SUBMITTED TO TCAD, DO NOT DISTRIBUTE !

Placement of Sleep Transistors in Power Supply Network

Changbo Long, Jinjun Xiong and Lei He
{longchb, jinjun, lhe}@ee.ucla.edu
EE department, University of California, Los Angeles, CA, 90095

Abstract

Considering the voltage drop constraint over a distributed model for power/ground (P/G) network, we study the following two problems for physical synthesis of sleep transistors: the min-area sleep transistor insertion (and sizing) (TIS) problem with respect to a fixed P/G network and the simultaneous sleep transistor insertion and P/G network sizing ($TIPGS$) problem to minimize the weighted area of sleep transistors and P/G networks. We show that all sleep transistor placement that disjoin gates from P/G pins lead to a same minimum area in the TIS and $TIPGS$ problems. We develop optimal algorithms to TIS and $TIPGS$ problems by modeling the circuit as a single current source, and then extend these algorithms to model the circuit as distributed current sources. Compared with the best known sizing approach based on sequential linear programming, our algorithms reduce area by up to 44.1% and 61.3% for TIS and $TIPGS$ problems, respectively. To the best of our knowledge, it is the first in-depth study on sleep transistor synthesis considering distributed model for P/G network.

I. Introduction

LEAKAGE current in nanometer devices has increased drastically due to reduction in threshold voltage, channel length and gate oxide thickness [1]. In addition, an increasing number of modules in a highly integrated system are idle at any given time. The high leakage devices and low activity rates both contribute to the growing significance of leakage power at the system level. The Intel Pentium IV processors running at 3GHz already have an almost equal amount of leakage and dynamic power [2]. To effectively reduce leakage power, sleep transistors (see Fig. 1) can be used to turn off supply voltage when modules are idle. However, sleep transistors introduce extra supply voltage drop, increase device delay and reduce device noise margin. Sleep transistors and power supply networks (in short, P/G networks) should be optimized to limit supply voltage drop.

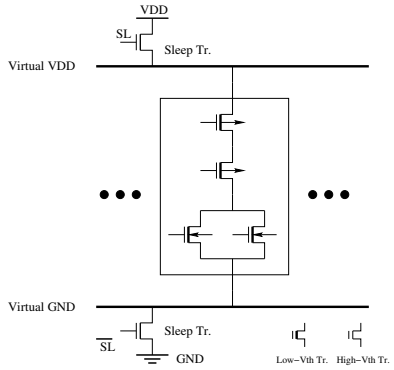

Fig. 1. Illustration for sleep transistors. They are turned off when the circuit is idle.

Most existing work optimizes sleep transistors and P/G networks separately. The following recent papers have studied the synthesis of sleep transistors for bounded supply voltage drop. In [3], the discharging pattern of the switching current is exploited to reduce sleep transistor area. In [4], circuits are divided into clusters and each cluster is connected to a sleep transistor. Bin-packing and set-partitioning are employed to reduce the simultaneous switching current in the clusters and therefore reduce the sleep transistor area. Take advantage of the discharge balancing property of switching current, a mesh of distributed sleep transistors is proposed to reduce the area of sleep transistors without clustering circuits for simultaneous switching current reduction [5]. However, all above work assumes ideal or fixed P/G networks. [6] employs a distributed P/G model and proposes two design styles to layout sleep transistors. They are inserted between each row of the standard cells and P/G network in one style and form an external ring between all gates and external power supply pins in the other. But, there is no automatic method presented in [6].

Analysis and optimization of P/G networks have also been studied without considering sleep transistors. Early work such as [7] applies an integrated package-level and chip-level power bus model to analyze the on-chip power supply noise, i.e., supply voltage drop. Recent work [8] proposes a random walk method for P/G network analysis and [9] incorporates P/G noise

into static timing analysis. The following papers minimize P/G area subject to P/G voltage drop. Assuming no correlation between current sources modeling logic gates, [10] proposes a sequential linear programming formulation. To consider the correlation between current sources, [11] introduces a nonlinear programming formulation, and [12] employs an sensitivity-based optimization based on an event-driven P/G network simulator. To reduce the complexity of P/G optimization, hierarchical approaches are employed in [12], [13]. Additionally, decoupling capacitance has been allocated to reduce P/G noise [14]. Among these work, DC analysis with time-invariant current models is used in [9], [10], AC analysis with time-variant current models is used in [7], [11], [12], [14], and both analysis models are used in [8]. However, none of the above work [3]- [14] considers simultaneous optimization of sleep transistors and P/G networks in an automatic fashion.

In this paper, we study simultaneous optimization of sleep transistors and P/G networks. Specifically, we study two problems: the sleep transistor insertion (and sizing) (TIS) problem with fixed P/G network, and simultaneous sleep transistor insertion and P/G network sizing ($TIPGS$) problem that sizes both sleep transistors and P/G network wires. We show that the optimal sleep transistor insertion solution to above two problems must form a cut-set that partitions the logic gates and power supply pins into two disconnected graphs. Modeling the circuit as a single current source, we show that there exist multiple optimal TIS or $TIPGS$ solutions with a same area, which offers extra design flexibility to consider other constraints such as routing congestion. We then extend this conclusion to the case modeling the circuit as distributed current sources.

The rest parts of the paper are organized as follows. We present modeling and problem formulations in Section II, and solve the TIS and $TIPGS$ problems in Section III and IV, respectively. We present the experiment results in Section V and conclude the paper in Section VI.

II. Modeling and problem formulations

We summarize the notations frequently used in this paper in Table I. We will first discuss current model and P/G model and then formulate the TIS and $TIPGS$ problems.

ρ_p	sheet resistance of P/G network.
ρ_s	effective sheet resistance of sleep transistors.
L_p	length of P/G branches.
L_s	channel length of sleep transistors.
W_p	width of P/G branches.
W_s	channel width of sleep transistors.
r_p	resistance of P/G branches.
r_s	channel resistance of sleep transistors.
TP	tapping points where gates connect to P/G network.
$\overline{V}$	upper bound of supply voltage drop. The default value is 10% VDD.
$\overline{V_p}$	upper bound of voltage drop for P/G network.
$\overline{V_s}$	upper bound of voltage drop for sleep transistors.
C_{TP}	cut-set of P/G branches disconnecting all gates from power supply.
$\overrightarrow{C_{TP}}$	C_{TP} with a uniform current direction.
A_p	area of P/G network.
A_p^*	optimal area of P/G network.
A_s	area of sleep transistors.
A_s^*	optimal area of sleep transistors.
TIS	min-area sleep transistor insertion and sizing problem.
$TIPGS$	simultaneous sleep transistor insertion and P/G network sizing problem.
SSN	single source network.
MSN	multiple source network.

TABLE I
Summary of notations.

A. Switching current model

The switching current of gates is time-variant and varies with respect to the input of the circuit. It has been modeled as time-invariant variable to reduce the complexity in [9], [10], [15], [16]. In this paper, we model the switching current as time-invariant maximum current and will extend to time-variant current model in the future.

B. P/G network model

P/G networks include power networks and ground networks. A power network can be transferred into a ground network by reversing the directions of currents. Therefore, in this paper we only consider the ground network without loss of generality.

The P/G network is modeled as an adjoint multi-port resistive network with one common-terminal, the ground(GND). The resistance of P/G branches is

$$r_p = \rho_p \cdot \frac{L_p}{W_p}, \tag{1}$$

where ρ_p, L_p and W_p are the sheet resistance, length, and width of P/G branches, respectively. We illustrate the modeling of P/G network in Fig. 2. As shown in the figure, gates are modeled as current sources and connect to the P/G network at tapping points (*TP*). P/G branches are modeled as resistors.

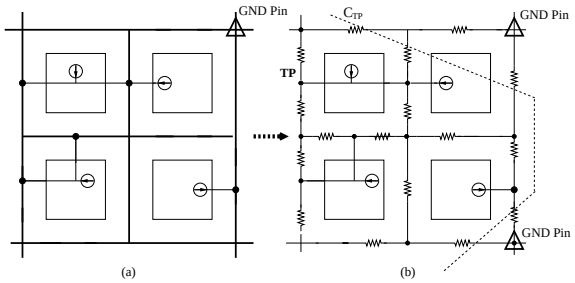

Fig. 2. An example of P/G network modeling.

A resistive network can be represented as a graph $\mathbf{\Gamma}(V, \mathbf{B})$, where V is the vertex set and $\mathbf{B}$ is the branch set. Of particular interests are special subsets of $\mathbf{B}$ called *cut-set* defined as follows.

Definition 1: A *cut-set* of $\mathbf{\Gamma}(V, \mathbf{B})$ is a set of branches $\mathbf{C} \subseteq \mathbf{B}$. Removing all branches in $\mathbf{C}$ causes the network unconnected, but the removal of any proper subset of this set keeps the network connected. Among all cut-sets, those disconnecting all TP from power supply pins are defined as *TP cut-set* and denoted as C_{TP} (see Fig. 2 for an example).

C. Sleep transistor insertion and sizing

We formulate the sleep transistor insertion problem as follows.

Formulation 1: Given a fixed P/G network $\mathbf{\Gamma}(V, \mathbf{B})$, the min-area sleep transistor insertion (and sizing) problem (TIS) finds a set of branches $\mathbf{C} \subseteq \mathbf{B}$ to insert sleep transistors such that all paths between TP and power pins are disjointed, the additional voltage drop introduced by inserted sleep transistors are bounded, and the total sleep transistor area is minimized.

Theorem 1: The optimal solution to the TIS problem must be a C_{TP}.

Proof: Assume $\mathbf{S}$ is a solution of TIS and $\mathbf{S} = C_{TP} \cup \mathbf{B}_e$. For purpose of contradiction, we assume $\mathbf{B}_e \neq \Phi$, i.e, there exists at least one branch $c_e \in \mathbf{B}_e$ and one sleep transistor st_e inserted on c_e. Removing st_e from c_e decreases the resistance of c_e. According to Corollary 2.22 in [17], decreasing the resistance of one or more edges in a finite resistive network never increases the effective resistance between any two vertexes. In other words, the effective resistance between any TP and GND never increases. In turn, the voltage on any TP never increases and it still satisfies the voltage drop constraint if it does before removing st_e. I.e., removing st_e results in another solution to TIS with smaller area than $\mathbf{S}$. So, $\mathbf{S}$ can not be an optimal solution of TIS. This leads to a contradiction. Therefore, if $\mathbf{S}$ is a solution of TIS, $\mathbf{S}$ must be a C_{TP}.
□

A C_{TP} divides V into two disjointed subsets where all TP are in one set V_1, and all external power pins in the other set V_2. Although the net current should flow from V_1 to V_2, the current directions in particular branches of C_{TP}, however, could be different. Intuitively, the non-uniform current directions in C_{TP} result in a larger sleep transistor area for the given voltage drop constraints. Therefore, we only consider a C_{TP} with the uniform current direction from V_1 to V_2. This kind of C_{TP} is denoted as $\overrightarrow{C_{TP}}$ in the rest of the paper.

D. Simultaneous sleep transistor insertion and P/G network sizing

Under a constant constraint for voltage drop over sleep transistors and P/G network, increasing the area of sleep transistors reduces voltage drop over sleep transistors. This allows us to reduce area on P/G network with a constant bound on the overall voltage drop, or vice versa. In this sense, there is a trade-off between the area of P/G network and that of sleep transistors. This trade-off can be used to reduce the total chip area. For example, in a design with small number of metal layers, the routing area may be the bottleneck to decide the size of the chip. In this case, budgeting a relatively large area to sleep transistors but a small routing area to P/G network can reduce the total chip area.

To provide a smooth trade-off between the area of P/G network and that of sleep transistors, we formulate the *simultaneous sleep transistor insertion and P/G network sizing* problem as follows:

Formulation 2: Simultaneous sleep transistor insertion and P/G network sizing ($TIPGS$): Given P/G network topology and voltage drop constraint over the P/G network with embedded sleep transistors, the $\overrightarrow{C_{TP}}$ to insert sleep transistors and determines the size of sleep transistors and P/G branches such that $\alpha A_p + \beta A_s$ is minimized, where α and β are given constants, and A_p and A_s are the area of P/G network and sleep transistors, respectively.

III. TIS PROPERTIES AND ALGORITHMS

We first solve TIS on *Single Source Network* (SSN), where all gates are modeled as a single current source and then extend the solution to *Multiple Source Network* (MSN), where gates are modeled as distributed current sources.

A. Single source network

SSN falls into the category of one-port two-terminal resistive network as shown in Fig. 3. The two terminals are TP and ground(GND). In this network, *driving-point impedance* is defined as

$$R = \frac{V}{I},$$

where V and I are the voltage and current between TP and GND, respectively. Regarding this network, TP is a single node and we have:

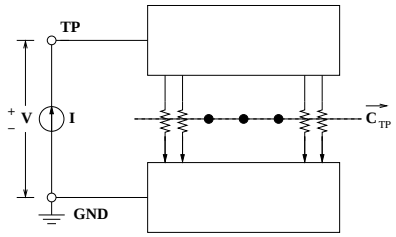

Fig. 3. Illustration of SSN.

Lemma 1: For an arbitrary $\overrightarrow{C_{TP}} = \{c_1, c_2, ..., c_k\}$ in a one-port two-terminal network $\mathbf{\Gamma}(V, \mathbf{B})$, if the resistance of the resistor in each branch c_i increases by $\Delta r_i > 0$, we have

$$\frac{1}{\Delta R} \leq \sum_{\overrightarrow{C_{TP}}} \frac{1}{\Delta r_i}, \tag{2}$$

where ΔR is the increase of the driving-point impedance.

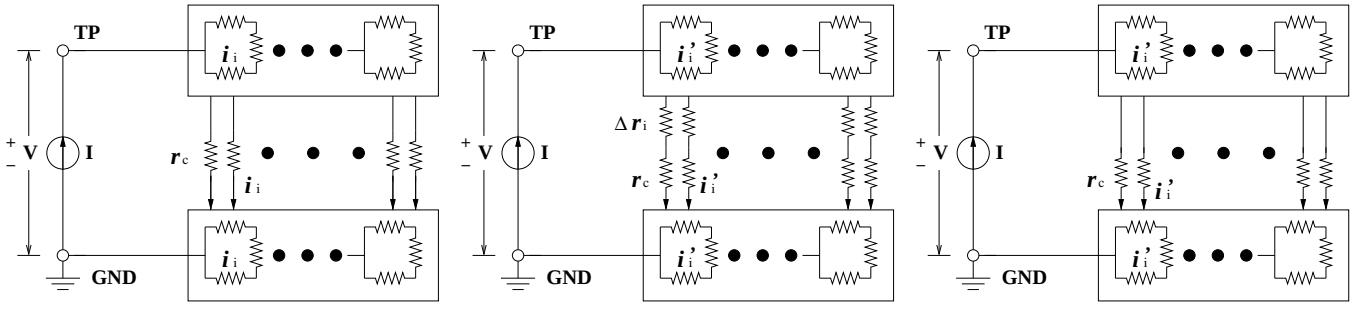

Fig. 4. Explanation of (3).

Proof: Suppose i_i and v_i are current and voltage drop for branch $c_i \in \mathbf{B}$ in the network, respectively. By the *law of the conservation of the energy* we have

$$I^2 R = \sum_{\mathbf{B}} i_i^2 r_i$$

and

$$I^2(R+\Delta R) = \sum_{\mathbf{B}} i_i'^2 r_i + \sum_{\overrightarrow{C_{TP}}} i_i'^2 \Delta r_i,$$

where i_i' is the current for branch c_i after the increase of the resistance. According to *Thomson's principle*[1] [17],

$$\sum_{\mathbf{B}} i_i'^2 r_i \geq \sum_{\mathbf{B}} i_i^2 r_i. \tag{3}$$

We explain (3) in Fig. 4. The original network is shown in (a), where the currents on branches are i_i. (b) is the network after the increase of resistance Δr_i on $\overrightarrow{C_{TP}}$. Note that the currents on branches change to i_i'. We build (c) by removing Δr_i from (b) but keeping the currents i_i' in each branch. One can see that (c) satisfies KCL equations but unnecessarily satisfies KVL equations. According to *Thomson's principle*, the power consumed on the resistors in (c) should be large or equal to that in (a). Therefore, we can obtain (3). Consequently,

$$I^2 \Delta R \geq \sum_{\overrightarrow{C_{TP}}} i_i'^2 \Delta r_i. \tag{4}$$

According to the *arithmetic-geometric mean inequality*, we have

$$(\sum_{\overrightarrow{C_{TP}}} \frac{1}{\Delta r_i})(\sum_{\overrightarrow{C_{TP}}} i_i'^2 \Delta r_i) \geq (\sum_{\overrightarrow{C_{TP}}} i_i')^2,$$

by expanding the two sides of the above inequality. Therefore, we can obtain

$$\begin{aligned} \sum_{\overrightarrow{C_{TP}}} \frac{1}{\Delta r_i} &\geq \frac{(\sum_{\overrightarrow{C_{TP}}} i_i')^2}{\sum_{\overrightarrow{C_{TP}}} i_i'^2 \Delta r_i} \\ &\geq \frac{I^2}{I^2 \Delta R} \\ &\geq \frac{1}{\Delta R}. \end{aligned}$$

□

Lemma 2: For an arbitrary $\overrightarrow{C_{TP}} = \{c_1, c_2, \ldots, c_k\}$, if the current on P/G branch c_i is i_i and $\sum_{\overrightarrow{C_{TP}}} 1/\Delta r_i$ is given, the following conditions minimize ΔV on TP (the increase of voltage after increasing the resistance):

$$\frac{1/\Delta r_i}{\sum_{\overrightarrow{C_{TP}}} 1/\Delta r_i} = \frac{i_i}{I}. \tag{5}$$

and the minimum ΔV is

$$\Delta V = \frac{I}{\sum_{\overrightarrow{C_{TP}}} 1/\Delta r_i}. \tag{6}$$

Proof: If the the current on c_i is I_i and Δr_i satisfies

$$\frac{1/\Delta r_i}{\sum_{\overrightarrow{C_{TP}}} 1/\Delta r_i} = \frac{I_i}{I}, \tag{7}$$

the current in the network remains unchanged after the increase of resistance in $\overrightarrow{C_{TP}}$. Therefore, we have

$$\Delta V = \frac{I}{\sum_{\overrightarrow{C_{TP}}} 1/\Delta r_i}. \tag{8}$$

By Lemma 1, ΔV is minimized.

□

Because Δr_i is the resistance of the sleep transistor inserted at branch c_i, Lemma 2 implies the following Lemma.

Lemma 3: All the sleep transistors have a same voltage drop in an optimal TIS solution.

Lemmas 1, 2 and 3 reveal the following solution to TIS in SSN.

[1] *Thomson's principle* states that the network satisfying KCL and KVL equations simultaneously has a smaller or equal power consumption than the network satisfying KCL only.

Theorem 2: For any $\overrightarrow{C_{TP}}$ in SSN, inserting sleep transistor into branch $c_i \in \overrightarrow{C_{TP}}$ with area of

$$A_i = \rho_s \cdot L_s^2 \cdot \frac{i_i}{\overline{V} - V_p} \tag{9}$$

leads to an optimal solution for TIS, where i_i is the current in c_i, $\overline{V}$ is the voltage constraint on TP, and V_p is the voltage on TP before the insertion of sleep transistors.

Proof: Let the width of the inserted sleep transistor ST_i be W_i and assume the length of the sleep transistors is uniform and denoted as L_s. The total area of the inserted sleep transistors is

$$A_s = L_s \cdot \sum_{\overrightarrow{C_{TP}}} W_i \tag{10}$$

On the other hand, the channel resistance of the sleep transistor in the linear region can be expressed as

$$R_s^i = \rho_s \cdot \frac{L_s}{W_i}, \tag{11}$$

where ρ_s is constant. Therefore, we have

$$A_s = \rho_s \cdot L_s^2 \cdot \sum_{\overrightarrow{C_{TP}}} \frac{1}{R_s^i}. \tag{12}$$

Note that R_s^i is Δr_i in (2), by (2) and (12), we have

$$A_s \geq \rho_s \cdot L_s^2 \cdot \frac{1}{\Delta R} \tag{13}$$

$$\geq \rho_s \cdot L_s^2 \cdot \frac{I}{\Delta V} \tag{14}$$

$$\geq \rho_s \cdot L_s^2 \cdot \frac{I}{\overline{V} - V_p}. \tag{15}$$

By Lemma 2, this minimum area for A_s of

$$\rho_s \cdot L_s^2 \cdot \frac{I}{\overline{V} - V_p} \tag{16}$$

can be achieved in any $\overrightarrow{C_{TP}}$ by the following conditions:

$$\frac{A_i}{A_s} = \frac{I_i}{I}, \tag{17}$$

where A_i is the area of the sleep transistor inserted in branch c_i. In other words, we have

$$A_i = \rho_s \cdot L_s^2 \cdot \frac{I_i}{\overline{V} - V_p}. \tag{18}$$

□

Theorem 3: Any $\overrightarrow{C_{TP}}$ leads to an optimal solution of TIS with the same area.

Proof: It can be directly derived from Theorem 2.

□

Note that Theorems 2 and 3 solve TIS optimally and indicate that the optimal solution of TIS is not unique. This design freedom could be used to optimize for other design constraints such as routing congestion.

B. Multiple source network

MSN belongs to m-terminal network as shown in Fig. 5 with $m-1$ nodes in TP. Similar to Lemma 1, we have

Hypothesis 1: For an arbitrary $\overrightarrow{C_{TP}} = \{c_1, c_2, ..., c_k\}$ in an m-terminal network, if the resistance of the resistor in branch c_i increases by $\Delta r_i > 0$, then

$$\sum_{i=1}^{m-1} \frac{I_i}{\Delta V_i} \leq \sum_{i=1}^{k} \frac{1}{\Delta r_i}, \tag{19}$$

where I_i is the current source between terminal i and GND, and ΔV_i is the increase of voltage at terminal i.

TIS of MSN can be solved based on Hypothesis 1. By Hypothesis 1, we have

$$\sum_{i=1}^{m-1} \frac{I_i}{\overline{V} - v_{p,i}} \leq \sum_{i=1}^{k} \frac{1}{r_{s,i}}, \tag{20}$$

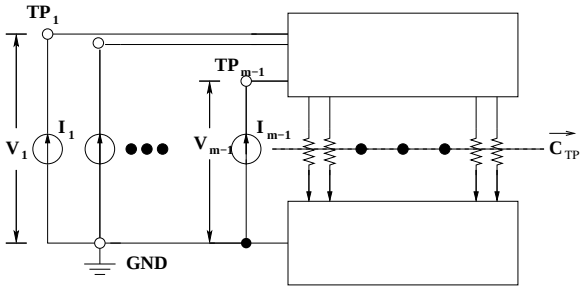

Fig. 5. Illustration of MSN.

where $\overline{V}$ is the voltage drop constraint on TP, I_i is the current on TP_i, $v_{p,i}$ is the voltage on TP_i with no sleep transistors inserted, and $r_{s,i}$ is the resistance of sleep transistors. Similar to Theorem 2 in SSN, we have

$$A_s \geq \rho_s \cdot L_s^2 \cdot \sum_{i=1}^{m-1} \frac{I_i}{\overline{V} - v_{p,i}}. \tag{21}$$

The right-hand side of (21) is the lower bound on the area of sleep transistors in MSN. One way to achieve the minimum area is to find a separable $\overrightarrow{C_{TP}}$, which is defined as follows.

Definition 2: A $\overrightarrow{C_{TP}}$ is separable if it can be partitioned to $m-1$ subset $\overrightarrow{C_{TP}}^{(1)}$, $\cdots$, $\overrightarrow{C_{TP}}^{(m-1)}$ such that 1) For any $1 \leq i, j \leq m-1$, $\overrightarrow{C_{TP}}^{(i)} \cap \overrightarrow{C_{TP}}^{(j)} = \Phi$. 2) Each subset $\overrightarrow{C_{TP}}^{(i)}$ is a $\overrightarrow{C_{TP}}$ for TP_i.

The simplest separable $\overrightarrow{C_{TP}}$ is to use all P/G branches directly connected to a current source as $\overrightarrow{C_{TP}}^{(i)}$.

Hypothesis 1 will be verified experimentally in Section V. Assuming a valid Hypothesis 1, we present an algorithm for the TIS problem (see figure 6) as follows: A separable $\overrightarrow{C_{TP}}$ is first obtained and then for each $\overrightarrow{C_{TP}}^{(t)}$ sleep transistors are inserted according to Theorem 2.

TIS algorithm for MSN
1. Find a separable $\overrightarrow{C_{TP}} = \overrightarrow{C_{TP}}^{(1)} \cup \cdots \cup \overrightarrow{C_{TP}}^{(m-1)}$. 2. For each $\overrightarrow{C_{TP}}^{(t)}$ For each $c_i \in \overrightarrow{C_{TP}}^{(t)}$, insert sleep transistor with $A_i = \rho_s \cdot L_s^2 \cdot \frac{i_i}{\overline{V} - v_{p,t}}$, where i_i is the current on c_i and $v_{p,t}$ is the voltage on TP_t before inserting sleep transistors.

Fig. 6. TIS algorithm for MSN.

IV. TIPGS PROPERTIES AND ALGORITHMS

As in Section III, we first solve $TIPGS$ in SSN and then extend the solution to MSN in this section.

A. Single source network

Let A_p be the area of the P/G network, we have

$$A_p = \sum_{\mathbf{B}} L_p \cdot W_p. \tag{22}$$

To solve the $TIPGS$ problem for SSN, we introduce the following lemmas first.

Lemma 4: If a min-area P/G network without sleep transistors satisfies voltage drop constraint $\overline{V_p}$ at tapping points, the product of the minimal P/G area A_p^* and $\overline{V_p}$ is a constant. We define the constant product as

$$K_p^* = A_p^* \cdot \overline{V_p}. \tag{23}$$

Proof: Considering two arbitrary voltage drop constraint $\overline{V_p}^{(1)}$ and $\overline{V_p}^{(2)}$, for purpose of contradiction we assume

$$K_p^{*(1)} = \overline{V_p}^{(1)} \cdot \sum_i L_{p_i} \cdot {W_{p_i}^*}^{(1)}, \tag{24}$$

$$K_p^{*(2)} = \overline{V_p}^{(2)} \cdot \sum_i L_{p_i} \cdot {W_{p_i}^*}^{(2)}, \tag{25}$$

and

$$K_p^{*(1)} > K_p^{*(2)} \tag{26}$$

Since the resistive network is linear, setting the width of wire p_i to be

$${W_{p_i}}^{(1')} = \frac{\overline{V_p}^{(2)}}{\overline{V_p}^{(1)}} \cdot {W_{p_i}^*}^{(2)}, \tag{27}$$

the voltage drop on the tapping points are $\overline{V_p}^{(1)}$. By (26) and (27),

$$\sum_i L_{p_i} \cdot {W_{p_i}}^{(1')} = \frac{\overline{V_p}^{(2)}}{\overline{V_p}^{(1)}} \cdot \sum_i L_{p_i} \cdot {W_{p_i}^*}^{(2)} \tag{28}$$

$$< \sum_i L_{p_i} \cdot {W_{p_i}^*}^{(1)} \tag{29}$$

Which contradicts to the assumption that ${W_{p_i}^*}^{(1)}$ is the optimal solution. Therefore,

$$K_p^{*(1)} \leq K_p^{*(2)}. \tag{30}$$

Similarly, we can prove that

$$K_p^{*(2)} \leq K_p^{*(1)}.$$

In other words,

$$K_p^{*(2)} = K_p^{*(1)}.$$

I.e., K_p^* must be constant.

□

Lemma 4 indicates that A_p^* is reversely proportional to $\overline{V_p}$ and shows that the optimal sizing solution under a voltage drop constraint $\overline{V_{p,1}}$ can be extended to another voltage drop constraint $\overline{V_{p,2}}$ by scaling the wire widths of all P/G grids using the ratio of $\overline{V_{p,1}}/\overline{V_{p,2}}$. Similar to Lemma 4, we have the following lemma for sleep transistors.

Lemma 5: For a given P/G network, we assume that sleep transistors inserted at an arbitrary $\overrightarrow{C_{TP}}$ have a voltage drop equal to or below $\overline{V_s}$. The product of the minimum sleep transistor area A_s^* and $\overline{V_s}$ is a constant. We define the constant product as

$$K_s^* = A_s^* \cdot \overline{V_s}. \tag{31}$$

Proof: According to (15), the minimum A_s should be

$$A_s^* = \rho_s \cdot L_s^2 \cdot \frac{I}{\overline{V_s}}. \tag{32}$$

Therefore

$$K_s^* = A_s^* \cdot \overline{V_s} \tag{33}$$

$$= \rho_s \cdot L_s^2 \cdot I \tag{34}$$

is constant.

□

Lemma 5 indicates the same property for sleep transistors as Lemma 4 for P/G network.

With a total voltage drop $\overline{V}$ over P/G network and sleep transistors, we denote the voltage drop constraint on sleep transistors as $\overline{V_s}$ and the voltage drop constraint on P/G network by removing sleep transistors as $\overline{V_p}$.

Lemma 6: Given the voltage drop constraint on the TP in SSN as $\overline{V}$, we have

$$\alpha A_p + \beta A_s \geq \frac{(\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*})^2}{\overline{V}}. \tag{35}$$

In other words, (35) provides a lower bound on the weighted area of P/G network and sleep transistors.

Proof: According to Lemma 4 and 5, we have

$$\alpha A_p + \beta A_s \geq \frac{\alpha K_p^*}{\overline{V_p}} + \frac{\beta K_s^*}{\overline{V_s}}, \tag{36}$$

where

$$\overline{V_p} + \overline{V_s} = \overline{V}. \tag{37}$$

Because

$$(\overline{V_p} + \overline{V_s}) \cdot (\frac{\alpha K_p^*}{\overline{V_p}} + \frac{\beta K_s^*}{\overline{V_s}}) = \alpha K_p^* + \beta K_s^* + \alpha K_p^* \frac{\overline{V_s}}{\overline{V_p}} + \beta K_s^* \frac{\overline{V_p}}{\overline{V_s}} \tag{38}$$

$$\geq \alpha K_p^* + \beta K_s^* + 2\sqrt{\alpha K_p^* \cdot \beta K_s^*} \tag{39}$$

$$= (\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*})^2 \tag{40}$$

Therefore

$$\alpha A_p + \beta A_s \geq \frac{(\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*})^2}{\overline{V}}.$$

□

Theorem 4: In an optimal $TIPGS$, $\overline{V_s^*}$ and $\overline{V_p^*}$ must be

$$\frac{\sqrt{\beta K_s^*}}{\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*}} \cdot \overline{V}, \tag{41}$$

and

$$\frac{\sqrt{\alpha K_p^*}}{\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*}} \cdot \overline{V}, \tag{42}$$

respectively (note that TP is a single node in $TIPGS$).

Proof: Obviously, the conditions to make (40) valid are:

$$\overline{V_s^*} = \frac{\sqrt{\beta K_s^*}}{\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*}} \cdot \overline{V}, \tag{43}$$

and

$$\overline{V_p^*} = \frac{\sqrt{\alpha K_p^*}}{\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*}} \cdot \overline{V}. \tag{44}$$

□

Theorem 5: Inserting sleep transistors at any $\overrightarrow{C_{TP}}$ leads to optimal $TIPGS$ solutions with the same weighted sum of P/G network and sleep transistor area.

Proof: This is a direct conclusion from Theorem 4.

□

Theorem 4 is a necessary condition to minimize the weighted sum of P/G network and sleep transistor area. To make it sufficient, additionally we need to (i) optimally size P/G network to minimize A_p under the voltage drop constraint $\overline{V_p^*}$ determined by (42) and (ii) follow the solution of TIS to insert sleep transistors under the voltage drop constraint $\overline{V_s^*}$ determined by (41).

B. Multiple source network

Similar to SSN, K_p^* and K_s^* can be defined for MSN. Then, the counterpart of Lemma 6 is presented as follows.

Hypothesis 2: Given the voltage drop constraint on TP in MSN as $\overline{V}$, we have

$$\alpha A_p + \beta A_s \geq \frac{(\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*})^2}{\overline{V}}. \tag{45}$$

In other words, (45) provides a lower bound on the weighted area of P/G network and sleep transistors in MSN.

If Hypothesis 2 holds, Theorem 4 and 5 hold for MSN, too. Therefore, an $TIPGS$ algorithm for MSN can be developed as in Fig. 7. We first determine $\overline{V_s^*}$ and (41) and (42), respectively, and then size P/G network and sleep transistors separately.

However, no algorithm has been proposed in the literature to optimally size P/G network (step 2 in Fig. 7). Nevertheless, we can minimize $\alpha A_p + \beta A_s$ based on the best known algorithm to size P/G network and sequential linear programming (SLP) from [10] is used in this paper.

$TIPGS$ algorithm for MSN
1. Determine $\overline{V_s^*}$ and $\overline{V_p^*}$ by (41) and (42), respectively. 2. Size P/G network under constraint $\overline{V_p^*}$ to minimize A_p. 3. Insert and size sleep transistors under constraint $\overline{V_s^*}$ using TIS algorithm in Fig. 6.

Fig. 7. TIS algorithm for MSN.

V. Experiment

In this section, we first verify Hypotheses 1 and 2 by experiments, and then compare the *Hypo1-based TIS algorithm* in Fig. 6 and *Hypo2-based $TIPGS$ algorithm* in Fig. 7 with alternative algorithms based on sequential linear programming.

A. Verification of Hypothesis 1

For the purpose of verifying Hypothesis 1, we define *effective area ratio* for TIS as

$$EAR_{TIS} = (\sum_{i=1}^{m-1} \frac{I_i}{\Delta V_i})/(\sum_{i=1}^{k} \frac{1}{\Delta r_i}). \tag{46}$$

where I_i, ΔV_i, and Δr_i are same as those in Hypothesis 1. If Hypothesis 1 holds, we have

$$EAR_{TIS} \leq 1. \tag{47}$$

To verify Hypothesis 1, we compute the EAR_{TIS} for nine mesh networks as shown in Table II under 100,000 random solutions. For each solution, the value of current sources , the $\overrightarrow{C_{TP}}$, and the size of sleep transistors are randomly chosen, and EAR_{TIS} is obtained by solving the networks with a linear solver integrated in SIS1.2 [18]. We report the computed EAR_{TIS} in column 4 of Table II.

1	2	3	4	5
			Max. EAR	
Mesh	# Node	# Branches	TIS	$TIPGS$
3×3	16	24	1.00	0.79
5×5	25	60	1.00	0.68
10×10	121	220	0.96	0.89
20×20	441	840	1.00	0.96
30×30	961	1,860	0.97	0.97
40×40	1,681	3,280	0.98	0.89
60×60	3,721	7,320	0.97	0.93
80×80	6,561	12,960	0.97	1.00
100×100	10,201	20,200	0.96	0.96

TABLE II

Random solutions(100,000 ×) to compute the maximum EAR.

According to column 4 of Table II, it clearly shows that the maximum EAR_{TIS} values in all networks are equal to or less than 1. This means that the solution of TIS by the algorithm in Fig. 6 has the smallest area among all these 100,000 random solutions. This strongly indicates the correctness of Hypothesis 1.

B. Verification of Hypothesis 2

To verify Hypothesis 2, we define *effective area ratio* for $TIPGS$ as

$$EAR_{TIPGS} = (\frac{(\sqrt{\alpha K_p^*} + \sqrt{\beta K_s^*})^2}{\overline{V}})/(\alpha A_p + \beta A_s). \tag{48}$$

If Hypothesis 2 holds, we have

$$EAR_{TIPGS} \leq 1. \tag{49}$$

We compute the EAR_{TIPGS} for $TIPGS$ in the same fashion as for TIS. For each circuit, we carry out 100,000× random solutions to find the maximum EAR_{TIPGS}. However, in $TIPGS$ K_p^* and K_s^* are needed to compute EAR_{TIPGS}. According to Lemma 4,

$$K_p^* = A_p^* \cdot \overline{V_p}. \tag{50}$$

Since A_p^* is unavailable in the experiments, we approximate K_p^* by

$$K_p^* = \min_S(A_p \cdot \overline{V_p}), \tag{51}$$

where S represents the set for all solutions. K_s^* is computed by

$$K_s^* = \rho_s \cdot L_s^2 \cdot \sum_{i=1}^{m-1} I_i. \tag{52}$$

We reported the computed EAR_{TIPGS} in column 5 of Table II. According to column 5 of Table II, the maximum EAR_{TIPGS} is always less or equal to 1 among 100,000 random solutions for all networks. This clearly implies the correctness of Hypothesis 2.

C. Comparison between algorithms for TIS and $TIPGS$

Circuit	# Block	# GND	*SLP-based* $A_s(\%)$	*Hypo1-based* $A_s(\%)$
apte	9	2	0.18	0.14 (-22.2%)
xerox	9	4	0.28	0.17 (-29.3%)
hp	10	3	0.25	0.14 (-44.0%)
a3	25	3	0.21	0.13 (-38.1%)
ami	33	3	0.19	0.13 (-31.2%)
playout	62	5	0.34	0.19 (-44.1%)
g2	241	4	0.15	0.10 (-33.3%)

TABLE III
COMPARISON BETWEEN SLP-BASED AND HYPO1-BASED ALGORITHM FOR TIS.

Circuit	# Block	# GND Pin	*SLP-based (%)*			*Hypo2-based (%)*		
			A_p	A_s	Weighted sum	A_p	A_s	Weighted sum
apte	9	2	2.55	0.18	2.73	1.79	0.30	2.09 (-23.4%)
xerox	9	4	3.14	0.28	3.42	1.94	0.26	2.20 (-35.7%)
hp	10	3	2.31	0.25	2.56	1.20	0.31	1.51 (-41.0%)
a3	25	3	2.08	0.21	2.29	1.37	0.25	1.62 (-29.3%)
ami	33	3	1.88	0.19	2.07	0.90	0.19	1.09 (-47.3%)
playout	62	5	4.96	0.34	5.30	4.19	0.38	4.57 (-13.8%)
g2	241	4	3.67	0.15	3.82	1.35	0.13	1.48 (-61.3%)

TABLE IV
COMPARISON BETWEEN SLP-BASED AND HYPO2-BASED ALGORITHM FOR $TIPGS$.

1) Algorithms: We have revised the sequential linear programming algorithm proposed in [10] to solve $TIPGS$ (denote as *SLP-based algorithm*) as the baseline case for comparison. The sequential linear programming algorithm in [10] is employed to size P/G network, where each branch of P/G network is modeled as a resistor. Because sleep transistors are also modeled as resistors, we are able to modify [10] to size both the P/G network and sleep transistors simultaneously. Specifically, in the *SLP-based algorithm*, we choose a separable $\overrightarrow{C_{TP}}$ as close as possible to *TP* and minimize weighted area of P/G network and sleep transistors.

In fact, the *SLP-based algorithm* provides a baseline case for comparison for both *Hypo1-based algorithm* to solve TIS and *Hypo2-based algorithm* to solve $TIPGS$. *Hypo1-based algorithm* follows the exact steps in Fig. 6. The Hypo2-based algorithm follows the steps in Fig. 7 but with minor modifications. Because there is no optimal algorithm available to minimize A_p, we employ the *SLP-based algorithm* to obtain the "optimal" P/G network under given voltage drop constraints.

For all algorithms in the experiments, we have chosen the same separable $\overrightarrow{C_{TP}}$ that is directly adjacent to the tapping points. Theorem 3 and 5 indicate that all $\overrightarrow{C_{TP}}$ have the same optimal value for both TIS and $TIPGS$, but experiment results have shown that this $\overrightarrow{C_{TP}}$ produces a relatively good result for *SLP-base algorithm*. Therefore, the experiment setting is favorable to the *SLP-base algorithm*.

2) Results: The *SLP-based*, *Hypo1-based*, and *Hypo2-based algorithm* have been applied to NCSU benchmarks [19]. Switching current is modeled as time-invariant and the current density is $300mA/mm^2$, which is similar to that of the Alpha microprocessor in [20]. We assume the P/G pitch as $50\mu m$ and present A_p and A_s in the percentage of chip area.

To compare the *SLP-based algorithm* with the *Hypo1-base algorithm*, we first apply the *SLP-based algorithm* to find the size of P/G network branches and the size of sleep transistors. Then, we fix the size of P/G network branches and re-size the sleep transistors by using the *Hypo1-base algorithm*. We compare the total area of sleep transistors obtained by the *SLP-based algorithm* and *Hypo1-base algorithm* in Table III. For TIS problem, we found that the *Hypo1-base algorithm* is consistently better than the *SLP-based algorithm* and it can reduce the transistor area by up to 44.1%. As shown in Table IV for $TIPGS$

problem, the *Hypo2-base algorithm* reduces the total area significantly (up to 61.3%) with α and β being set as 1.0. Also, it is observed in the experiment that SLP-based algorithm has a different area distribution among sleep transistors compared to that calculated by Hypothesis 1. We believe this is the main reason that *SLP-based* algorithm introduces more area.

VI. Discussion and Conclusion

Under a distributed P/G network model, we have studied the sleep transistor insertion (and sizing) problem (TIS) and simultaneous sleep transistor insertion and P/G sizing problem ($TIPGS$). We have developed effective algorithms to solve these two problems by revealing the optimal solutions to them. Compared with the best known approach using sequential linear programming, our algorithms reduce area by up to 44.1% and 61.3% for TIS and $TIPGS$, respectively. Our TIS and $TIPGS$ algorithms are extremely efficiently too, as all steps are based on closed-form formulas. We have shown that there exist multiple optimal solutions with a same area to these problems. This offers extra design freedoms to consider other design constraints such as routing congestion and the layout placement for gates.

In this paper, the time-invariant current model is assumed and current density constraint for P/G network is not considered. In the future, we intend to extend our problem formulations and algorithms to time-variant current model and include the current density constraint in our study.

References

[1] A. Agarwal, C. H. Kim, S. Mukhopadhyay, and K. Roy, "Leakage in nano-scale technologies: mechanisms, impact and design considerations," in *Proc. Design Automation Conf*, pp. 6–11, June 2004.

[2] A.S.Grove, "Changing vectors of moore's law," *Keynote speech, International Electron Devices Meeting*, Dec. 2002.

[3] J. Kao, S. Narendra, and A. Chandrakasan, "MTCMOS hierarchical sizing based on mutual exclusive discharge patterns," in *DAC*, 1998.

[4] M. Anis, S. Areibi, and M. Elmasry, "Dynamic and leakage power reduction in MTCMOS circuits using an automated efficient gate clustering technique," in *DAC*, 2002.

[5] C. Long and L. He, "Distributed sleep transistor network for power reduction," in *Proc. Design Automation Conf*, pp. 181–186, 2003.

[6] S. Kosonocky, M. Irnmediato, P. Cottrell, T. Hook, R. Mann, and J. Brown, "Enhanced multi-threshold (mtcmos) circuits using variable well bias," in *Proc. Int. Symp. on Low Power Electronics and Design*, pp. 165–169, 2001.

[7] H. H. Chen and D. D. Ling, "Power supply noise analysis methodology for deep-submicron VLSI chip design," in *Proc. Design Automation Conf*, pp. 638–648, June 1997.

[8] H. Qian, S. R. Nassif, and S. S. Sapatnekar, "Random walks in a supply network," in *DAC*, pp. 93–98, 2003.

[9] D. Kouroussis and F. N. Najm, "A static pattern-independent technique for power grid voltage integrity verification," in *DAC*, pp. 99–104, 2003.

[10] X. D. Tan and C. J. Shi, "Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programmings," in *Proc. Design Automation Conf*, pp. 78–83, 1999.

[11] S. Boyd, L. Vandenberghe, A. E. Gamal, and S. Yun, "Design of robust global power and ground network," in *Proc. Int. Symp. on Physical Design*, pp. 60–65, April 2001.

[12] H. Su, K. Gala, and S. Sapatneka, "Fast analysis and optimization of power/ground networks," in *Proc. Int. Conf. on Computer Aided Design*, pp. 477–480, Nov 2000.

[13] M. Zhao, R. Panda, S. S. Sapatnekar, T. Edwards, R. Chaudhry, , and D. Blaauw, "Hierarchical analysis of power distribution networks," in *Proc. Design Automation Conf*, 2000.

[14] S. Zhao, K. Roy, and C.-K. Koh, "Decoupling capacitance allocation and its application to power-supply noise-aware floorplanning," in *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pp. 81–92, Jan. 2002.

[15] T. Mitsuhashi and E. Kuh, "Power and ground network topology optimization for cell based VLSIs," in *Proc. Design Automation Conf*, pp. 524–529, 1992.

[16] S. X. D. Tan and C. J. Shi., "Efficient vlsi power/ground network sizing based on equivalent circuit modeling," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, March 2003.

[17] P. M. Soardi, *Potential theory on infi nite networks.* New York, Springer-Verlag, 1994.

[18] E. M. Sentovich, K. J. Singh, L. Lavagno, and etc., "Sis: a system for sequential circuit synthesis," *Memorandum NO. UCB/ERL M92/41*, May 1992.

[19] http://www.cbl.ncsu.edu.

[20] A. Jain, W. Anderson, T. Benninghoff, and D. e. Berucci, "A 1.2 ghz alpha microprocessor with 44.8 gb/s chip pin bandwidth," in *Proc. IEEE Int. Solid-State Circuits Conf.*, pp. 240–241, 2001.