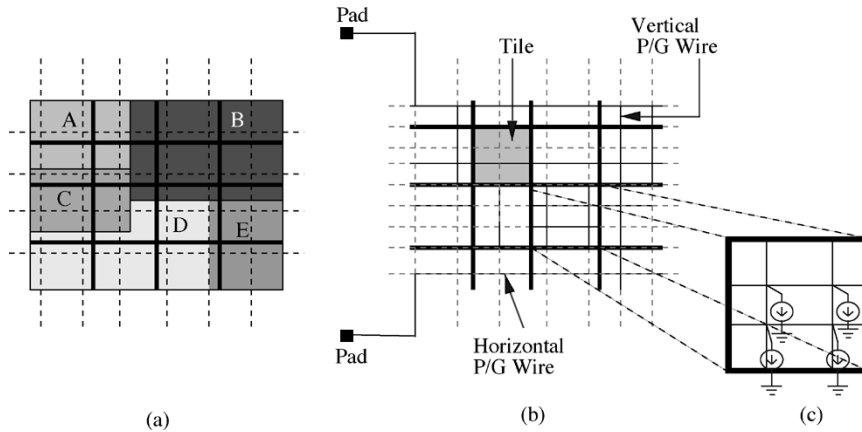


Fig. 2. (a) Illustration of the layout with functional blocks A–E. The dashed lines represent the skeleton grid. (b) Abstraction of the grid into tiles with orthogonal wires. The thick dark lines represent the edge boundaries, along which our model assumes the P/G wires that are always present. These P/G wires at the edge boundaries divide the chip into rectangular *tiles*, represented by the shaded region. The thin dark lines constitute the P/G wires inside a tile, and their density can be different in different tiles. The dashed lines represent the skeleton grid. (c) Internal view of a single tile showing the placement of current sources at internal nodes.

structured grid with regularly sized elements with the same number of wires in the four tiles, i.e., four wires in each tile. This design uses more wire resources than required since the wire sizes and the minimum number of wires in each tile are chosen according to the region with the worst-case voltage drop. The design in Fig. 1(c) meets the design constraints by employing three wires in each tile but the wires in the upper half are sized individually and irregularly to decrease the resistance and reduce the voltage drop. Such a design makes the task of the signal router more difficult, since it must keep track of the variable amount of space available in each region. The design in Fig. 1(d) utilizes the lowest wiring area by using variable pitches and by sizing individual wires separately. However, besides the fact that this design would make the routing of signal nets very difficult, the optimization itself involves numerous design variables and is therefore computationally intensive. The design in Fig. 1(e) is essentially the design we propose and optimize in this paper. This design is piecewise uniform as within a tile it employs a near-constant pitch and uses the same wire sizing throughout the chip. The wires are sized uniformly throughout the chip so as to maintain regularity and meet the IR drop needs. Such a design is more economical in utilization of wiring resources than designs in Fig. 1(b) and (c), and has the desirable property of regularity that Fig. 1(d) lacks, and does not aggravate the congestion problem for signal nets. Moreover, due to an inherent structure in the design, it is easy to optimize.

### III. SOLUTION TECHNIQUE

The sequence of entire optimization procedure is summarized in the following steps.

- 1) The P/G network is abstracted with an equivalent circuit model. Section III-A explains the P/G net circuit model.
- 2) The chip is divided into  $k$  rectangular tiles. An imaginary *skeleton grid* (to be defined in Section III-B) is superimposed on the chip area on which the actual supply grid is built, to maintain wire alignments across tile boundaries. Starting with an equal number of wires in

- all tiles in both horizontal and vertical directions, an initial sparse actual grid is formed on the skeleton grid.
- 3) Each tile is further divided into  $b$  bins or smaller rectangular regions. In this way, an initial congestion value is assigned to each bin in both horizontal and vertical directions. Such congestion values could be obtained from probabilistic congestion estimation techniques such as [17] and [18].
- 4) The grid is analyzed using the macromodeling technique after doing the required port approximations, as described in Section III-C, and the most critical node  $x$  in tile  $i$ , having the maximum voltage drop from  $V_{dd}$ , is determined.
- 5) The voltage sensitivity of the most critical node  $x$ , with respect to increase in area in tile  $i$ , due to the addition of  $l$  wires, is computed using the sensitivity calculation method, as described in Section III-D.
- 6) The increase in congestion of bins of tile  $i$ , due to the addition of  $l$  wires is computed, as explained in Section III-E. The cost function is calculated as the weighted sum of the voltage sensitivity term and the congestion term.
- 7) The number of horizontal or vertical wires in the tile having the minimum cost, is increased by  $l$ . The current sources to internal nodes of the tile are reassigned, so that the sum of the current sources at all internal nodes is the total current drawn by the P/G buses in that tile. The congestion values of the bins in the tile chosen for wire additions are updated.
- 8) Steps 4–7 are repeated until the voltage of the most critical node is greater than a specified value and the EM constraints for all wires in the grid are met.

The following sections explain each of these steps in details.

#### A. P/G Circuit Model

Our work is applicable in a typical design flow, after the floor-planning or placement stage, where some information about the switching patterns of various blocks is available.

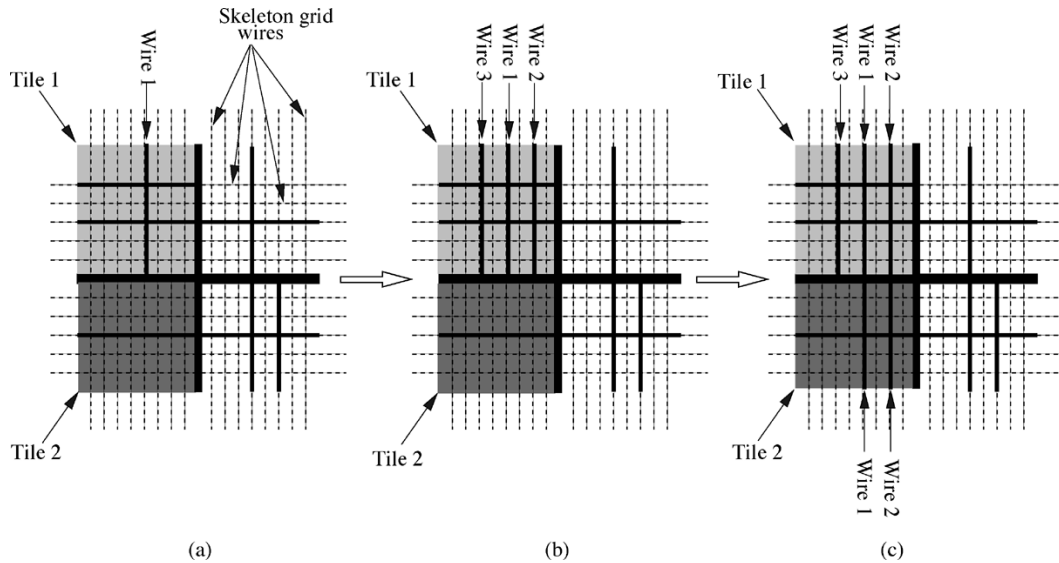


Fig. 3. Illustration of the procedure to build the power grid on a skeleton grid. The P/G wires must lie on the skeleton grid, shown with light lines. The positions of actual P/G wires are shown with dark lines. (a) The initial structure of the grid. (b) The structure of the grid after two wires are added in tile 1. The wires are added to maintain a near-constant pitch within the tile. (c) The grid structure after addition of two wires in tile 2. The wires in tiles 1 and 2 are added at the same local positions so that they are aligned with each other.

Fig. 2(a) shows a layout with a set of placed functional blocks, through letters A–E. The dashed lines show the super-imposition of a skeleton grid, described in the next section, over the chip area.

Fig. 2(b) describes the circuit model used for the P/G net. The entire layout is divided into a regular grid, with power grid wires, shown as the darkest wires in Fig. 2(b), at the boundaries between each grid region. Each subgrid referred to as a *tile* has  $m_i$  horizontal and  $n_i$  vertical wires, shown with thin dark lines.

It is possible to determine the maximum current drawn by a VLSI circuit by using the current estimation techniques such as [19]. Given the worst case currents drawn by P/G wires in rectangular tiles, the power grid is modeled as a resistive mesh, with constant current sources to ground placed at each node inside the tiles as shown in Fig. 2(c). The value of each current source within a tile is identical and is equal to the total current that flows through the P/G buses in the tile, divided by the number of nodes inside a tile.

Clean  $V_{dd}$  and ground pads are assumed to be distributed either on the periphery of the chip [two such pads are shown in Fig. 2(b)] or uniformly distributed over the chip area.

### B. Building the Power Grid

Our optimization procedure builds a nonuniform grid, with different densities of wires in different tiles. We use the concept of a *skeleton grid*, illustrated in Fig. 3(a), to ensure that the wires in the adjacent tiles are not misaligned. The skeleton grid is an imaginary uniform grid superimposed over the layout area, and the wires of the nonuniform real grid are placed on this grid. The pitches of the skeleton grid are chosen such that if the real grid were to completely occupy the skeleton grid, it would have enough wires to meet the voltage drop constraints.

Fig. 3 depicts how the actual nonuniform grid built on a uniform skeleton grid. In this example, the layout is divided into four tiles. The thick shaded lines are the power grid wires that

demarcate the tile boundaries. The initial structure of a nonuniform power grid, built on a skeleton grid, is shown in Fig. 3(a). One way to achieve perfect local regularity in the grid structure is to use a constant wire pitch inside a tile. However, if the wire densities of adjacent tiles are different, a constant pitch would result in the nonalignment of wires across tile boundaries. Hence, we place the wires inside a tile to achieve a near-constant pitch. As shown in Fig. 3(b), the wires additions inside a tile are distributed to maintain a near-constant pitch, as per the idea of local regularity.

Our method maximizes wire alignment across tile boundaries by adding the wires on the skeleton grid in the same predetermined order in all tiles in each iteration. For example, in Fig. 3(b) and (c), the wires are added in the tiles in the following order: wire 1 first, and then wire 2 and wire 3. Since, the wires in tiles are always added in the same order at identical local positions inside a tile, the wires in adjacent tiles are aligned with each other. As seen in Fig. 3(c), wires 1 and 2 in adjacent tiles, tile 1 and tile 2 are aligned with each other. Such a structure aids the routing of signal nets as the only book-keeping that is required is related to the presence or absence of wires on the skeleton grid in the given tile.

### C. Circuit Analysis by Macromodeling Approach

1) *Hierarchical Method of [16]*: For the purposes of determining the most critical nodes in the circuit, our work uses the hierarchical circuit analysis technique of [16]. This section discusses the adaptation of the macromodeling method to this work.

Referring to our model of the power grid in Fig. 2, the hierarchical modeling idea can be efficiently applied to our system since we have partitioned the entire power network into a set of rectangular tiles. As we are dealing with a resistive mesh model and dc excitations, each rectangular tile can be abstracted as a multiport electrical element which has a linear current-voltage

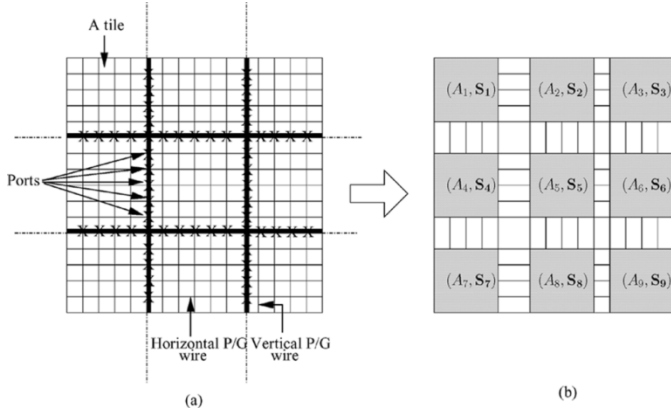


Fig. 4. Converting the P/G network to a system of macromodels. (a) P/G network with port nodes at the tile boundaries. (b) P/G network changed to a system of macromodels connected to each other through the port nodes.

relationship. These multiport elements are referred to as *macromodels*. If  $m$  is the number of wires in the horizontal direction in a tile and  $n$  is the number of vertical wires in a tile then each macromodel is a  $q$ -port linear element, where  $q = 2(m + n)$ , with the transfer characteristic given by the following equation:

$$\mathbf{I} = \mathbf{A}\mathbf{V} + \mathbf{S} \quad (1)$$

where  $\mathbf{I} \in \mathbf{R}^q$ ,  $\mathbf{A} \in \mathbf{R}^{q \times q}$ ,  $\mathbf{V} \in \mathbf{R}^q$ ,  $\mathbf{S} \in \mathbf{R}^q$ ,  $\mathbf{A}$  is the port admittance matrix,  $\mathbf{V}$  is the vector of voltages at the ports corresponding to the voltages at the nodes on edges of the tiles,  $\mathbf{I}$  is the current through the interface between the tiles, and  $\mathbf{S}$  is a vector of current sources between each port and ground.

We omit the mathematics involved in computing the macromodel parameters  $(\mathbf{A}, \mathbf{S})$  as it is explained in detail in [16]. The macromodel parameters  $(\mathbf{A}, \mathbf{S})$  of each tile are stamped into the modified nodal analysis (MNA) equation in the global system given by

$$\mathbf{M}\mathbf{X} = \mathbf{b} \quad (2)$$

where  $\mathbf{M}$  is the matrix containing the conductance links between global nodes and the tiles, the conductance links between the tiles and the stamps of  $\mathbf{A}$  for each tile;  $\mathbf{X}$  is the vector of voltages of global nodes and ports; and  $\mathbf{b}$  is the vector of current sources at global nodes and stamps of  $\mathbf{S}$  for each tile. The global system corresponding to the example of Fig. 3, where the chip is partitioned into four tiles, is described by

$$\mathbf{M} = \begin{bmatrix} G_{00} & G_{01} & G_{02} & G_{03} & G_{04} \\ G_{01}^T & A_1 & G_{12} & G_{13} & G_{14} \\ G_{02}^T & G_{12}^T & A_2 & G_{23} & G_{24} \\ G_{03}^T & G_{13}^T & G_{23}^T & A_3 & G_{34} \\ G_{04}^T & G_{14}^T & G_{24}^T & G_{34}^T & A_4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \mathbf{I}_0 \\ -\mathbf{S}_1 \\ -\mathbf{S}_2 \\ -\mathbf{S}_3 \\ -\mathbf{S}_4 \end{bmatrix}.$$

The entries corresponding to the index 0 are associated with the global nodes and the other indexes correspond to the tile numbers. Fig. 4 shows the conversion of tiles of the power grid into macromodels. In this example, the power grid which is divided into nine tiles, is reduced to a system of nine multiport elements given by the macromodel parameters  $(\mathbf{A}, \mathbf{S})$ . The macromodels are connected to each other through port connections.

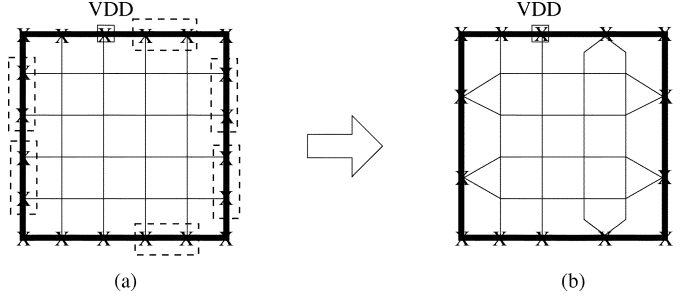


Fig. 5. Reducing the number of ports of a tile. (a) Tile of the original P/G network with 20 port nodes. (b) The port nodes of the tile reduced to 14 by combining some nodes.

For the purposes of efficiency, we also implement the step of sparsification of  $\mathbf{A}$  matrix as proposed in [16]. This increases the sparsity of the global matrix  $\mathbf{M}$  at the cost of reasonable simulation errors.

2) *Port Approximation: Approach and Validation:* As mentioned in Section III-A, the nodes on the edges of the tiles are on the same conducting wire, and so it is reasonable to make an approximation that collapses some of the nodes<sup>1</sup> at the edge.<sup>2</sup> We use this observation to reduce the size of the macromodels, making an approximation to reduce the number of ports of a tile from  $2(m_i + n_i)$  to  $2(k + p)$  ports, where  $k \leq m_i$  and  $p \leq n_i$ . Each of the rectangular tiles is considered as a  $2(k + p)$ -port linear element by making an approximation that the voltage variation of some nodes on the edges of the rectangular tiles is small. To further maintain the accuracy of the circuit solution, during the reduction of the edge nodes of the tiles, any  $V_{dd}$  and ground pad nodes that lie on the tile edges are always preserved and so are their immediate neighbors. Also, for every removed node, its immediate neighbor is always preserved, thus approximating for small voltage variations on the edges, but not for large ones. The corner nodes of the tiles are never removed. Fig. 5 illustrates this process where a 20 port tile is reduced to a 14 port tile. To validate that this reduction from  $2(m + n)$  ports to  $2(k + p)$  ports is a reasonable approximation, we perform simulations of power grid circuits and empirically choose the value of  $k$  and  $p$  so that there is a reasonable upper bound on the error in the solution of the original and the approximated systems. Table I shows the orders of errors and run time improvements for the approximations for the runs for power delivery to a  $2 \times 2$ -cm chip with a  $V_{dd}$  value of 1.2 V and pads distributed throughout the chip. The chip is divided into  $10 \times 10$  subgrids, i.e., 100 tiles with each tile having ten horizontal and ten vertical wires. The exact number of ports without any port approximations is 44 and this corresponds to the simulation data on the last row. The range of voltages for the exact simulation without any port approximations is 0.88–1.20 V.

The table shows the orders of the average and maximum errors for simulations with the port approximations as compared to a simulation without removing any ports. As seen in Table I, we gain significant runtime improvement from these approximations while ensuring that the errors remain within reasonable

<sup>1</sup>This is a coarser and faster approximation than that used in multigrid-based methods.

<sup>2</sup>Even if the grid boundaries do not have these wires, the assumption is likely to be valid for reasonably dense grids.

TABLE I  
ERRORS FOR PORT APPROXIMATIONS FOR A  $10 \times 10$  GRID

# of Ports Kept	Avg Error %	Max Err %	Runtime (sec)
4	9.21%	13.24%	0.54
8	7.52%	10.02%	0.67
10	5.89%	9.50%	0.78
12	5.38%	8.08%	0.87
20	3.19%	6.58%	1.88
24	2.60%	5.08%	2.79
28	2.41%	5.06%	3.95
32	1.57%	3.51%	5.46
36	1.27%	2.71%	7.19
40	0.07%	1.02%	9.49
44	0.00%	0.00%	11.60

bounds. For a power grid design problem in which the bound on the worst case voltage drop is to be kept within 8%–10% of  $V_{dd}$ , the level of accuracy given by an average error of about 1%–3% is adequate. Since we are at the early design level, there is a good amount of uncertainty involved in various design parameters like switching current waveforms and exact placement of the underlying functional blocks. Hence, it makes more sense to work with an efficient and reasonably accurate model of the power grid as opposed to a completely accurate but inefficient model.

#### D. Voltage Sensitivity Calculation

For our problem, the sensitivity calculations in matrix form provide the gradient information to guide the direction of optimization. The analysis in Section III-C determines the most critical node that has the greatest voltage drop. The task of sensitivity computation is to determine the identity of the tile to which the addition of  $l$  horizontal or vertical wires would have the greatest impact, in terms of improving the worst case voltage drop. To explain the sensitivity calculation method, the following notations will be used.

$k$	Number of rectangular tiles into which the power grid is divided.
$V_{xi}$	Voltage of the most critical node $x$ which is found in tile $i$ .
$p_i$	Number of wires in tile $i$ .
$m_i$	Number of wires in tile $i$ in the horizontal direction.
$n_i$	Number of wires in tile $i$ in the vertical direction.
$W$	The wire width for the power grid, which is assumed to be constant for the entire layout.
$L_{hi}$	The length of the horizontal wire in tile $i$ .
$L_{vi}$	The length of the vertical wire in tile $i$ .
$Ar_i$	Wiring area used in tile $i$ .
$V_i$	The vector of port voltages of tile $i$ which contains the most critical node $x$ .
$t$	Numbers of ports kept for tiles after the port approximations.
$l$	The number of wires added in any tile in each iteration.

The following relations exist between the above defined terms:

$$Ar_i = W(m_i L_{hi} + n_i L_{vi}) \quad (3)$$

$$p_i = m_i + n_i. \quad (4)$$

Using the chain rule we can make the following calculations:

$$\frac{\delta V_{xi}}{\delta Ar_i} = \frac{\delta V_{xi}}{\delta p_i} \cdot \frac{\delta Ar_i}{\delta p_i}. \quad (5)$$

From (3) and (4)

$$\frac{\delta Ar_i}{\delta p_i} = \frac{\delta Ar_i}{\delta m_i} + \frac{\delta Ar_i}{\delta n_i} \quad (6)$$

$$\frac{\delta Ar_i}{\delta p_i} = W(L_{hi} + L_{vi}). \quad (7)$$

To calculate the numerator of the right-hand side (RHS) of (5), i.e.,  $\delta V_{xi}/\delta p_i$ , we use the following procedure. Consider the global system,  $\mathbf{MX} = \mathbf{b}$  as described by (2). Let us assume that by the process of adding wires in tile  $i$ ,  $M$  changes to  $M + \delta M$  and  $\mathbf{b}$  changes to  $\mathbf{b} + \delta \mathbf{b}$ . It can be easily seen from (3) that the changes,  $\delta M$  in the global matrix  $M$ , are in the entries corresponding to the stamp of the first macromodel parameter  $A_i$ , and the changes,  $\delta \mathbf{b}$  in the vector  $\mathbf{b}$ , take place in the entries corresponding to stamps of second macromodel parameter  $\mathbf{S}_i$ . The sensitivities of particular response variables, i.e., the port voltages entries in the  $\mathbf{X}$  vector in (2), can be calculated by the following equations [12]:

$$(M + \delta M)(\mathbf{X} + \delta \mathbf{X}) = \mathbf{b} + \delta \mathbf{b}. \quad (8)$$

Simplifying (8) by substituting from (2) and neglecting the second-order variation, i.e.,  $\delta M \delta \mathbf{X}$ , we obtain

$$\begin{aligned} M \delta \mathbf{X} &= -\delta M \mathbf{X} + \delta \mathbf{b} \\ \delta \mathbf{X} &= M^{-1}(-\delta M \mathbf{X} + \delta \mathbf{b}). \end{aligned} \quad (9)$$

If we are concerned only with the  $j$ th response variable, i.e., the voltage sensitivity of  $j$ th port variable then (9) can be written as

$$\delta \mathbf{X}_j = [j\text{th row of } M^{-1}](-\delta M \mathbf{X} + \delta \mathbf{b}). \quad (10)$$

The  $j$ th row of  $M^{-1}$  can be calculated by solving the system

$$M^T \xi_j = e_j \quad (11)$$

where  $\xi_j$  is a column vector representing the  $j$ th row of  $M^{-1}$  and  $e_j$  is a column vector corresponding to the  $j$ th column of the identity matrix. Equation (10) can be rewritten as

$$\delta \mathbf{X}_j = \xi_j^T (-\delta M \mathbf{X} + \delta \mathbf{b}). \quad (12)$$

Referring to (12) the following relations can be found:

$$\frac{\delta \mathbf{X}_j}{\delta \mathbf{b}_k} = \xi_{jk} \quad (13)$$

where  $\xi_{jk}$  is the  $k$ th component of the  $\xi_j$ , and

$$\frac{\delta \mathbf{X}_j}{\delta M_{mn}} = -\xi_{jm} \mathbf{X}_n \quad (14)$$

where  $\xi_{jm}$  is the negative of  $m$ th component of column vector  $\xi_j$  multiplied by  $n$ th component of the original solution vector  $\mathbf{X}$ . Using the (13) and (14) and applying the chain rule, we can find out the sensitivities of voltages at the ports of the tile which

TABLE II  
TABLE LOOKUP EXAMPLE

Index	# of Horizontal Wires	# of Vertical Wires	$A_{t \times t}$	$S_{t \times 1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
q	3	3	$A_1$	$S_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
p	3	7	$\hat{A}_1$	$\hat{S}_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
100	10	10	$A_{100}$	$S_{100}$

contains the most critical node  $x$ , with respect to an addition of  $l$  wires in tile  $i$ , by the following equation:

$$\frac{\delta \mathbf{X}_j}{\delta p_i} = \sum_{m=1}^t \sum_{n=1}^t \frac{\delta \mathbf{X}_j}{\delta M_{mn}} \frac{\delta M_{mn}}{\delta p_i} + \sum_{k=1}^t \frac{\delta \mathbf{X}_j}{\delta \mathbf{b}_k} \frac{\delta \mathbf{b}_k}{\delta p_i} \quad (15)$$

where  $t$  is the number of kept ports for the tiles after the port approximations and the summation indexes in the terms  $\sum_{m=1}^t \sum_{n=1}^t$  and  $\sum_{k=1}^t$  arise due to the change in the macromodel  $(A_{t \times t}, S_{t \times 1})$  stamps for tile  $i$  where the extra wires are added. For the purposes of efficiency, the two unknown quantities in (15)  $(\delta M_{mn}/\delta p_i)$  and  $(\delta \mathbf{b}_k/\delta p_i)$  are calculated using a table lookup scheme described in the following section.

**Table Lookup:** The number of wires in a tile in horizontal and vertical directions can assume a finite set of values starting from the initial number of wires, to a maximum number that corresponds to the number of wires on the skeleton grid inside a tile.

The macromodel matrix  $A$  depends only on the number of wires in a tile and the vector  $\mathbf{S}$  depends on the number of wires and the value of current sources in a tile. However, since our model assumes equal valued current sources placed at the internal nodes of the tile as described in Section III-A, the  $\mathbf{S}$  vector can be calculated for a current source of unit value at the internal nodes, and then a new  $\tilde{\mathbf{S}}$  vector can be computed as a scalar multiple of the  $\mathbf{S}$  vector by the following relation:

$$\tilde{\mathbf{S}} = c\mathbf{S} \quad (16)$$

where  $\tilde{\mathbf{S}}$  is the vector for the tile with a current source of magnitude  $c$  placed at the internal nodes and  $\mathbf{S}$  is the vector for the tile with a current source of unit magnitude placed at the internal nodes.

Given that macromodel parameters  $(A, \mathbf{S})$  depend on only the number of wires in a tile, we construct, in advance, a table that contains the corresponding macromodel parameters for a set of values of horizontal and vertical wires in a tile.

The structure of the table is shown in Table II. We can now make the following computations from the table lookup:

$$\frac{\delta M_{mn}}{\delta p_i} \approx \frac{\Delta M_{mn}}{\Delta p_i} = (A_{uv})_p - (A_{uv})_q \quad (17)$$

$$\frac{\delta \mathbf{b}_k}{\delta p_i} \approx \frac{\Delta \mathbf{b}_k}{\Delta p_i} = c((\mathbf{S}_r)_p - (\mathbf{S}_r)_q) \quad (18)$$

where  $(u, v)$  are the rows and columns of  $A_{t \times t}$  and  $(m, n)$  are the rows and columns of  $M$  corresponding to the stamp of  $A$ . Similarly,  $r$  is the row of  $S_{t \times 1}$  and  $k$  is the row in  $\mathbf{b}$  corresponding to the stamp of  $\mathbf{S}$ . The index  $q$  is the index in the table

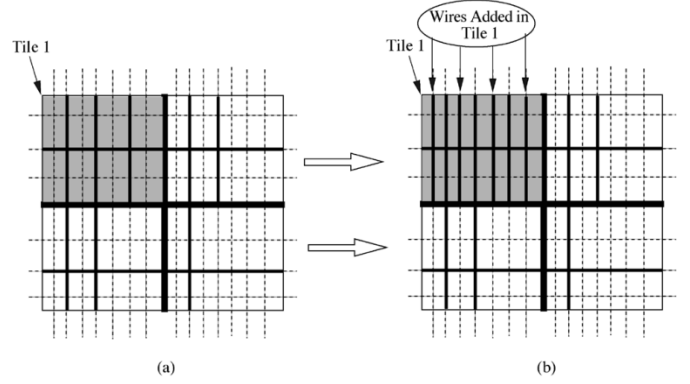


Fig. 6. Change in the grid structure by addition of wires in tile 1. (a) Tile 1 with three wires initially. (b) Four more wires added in tile 1. The corresponding change in macromodel parameters  $(A_1, S_1)$  can be calculated using Table II as a lookup table.

corresponding to the current numbers of horizontal and vertical wires in a tile and  $p$  is the index in the table corresponding to the number of horizontal and vertical wires after an addition of  $l$  wires in either of the directions. These equations hold because any change in the  $M$  matrix and the  $\mathbf{b}$  vector due to an addition of a wire in the tile would only be in the stamps of  $A$  and  $\mathbf{S}$ . The table lookup procedure can be better understood with the help of a small example. Let us consider a chip area divided into four tiles. We wish to compute the terms in the left-hand side (LHS) of (17) and (18) with respect to addition of wires in tile 1. Fig. 6 illustrates the change in the grid structure by wire additions in tile 1.

As seen in (3), after the addition of two wires in tile 1, the only changes in  $M$  are in the entries corresponding to  $A_1$  and the only changes in  $\mathbf{b}$  are in the entries corresponding to  $S_1$ . After the wire additions, the new macromodel parameters are  $(\hat{A}_1, \hat{S}_1)$ . Since we already have a preconstructed table of the form of as shown in Table II, all we need to do is search in the table for indexes  $q$  corresponding to the number of wires in tile 1 before the wire addition, and  $p$  corresponding to the number of wires in tile 1 after the wire addition. The entries stored in the table for index  $q$  are  $(A_1, S_1)$  and the entries stored in the table for index  $p$  are  $(\hat{A}_1, \hat{S}_1)$ . Hence, we can now easily evaluate (17) and (18).

All the terms in the RHS of (15) are now known from the solutions of (13), (14), (17), and (18). The evaluation of (15) yields the sensitivities of the port voltages with respect to the wire addition in the tile.

We now need to calculate the sensitivity of the most critical [i.e., LHS of (5)] node, which could be an internal node of a tile. Since our model of the power grid as shown in Fig. (2) is purely linear, we can relate the port voltages to the most critical node by the following equation:

$$\mathbf{V}_{xi} = \mathbf{C}^T \mathbf{V} + D_x \quad (19)$$

where  $\mathbf{V}$  is the vector of port voltages,  $\mathbf{C}^T$  is a row vector, and  $D_x$  is a constant. Referring to [16], these terms are calculated as

$$\begin{aligned} \mathbf{C}^T &= x\text{th row (for the internal node } x) \text{ of matrix } -G_{11}^{-1}G_{12}. \\ D_x &= x\text{th component of vector } G_{11}^{-1}\mathbf{J}_1. \end{aligned}$$

If the most critical node is not in the same tile in which a wire is being added then the terms  $\mathbf{C}^T$  and  $D_x$  are constant. So, from (19)

$$\frac{\delta \mathbf{V}_{xi}}{\delta p_j} = \mathbf{C}^T \frac{\delta \mathbf{V}}{\delta p_j}. \quad (20)$$

If the most critical node is indeed in the same tile in which a wire is being added

$$\frac{\delta \mathbf{V}_{xi}}{\delta p_i} = \mathbf{C}^T \frac{\delta \mathbf{V}}{\delta p_i} + \mathbf{V} \frac{\delta \mathbf{C}^T}{\delta p_i} + \frac{\delta D_x}{\delta p_i}. \quad (21)$$

Thus, from (20) and (21), we get the LHS of (5), which is simply the change in voltage of the most critical node by making an addition of  $l$  wires in tile  $i$ . This is the gradient information we require to guide the optimization process described in the next section. The table-lookup scheme is implemented as a file read and thus does not increase the memory requirements of the optimization procedure.

We empirically choose to keep the values of  $l$  to be from 5–7 wires added in a tile in every iteration. For larger values of  $l$ , the orders of errors in the sensitivity computations increase, as the assumptions made to neglect the second order variations, i.e.,  $\delta M \delta \mathbf{b}$  in the derivation of (13) and (14) no longer hold.

### E. Congestion-Aware Design

The constraint of maintaining local regularity may come at the expense of worsening the congestion problem where signal nets are unable to find sufficient routable resources to complete their routing and are forced to take detours. If structured regularity enforces placement of power grid wires in regions where the demand for routing resources from signal nets is high, it would clearly aggravate congestion.

To overcome this potential problem of congestion arising due to local regularity in the grid structure, we follow a preemptive strategy without compromising the property of structured local regularity in the power grid design. We introduce a term in the cost function which penalizes addition of power grid wires in high congestion regions. Since the input to our power grid design problem is a floorplan or a placed net-list, probabilistic congestion estimation techniques such as [17] and [18] can be used to assign congestion numbers to different regions of the chip. Dividing the chip into rectangular tiles for the purposes of building the power grid, produces tiles with fairly large areas from the perspective of signal net routing. Hence, we further tessellate each tile into smaller rectangular regions called *bins*. As shown in Fig. 7(a) and (b), after the chip area is divided into rectangular subgrids or tiles, the tiles are further divided into smaller rectangular bins. Considering the bins in tile  $i$  arranged in  $y_i$  rows and  $z_i$  columns, probabilistic congestion estimation techniques can be used to generate two matrices  $U_{i_h}$  and  $U_{i_v}$ , each of size  $y_i \times z_i$ , whose entries correspond, respectively, to the associated horizontal and vertical usage of each bin due to the signal net routing. The horizontal and vertical capacities,  $c_{i_h}$  and  $c_{i_v}$ , respectively, of each bin in tile  $i$  is given by the following relations:

$$c_{i_h} = \frac{h_i}{\min P_h} \quad (22)$$

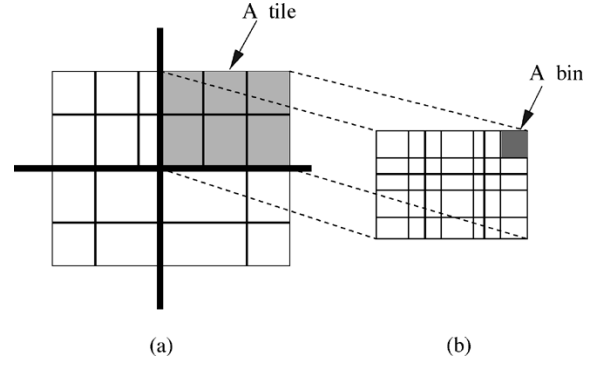


Fig. 7. (a) Example of a chip divided into four tiles demarcated with thick dark lines representing P/G wires. The P/G wires inside a tile are shown with thin dark lines. (b) Tile is further divided into 16 bins. The dark lines are P/G wires inside the tile.

$$c_{i_v} = \frac{w_i}{\min P_v} \quad (23)$$

where  $(\min P_h)$  and  $(\min P_v)$  are the minimum horizontal and vertical pitches, respectively, of the metal layer, and,  $h_i$  and  $w_i$  are the height and the width, respectively, of each bin in tile  $i$ . The horizontal and vertical congestions of each bin are calculated and stored in matrices  $C_{i_h}$  and  $C_{i_v}$ , respectively. For a bin indexed by row  $a$  and column  $b$ , the congestion value is given by the fraction of the total bin capacity utilized by the following relation:

$$C_{i_h}[a, b] = \frac{U_{i_h}[a, b]}{c_{i_h}} \quad (24)$$

$$C_{i_v}[a, b] = \frac{U_{i_v}[a, b]}{c_{i_v}}. \quad (25)$$

In each iteration of the optimization loop, addition of  $l$  wires in tile  $i$  change the values in congestion and usage matrices for exactly  $l$  columns or  $l$  rows, depending on whether the wires added were in the vertical or the horizontal direction. For instance, addition of one vertical wire in tile  $i$ , at a position corresponding to all bins indexed by column  $b$ , changes the values of the usage matrix  $U_{i_v}$  in the following way:

$$U_{i_v}[j, b]_{\text{new}} = U_{i_v}[j, b]_{\text{old}} + \frac{W}{\min P_v}, \quad j = 1 \cdots y_i \quad (26)$$

where  $W$  is the wire width which is assumed constant for the entire layout. Using (24) and (25), the new congestion values, after wire additions, can be calculated.

We define  $\text{CongV}_{\text{avg}_i}$  as the average vertical and  $\text{CongV}_{\text{max}_i}$  as the maximum vertical congestion of all bins indexed by column  $b$  as

$$\begin{aligned} \text{CongV}_{\text{avg}_i}[b] &= \sum_{j=1}^{y_i} \frac{C_{i_v}[j, b]}{y_i} \\ \text{CongV}_{\text{max}_i}[b] &= \max C_{i_v}[j, b], \quad j = 1 \cdots y_i. \end{aligned} \quad (27)$$

The average horizontal,  $\text{CongH}_{\text{avg}_i}$  and the maximum horizontal  $\text{CongH}_{\text{max}_i}$  congestion of all bins indexed by row  $a$  is similarly defined as

$$\begin{aligned} \text{CongH}_{\text{avg}_i}[a] &= \sum_{j=1}^{z_i} \frac{C_{i_h}[a, j]}{z_i} \\ \text{CongH}_{\text{max}_i}[a] &= \max C_{i_h}[a, j], \quad j = 1 \cdots z_i. \end{aligned} \quad (28)$$



The average vertical and horizontal congestion for all of the bins in tile  $i$  is

$$\begin{aligned} \text{CongV}_{\text{avg}_i} &= \sum_{j=1}^{z_i} \frac{\text{CongV}_{\text{avg}_i}[j]}{z_i} \\ \text{CongH}_{\text{avg}_i} &= \sum_{j=1}^{y_i} \frac{\text{CongH}_{\text{avg}_i}[j]}{y_i} \end{aligned} \quad (29)$$

The maximum vertical and horizontal congestion for all the bins in tile  $i$  is given by

$$\begin{aligned} \text{CongV}_{\text{max}_i} &= \max(\text{CongV}_{\text{max}_i}[1], \dots, \text{CongV}_{\text{max}_i}[z_i]) \\ \text{CongH}_{\text{max}_i} &= \max(\text{CongH}_{\text{max}_i}[1], \dots, \text{CongH}_{\text{max}_i}[y_i]). \end{aligned} \quad (30)$$

The penalty term introduced for congestion is the following:

$$\begin{aligned} \text{Cong}_i &= \gamma(\text{CongV}_{\text{avg}_i} + \text{CongH}_{\text{avg}_i}) + (1 - \gamma) \\ &\quad \times (\text{CongV}_{\text{max}_i} + \text{CongH}_{\text{max}_i}) \end{aligned} \quad (31)$$

where,  $\gamma$  is a parameter  $\in [0, 1]$  to appropriately penalize the average and the maximum congestion.

The cost function associated with adding wires in tile  $i$  is a weighted sum of the voltage sensitivity term,  $\delta V_x / \delta A r_i$ , computed using the procedure described in Section III-D, and the congestion penalty term,  $\text{Cong}_i$ .

$$\text{Cost}_i = \beta \text{Cong}_i - \alpha \frac{\delta V_x}{\delta A r_i} \quad (32)$$

where  $\alpha$  and  $\beta$  are normalized weight parameters. The voltage sensitivity term represents the cost to benefit ratio of voltage drop reduction with increase in area, and the congestion term penalizes for aggravating congestion by wire additions in the congested regions. The cost function guides the optimization loop as the tile having the minimum value of the cost function is selected for wire additions.

#### F. Optimization Heuristic

We use a greedy optimization heuristic based on the information obtained from the sensitivity and congestion computations. Using some of the definitions in the previous sections, the optimization problem is formulated as follows:

$$\begin{aligned} &\text{Minimize } \sum_{i=1}^k \text{Cost}_i \\ &\text{Subject to} \\ &\quad \text{Min}(\mathbf{V}_x) \geq V_{\text{spec}} \\ &\quad P_{vi} \approx c_1 \lambda_v, \forall i : 1 \text{ to } k \\ &\quad P_{hi} \approx c_2 \lambda_h, \forall i : 1 \text{ to } k \\ &\quad \text{EM constraints:} \\ &\quad \frac{I_{\text{wire}_j}}{W} \leq \sigma \\ &\quad \frac{|V_p - V_q|}{\rho_s P} \leq \sigma \quad \forall p, q \text{ nodes on wire } j \quad \forall j \end{aligned} \quad (33)$$

where  $k$  is the number of tiles in the P/G grid,  $P_{hi}$  and  $P_{vi}$  are the wire pitches in the horizontal and vertical direction in tile  $i$ ,  $\lambda_h$  and  $\lambda_v$  are the pitches in the horizontal and vertical directions of the *skeleton grid*, and  $c_1$  and  $c_2$  are some integer constants. These approximation constraints enforce the wires in a tile to be placed on the skeleton grid with a near-constant pitch within a tile  $i$ . As described in Section III-B, these constraints aid in maximum alignment of wires in the adjacent tiles. The wire pitches are related to the line resistances by the following relations

$$R_{vi} = \frac{\rho_s P_{hi}}{W} \quad \text{and} \quad R_{hi} = \frac{\rho_s P_{vi}}{W} \quad (34)$$

where  $R_{vi}$  and  $R_{hi}$  are the line resistances of the vertical and horizontal wires in tile  $i$ ,  $\rho_s$  is the sheet resistivity, and  $W$  is the wire width which is assumed to be constant for the entire layout.

The last two constraints are the EM constraints, where  $\sigma$  is the current density for fixed thickness of the metal layer and  $I_{\text{wire}_j}$  is the current flowing through wire  $j$ .

#### G. Speedup Techniques

In the optimization procedure, we can generate the following savings in the some of the steps to achieve significant speed up.

- 1) Instead of calculating the voltage sensitivities and congestion values for all the tiles, we can define an *active sensitivity window* around the most critical node so that there are enough pads within the window and then make the sensitivity and congestion calculations for only the tiles inside that window.
- 2) By adding a few wires in only one of the tiles, we need to Cholesky factor the  $G$  matrix for only the changed tile. The factors of other unchanged tiles are reused.
- 3) In any analysis step after the initial one, we do not necessarily have to solve all the tiles to determine the most critical node. After solving for the tile  $j$ , that had the worst drop, all the tiles having greater minimum voltages in the previous iteration than the minimum of voltage of tile  $j$  in the current iteration, need not be solved.

*Extension to Multiple Metal Layers:* The proposed optimization technique can be easily extended to design a power grid for chips having multiple layers of metal. Typically, the upper metal layers use wider wire widths than the lower layers. Also, adding wires in upper layers would yield greater reduction in the IR drop, since the upper layer wires affect the voltages at larger number of sinks. The chip area can be divided into rectangular tiles in each of the layers. In any iteration, to determine the layer and the tile for wire additions that would result in the greatest IR drop reduction, with minimum wire area increase, we need to calculate the voltage sensitivity, i.e.,  $\delta V_{xi} / \delta A r_i$  for a few tiles in each layer. The penalty for congestion aggravation can also be set appropriately for different layers. Since the top metal layers are not much used for the signal net routing, they can be assigned a lesser congestion penalty term. Hence, the power grid is designed by iteratively adding wires in the tile and layer combination which yields the least cost.

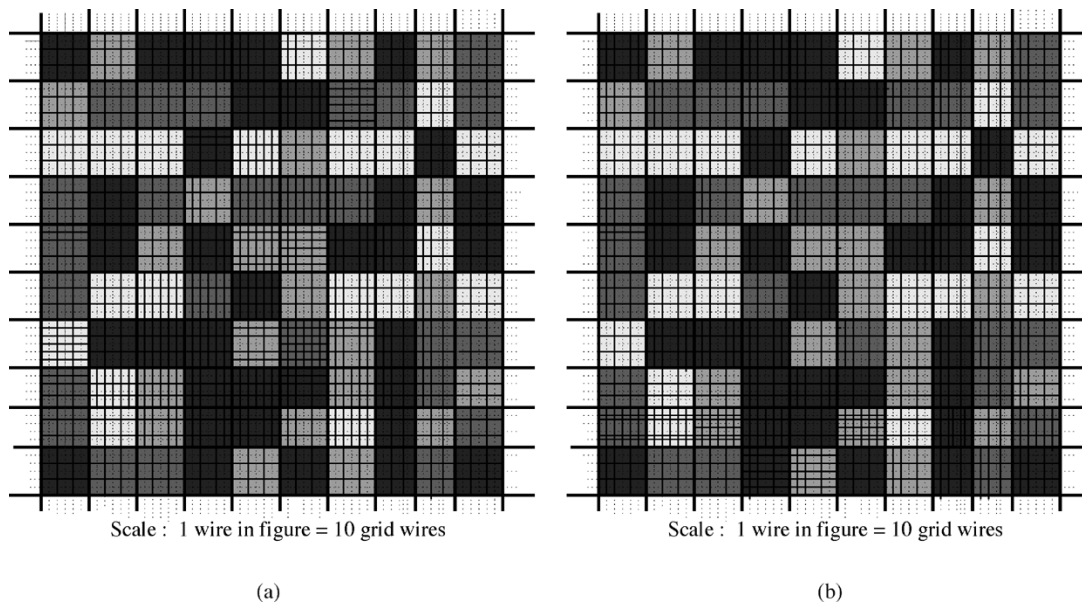


Fig. 8. Wire density pattern of power grids constructed by the proposed optimization for a  $2 \times 2$ -cm chip divided into 100 tiles, superimposed over the current density patterns. The regions with darker shades have higher current densities. (a) Optimized nonuniform power grid for a wire-bonded package. (b) Optimized nonuniform power grid for a flip-chip package.

#### IV. EXPERIMENTAL RESULTS

The proposed optimization scheme was implemented in C using a sparse matrix library [20], and results on several power networks were tested. Due to the unavailability of benchmark circuits for power networks, the circuits were randomly generated but with real circuit parameter values. The circuit parameter values, sheet resistivity ( $\rho_s$ ), wire width ( $W$ ), current density ( $\sigma$ ), minimum wire pitches ( $\min P_h$ ) and ( $\min P_v$ ), and the ranges of worst-case current sources were taken from [21] and [22] for power delivery to a  $2 \times 2$ -cm chip in 130-nm technology with  $V_{dd} = 1.2$  V. The voltage constraints for the power grids, i.e.,  $V_{spec}$  was 1.08 V, i.e., 90% of  $V_{dd}$ . Also, due to the unavailability of large block-level benchmark circuits, we could not use the probabilistic technique to estimate congestion values for different regions of the chip. Instead, we randomly generated congestion numbers to model the initial horizontal and vertical usages of the bins due to the assumed signal net routing. The congestion values were generated such that the signal nets consumed between 30% and 60% of the bin capacities. The bin size was assumed to be  $20 \mu\text{m} \times 20 \mu\text{m}$ . The experiments were performed on Pentium IV processor Linux machines with a clock speed of 2.4 GHz.

Fig. 8 illustrates the nonuniform grids obtained at the end of the optimization heuristic for a  $2 \times 2$ -cm chip divided into 100 tiles. Fig. 8(a) refers to the case when the  $V_{dd}$  pads are distributed at the periphery of the chip, as for a wire-bonded package. Fig. 8(b) shows the grid obtained when the pads are distributed throughout the chip, as for a flip-chip package. The grids are superimposed on the current density patterns, where darker colors represent higher current density regions. The light dashed lines represent the skeleton grid and the dark solid lines represent the actual grid. According to the scale chosen, one solid line is a substitute for ten wires in the actual grids. The highest current density regions, represented by the darkest shade

in the figure, has a range of currents of about five to six times greater than the range of currents in the lowest current density regions, represented by the lightest shade in the figure. As seen from the figure, it is not always true that the densest grids would lie in tiles with the highest current densities. The requirement to maintain sufficient voltage levels, in regions which draw high currents, can be met with fewer wires, provided there are enough pad connections to the P/G wires in that region. Hence, factors such as the pad locations also affect the density of the power grids in a given region.

Two sets of experiments were conducted that intended to compare our approach with the following.

- 1) The wire area utilized by a uniformly structured grid and constant wire width throughout the chip. Table III refers to this comparison.
- 2) The wire area utilized by a uniform grid for which wires are sized differently in different regions of the chip. Table IV refers to this comparison. The number of wires for the uniform grids of Table IV is less than that of the uniform grids of Table III, but the wire widths are greater for wires in some of the high current density tiles of the uniform grids of Table IV.

For the power grids of Tables III and IV, the  $V_{dd}$  pads were distributed throughout the chip. The weights in the cost function of (32),  $\alpha$  and  $\beta$ , were adjusted to give approximately equal importance to the voltage sensitivity and the congestion term. The value of  $\gamma$ , in (31), was set to 0.5 to equally penalize the average and the maximum congestion of all bins in a tile. As listed in Table III, the number of tiles into which the chip was divided for optimization are given in the second column. The third column ( $V_{init_x}$ ) refers to the voltage of the most critical node before the optimization, when starting from an initial grid with equal number of wires in all tiles. Column four ( $V_{opt_x}$ ) indicates the voltage at the end of the optimization process. This voltage is measured using an exact simulation of the power grid

TABLE III  
WIRING AREA COMPARISONS WITH REGULAR GRID AND CONSTANT WIDTH TOPOLOGY

Ckt	# of Tiles	$V_{init_x}$	$V_{opt_x}$	# of Ports	# of Wires	Wire Area	Wire Area	% Reduction in	CPU Time
		(V)	(V)	Per Tile	Per Iteration	Regular Grid ( $cm^2$ )	Proposed Design ( $cm^2$ )	Wire Area	
1	80	0.852	1.081	52	5	0.962	0.739	23.12%	86.2
2	100	0.847	1.083	44	5	0.840	0.699	16.82%	67.4
3	144	0.852	1.087	36	6	0.922	0.764	17.14%	55.6
4	160	0.858	1.083	32	7	0.984	0.861	12.46%	48.3

TABLE IV  
WIRING AREA COMPARISONS WITH THE WIRE SIZING METHOD

Ckt	# of	# of Sized	Wire Area	Wire Area	% Reduction
	Tiles	Tiles	Sizing ( $cm^2$ )	Proposed Design ( $cm^2$ )	in Wire Area
1	80	32	0.843	0.739	14.07%
2	100	38	0.883	0.699	20.84%
3	144	48	0.894	0.764	14.54%
4	160	62	1.096	0.861	21.44%

circuit, i.e., without the approximation of reducing the number of ports and sparsification of  $A$  matrix. The optimization is run for about two to five additional iterations even after meeting the voltage drop constraint, so that the errors introduced by the port approximations and sparsification can be accounted for by a bit of overdesign. Also, if the grid meets the IR drop constraints but the EM constraints are not yet satisfied, the optimization is continued by iteratively adding wires in the tiles where EM constraints are violated, until all wires have a current density less than the specified current density. The number of ports retained for each tile after the process of *Port Approximation* is shown in the fifth column. Column six shows the number of wires that were added in every iteration in the tile having the least cost for wire addition. The seventh column shows the amount of wiring area that would be utilized by a P/G network topology having a constant pitch of wires throughout the chip and constant wire width. By enumeration, a minimum number of wires that satisfy the voltage drop constraint, was chosen to construct this regular grid. The eighth column (Wire Area Proposed Design) shows the amount of wire area used by the proposed design at the end of the optimization. The wire widths used by the proposed optimization is the same as the one used for the regular grid for a fair comparison. As seen from the table, there is a saving of about 12% to 23% in wire area by the proposed optimization scheme over the topology having a regular grid and constant sizing with same number of wires in all tiles.

Table IV shows the comparison of wire area of the proposed P/G network structure with a design that employs wire width sizing, starting from an initial uniform grid with same number of wires in all tiles and uniform wire widths. It should be emphasized that the number of wires in the tiles for the regular grid of Table IV is less than that (about 0.6–0.7  $\times$ ) of the regular grid design that is listed in column seven of Table III and the minimum wire width of the regular grid design of Table IV is the same as that of the proposed design. The third column in Table IV refers to the number of tiles in which the wire widths were incrementally sized. The tiles with high current densities

are identified and the wire widths in those tiles are incrementally sized until the voltage drop constraints are met. We compare the sizing solution against the four optimized power networks of Table III. The fourth column (Wiring Area Sizing) lists the wire area used of various chips by the wire sizing solution. The fifth column (Wire Area Proposed Design) lists the wire area used by the proposed optimization. The last column indicates the percentage change in the wire area. There is a reduction of 14%–21% in wire area by the proposed optimization scheme over the design with the wire width sizing technique.

To verify that the proposed design technique of producing locally regular and globally nonregular power grids is congestion-aware, we perform another series of experiments. As shown in Table V, we design the piecewise uniform power grid proposed in this paper, both with and without the congestion penalty term in the cost function. According to the cost function (32), the values in the table corresponding to the parameters,  $\alpha = 1$  and  $\beta = 0$ , refer to the design of the proposed locally regular and globally regular supply grid with no penalty for congestion. The values in Table V corresponding to the parameters,  $\alpha = 1$  and  $\beta = 1$ , refer to the congestion-aware design of the piecewise uniform supply grid. In this design, the normalized weight parameters,  $\alpha$  and  $\beta$ , are used in the cost function to balance the objectives of 1) reducing the IR drop with minimum increase in area and 2) minimizing congestion aggravation. As given by (31), the value of the parameter  $\gamma$  is varied to relatively penalize the maximum and the average congestion of all bins. The value of  $\gamma < 0.5$  penalizes the maximum congestion more than average congestion. As seen in Table V, the congestion-aware design achieves substantial reduction in the maximum and the average congestion values over a design with no congestion penalty. However, the congestion-aware design utilizes more wiring area than the noncongestion-aware design. This is due to the fact that in a congestion aware design, in every iteration, the wires are placed in suboptimal locations from the point of view of IR drop reduction. Hence, more wires are needed to meet the IR drop constraint. The tradeoff between congestion reduction and wire area increase can be explored by changing the relative weights,  $\alpha$  and  $\beta$ , to penalize one objective more than the other in the cost function of (32).

The runtime of the proposed design of the power grid is a function of the number of wires added in each iteration step and the number of ports that are removed to make the macromodels. To better understand the runtime, accuracy, and area-reduction tradeoffs, we perform a series of experiments in which we construct a grid using the proposed optimization scheme for power delivery to  $2 \times 2$ -cm chip divided into 100 tiles. The runtime,

TABLE V  
COMPARISONS OF NONCONGESTION-AWARE AND CONGESTION-AWARE, PIECEWISE UNIFORM POWER GRID DESIGN

Ckt	# of Tiles	$\gamma = 0.25$			$\gamma = 0.5$			$\gamma = 0.75$					
		$\alpha = 1, \beta = 0$			$\alpha = 1, \beta = 1$			$\alpha = 1, \beta = 1$					
		Area ( $cm^2$ )	Congestion		Area ( $cm^2$ )	Congestion		Area ( $cm^2$ )	Congestion				
			Avg	Max		Avg	Max		Avg	Max			
1	120	0.753	0.615	2.039	0.788	0.565	1.312	0.795	0.545	1.395	0.784	0.512	1.476
2	150	0.712	0.593	1.895	0.734	0.572	1.214	0.742	0.561	1.314	0.734	0.534	1.371
3	180	0.789	0.658	1.952	0.809	0.591	1.118	0.816	0.572	1.232	0.812	0.567	1.291

TABLE VI  
VARIOUS TRADEOFFS IN POWER GRID DESIGN

Design	# of Ports	# Wires per Iteration	% Reduction in Wire Area	Avg Err in Opt Design	CPU time (mins)
1	44	5	19.41%	2.14%	63.4
2	44	6	16.23%	2.32%	60.6
3	44	7	13.34%	2.64%	58.9
4	40	5	19.09%	3.67%	56.2
5	40	6	16.50%	3.95%	54.5
6	40	7	14.41%	4.23%	53.4
7	36	5	19.52%	5.81%	41.6
8	36	6	17.12%	6.04%	40.4
9	36	7	13.21%	6.23%	39.2
10	28	5	18.98%	9.02%	19.7
11	28	6	16.47%	9.55%	18.6
12	28	7	13.59%	10.17%	17.1

accuracy and wire area reduction tradeoff is explored by varying the number of ports kept for the tiles and the number of wires added per iteration.

As shown by Table VI, on the one hand, the runtime significantly reduces by removing more ports while forming the macromodels, but on the other hand, the errors in measurement of voltages increase. If we add more wires per iteration for the same number of kept ports, the runtime slightly improves but at the cost of accuracy and savings in the wire area. By adding more wires per iteration, we are increasing the amount of overdesign and hence reducing the improvement in area reduction over other topologies. Thus, according to the level of accuracy and the order of runtime required, we can choose the appropriate parameters to guide the optimization.

## V. CONCLUSION

In this paper, we have proposed a new optimization scheme for the design of structured P/G networks that are locally regular and global irregular. Our approach maximizes alignment between wires in each tile using the idea of the skeleton grid. Experimental results on randomly generated P/G networks of practical chip size and circuit parameters show significant reduction in the wire area used by the proposed scheme when compared to the area consumed by the wire sizing solution and the uniform grid topologies. This design also aids in an easier routing scheme for the signal nets later in the design as minimal bookkeeping needs to be done for the proposed P/G architecture. Including a congestion penalty term in the cost function helps in controlling the aggravation in congestion without compromising the local structured regularity of the supply grid.

Including a simulator in the optimization loop ensures the accuracy of the optimized solution but our results show that the runtimes are reasonable.

## REFERENCES

- [1] X. D. S. Tan, C. J. R. Shi, and J. C. Lee, "Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programmings," in *IEEE Trans. Computer-Aided Design Integr. Circuits Syst. Conf.*, vol. 22, Dec. 2003, pp. 156–161.
- [2] T. Wang and C. C. Chen, "Optimization of the power/ground network wire-sizing and spacing based on sequential network simplex algorithm," in *Proc. IEEE Int. Symp. Quality Electron. Design*, 2002, pp. 157–162.
- [3] X. D. S. Tan and C. J. R. Shi, "Fast power/ground network optimization based on equivalent circuit modeling," in *Proc. ACM/IEEE Design Automation Conf.*, 2001, pp. 550–554.
- [4] R. Dutta and M. Marck-Sadowska, "Automatic sizing of Power Ground (P/G) networks in VLSI," in *Proc. ACM/IEEE Design Automation Conf.*, 1989, pp. 783–786.
- [5] X. Wu, X. Hon, Y. Ca, C. K. Cheng, J. Gu, and W. Dai, "Area minimization of power distribution network using efficient nonlinear programming techniques," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, 2001, pp. 153–157.
- [6] S. Boyd, L. Vandenberghe, A. E. Gamal, and S. Yun, "Design of robust global power and ground networks," in *Proc. ACM Int. Symp. Phys. Design*, 2001, pp. 60–65.
- [7] H. Su, K. H. Gala, and S. S. Sapatnekar, "Fast analysis and optimization of power/ground networks," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, 2000, pp. 477–480.
- [8] H. H. Chen and D. D. Ling, "Power supply noise analysis methodology for deep-sub-micron VLSI chip design," in *Proc. ACM/IEEE Design Automation Conf.*, 1997, pp. 638–643.
- [9] T. Mitsuhashi and E. S. Kuh, "Power and ground network topology optimization for cell based VLSIs," in *Proc. ACM/IEEE Design Automation Conf.*, 1992, pp. 524–529.
- [10] H. Cai, "Multi-pads single layer power net routing in VLSI circuit," in *Proc. ACM/IEEE Design Automation Conf.*, 1988, pp. 183–188.
- [11] J. Oh and M. Pedram, "Multi-pad power/ground network design for uniform distribution of ground bounce," in *Proc. ACM/IEEE Design Automation Conf.*, 1998, pp. 287–290.
- [12] L. T. Pillage, R. A. Rohrer, and C. Visweswariah, *Electronic Circuit and System Simulation Methods*. New York: McGraw-Hill, 1995.
- [13] P. Saxena and S. Gupta, "On integrating power and signal routing for shield count minimization in congested regions," *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, vol. 24, pp. 437–445, Apr. 2003.
- [14] H. Su, J. Hu, S. S. Sapatnekar, and S. Nassif, "Congestion-driven code-sign of power and signal networks," in *Proc. ACM/IEEE Design Automation Conf.*, 2002, pp. 64–69.
- [15] J. Singh and S. S. Sapatnekar, "Topology optimization of structured power/ground networks," in *Proc. ACM Int. Symp. Phys. Design*, 2004, pp. 116–123.
- [16] M. Zhao, R. V. Panda, S. S. Sapatnekar, T. Edwards, R. Chaudhry, and D. Blaauw, "Hierarchical analysis of power distribution networks," in *Proc. ACM/IEEE Design Automation Conf.*, 2000, pp. 482–486.
- [17] J. Lou, S. Thakur, S. Krishnamoorthy, and H. S. Sheng, "Estimating routing congestion using probabilistic analysis," *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, vol. 21, no. 1, pp. 32–41, Jan. 2002.
- [18] J. Westra, C. Bartels, and P. Groeneveld, "Probabilistic congestion prediction," in *Proc. ACM Int. Symp. Phys. Design*, 2004, pp. 204–209.

- [19] H. Kriplani, F. Najm, and I. Hajj, "Pattern independent maximum current estimation in power and ground buses of CMOS VLSI circuits algorithms, signal correlations, and their resolution," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, 1995, pp. 998–1012.
- [20] Available: <http://www.netlib.org/c/meschach> [Online]
- [21] International Technology Roadmap for Semiconductors, Semiconductor Industry Association. (2001). <http://public.itrs.net> [Online]
- [22] J. Cong, "An interconnect-centric design flow for nanometer technologies," *Proc. IEEE*, vol. 89, no. 4, pp. 477–480, Apr. 2001.



**Jaskirat Singh** received the B.E. degree in electronics and communication engineering from the Indian Institute of Technology, Roorkee, India, in 2000 and the M.S. degree in electrical engineering in 2003 from the University of Minnesota, Minneapolis, where he is currently pursuing the Ph.D. degree.

His research interests include design and analysis of power/ground networks, circuit simulation, and circuit optimization.



**Sachin S. Sapatnekar** (S'86–M'88–SM'99–F'03) received the B.Tech. degree from the Indian Institute of Technology, Bombay, in 1987, the M.S. degree from Syracuse University, Syracuse, NY, in 1989, and the Ph.D. degree from the University of Illinois, Urbana-Champaign, in 1992.

From 1992 to 1997, he was an Assistant Professor in the Department of Electrical and Computer Engineering at Iowa State University, Ames. He is currently a Professor in the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis. He is the author of four books and a coeditor of one volume, and has published mostly in the areas of timing and layout.

Dr. Sapatnekar received the National Science Foundation Career Award and best paper awards at the ACM/IEEE Design Automation Conference in 1997, 2001, and 2003, and at the IEEE International Conference on Computer Design in 1998. He has been an Associate Editor for the IEEE TRANSACTIONS ON VLSI SYSTEMS, the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN., and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. He has served on the Technical Program Committee for various conferences, and as Technical Program and General Chair for the Tau workshop and the International Symposium on Physical Design.