

Two-Phase Fine-Grain Sleep Transistor Insertion Technique in Leakage Critical Circuits

Yu Wang, *Member, IEEE*, Ku He, Rong Luo, *Member, IEEE*, Hui Wang, and Huazhong Yang, *Senior Member, IEEE*

Abstract—Sleep transistor (ST) insertion is a valuable leakage reduction technique in circuit standby mode. Fine-grain sleep transistor insertion (FGSTI) makes it easier to guarantee circuit functionality and improve circuit noise margins. In this paper, we introduce a novel two-phase FGSTI technique which consists of ST placement and ST sizing. These two phases are formally modeled using mixed integer linear programming (MILP) models. When the circuit timing relaxation is not large enough to assign ST everywhere, leakage feedback (LF) gates, which are used to avoid floating states, induce large area and dynamic power overhead. An extended multi-object ST placement model is further proposed to reduce the leakage current and the LF gate number simultaneously. Finally, heuristic algorithms are developed to speed up the ST placement phase. Our experimental results on the ISCAS'85 benchmarks reveal that: 1) the two-phase FGSTI technique achieves better results than the simultaneous ST placement and sizing method; 2) when the circuit timing relaxation varies from 0% to 5%, the multi-object ST placement model can achieve on average $4 \times 9 \times$ LF gate number reduction, while the leakage difference is only about 8% of original circuit leakage; 3) our heuristic algorithm is $1000 \times$ faster than the MILP method within an acceptable loss of accuracy.

Index Terms—Fine-grain sleep transistor insertion, leakage feedback gate, leakage reduction, mixed integer linear programming (MILP).

I. INTRODUCTION

POWER consumption is becoming a first-order design criterion. The total power dissipation consists of dynamic power, short circuit power, and leakage power, thus, can be expressed as

$$P_{\text{total}} = P_{\text{dynamic}} + P_{\text{leakage}} + P_{\text{shortcircuit}}$$

$$= \sum_{i=1}^N \left(\frac{1}{2} \alpha_i f C_i V_{\text{DD}}^2 + I_{L,i} V_{\text{DD}} + \alpha_i f Q_{\text{short},i} V_{\text{DD}} \right)$$

where f is the operation frequency, V_{DD} is the supply voltage, and N is total gate number. α_i , C_i , $I_{L,i}$, and $Q_{\text{short},i}$ are the transition probability, load capacitance, leakage current, and short circuit charge of the i th gate, respectively. Leakage power dissipation has become comparable to switching power dissipation

Manuscript received September 6, 2006; revised July 10, 2007. Published August 20, 2008 (projected). This work was supported in part by National Natural Science Foundation of China under Contract 90207002 and Contract 60506010 and by the National 863 Project of China under Contract 2006AA01Z224. This is a revised and expanded version based on two papers accepted by ISLPED 2006 and APCCAS 2006.

The authors are with the Circuits and Systems Division, Electronics Engineering Department, Tsinghua University, Beijing, 100084, China (e-mail: yu-wang@tsinghua.edu.cn).

Digital Object Identifier 10.1109/TVLSI.2008.2000523

[1], while the behavior of the short circuit power dissipation remains at around 10% of the total power dissipation [2].

As technology scales, leakage power is becoming a major portion of the total power consumption, for example, leakage power is reported to make up about 40% of total power at the 90-nm technology node [3]. Large leakage current will degrade noise immunity in dynamic circuits, increase the standby power dissipation to unacceptable levels, and lead to excessive heating; all of these may cause circuits fail to function properly [4]. Inevitably, many techniques have been proposed to reduce the increasing leakage power. Besides architecture level techniques [4], leakage control techniques can be broadly categorized into two main categories: process level and circuit level techniques [5]. At the process level, leakage reduction can be achieved by controlling the dimensions (length, oxide thickness, junction depth, etc.) and doping profile in transistors. Circuit level techniques include transistor stacking [6], input vector control [7], body biasing [8], multi- V_t assignment [9]–[11], dynamic V_t scaling (DVTS) [12], variable supply voltage [13], cell re-sizing [14], and leap transistor insertion. Among these circuit level techniques, sleep transistor insertion is the most effective one when the circuit is standby [15].

A. Multi-Threshold CMOS Techniques

Sleep transistor (ST) insertion technique is essentially placing an ST between the gates and the power/ground (P/G) net. In burst mode circuits, where the system spends the majority of the time in an idle standby state, ST insertion is proven to be a very effective technique for leakage current reduction during the standby mode [15]–[24]. STs can be inserted into the circuit by two different manners: global and local [22]. Thus, we classify ST insertion techniques into “block-based ST insertion” [as shown in Fig. 1(a)] and “fine-grain ST insertion” (FGSTI) [as shown in Fig. 1(b)] according to different ST insertion manners.

1) *Block-Based ST Insertion (BBSTI)*: The most popular ST insertion technique is gating the power supply of sizable blocks using large STs [15]. This is concluded as *BBSTI* technique. In a *BBSTI* technique, all the gates in one block have a fixed timing relaxation, so it is also called fixed slowdown method. The previous works on *BBSTI* techniques [16]–[20] presented some methods on clustering gates into blocks in order to optimize the leakage current and ST sizes. All these methods focus on how to reduce the ST area penalty along with a remarkable leakage saving.

J. Kao *et al.* [16] presented the first method to automatically size ST of a large block based on mutual exclusive discharge patterns of internal blocks. M. Anis *et al.* [17], [18] developed a cluster-based design structure to avoid putting a larger ST in the center of a large block: the whole circuit is divided into small

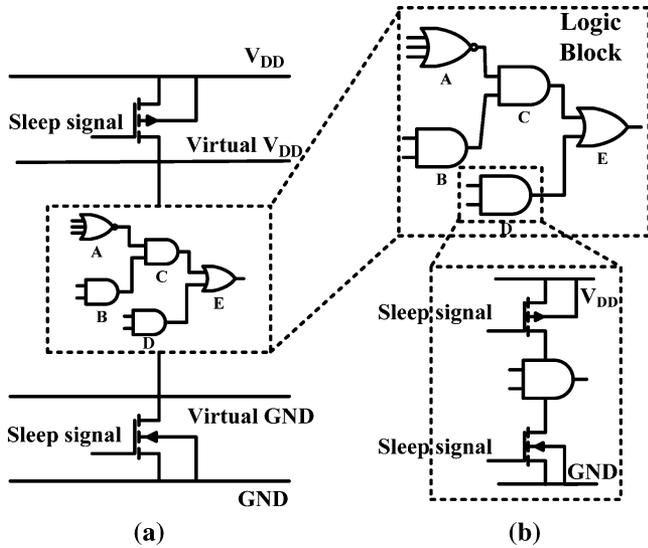


Fig. 1. FGSTI versus BBSTI.

blocks with several gates in one block; they presented several fast heuristic techniques for efficient gate clustering and optimized the ST size for each block according to the current of the block. C. Long *et al.* [19], [20] proposed a distributed sleep transistor network (DSTN) approach in which all the STs are connected to further reduce the area penalty and improve the circuit performance.

BBSTI techniques greatly reduce the area penalty, but in the P/G network they induce large ground bounce which has adverse effects on circuit speed and noise immunity [23]. Additional delay, about 5% circuit propagation time, must be suffered from because of ST insertion. Furthermore, for each block, ST size is decided by the worst case current which is quite difficult to determine without comprehensive simulation [17]. So it is harder to guarantee circuit functionality for large blocks with only one ST [22].

2) *Fine-Grain ST Insertion (FGSTI)*: In recent years, the *FGSTI* technique [21]–[24], which can be also called the gate level ST insertion, shows some advantages over the *BBSTI* technique. It is easier to guarantee circuit functionality in the *FGSTI* technique because ST sizes are not determined by the worst case current of large circuit blocks. The *FGSTI* technique leads to a smaller simultaneous switching current when the circuit mode changes between standby and active, thus improves circuit noise margins. Furthermore, better circuit slack utilization is achieved because the slowdown of each gate is not fixed, and this leads to further leakage and area reduction [23]. V. Khandelwal *et al.* [23] pointed out that the *FGSTI* technique corresponded to an area penalty of roughly only 5% using standard cell placement.

When the circuit slowdown is not enough to assign STs everywhere in the *FGSTI* technique, a large amount of leakage feedback (LF) gates may be used to avoid floating states [21]. As it will be shown in our results, the LF gate number may exceed as much as 80% of the gates with ST for certain circuits when LF gate is not considered in the *FGSTI* technique; the additional inverters in the LF gates will induce large area and dynamic power penalty.

B. H. Calhoun *et al.* [22] proposed a fine-grain ST insertion design methodology and several design rules. The authors also

make a detailed comparison between local and global ST insertion. Recently, V. Khandelwal *et al.* [23] presented a selectively ST insertion methodology with better utilization of circuit slack. They used a one-shot algorithm to determine where to put ST in the *FGSTI* design considering leakage feedback gates, but they did not address how to perform the *FGSTI* technique when the circuit slowdown is 0%. Also, the one-shot algorithm may easily lead to a local optimal result. Our previous work [24] introduced a mixed integer linear programming (MILP) model for the *FGSTI* technique to determine ST placement and sizing simultaneously without considering LF gate. The MILP model leads to an accurate result, but its computation time is considerably long.

B. Our Contributions

In this paper, we proposed a two-phase *FGSTI* technique that has the following contributions.

- 1) Simple leakage current and delay models of a single gate are presented. To the best of our knowledge, our leakage current model analysis is the first to provide the designer the negligible dependence of ST size on the amount of leakage saving, and makes the two-phase *FGSTI* reasonable.
- 2) The two phases of our *FGSTI* technique: a) ST placement and b) ST sizing, are modeled using MILP and LP models, respectively. Fewer variables and constraints with less approximation are used in the models, so that our two-phase *FGSTI* technique is more accurate and faster comparing with our previous simultaneous ST placement and sizing method using MILP [24]. The ST placement phase can achieve an impressive leakage saving when the conventional fixed slowdown method can not be performed. Furthermore, if the circuit timing relaxation is large enough to use conventional fixed slowdown method, our ST sizing still leads to a much smaller total ST size.
- 3) LF gate and normal ST gate are compared to prove that a carefully sized LF gate can substitute for a normal ST gate without affecting the circuit performance. An extended multi-object ST placement model is presented to provide the designer the relationship between LF gate number and the leakage current reduction rate. Our experimental results show that, when the circuit slowdown is 0%, comparing with the method only considering the leakage current reduction, on average 4× LF gate number reduction can be achieved; meanwhile the leakage current difference between our method and the method only considering leakage current reduction is only about 7.9% of the original circuit leakage.
- 4) Since the computation time for solving MILP model is not stable and may be considerably long, fast heuristic algorithms are developed for ST placement phase with simultaneous LF gate reduction. Our simple algorithms are investigated from intuition thinking to detailed implementation to show their effectiveness. On average 1000× speed up can be achieved using our heuristic algorithm compared with an MILP solver, while the loss of accuracy is acceptable.

This paper is organized as follows. In Section II, our leakage current and delay models are first presented and then analyzed to prove the rationality of our two-phase *FGSTI* technique; LF

TABLE I
LEAKAGE CURRENT IN NOR2XL AND NAND4XL (fA). (180-nm TECHNOLOGY)

	I_{leak} in NOR2XL $A(v) = 1.60495$			I_{leak} in NAND4XL $A(v) = 2.97335$		
	Hspice	Our Model	Error	Hspice	Our Model	Error
w/o ST	14606.8	N/A	N/A	12261.3	N/A	N/A
$(W/L) = 2$	5.39	3.21	-40.4%	10.20	5.95	-41.7%
$(W/L) = 4$	5.95	6.42	8.0%	10.97	11.89	8.4%
$(W/L) = 8$	8.59	12.84	49.4%	15.58	23.79	52.6%
$(W/L) = 16$	27.65	25.68	-7.1%	51.37	51.37	-7.4%

gate and normal ST gate are compared in Section II-D. The two-phase *FGSTI* technique formulated using MILP model is proposed in Section III. Section IV introduces our heuristic algorithms for ST placement phase. The implementation and experimental results are presented and analyzed in Section V. In Section VI, we conclude this paper.

II. PRELIMINARIES

In this section, leakage current and delay models used in our two-phase *FGSTI* technique are first proposed. The models are examined to prove that an *FGSTI* design can be performed in two phases. Finally, LF gate and normal ST gate are compared to prove that an LF gate can substitute for a normal ST gate without affecting the circuit performance. ST with variable size decided by the process technology are used in our two-phase *FGSTI* design.

In this paper, a combinational circuit is modeled by a directed acyclic graph (DAG) $G = (V, E)$. A vertex $v \in V$ represents a CMOS gate from the given library, while an edge $(i, j) \in E$, $i, j \in V$ represents a connection from vertex i to vertex j . (W/L) is used to measure the size of a ST, because L is a constant that equals the minimum transistor channel length.

A. Leakage Model

For the gates without ST, a leakage lookup table is created by simulating all the gates in the standard cell library under all possible input patterns. Thus the leakage current $I_{w/o}(v)$ can be expressed as

$$I_{w/o}(v) = \sum_{\text{IN}} I_l(v, \text{IN}) \times \text{Prob}(v, \text{IN}) \quad (1)$$

where $I_l(v, \text{IN})$ and $\text{Prob}(v, \text{IN})$ are the leakage current and the probability of gate v under input pattern IN. A linear model is proposed to represent leakage current $I_w(v)$ based on HSPICE simulation results

$$I_w(v) = A(v) \times (W/L)_v \quad (2)$$

where $A(v)$ is a constant and decided by the gate type. We assume all the input patterns have the same probability and estimate every $A(v)$ for all the standard cells in the library. Considering two standard cells: NOR2XL and NAND4XL in the TSMC 0.18- μm standard cell library, the largest error is about

52% as shown in Table I. The error of linear approximation may be neglected in the *FGSTI* technique due to *law of large numbers* [25] with the growing circuit size. It will be disclosed in Section II-C that the influence of this linear model error on the *FGSTI* technique will be diminished by the large difference between leakage current of a gate with or without ST.

B. Delay Model

The gate delay is influenced by the ST insertion [15]. The load dependent delay $d_{w/o}(v)$ of gate v without ST is given by

$$d_{w/o}(v) = \frac{KC_L V_{DD}}{(V_{DD} - V_{THlow})^\alpha} \quad (3)$$

where C_L , V_{THlow} , α , and K are the load capacitance at the gate output, the low threshold voltage, the velocity saturation index, and the proportionality constant, respectively. The propagation delay $d_w(v)$ of gate v with ST can be expressed as

$$d_w(v) = \frac{KC_L V_{DD}}{(V_{DD} - 2V_x - V_{THlow})^\alpha} \quad (4)$$

where V_x is the V_{ds} of the ST. $\Delta d(v)$ is derived from the previous equations

$$\begin{aligned} \Delta d(v) &= d_w(v) - d_{w/o}(v) \\ &= \left(\left(1 - \frac{2V_x}{V_{DD} - V_{THlow}} \right)^{-\alpha} - 1 \right) d_{w/o}(v). \end{aligned} \quad (5)$$

$I_{ON}(v)$ is the current flowing through ST in gate v during the active mode, which can be expressed as given by [23]

$$\begin{aligned} I_{ON}(v) &= \mu_n C_{ox} (W/L)_v \left((V_{DD} - V_{THhigh}) V_x - \frac{V_x^2}{2} \right) \\ &= \mu_n C_{ox} (W/L)_v (V_{DD} - V_{THhigh}) V_x. \end{aligned} \quad (6)$$

Thus, the voltage drop V_x in gate v due to ST insertion can be expressed as

$$V_x = \frac{I_{ON}(v)}{\mu_n C_{ox} (V_{DD} - V_{THhigh})} \times \frac{1}{(W/L)_v}. \quad (7)$$

Refer to (3) and (4), V_x in gate v due to ST insertion can also be given out as

$$V_x = \frac{1}{2} \left(1 - \left(\frac{d_{w/o}(v)}{d_w(v)} \right)^{1/\alpha} \right) (V_{DD} - V_{THlow}). \quad (8)$$

TABLE II
LEAKAGE CURRENT COMPARISON OF STANDARD CELLS (fA). ($V_{DD} = 1.8$ V, $V_{THhigh} = 500$ mV FOR ST, $V_{THlow} = 300$ mV FOR LOGIC CELLS)

Cell Name	$I_{w/o}$	I_w	$I_{w/o}/I_w$	Cell Name	$I_{w/o}$	I_w	$I_{w/o}/I_w$
NAND2XL	14076.3	45.0	313	AND3X4	54900.3	53.4	1028
NAND2X4	84392.0	45.5	1854	BUFX4	80876.7	53.4	1513
INVXL	14213.2	36.9	388	NOR2X1	16261.1	27.6	516
NOR2XL	14606.8	27.6	528	CLKINVX4	38763.4	37.2	1043
XOR2XL	95853.9	53.4	1794	NAND4X4	72554.2	51.4	1411
NAND4XL	12261.2	51.4	239	AND2XL	26956.7	53.4	505
NAND3XL	14186.1	49.3	288	AND4XL	13768.6	53.4	258
AND2X4	60305.1	53.4	1129	OR4XL	33827.5	53.4	633
AND4X4	48899.1	53.4	915	CLKINV8	69175.0	37.7	1833

C. Rationality of Our Two-Phase FGSTI

From Table I, a linear leakage current model may have an error larger than 50% compared with the HSPICE simulation results. Referring to [23], the leakage current for a gate with ST is also modeled as a linear function from [26]

$$I_w(v) = \mu_n C_{ox} (W/L)_v e^{1.8V_T^2} \left(1 - e^{-\frac{V_{gs} - V_{THhigh}}{nV_T}} \right) \left(1 - e^{-\frac{V_{ds}}{V_T}} \right) \quad (9)$$

where μ is the N -mobility, C_{ox} is the oxide capacitance, V_{THhigh} is the high threshold voltage, V_T is the thermal voltage, n is the subthreshold swing parameter. Notice that their model is also linear by assuming all parameters except $(W/L)_v$ are constants which are decided by process information and gate structure. Such a linear model will also endorse comparable error as our leakage current model.

From Table I, the leakage current of a gate without ST is much larger than that of a gate with ST, so that the error of the linear model can be neglected in the *FGSTI* procedure. In Table II, we compare the leakage current of cells in the TSMC 0.18- μm standard cell library under two different ST conditions: with ST or without ST. Because the leakage current of a gate with ST becomes larger with a larger ST, in Table II, (W/L) of a ST is set to 16, which is the maximum ratio of ST in our *FGSTI* technique, in order to get the largest leakage current.

As shown in Table II, the leakage current difference under different ST conditions is at least 238 \times , because of the large difference between the threshold voltage of ST and logic cells. Referring to (5) and (7), the delay difference is less than 20% of the original gate delay between delays of a gate with and without ST. However, the delay difference of a gate with different ST sizes is much larger; for example, setting (W/L) of a ST to 1 will lead to about 140% additional delay comparing with the original gate without ST. Also, we can see from Table I, the leakage current difference of a gate with different ST size is less than 1% of the original gate leakage. Hence, the leakage current variation range due to the change of ST size can be neglected, because it is much smaller comparing with the leakage saving of changing a gate's ST condition. Therefore, ST placement is not affected by ST sizing owing to the large gap between their effects on leakage saving.

With technology scaling, the leakage current difference may be smaller under different ST conditions, but it will still be very large due to the use of high V_{TH} ST and the stacking effect. We can draw a conclusion that *the leakage reduction depends on where to insert ST and the leakage difference of each gate*

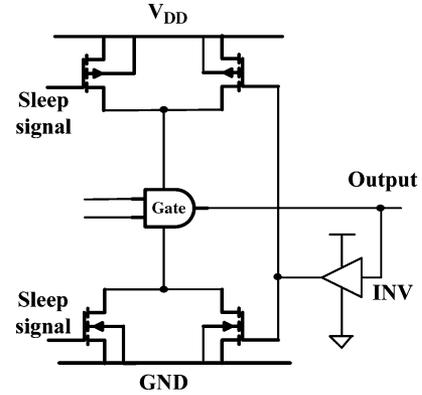


Fig. 2. Leakage feedback gate [21].

under different ST conditions; while the area penalty is decided by the ST sizing procedure.

We further assume that ST placement and sizing are independent in an *FGSTI* design. Therefore, a two-phase *FGSTI* technique is developed: first, ST placement is performed to decide which gate will be assigned with ST in order to achieve most of the leakage saving; and ST sizing is used to reduce the area overhead along with further leakage current reduction.

D. Leakage Feedback Gate

During the ST placement phase, when the circuit slowdown is not large enough to assign ST to every gate, the *FGSTI* technique can cause a gate with ST to drive a gate without ST. This will lead to floating state at the output of the gate with ST and may cause large power dissipation due to the short circuit current in the gate without ST. In this subsection, the circuit scheme of LF gate is first reviewed; and then a comparison is made between an LF gate and a normal ST gate to prove that a specialized LF gate can substitute for a normal ST gate.

1) *Circuit Scheme:* As mentioned in [23], the LF gate structure [21] shown in Fig. 2 should be used in order to avoid the floating states. The important characteristic of an LF gate is that depending on the state of the latest output, one but not both helper ST's (P_H or N_H), is turned on by the feedback inverter, thus, the output state of the LF gate are set to "1" or "0".

2) *Comparison With Normal ST Gate:* During the standby mode, both high V_{TH} ST's P_S and N_S are turned off, only one of the helper ST's will be kept on to drive the output signal to the appropriate rail. On the other hand, when the circuit is active, both high V_{TH} ST's P_S and N_S are turned on. One and only one of the helper ST's will be turned on to accelerate the circuit

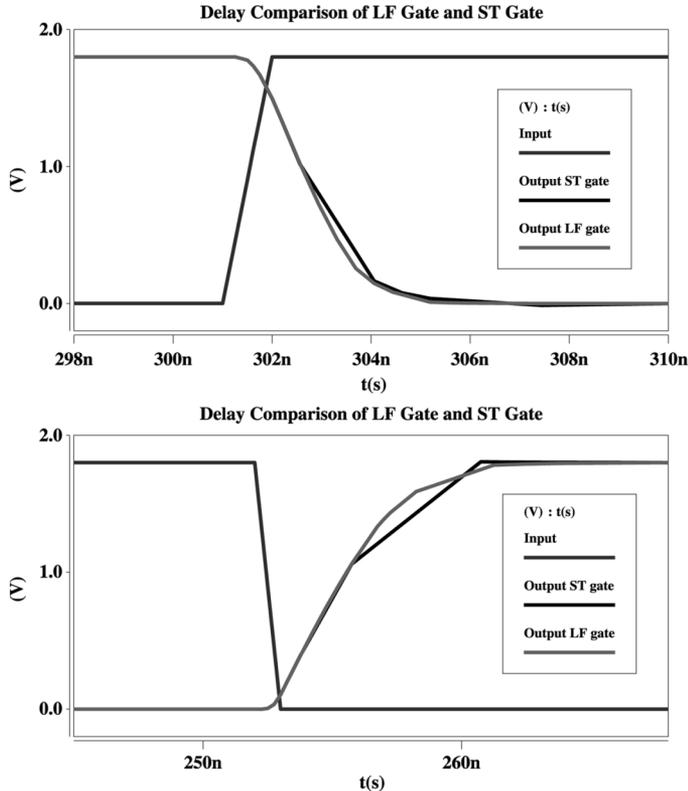


Fig. 3. Delay comparison: an LF gate and a normal gate with ST.

speed, because the feedback inverter is sensitive to the change of the output signal.

The signal propagation delay of an inverter with ST and an LF gate for an inverter are compared under same load capacitance and shown in Fig. 3. The sizes of helper ST's are the same as those of the original sleep transistors. As we can see, the rise and fall slope of an LF gate is steeper than that of a normal ST gate. Therefore, we conclude that every gate with ST can be replaced with a carefully sized LF gate without affecting the circuit delay constraints.

An LF gate will certainly bring two extra helper ST's and a feedback inverter, hence it leads to area and dynamic power overhead. Since all the extra transistors can be high threshold transistors, the leakage overhead can be neglected.

III. TWO-PHASE *FGSTI* TECHNIQUE

In this section, our two-phase *FGSTI* technique is modeled using linear programming methods. First, we show how to place the ST as many as possible in order to reduce the total leakage; an extended MILP model considering LF gate is also proposed. Then an optimal sizing method is proposed to reduce the area overhead based on the ST placement information from the first phase. At the end of this section, we briefly review the simultaneous placement and sizing method [24] using MILP for comparison.

A. Phase I: ST Placement

1) *MILP Model*: We propose a novel ST placement method that tries to maximize the leakage saving in the circuits through MILP model.

First, we construct the object function which is the total leakage current as follows:

$$I_{\text{leak}}(G) = \sum_{v \in V} (I_{w/o}(v) \times (1 - ST(v)) + I_w(v) \times ST(v)) \quad (10)$$

where $ST(v)$ is a binary variable to represent gate v 's ST condition, $ST(v) = 1$ means gate v has ST inserted and $ST(v) = 0$ means gate v is without ST. As ST size is not considered, we choose the largest ST size $(W/L)_{\text{max}}$ to obtain the minimum delay overhead. The leakage current of gate v with ST is given by

$$I_w(v) = A(v) \times (W/L)_{\text{max}}. \quad (11)$$

The timing constraints of $G(V, E)$ can be expressed as

$$t_a(m) = 0 \quad m \in PI \quad (12)$$

$$t_a(n) + d(n) \leq T_{\text{req}} \quad n \in PO \quad (13)$$

$$t_a(i) + d(i) \leq t_a(j) \quad \forall (i, j) \in E, i, j \in V \quad (14)$$

where PI and PO refer to the primary input and primary output gates of the circuit; $t_a(v)$ represents the arrival time of gate v , T_{req} is the overall circuit delay; $d(v)$ represents the gate delay, which can be expressed as referring to (5) and (7)

$$\begin{aligned} d(v) &= d_{w/o}(v) + \Delta d(v) \times ST(v) \\ &= d_{w/o}(v) \\ &\quad + \left(\left(1 - \frac{\frac{2I_{\text{ON}}(v)}{\mu_n C_{\text{ox}}(V_{\text{DD}} - V_{\text{THhigh}})} \times \frac{1}{(W/L)_{\text{max}}}}{V_{\text{DD}} - V_{\text{THlow}}} \right)^{-\alpha} - 1 \right) \\ &\quad \times d_{w/o}(v) \times ST(v) \\ &= d_{w/o}(v) + \varphi((W/L)_{\text{max}}) \times d_{w/o}(v) \times ST(v) \end{aligned} \quad (15)$$

where $d_{w/o}(v)$ is a constant, which is extracted from the technology library. As $(W/L)_{\text{max}} = 16$, $\varphi((W/L)_{\text{max}})$ is also a constant for each gate.

ST placement phase is similar to dual V_{TH} assignment with fixed high and low V_{TH} values, thereby it can also be solved by sensitive-based heuristic algorithms which are previously used to deal with dual V_{TH} assignment [9]–[11].

2) *MILP Model Considering LF Gate*: As we mentioned before, when the circuit slowdown is not large enough to assign ST to each gate in the circuit, LF gates are used to avoid floating states. The LF gate number may exceed as much as 80% of the gates with ST when the effect of LF gate is not considered in *FGSTI* technique. The additional inverters in the LF gates will induce large area and dynamic power penalty. Thus, LF gate number is as important as the leakage reduction rate. In this subsection, an extended MILP model to simultaneously maximize the leakage saving in the circuits and minimize the LF gate number is proposed.

First, the original object function for ST placement (10) is amended to consider the LF gate number N_{LF}

$$\begin{aligned} I_{\text{leak}}(G) + \gamma N_{\text{LF}}(G) &= \sum_{v \in V} (I_{w/o}(v) \times (1 - ST(v)) \\ &\quad + I_w(v) \times ST(v)) + \gamma \sum_{v \in V} (\text{LF}(v)) \end{aligned} \quad (16)$$

Minimize:

$$I_{leak}(G) + \gamma N_{LF}(G) \\ = \sum_{v \in V} (I_{w/o}(v) \times (1 - ST(v)) + I_w(v) \times ST(v)) + \gamma \sum_{v \in V} (LF(v))$$

Subject to:

{Timing constraints}

$$t_a(m) = 0 \quad m \in PI \\ t_a(n) + d(n) \leq T_{req} \quad n \in PO \\ t_a(i) + d(i) \leq t_a(j) \quad \forall (i, j) \in E, i, j \in V \\ d(v) = d_{w/o}(v) + \varphi((W/L)_{max}) \times d_{w/o}(v) \times ST(v) \quad v \in V$$

{Variable constraints}

$$ST(v) \text{ and } LF(v) \text{ are binary variables} \\ LF(i) \geq ST(i) - ST(j), \forall (i, j) \in E, i, j \in V$$

Fig. 4. MILP model for multi-object ST placement.

where I_{leak} is the total leakage current; N_{LF} is the LF gate number in the circuit; $LF(v)$ is also a binary variable to represent gate v 's LF gate condition, $LF(v) = 1$ means gate v is an LF gate and $LF(v) = 0$ means gate v is not an LF gate; γ is a weight value that can be modified by the circuit designer. The performance and delay constraints are still the same with the MILP model without considering LF gate, because of the assumption in Section II-D, every gate with ST can be changed into a carefully sized LF gate without affecting the circuit delay constraints.

A gate v must be changed into LF gate if $ST(v) = 1$ and one of its fan-out gate is a gate without ST. Thus the binary variable $LF(v)$ should satisfy the following constraint:

$$LF(i) \geq ST(i) - ST(j) \quad \forall (i, j) \in E, i, j \in V. \quad (17)$$

If γ is not zero, the LF gate is considered during the optimization; on the other hand, if γ is zero, the LF gate is not considered, the variable and constraints related to $LF(v)$ can be deleted in the model. To sum up, the general form of our MILP model for ST placement phase is shown in Fig. 4.

B. Phase II: Optimal ST Sizing

After ST condition for each gate v is decided, the optimal ST sizing is derived using a linear programming model. The object function for optimal ST sizing is given as follows:

$$\text{Area}(ST) = \sum_{v \in V} ((W/L)_v \times ST(v)) \quad (18)$$

where $ST(v)$ is a binary value decided in the first phase: ST placement phase; $(W/L)_v$ is a continuous variable. Moreover, the expression for $(W/L)_v$ from (7) and (8) can be derived as follows:

$$\begin{aligned} (W/L)_v &= \frac{I_{ON}(v)}{\mu_n C_{ox}(V_{DD} - V_{THhigh})} \times \frac{1}{V_x} \\ &= \frac{I_{ON}(v)}{\mu_n C_{ox}(V_{DD} - V_{THhigh})} \\ &\quad \times \left(\frac{1}{2} \left(1 - \left(\frac{d_{w/o}(v)}{d_w(v)} \right)^{1/\alpha} \right) \right. \\ &\quad \left. \times (V_{DD} - V_{THlow}) \right)^{-1}. \end{aligned} \quad (19)$$

Minimize:

$$\text{Area}(ST) = \sum_{v \in V} \left(\left(\frac{1}{2} \left(1 - \left(\frac{d_{w/o}(v)}{d_w(v)} \right)^{1/\alpha} \right) (V_{DD} - V_{THlow}) \right)^{-1} \times ST(v) \right)$$

Subject to:

{Timing constraints}

$$t_a(m) = 0 \quad m \in PI \\ t_a(n) + d(n) \leq T_{req} \quad n \in PO \\ t_a(i) + d(i) \leq t_a(j) \quad \forall (i, j) \in E, i, j \in V \\ d(v) = d_{w/o}(v) + (d_w(v) - d_{w/o}(v)) \times ST(v) \quad v \in V$$

{Variable constraints}

$$d_w^{min}(v) \leq d_w(v) \leq d_w^{max}(v)$$

Fig. 5. LP model for optimal ST sizing.

The timing constraints can also be expressed as (12)–(14). The propagation delay $d_w(v)$ of gate v with ST can be rewrite using (5) and (7) as

$$\begin{aligned} d_w(v) &= d_{w/o}(v) + \Delta d(v) \\ &= d_{w/o}(v) \\ &\quad + \left(\left(1 - \frac{\frac{2I_{ON}(v)}{\mu_n C_{ox}(V_{DD} - V_{THhigh})} \times \frac{1}{(W/L)_v}}{V_{DD} - V_{THlow}} \right)^{-\alpha} - 1 \right) \\ &\quad \times d_{w/o}(v) \\ &= d_{w/o}(v) + \varphi((W/L)_v) \times d_{w/o}(v). \end{aligned} \quad (20)$$

With a given boundary of $(W/L)_v$: $[(W/L)_{min}, (W/L)_{max}]$, we can easily derive the boundary of $d_w(v)$: $[d_w^{min}, d_w^{max}]$ using (20). Consequently, the general form of our LP model for ST sizing is shown in Fig. 5.

C. Simultaneous ST Placement and Sizing

In this subsection, the simultaneous ST placement and sizing method [24] is briefly reviewed. The object function is very similar to ST placement as shown in (10)

$$I_{leak}(G) = \sum_{v \in V} (I_{w/o}(v) \times (1 - ST(v)) + A(v) \times (W/L)_v \times ST(v)) \quad (21)$$

where $ST(v)$ and $(W/L)_v$ are variables that decide where to put ST and how to size ST, respectively. The timing constraints also follow (12)–(14). Referring to (15), gate delay $d(v)$ for gate v can be derived as

$$\begin{aligned} d(v) &= d_{w/o}(v) + \Delta d(v) \times ST(v) \\ &= d_{w/o}(v) \\ &\quad + \left(\left(1 - \frac{\frac{2I_{ON}(v)}{\mu_n C_{ox}(V_{DD} - V_{THhigh})} \times \frac{1}{(W/L)_v}}{V_{DD} - V_{THlow}} \right)^{-\alpha} - 1 \right) \\ &\quad \times d_{w/o}(v) \times ST(v) \\ &= d_{w/o}(v) + d_{w/o}(v) \times \varphi((W/L)_v) \times ST(v). \end{aligned} \quad (22)$$

As we can see from (21) and (22), this problem is actually a nonlinear programming model. In [24], Taylor series expansion

and piece wise linear approximation technique are used to get an MILP model. Some dummy variables are needed for linear approximation and corresponding linearization constraints are added in the MILP model for each dummy variable. Unfortunately, the model size becomes extremely large with the increasing gate number in the circuit.

IV. HEURISTIC ALGORITHM FOR ST PLACEMENT

In this section, a heuristic algorithm is introduced to solve the ST placement phase explained in Section III-A. One of the major bottlenecks of using MILP method is the unpredictable computation time when the circuit slowdown is not large enough to assign ST to every gate. Although the MILP model leads to optimal result, it can not be used in a real design cycle because of the unaffordable runtime cost for reiterations. A fast and accurate heuristic algorithm is needed to speed up the ST placement phase with a near optimal result. The LF gate number depends mainly on the ST placement information and the topology of the circuit; while the leakage reduction rate depends on ST gate number under the circuit performance constraints. It is time consuming to consider both ST gate number and LF gate number simultaneously during ST placement phase.

In our heuristic algorithm, the DAG of the circuit is pruned to reduce the problem size in the first step; then, in the second step, a greedy algorithm is used to assign ST as many as possible without affecting the performance constraints; finally, in the third step, the LF gate number is reduced according to different weight value: γ in (16).

A. Step I: DAG Pruning

The MILP model size for ST placement is decided by the gate number and the interconnect number in the circuit, that is, the vertex number and the edge number in the DAG. Hence, the DAG is first pruned in our heuristic algorithm to reduce the problem size. This problem is actually how to find the gates which must be assigned with ST in the circuit, so that the corresponding vertexes can be deleted from the DAG.

Definition 1 (Signal Path): Signal path in a DAG $G = (V, E)$ from a vertex u to a vertex u' is a sequence $\langle v_1, v_2, \dots, v_k \rangle$ of vertexes, such that $u = v_1 \in PI$, $u' = v_k \in PO$ and $(v_{i-1}, v_i) \in E$ for $i = 1, 2, \dots, k$.

Definition 2 (Critical Path Delay): If ST is assigned to each gate, critical path delay of vertex v , $T_{\text{path}}(v)$, in a DAG $G = (V, E)$, is the longest path delay of all the signal paths which contain vertex v .

If ST is assigned to each gate, the leakage current of the circuit is minimized, and the overall delay T_{circuit} will exceed the requested delay T_{req} . However, there may be some signal paths in the circuit still satisfy the performance constraints. Some gates in such signal paths can be deleted without affecting the timing constraints. That is to say, if $T_{\text{path}}(v)$ in the circuit still do not exceed T_{req} when ST is assigned to each gate in the circuit. Gate v does not affect the circuit performance constraints whether gate v is assigned with ST or not.

$T_{\text{path}}(v)$ can be derived directly according to Definition 2: 1) calculate delay of all the signal paths which contain vertex

v and 2) the largest delay of these signal paths is picked up to be $T_{\text{path}}(v)$. However, this direct computation is impractical, because it is very complicated to find all the signal paths which contain vertex v .

Lemma 1 (Subpaths of Largest Delay Paths are Largest Delay Paths): Given a DAG $G = (V, E)$, let $p = \langle v_1, v_2, \dots, v_k \rangle$ be the largest delay path from vertex v_1 to vertex v_k and, for any i and j such that $1 \leq i \leq j \leq k$, let $p_{ij} = \langle v_i, v_{i+1}, \dots, v_j \rangle$ be the subpath of p from vertex v_i to vertex v_j ; then p_{ij} is the largest delay path from vertex v_i to vertex v_j .

Proof: Path p can be decomposed into three parts: $v_1 \xrightarrow{p_{1i}} v_i \xrightarrow{p_{ij}} v_j \xrightarrow{p_{jk}} v_k$, thus, the path delay of p can be denoted as $\text{Delay}(p) = \text{Delay}(p_{1i}) + \text{Delay}(p_{ij}) + \text{Delay}(p_{jk})$. Assuming there is a path p'_{ij} from v_i to v_j with $\text{Delay}(p'_{ij}) > \text{Delay}(p_{ij})$, then there is a path $p'v_1 \xrightarrow{p_{1i}} v_i \xrightarrow{p'_{ij}} v_j \xrightarrow{p_{jk}} v_k$ with total delay $\text{Delay}(p') = \text{Delay}(p_{1i}) + \text{Delay}(p'_{ij}) + \text{Delay}(p_{jk}) > \text{Delay}(p)$, so that contradicts the assumption that p is a largest delay path from vertex v_1 to vertex v_k . ■

Referring to Lemma 1, the critical path delay of gate v : $T_{\text{path}}(v)$, essentially, consists of three parts

$$T_{\text{path}}(v) = \text{Max}(\text{Delay}(PI, v)) + \text{Max}(\text{Delay}(v, PO)) + d_w(v) \quad (23)$$

where $\text{Max}(\text{Delay}(PI, v))$ is the maximum delay of the all possible paths from primary inputs to v ; $\text{Max}(\text{Delay}(v, PO))$ is the maximum delay of all the possible paths from v to primary outputs. As we all know, the arrival time of v : $t_a(v)$ is defined as the worst case delay from primary inputs to vertex v : $\text{Max}(\text{Delay}(PI, v))$

$$t_a(v) = \begin{cases} 0, & v \in PI \\ \max_{i \in \text{fanin}(v)} \{t_a(i) + d(i)\}, & \text{otherwise} \end{cases} \quad (24)$$

where $d(i)$ is the delay of vertex i shown in (15).

Furthermore, in a DAG, once a path p_1 is chosen in path set from vertex v to PO , we can obtain a corresponding path p_2 in a reverse order by simply reversing the vertex sequence order. p_2 belong to path set from PO to vertex v , and $\text{Delay}(p_1) = \text{Delay}(p_2)$, $\text{Max}(\text{Delay}(v, PO)) = \text{Max}(\text{Delay}(PO, v))$.

Therefore, $T_{\text{path}}(v)$ can be expressed as

$$T_{\text{path}}(v) = t_a(v) + t_{ba}(v) + d_w(v) \quad (25)$$

where $t_a(v)$ is the arrival time of v ; $d_w(v)$ is the delay of gate v with ST; $t_{ba}(v)$ is the reverse arrival time of vertex v , and can be derived as $t_a(v)$ of a DAG in a reverse direction. The arrival time and the reverse arrival time can be simply derived by Breadth-first search (BFS) [27].

Theorem 1: Given a DAG $G = (V, E)$ with timing information, if $T_{\text{path}}(v) < T_{\text{req}}$, vertex v can be deleted in the DAG, which means that v will not affect the performance constraints.

Proof: Assuming v will affect the performance constraints, that means the delay of at least one signal path $SP(i, j)$ contain gate v is larger than T_{req}

$$\text{Delay}(p_{iv}) + \text{Delay}(p_{vj}) + d_w(v) > T_{\text{req}}. \quad (26)$$

DAG Pruning Algorithm
 1. Set $ST(v) = 1, v \in V$
 2. Calculate $T_{path}(v) = t_a(v) + t_{ba}(v) + d_w(v)$,
 $v \in (V - PI - PO)$
 3. $G_{ker} = (V_{ker}, E_{ker}), V_{ker} = \{v | T_{path} \geq T_{req}, v \in V\}$

Fig. 6. DAG pruning algorithm.

Referring to (25)

$$\begin{aligned} T_{req} > T_{path}(v) \\ &= \text{Max}(\text{Delay}(PI, v)) + \text{Max}(\text{Delay}(v, PO)) + d_w(v) \\ &> \text{Delay}(p_{iv}) + \text{Delay}(p_{vj}) + d_w(v). \end{aligned} \quad (27)$$

Thus, (26) and (27) are contradiction according the deduction above, this fact demonstrates the theorem. ■

Apparently, the vertex set V can be divided into two parts: $V' = \{v | T_{path}(v) < T_{req}, v \in V\}$ and $V_{ker} = \{v | T_{path}(v) \geq T_{req}, v \in V\}$. $ST(v) \equiv 1$ if $v \in V'$. Therefore, the original DAG $G = (V, E)$ is pruned to $G_{ker} = (V_{ker}, E_{ker})$ in which the ST condition of each vertex needs to be decided. The DAG Pruning Algorithms is shown in Fig. 6.

B. Step II: ST Assignment

After DAG pruning, we should find a way to assign ST in G_{ker} as many as possible while minimizing the LF gate number. Every time $ST(v)$ is switched from 0 to 1 or from 1 to 0, the potential LF gate number should be recalculated, which will greatly slow down the optimization procedure. Thus, it is rather time consuming to optimize the LF gate number during ST assignment. One intuitional solution to tackle this difficulty is to separate the optimization of ST gate number and LF gate number. The ST assignment procedure is carried out first to assign as many ST as possible to G_{ker} .

The ST assignment problem is how to minimize the leakage current through assigning ST to gates in G_{ker} while satisfying the performance constraints. During our ST assignment algorithm, all $ST(v)$'s for gates in G_{ker} are first set to 1. The longest delay path is then picked out; $ST(v)$ for some typical gates in this path are set back to zero in order to reduce the path delay. This step iterates until all the signal paths in the G_{ker} satisfy the performance constraints. A near optimal ST assignment solution can be derived from our algorithm.

Intuitionally, the LF gate is caused by the “1 → 0” mode (which means a gate without ST is driven by a gate with ST, 1 and 0 are corresponding value for $ST(v)$). If the set of gates with $ST(v) = 0$ and the set of gates with $ST(v) = 1$ only have one single boundary, the LF gate number will be definitely small. For example, on a certain path, the continuous $ST(v)$ pattern $1 \rightarrow 0 \rightarrow 1 \rightarrow 0 \rightarrow 1$ (four boundaries, 2 LF gates) or $1 \rightarrow 0 \rightarrow 0 \rightarrow 1 \rightarrow 1$ (two boundaries, 1 LF gate) cause more LF gates than the pattern $0 \rightarrow 0 \rightarrow 1 \rightarrow 1 \rightarrow 1$ (single boundary, without LF gate). Furthermore, there are two different conditions for single boundary. For example, there may be two continuous $ST(v)$ patterns: $0 \rightarrow 0 \rightarrow 1 \rightarrow 1 \rightarrow 1$ and $1 \rightarrow 1 \rightarrow 1 \rightarrow 0 \rightarrow 0$; both of them have only one boundary, however, the LF gate number is not the same. The former one will lead to no LF gate while the other one will cause one LF gate. It can be inferred that, if a gate with $ST(v) = 0$ is placed

ST Assignment Algorithm
 1. Set $ST(v) = 1, v \in V_{ker}$
 2. Find a critical path $p = \langle v_1, v_2, \dots, v_k \rangle$ with delay $T_{critical}$
 If $T_{critical} > T_{req}$
 $\Delta T = T_{critical} - T_{req}$
 /*Forward Selection*/
 $i = 1$
 While $\Delta T > 0$
 If $ST(v_i) = 1$
 $ST(v_i) = 0; \Delta T = \Delta T - d_w(v_i) + d_{w/o}(v_i)$
 $i = i + 1$
 /*End of Forward Selection*/
 jump to step 2
 Else
 get $ST(v)$ for $v \in V_{ker}$
 3. Combine with $ST(v) = 1, \forall v \in V'$

Fig. 7. ST assignment algorithm.

on the input node when there is only one boundary, there is no LF; it enlightens us to place the gate with $ST(v) = 0$ close to each other and close to the input vertex to reduce the potential LF gates during the ST assignment step.

In ST assignment, Forward Selection method is used to select gates from a longest signal path: assign gate with $ST(v) = 0$ close to input vertex. Fig. 7 shows the ST assignment algorithm, the Forward Selection part can be easily changed into other selection method, such as: random selection (randomly select gates from the signal path).

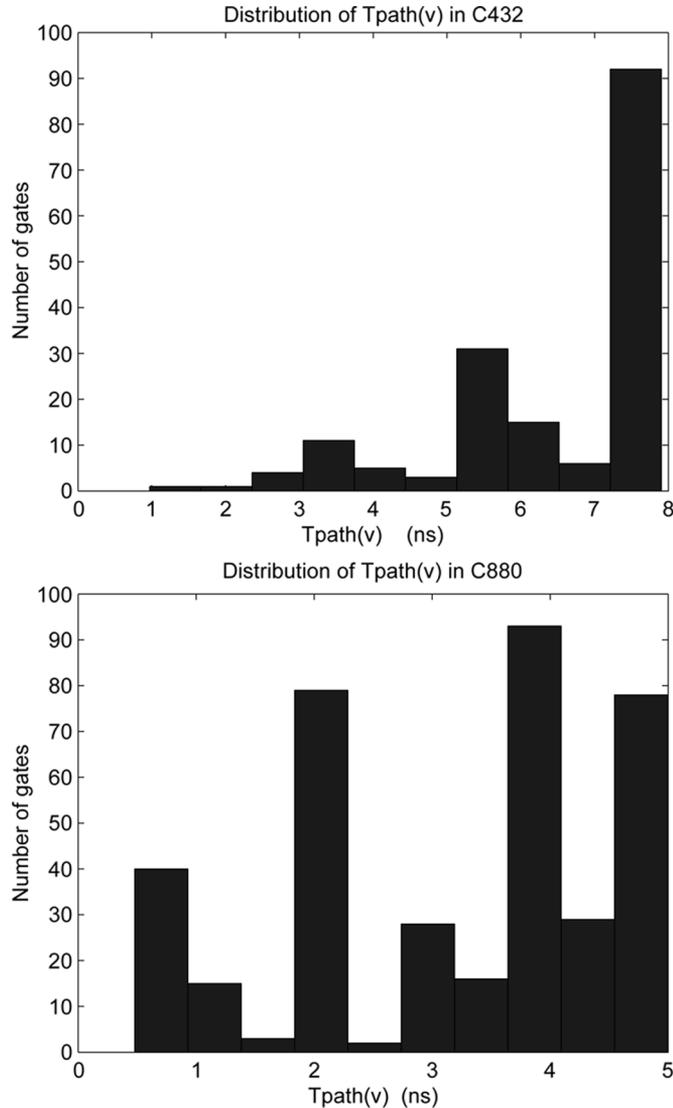
The $T_{path}(v)$ distribution graph can give us an intuitional evaluation about the percentage of gates that can be optimized in a given circuit. For example, in Fig. 8, $T_{path}(v)$ distribution graph for C432 shows that a high percentage of gates are with large $T_{path}(v)$, it can be inferred that the total gates with ST in the final optimal result will be small; on the contrary, $T_{path}(v)$ distribution graph for C880 shows that a small percentage of gates are with large $T_{path}(v)$, so that the total number of gates with ST in the final optimal result will be large. These deductions are well supported by the simulation results shown in Section V.

C. Step III: LF Reduction

After we have the ST assignment information, an LF reduction algorithm is performed following the different weight value: γ in (16). First, we run a breadth-first search in DAG from PI , for each vertex v with $ST(v) = 1$, if there is a fan-out gate u of vertex v with $ST(u) = 0$, $LF(v)$ is set to 1 and gate v is add into queue Q_{LF} . We examine each LF gate in Q_{LF} to find out whether it is worth to change this LF gate back into a gate without ST. The weight value γ can be considered as a kind of “leakage current overhead”. Thus, (16) is rewritten as

$$\begin{aligned} I'_{leak} = \sum_{v \in V} (I_{w/o}(v) \times (1 - ST(v)) + (\gamma + I_w(v)) \\ \times LF(v) + I_w(v) \times (ST(v) - LF(v))). \end{aligned} \quad (28)$$

If one LF gate v is changed back into gate without ST, the LF gate distribution in the area looking back from gate v to PI should be re-examined. There may be some gates with ST should be changed back into gate without ST to avoid additional LF gates in this fan-shaped part of DAG which connected with

Fig. 8. $T_{\text{path}}(v)$ distribution in C432 and C880.

gate v . This operation leads to additional leakage current. Meanwhile, some of these gates with ST are LF gates; the change of LF gates back into gates without ST will also lead to “ γ ” leakage current saving. Therefore, we define another weight value for each LF gate as

$$\Delta P(v) = \sum_{i \in V_1(v)} (I_{w/o}(i) - I_w(i)) - \gamma \sum_{j \in V_2(v)} \text{LF}(j) \quad (29)$$

where $V_1(v)$ is the vertex set in which should be changed back into gate without ST due to the change of gate v ; $V_2(v)$ belongs to $V_1(v)$, and all gates in $V_2(v)$ are LF gates. Notice that, gate v belongs to both $V_1(v)$ and $V_2(v)$. $V_1(v)$ and $V_2(v)$ are derived by examining each gate during breadth-first search from v back to PI . The LF Reduction Algorithm is shown in Fig. 9.

V. IMPLEMENTATION AND EXPERIMENTAL RESULTS

A. Implementation

All ISCAS’85 benchmark circuit netlists are synthesized using Synopsys Design Compiler and a TSMC 0.18- μm

LF Reduction Algorithm

1. Run BFS, add each LF gate in Queue Q_{LF}
2. For each LF gate v in Q_{LF} ,
 - If $\Delta P > 0$
 - $LF(v)$ remains 0
 - Else
 - $ST(i) = 0, LF(i) = 0, i \in V_1(v)$
3. Calculate N_{LF} and I_{leak}

Fig. 9. LF reduction algorithm.

standard cell library. A leakage current look up table of all the standard cells without ST is generated using HSPICE. In addition, every $A(v)$ in (2) for all the standard cells is estimated using the HSPICE simulation results under different $(W/L)_v$. The values of various transistor parameters have been taken from the TSMC 0.18- μm process library. $V_{DD} = 1.8$ V, $V_{TH\text{high}} = 500$ mV, $V_{TH\text{low}} = 300$ mV, and $I_{ON} = 200$ μA are set for all the gates in the circuit. The timing constraints are set up with a static timing analysis (STA) tool [10], and the MILP and LP models for ST placement phase and ST sizing phase are automatically generated. An LP solver named *lp_solve*¹ is used to solve the models. The heuristic algorithm for ST placement phase is implemented using MATLAB. The simulations are conducted on a 1.83 GHz CPU, 1.5 G memory computer.

We assume $1 \leq (W/L)_v \leq 16$, corresponding to a least delay variance of 6% if ST is assigned to all the gates in the circuit. We perform our two-phase *FGSTI* technique by first using the MILP model to get value of $ST(v)$ for all the gates in the circuit and then solving the LP model to get the optimal $(W/L)_v$ based on the results of $ST(v)$. The MILP model to simultaneously determine ST placement and sizing are also solved using the same LP solver under the same set of parameters in order to compare the results with our two-phase *FGSTI* technique. At last our heuristic algorithms for ST placement phase are performed to show the runtime merit over MILP method with an acceptable loss of accuracy.

B. MILP Model for ST Placement Phase

1) *ST Placement Without Considering LF Gate*: For 0%, 3%, 5% circuit slowdown, a valid solution can not be derived from conventional fixed slowdown method. Thus, the leakage current saving for 0%, 3%, 5% circuit slowdown are compared between our two-phase *FGSTI* technique and MILP method [24]. As shown in Table III, our two-phase *FGSTI* technique can achieve 78.91% leakage saving even if the circuit slowdown is not influenced. When the circuit slow down is 3%, 5%, the leakage saving of our two-phase *FGSTI* technique is 92.55%, 97.97%, respectively. Because of less approximation in MILP model of ST placement phase, more leakage saving is achieved comparing with the simultaneous ST placement and sizing method [24]. The leakage saving is about on average 2% more than the MILP method.

In Table IV, we show that our two-phase *FGSTI* technique for some circuits can achieve impressive runtime savings compared with the simultaneous ST placement and sizing method [24]. The runtime saving is largely caused by two reasons: one is the two-phase procedure of *FGSTI* technique and the other is less

¹[Online]. Available: http://groups.yahoo.com/group/lp_solve/

TABLE III
LEAKAGE CURRENT COMPARISON BETWEEN TWO-PHASE *FGSTI* AND MILP METHOD [24] (pA)

ISCAS85 Benchmark Circuits	Original I_{leak}	Total Gate Num.	0% circuit slowdown				3% circuit slowdown				5% circuit slowdown			
			Two-phase <i>FGSTI</i>		MILP		Two-phase <i>FGSTI</i>		MILP		Two-phase <i>FGSTI</i>		MILP	
			I_{leak}	ST	I_{leak}	ST	I_{leak}	ST	I_{leak}	ST	I_{leak}	ST	I_{leak}	ST
C432	4609.4	169	1759.3	130	1964.9	127	463.7	151	463.7	151	205.5	157	205.5	157
C499	21374.9	204	14479.8	97	14587.3	101	805.9	189	1451.8	164	105.3	200	757.4	189
C880	9261.3	383	619.2	352	619.2	352	232.2	370	364.9	365	126.1	375	126.1	375
C1355	11874.5	548	6417.4	308	6712.3	287	5099.4	402	5220.4	386	945.2	535	4382.8	417
C1908	23418.2	911	2498.8	830	3177.8	831	590.0	878	1296.4	882	224.5	900	258.2	900
C2670	35191.3	1279	1356.6	1235	1382.0	1235	364.8	1264	667.6	1260	161.9	1274	269.8	1270
C3540	40369.7	1699	2060.4	1617	2251.2	1612	1020.4	1658	1558.7	1637	270.4	1690	611.1	1675
C5315	56292.2	2329	1660.9	2253	1841.9	2254	788.6	2293	1025.5	2283	433.8	2312	593.6	2305
C6288	40968.8	2447	7427.8	1948	8083.8	1903	2545.6	2282	3042.1	2248	977.7	2385	1088.5	2382
C7552	85523.9	3566	3012.4	3415	4190.7	3385	1320.2	3504	2004.9	3471	682.4	3539	975.9	3519
Saving	N/A	N/A	78.91%	N/A	77.49%	N/A	92.55%	N/A	91.24%	N/A	97.97%	N/A	94.55%	N/A
Additional Leakage Saving: (MILP - Two-phase)/MILP			6.31%				14.95%				62.75%			

TABLE IV
RUNTIME COMPARISON BETWEEN TWO-PHASE *FGSTI* AND MILP METHOD [24] (TIME IN SECONDS)

ISCAS85 Benchmark Circuits	0% circuit slowdown						3% circuit slowdown						5% circuit slowdown					
	Two-phase <i>FGSTI</i>			MILP	Two-phase <i>FGSTI</i>			MILP	Two-phase <i>FGSTI</i>			MILP						
	Placement	Sizing	Total		Placement	Sizing	Total		Placement	Sizing	Total							
C432	0.49	1.23	1.73	32.05	1.50	1.59	3.10	1905.09	0.55	1.63	2.18	635.02						
C499	3.86	1.45	5.31	75.0	2400.80	2.39	2403.19	154825.17	23.30	2.94	26.24	99610.59						
C880	0.35	7.62	7.97	134.11	0.50	6.58	7.08	2973.73	0.321	6.34	6.67	958.77						
C2670	43.61	107.80	151.41	22121.92	42.53	221.73	264.27	27438.11	4.25	176.45	180.70	57462.64						

TABLE V
COMPARISON BETWEEN MULTI-OBJECTIVE ST PLACEMENT (M-STP) AND ST PLACEMENT WITHOUT CONSIDERING THE LF GATE (STP-WO)

ISCAS85 Bench Circuits	0% circuit slowdown				3% circuit slowdown				5% circuit slowdown			
	I_{leak}		LF/ ST		I_{leak}		LF/ ST		I_{leak}		LF/ ST	
	M-STP	STP-WO	M-STP	STP-WO	M-STP	STP-WO	M-STP	STP-WO	M-STP	STP-WO	M-STP	STP-WO
C432	2954.6	1759.3	10/59	37/130	1243.37	463.7	4/111	31/151	586.6	205.5	3/132	19/157
C499	15340.9	14479.8	25/71	78/97	1437.9	805.9	12/159	48/189	636.4	105.3	6/182	32/200
C880	1292.3	619.2	11/315	105/352	541.4	232.2	6/359	68/370	255.7	126.1	3/370	46/375
C1355	7827.5	6417.5	40/248	206/308	5604.5	5099.4	51/381	134/402	1570.5	945.2	8/512	64/535
C1908	4535.9	2498.8	24/782	261/830	1040.2	590.0	8/862	171/878	357.6	224.5	5/895	83/900
C2670	1900.8	1356.6	22/1214	293/1235	833.1	364.6	3/1248	193/1264	328.0	161.8	2/1265	98/1274
C3540	2680.0	2060.4	41/1588	368/1617	1686.9	1020.4	11/1628	246/1658	521.8	270.4	4/1683	122/1690
C5315	7122.1	1660.9	14/2193	428/2253	5191.1	788.6	9/2270	294/2293	4284.5	433.8	7/2302	148/2312
C6288	10371.5	7427.8	170/1752	800/1948	3709.7	2545.6	81/2203	485/2282	1598.1	977.7	29/2350	213/2385
C7552	3877.0	3012.4	120/3378	937/3415	2067.8	1320.2	37/3461	577/3504	945.6	682.4	21/3524	256/3539
Average	71.0%	78.9%	9.3%	37.1%	88.2%	92.5%	3.3%	19.8%	95.2%	98.0%	1.1%	9.9%

variables and constraints used in MILP model for ST placement. For example, in circuit C432, there are only 271 constraints and 338 variables in our MILP model for ST placement; however, in [24], the MILP model has 2975 constraints and 1183 variables. Although MILP problems need a long time to solve, some of the benchmark, especially the small ones, at least $10\times$ runtime saving can be achieved using our two-phase *FGSTI* method. We only list the results of four benchmarks, because other benchmarks take hours to get the optimal results. The stopping time criteria is set to four hours for larger circuits.

2) *ST Placement Considering LF Gate*: As we mentioned before, $(W/L)_v = 16$ corresponds to a delay variance of 6% if ST is assigned to all the gates in the circuit [24]. Thus, when the circuit slowdown varies in the range of 6% circuit original delay, ST cannot be assigned to every gate in the circuit. The LF gate should be used when a gate with ST is driving a gate without ST. The results of our multi-object ST placement (M-STP) and the ST placement without considering the LF gate (STP-WO) are compared in Table V. The weight value γ is assumed to be 100.

In Table V, if the LF gate is not considered during ST placement, on average 37.1% of the gates with ST should be changed into leakage feed back structure if there is no circuit slowdown. When circuit slowdown is 3% and 5%, on average 19.8% and 9.9% of the gate with ST should be changed into LF gate, respectively. When the circuit slowdown is 0%, some of the benchmarks, such as C499, C1355, need to change 80.4% and 66.4% of normal ST gates into LF gates. This will lead to a large area increasing due to large number of high V_{TH} feedback inverters and help ST's. As the LF gate is considered during the multi-object ST placement, the LF gate number is about 9.3%, 3.3%, and 1.1% of the total gates with ST when the circuit slowdown is 0%, 3%, and 5%, respectively. Meanwhile, the difference of leakage reduction rate is only 7.9%, 4.3%, and 2.8%. For the two typical benchmarks mentioned previously: C499 and C1355, the LF gate becomes 35.2% and 16.1% of the gates with ST, respectively, when the circuit slowdown is 0%.

Furthermore, the weight value γ can be used to control the tradeoff between leakage reduction rate and the LF gate number.

TABLE VI
DIFFERENT WEIGHT VALUE γ FOR C880 (pA)

C880	0% circuit slowdown			3% circuit slowdown			5% circuit slowdown		
	I_{leak}	ST	LF	I_{leak}	ST	LF	I_{leak}	ST	LF
$\gamma = 0$	619.2	352	105	232.2	370	68	126.1	375	46
$\gamma = 10$	630.7	352	25	252.2	369	14	157.5	375	7
$\gamma = 50$	723.6	350	21	365.6	366	10	199.7	373	4
$\gamma = 100$	1292.3	315	11	541.4	359	6	255.7	370	3
$\gamma = 200$	2415.9	263	2	1034.1	330	2	479.5	357	1

TABLE VII
RESULTS OF LEAKAGE AND LF GATE USING HEURISTIC ALGORITHM DURING ST PLACEMENT PHASE

ISCAS85 Benchmark Circuits	0% circuit slowdown						3% circuit slowdown						5% circuit slowdown					
	Step I N_R	Step II: $\gamma = 0$		Step III: $\gamma = 100$		Step I N_R	Step II: $\gamma = 0$		Step III: $\gamma = 100$		Step I N_R	Step II: $\gamma = 0$		Step III: $\gamma = 100$				
		I_{leak}	LF/ST	I_{leak}	LF/ST		I_{leak}	LF/ST	I_{leak}	LF/ST		I_{leak}	LF/ST	I_{leak}	LF/ST	I_{leak}	LF/ST	
C432	89	2304	35/120	4042	2/28	88	1847	23/134	2990	1/78	70	735	10/150	1304	3/134			
C499	164	4038	34/157	6270	8/113	164	2971	17/173	4208	5/161	131	863	4/195	1250	1/192			
C880	75	682	26/349	1951	3/297	55	396	17/361	1420	4/328	46	278	8/368	707	0/347			
C1355	436	3607	55/348	5575	8/296	436	2422	30/376	3634	8/354	434	1422	8/447	1547	0/439			
C1908	296	2914	52/809	5616	8/727	230	1302	28/848	2446	1/794	160	823	19/866	1491	0/832			
C2670	217	1472	32/1321	3202	9/1198	126	675	13/1251	1349	4/1239	67	274	3/1267	406	1/1265			
C3540	376	2543	80/1600	6152	11/1476	241	1158	40/1657	3444	10/1596	157	716	18/1675	1565	2/1641			
C5315	339	2216	73/2220	4890	5/2124	192	1101	31/2273	2216	1/2230	109	525	15/2304	1205	1/2279			
C6288	1565	9215	397/1828	17115	0/1374	1274	8068	333/1902	14396	0/1548	943	3782	156/2193	6818	0/2023			
C7552	585	4115	107/3381	8233	8/3237	376	1725	42/3489	3448	3/3426	222	841	15/3533	1328	0/3510			
Average	60.1%	83.9%	11.6%	70.7%	2.1%	65.4%	88.7%	6.7%	80.4%	0.9%	72.3%	94.8%	2.4%	91.3%	0.3%			

Four different weight values: 10, 50, 100, 200 are used in our MILP model for C880. Table VI shows the leakage current and LF number under different weight value. As in Table VI, when the circuit slowdown is 0%, a larger weight value γ should be chosen to reduce the LF gate number; when the circuit slowdown is becoming larger, the original LF gate number without any optimization reduces to a low level, thus a smaller weight value γ can be used to get a larger leakage reduction rate with an acceptable LF gate number.

The runtime for solving the previous MILP model of ST placement is not stable, it will be time consuming for many circuits; hence heuristic algorithms are needed to get near optimal results with a very fast speed. However, the heuristic may lead to local optimal and can not guarantee the optimality of the result; thus, the results of MILP models can be used as a reference.

C. Heuristic Algorithm for ST Placement Phase

Our three-step heuristic algorithm is developed to accelerate the computation speed within acceptable loss of accuracy. As shown in Table VII, after Step I, DAG pruning, when the circuit slowdown is 0%, 3%, and 5%, about 60.1%, 65.45%, and 72.3% of total gates are deleted from the original circuits, thus, problem size is greatly reduced. N_R in Table VII represents the gate number in V_{ker} . The second step ST Assignment can be considered as the MILP method without considering LF gate and is explained in Section III-A1. Our second step can achieve similar leakage reduction rate compared with the results of two-step MILP method shown in Table V. For C499 and C1355, our algorithm leads to even smaller leakage, especially when the circuit slowdown is 0%; meanwhile the average difference of leakage reduction rate for other circuits is about 4% of original leakage current. The MILP models for C499 and C1355 may not converge well using the MILP solver. Furthermore, because potential LF reduction is considered by using Forward Selection during ST Assignment Algorithm, the LF gate number is much less than the MILP method in Table V. When the circuit slowdown is 0%, 3%,

and 5%, the LF gate is only 11.6%, 6.7%, and 2.4% of total gate with ST.

In Step III, the weight value γ is also assumed to be 100. The LF gate number is reduced to below 3% of the total gate with ST while the difference of leakage reduction rate is only 13.2%, 8.3%, and 3.5% when the circuit slowdown is 0%, 3%, and 5%, respectively.

As shown in Table VIII, the computation time of our heuristic algorithm is linear with the circuit size. As the MILP method is very time-consuming, our heuristic algorithm appears controllable and promising for a larger circuit design.

Finally, as discussed in Section IV-B, the leakage reduction rate of each circuit relies on its $T_{path}(v)$ distribution, which gives us an intuitional evaluation about the percentage of gates that can be optimized. For C432, because of a high percentage of gates are with large $T_{path}(v)$, only 50% leakage current can be reduced when the circuit slowdown is 0%; on the other hand, C880 shows that a small percentage of gates are with large $T_{path}(v)$, so that the leakage reduction rate can achieve 92% which is very high when the circuit slowdown is also 0%. Therefore, the $T_{path}(v)$ distribution shows the potential of leakage reduction rate in a certain circuit to the designers.

D. Linear Programming Model for Optimal ST Sizing

When the circuit slowdown is less than 6%, ST sizing phase is performed using the ST information of each gate decided in the ST placement phase. The optimal ST sizing method is used to get all the leakage current values in Tables III–VII.

When the circuit slowdown is larger than 6%, ST can be assigned to all the gates in the circuits, the two-phase procedure of *FGSTI* technique is reduced into one phase: ST sizing. The LP model for ST sizing can be solved to get the same result as optimal sizing method in [25]. We compare the area penalty with the fixed slowdown method and the MILP method in Table IX. With 7% circuit slowdown, our ST sizing LP model causes 75.48% ST area saving compared to fixed slowdown method and the result

TABLE VIII
RUNTIME OF HEURISTIC ALGORITHM FOR ST PLACEMENT PHASE (TIME IN SECONDS)

ISCAS85 Benchmark Circuits	0% circuit slowdown			3% circuit slowdown			5% circuit slowdown		
	Heuristic Algorithm $\gamma = 100$			Heuristic Algorithm $\gamma = 100$			Heuristic Algorithm $\gamma = 100$		
	Step I	Step II	Total	Step I	Step II	Total	Step I	Step II	Total
C432	0.016	0.328	0.422	0	0.297	0.36	0.016	0.14	0.219
C499	0	0.875	0.969	0	0.828	0.922	0	0.375	0.469
C880	0.015	0.422	0.765	0	0.39	0.703	0.016	0.359	0.688
C1355	0.016	12	12.672	0.016	15.375	16.047	0.032	14.406	15.094
C1908	0.047	4.016	5.891	0.046	2.563	4.437	0.047	2.141	4
C2670	0.078	4.172	7.781	0.062	3.86	7.422	0.063	3.703	7.313
C3540	0.109	9.578	15.937	0.125	7.359	13.703	0.125	6.75	13.079
C5315	0.204	15.156	27.079	0.203	13	24.906	0.204	12.312	24.204
C6288	0.391	103.437	116.844	0.391	90.641	103.953	0.375	34.656	47.906
C7552	0.407	47.828	75.579	0.406	32.782	60.484	0.422	29.36	57.078

TABLE IX
ST SIZING RESULTS COMPARED WITH MILP AND FIXED SLOWDOWN METHOD

ISCAS85 Benchmark Circuits	7% circuit slowdown			9% circuit slowdown		
	ST Sizing	MILP	Fixed Slowdown	ST Sizing	MILP	Fixed Slowdown
C432	714	714	2317.72	596	597	1802.67
C499	1146	1146	2797.72	959	959	2176.01
C880	876	876	5252.58	780	780	4085.35
C1355	3365	3364	7515.44	2719	2720	5845.35
C1908	2355	2355	12493.73	2081	2081	9717.36
C2670	2087	2088	17540.59	1937	1937	13642.71
C3540	3371	3371	23300.60	3160	3160	18122.72
C5315	4292	4293	31940.60	3917	3918	24842.74
C6288	11733	11733	33558.89	9610	9611	26101.41
C7552	8980	8981	48905.19	8197	8197	38037.45
Area saving	75.48%	N/A	N/A	73.0%	N/A	N/A

is almost the same with MILP method. In Table IX, ST area is calculated using (18), just summing up all the $(W/L)_v$, since the transistor channel length of ST is a constant.

VI. CONCLUSION

In this paper, we present a novel two-phase *FGSTI* technique to reduce the leakage current. Simple leakage current and delay models for our two-phase *FGSTI* technique are proposed and analyzed to prove the rationality of our method. ST placement and sizing are modeled using MILP and LP models respectively. The LF gate number is reduced during the ST placement phase. Both LP-solver and heuristic algorithms are used to solve the MILP model. Our experimental results show that our two-phase *FGSTI* technique leads to 2% more leakage saving and at least 10× runtime saving compared with simultaneous ST placement and sizing method using MILP. The LF gate number can be reduced and controlled using our multi-object model when the circuit slowdown is below 6%. Our heuristic algorithm is much faster and more stable than the MILP method. When the circuit slowdown is larger than 6%, the two-phase *FGSTI* can achieve 75.48% ST area saving comparing with fixed slowdown method. In conclusion, two-phase *FGSTI* technique is reasonable from

our results. For the future work, the detailed comparison between the *FGSTI* and *BBSTI* techniques should be carefully examined in the physical level, such as place and routing penalty.

REFERENCES

- [1] G. Moore, "No exponential is forever: But forever can be delayed," in *IEEE ISSCC Dig. Tech. Papers*, 2003, pp. 20–23.
- [2] D. Duarte, N. Vijaykrishnan, M. J. Irwin, and M. Kandemir, "Formulation and validation of an energy dissipation model for the clock generation circuitry and distribution networks," in *Proc. Int. Conf. VLSI Des.*, 2001, pp. 248–253.
- [3] J. Kao, S. Narendra, and A. Chandrakasan, "Subthreshold Leakage modeling and reduction techniques," in *Proc. Int. Conf. Comput.-Aided Des.*, 2002, pp. 141–149.
- [4] W. M. Elgharbawy and M. A. Bayoumi, "Leakage sources and possible solutions in nanometer CMOS technologies," *IEEE Circuits Syst. Mag.*, vol. 5, no. 4, pp. 6–17, 4th Quarter, 2005.
- [5] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305–327, Feb. 2003.
- [6] S. Mukhopadhyay, C. Neau, R. T. Cakici, A. Agarwal, C. H. Kim, and K. Roy, "Gate leakage reduction for scaled devices using transistor stacking," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 11, no. 4, pp. 716–729, Aug. 2003.
- [7] A. Abdollahi, F. Fallah, and M. Pedram, "Leakage current reduction in CMOS VLSI circuits by input vector control," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 2, pp. 140–154, Feb. 2004.

- [8] S. Narendra, A. Keshavarzi, B. A. Bloechel, S. Borkar, and V. De, "Forward body bias for microprocessors in 130-nm technology generation and beyond," *IEEE J. Solid-State Circuits*, vol. 38, no. 5, pp. 696–701, May 2003.
- [9] L. Wei, Z. Chen, and K. Roy, "Design and optimization of dual threshold circuits for low voltage, low power applications," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 7, no. 1, pp. 16–24, Mar. 1999.
- [10] Y. Wang, H. Z. Yang, and H. Wang, "Signal-path level dual-Vt assignment for leakage power reduction," *J. Circuits, Syst. Comput.*, vol. 15, no. 2, pp. 197–216, 2006.
- [11] Q. Wang and S. B. K. Vrudhula, "Algorithms for minimizing standby power in deep submicrometer, dual-Vt CMOS circuits," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 21, no. 3, pp. 306–318, Mar. 2002.
- [12] C. H. Kim and K. Roy, "Dynamic VTH scaling scheme for active leakage power reduction," in *Proc. DATE*, 2002, pp. 163–167.
- [13] A. J. Bhavnagarwala, B. L. Austin, K. A. Bowman, and J. D. Meindl, "A minimum total power methodology for projecting limits on CMOS GSI," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 3, pp. 235–251, Jun. 2000.
- [14] S. Shah, A. Srivastava, D. Sharma, D. Sylvester, D. Blaauw, and V. Zolotov, "Discrete Vt assignment and gate sizing using a self-snapping continuous formulation," in *Proc. ICCAD*, 2005, pp. 705–712.
- [15] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.
- [16] J. Kao, S. Narendra, and A. Chandrakasan, "MTCMOS hierarchical sizing based on mutual exclusive discharge patterns," in *Proc. Des. Autom. Conf.*, 1998, pp. 495–500.
- [17] M. Anis, S. Areibi, and M. Elmasry, "Dynamic and leakage power reduction in MTCMOS circuits using an automated efficient gate clustering technique," in *Proc. Des. Autom. Conf.*, 2002, pp. 480–485.
- [18] W. Wang, M. Anis, and S. Areibi, "Fast techniques for standby leakage reduction in MTCMOS circuits," in *Proc. IEEE Int. Conf. SOC*, Sep. 2004, pp. 21–24.
- [19] C. Long and L. He, "Distributed sleep transistors network for power reduction," in *Proc. Des. Autom. Conf.*, Jun. 2003, pp. 181–186.
- [20] C. Long and L. He, "Distributed sleep transistor network for power reduction," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 9, pp. 937–946, Sep. 2004.
- [21] J. Kao and A. Chandrakasan, "MTCMOS sequential circuits," in *Proc. 27th Eur. Solid-State Circuits Conf.*, Sep. 2001, pp. 317–320.
- [22] B. H. Calhoun, F. A. Honor, and A. P. Chandrakasan, "A leakage reduction methodology for distributed MTCMOS," *IEEE J. Solid-State Circuits*, vol. 39, no. 5, pp. 818–826, May 2004.
- [23] V. Khandelwal and A. Srivastava, "Leakage control through fine-grained placement and sizing of sleep transistors," in *Proc. Int. Conf. Comput.-Aided Des.*, 2004, pp. 533–536.
- [24] Y. Wang, H. Lin, R. Luo, H. Z. Yang, and H. Wang, "Simultaneous fine-grain sleep transistor placement and sizing for leakage optimization," in *Proc. Int. Symp. Quality Electron. Des.*, Mar. 2006, pp. 723–728.
- [25] W. Feller, "The strong law of large numbers," in *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: Wiley, 1968, vol. 1, pp. 243–245.
- [26] S. Mukhopadhyay and K. Roy, "Modeling and estimation of total leakage current in nano-scaled CMOS devices considering the effect of parameter variation," in *Proc. Int. Symp. Low Power Electron. Des.*, Aug. 2003, pp. 172–175.
- [27] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA: The MIT Press, pp. 531–540.



Yu Wang (M'07) received the B.S. and Ph.D. degrees with honors in electronics engineering from Tsinghua University, Beijing, China, in 2002 and 2007, respectively.

He is currently with the Electrical Engineering Department, Tsinghua University. His research mainly focuses on electronic design automation area, including fast circuit analysis, low power circuit design methodology, reliability-aware circuit design methodology, application specific FPGA design, and MPSoC design methodology. He has authored

or coauthored over 10 journal and conference papers in this area and holds 1 China patent.



Ku He received the B.S. degree in electronic engineering and the M.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004 and 2007, respectively. He is currently pursuing the Ph.D. degree in computer engineering from the University of Texas, Austin.

His research interests include statistical and robust circuit optimization.



Rong Luo (M'05) received the double B.S. degree in engineering physical and electronic engineering and the Ph.D. degree from Tsinghua University, Beijing, China, in 1992 and 1997, respectively.

Currently, she is an Associate Professor with the Department of Electronic Engineering, Tsinghua University. Her current research work is mainly on SoC design technology, VLSI design, and embedded system design technology.



Hui Wang received the B.S. degree in radio electronics from Tsinghua University, Beijing, China, in 1970.

She has been with Tsinghua University since 1970, where she served as Vice Director of the Department of Electronics Engineering, from 1996 to 1999, Deputy Dean of Academic Affairs Office, from 1999 to 2005, and has been the Vice Dean of School of Information Science and Technology, since 2005. She was a visiting Scholar at Stanford University, Stanford, CA, from 1991 to 1992.

Dr. Wang was a recipient of awards including the Science and Technology Progress Award (I) from Chinese Ministry of Electronics, the National Science and Technology Progress Award (III) from National Science and Technology Communication and Teaching Achievement Award (I) from Chinese Ministry of Education. She participated in many projects from the Natural Science Foundation of China (NSFC), "863" program and some key programs of fundamental research. received some national awards,



Huazhong Yang (M'97–SM'00) received the B.S. degree in microelectronics and the M.S. and Ph.D. degrees in circuits and systems from Tsinghua University, Beijing, China, in 1989, 1993, and 1998, respectively.

Since 1993, he has been with the Department of Electronic Engineering, Tsinghua University, where he has been a Full Professor since 1998. His research interests include CMOS radio-frequency integrated circuits, VLSI system structure for digital communications and media processing, low-voltage and low-

power integrated circuits, and computer-aided design methodologies for system integration. He has authored or coauthored 6 books and more than 180 journal and conference papers in this area and holds 9 China patents. He is also a coeditor of the research monograph *High-speed Optical Transceivers-Integrated Circuits Designs and Optical Devices Techniques* (World Scientific, 2006).

Dr. Yang was a recipient of the fund for Distinguished Young Scholars from NSFC in 2000, the outstanding researcher award of the National Keystone Basic Research Program of China in 2004, and the Special Government Allowance from the State Council of China in 2006. served as a TPC member of the Asia-Pacific Conference on Circuits and Systems, the International Conference on Communications, Circuits and Systems, and the Asia and South Pacific Design Automation Conference. He is an Associate Editor of the *International Journal of Electronics*.