Two-phase Fine-grain Sleep Transistor Insertion Technique in Leakage Critical Circuits

Yu Wang, Yongpan Liu, Rong Luo, Huazhong Yang, Hui Wang
E.E. Dept., Tsinghua University, Beijing, P.R.China
(8610)62772966
{wangyu99, ypliu99}@mails.tsinghua.edu.cn

ABSTRACT
Multi-threshold CMOS is a valuable leakage reduction method in circuit standby mode. Reducing leakage current through fine-grain sleep transistor insertion (FGSTI) makes it easier to guarantee circuit functionality and improves circuit noise margins. In this paper, we first indicate the negligible dependence of ST size on the amount of leakage saving which makes the two-phase FGSTI reasonable based on our leakage current and delay models. Then we introduce a novel two-phase FGSTI technique: a) ST placement and b) ST sizing, which are formally modeled as two linear programming (LP) models respectively. Our experimental results show that the two-phase FGSTI technique can achieve 78.91%, 92.55%, 97.97% leakage saving when the circuit slowdown is 0%, 3%, 5% respectively. Comparing to the simultaneous ST placement and sizing method using mix integer linear programming (MLP) [1], our technique leads to an average 2% more leakage current reduction while at least 10X runtime saving. Since fewer variables and constraints with less approximation are used in the LP models. When the circuit slowdown is large enough to perform conventional fixed slowdown method, our technique can still achieve 75.48% ST area saving. Moreover, we show that when the circuit slowdown is 0%, it should be carefully considered to use FGSTI technique due to a large amount of leakage feedback gates.

Categories and Subject Descriptors:
J.6 [Computer Aided Engineering]: Computer aided design (CAD), B.6.3 [Design Aids]: Optimization

General Terms
Algorithms, Design

Keywords
Leakage current reduction, two-phase fine-grain sleep transistor insertion, mixed integer linear programming.

1. INTRODUCTION
With the development of fabrication technology, leakage power dissipation has become comparable to switching power dissipation [2]. As we all know, the total power dissipation consists of dynamic power, short circuit power and leakage power. The behavior of the short circuit power dissipation remains at around 10% of the total power dissipation [3]. At the 90nm technology node, leakage power may make up 42% of total power [4]. Leakage power reduction techniques can be broadly categorized into two main categories [5]: process level and circuit level techniques. The circuit level techniques consist of adapt body bias [6], DVTs [7], input vector control [8], dual-\(V\_t\) assignment [9-11] and Multi-Threshold CMOS (ST insertion) [1] [12-18]. Among these, Multi-Threshold CMOS (MTCMOS) technique is essentially placing a ST between the gates and the power/ground (P/G) net in a circuit in order to put it into sleep mode when the circuit is standby.

The most popular MTCMOS technique is gating the power of sizable blocks using large sleep transistors which is concluded as block based ST insertion (BBSTI) technique. In BBSTI techniques, all the gates in the block are assumed to have a fixed slowdown, so it is also called fixed slowdown method. The existing literatures on BBSTI techniques [12-16] present some details in clustering gates into blocks in order to optimize the leakage current and ST size. All these literatures focus on how to reduce the ST area penalty along with a remarkable leakage saving: [12] first gives out a mutual exclusion method; [13] [14] present several fast heuristic techniques for efficient gate clustering; [15] [16] propose a distributed sleep transistor network (DSTN) approach which assumes that all the sleep devices are connected to further reduce the area penalty.

Although BBSTI techniques greatly reduce the area penalty, they induce large ground bounce in the P/G network which has adverse effects on circuit speed and noise immunity [18]. What is more, ST size is determined by the worst case current of the clustering block which is quite difficult to determine without comprehensive simulation [12]. Thus it is harder to guarantee circuit functionality for large blocks with only one ST [17].

In recent years gate level ST insertion, which can be also called fine-grain ST insertion (FGSTI) technique [1] [17] [18] (as shown in figure 1 (a)) shows some advantages over the BBSTI technique (as shown in figure 1 (b)). It is easier to guarantee circuit functionality in an FGSTI technique as ST sizes are not determined by the worst case current of large circuit blocks. And FGSTI technique leads to a smaller simultaneous switching current when the circuit changes between standby mode and active mode, thus improves circuit noise margins. Furthermore, better circuit slack utilization can be achieved as the slowdown of each gate is not fixed, and then leads to a further reduction of leakage and area. As shown in [18], FGSTI technique corresponds to an area penalty of roughly 5% using standard cell placement.
2. PRELIMINARIES

In this section, the leakage current and delay models used in our two-phase FGSTI technique are given out and analyzed to prove that a FGSTI design can be performed in two phases. ST is used with variable size which is decided by the process technology in our two-phase FGSTI design. A combinational circuit is represented 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 $j$.

2.1 Leakage current 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 $I_{\text{leak}}(v)$ can be expressed as:

$$I_{\text{leak}}(v) = \sum I_i(v, \text{IN}) \times PB(v, \text{IN})$$  \hspace{1cm} (1)

Where $I_i(v, \text{IN})$ and $PB(v, \text{IN})$ are the leakage current and the probability of gate $v$ under input pattern $IN$.

We simply use a linear model to represent leakage current $I_{\text{leak}}^L(v)$ based on HSPICE simulation results:

$$I_{\text{leak}}^L(v) = A(v)(W/L)$$  \hspace{1cm} (2)

where $A(v)$ is constant and decided by the gate type. Here we assume all the input patterns have same probability and estimate every $A(v)$ for all the standard cells in the library. Consider two standard cells: NOR2XL and NAND4XL in the TSMC 0.18um standard cell library, the largest error is about 52% as shown in table 1. The error of linear approximation may be neglected in FGSTI due to Law of large numbers [19] with the growing circuit size. As we will mention in Section 2.3, the influence of the linear model error on the FGSTI technique will be diminished by the large difference between leakage current of a gate with or without ST.

2.2 Delay model

As shown in [20], gate delay is influenced by the ST insertion. The load dependent delay $d^{\text{load}}(v)$ of gate $v$ without ST is given by:

$$d^{\text{load}}(v) = \frac{K_C V_{DD}}{(V_{DD} - V_{THlow})^\alpha}$$  \hspace{1cm} (3)

where $C_L, V_{THlow}, \alpha, K$ are the load capacitance at the gate output, the low threshold voltage, the velocity saturation index and the proportionality constant respectively.

The propagation delay $d^P(v)$ of gate $v$ with ST can be expressed as:

$$d^P(v) = \frac{K_C V_{DD}}{(V_{DD} - 2V_s - V_{THlow})^\alpha}$$  \hspace{1cm} (4)

where $V_s$ is the $V_s$ of the ST, that is to say the voltage drop from $V_{DD}$ to the virtual $V_{DD}$. $AD(v)$ is derived from the above equations:

$$AD(v) = d^P(v) - d^{\text{load}}(v) = \left(\frac{1 - 2V_s}{V_{DD} - V_{THlow}}\right)^\alpha - 1 \times d^{\text{load}}(v)$$  \hspace{1cm} (5)

The paper is organized as follows. In Section 2, our leakage current and delay models are given out and analyzed to prove the rationality of our two-phase FGSTI technique. The two-phase FGSTI technique is proposed in Section 3. The implementation and experimental results are presented and analyzed in Section 4. In Section 5, we conclude this paper.
\( I_{\text{on}}(v) \) is the current flowing through ST in gate \( v \) during the active mode, which can be expressed as given by [18]:

\[
I_{\text{on}}(v) = \mu_C(W/L),(V_{\text{co}} - V_{\text{th}})W_L - \frac{V^2}{2}
\]

(6)

Thus the voltage drop \( V_s \) in gate \( v \) due to ST insertion can be expressed as:

\[
V_s = \frac{I_{\text{on}}(v)}{\mu_C(W/L),(V_{\text{co}} - V_{\text{th}})W_L} \times \frac{1}{W/L}
\]

(7)

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

\[
V_s = \frac{1}{2} \left( 1 - e^{-\frac{V_s}{\sqrt{2V_s}}} \right)(V_{\text{co}} - V_{\text{th}})
\]

(8)

2.3 Relationship between ST placement and sizing

From previous part, we find that a linear leakage current model may have an error as large as 50% comparing with the HSPICE simulation results. Refer to [18], the leakage current for a gate with ST is also modeled as a linear function from [21]:

\[
I_{\text{on}}^\text{ST}(v) = \mu_C(W/L)\left(V_{\text{co}} - V_{\text{th}}\right)W_L - \frac{V^2}{2}
\]

(9)

where \( \mu_C \) is the N-mobility, \( C_{ox} \) is the oxide capacitance, \( V_{\text{th}} \) is the high threshold voltage, \( V_C \) is the thermal voltage, \( n \) is the sub-threshold swing parameter, \( (W/L) \) represents the ST size of gate \( v \). Notice that their model is also linear by assuming parameters except \( (W/L) \) are constant decided by process information and gate structure. Such a linear model will also consume comparative error as our leakage current model.

However, as we explore the leakage current model further, the leakage current of a gate without ST is much larger than that of a gate with ST as shown in Table 2, so the error of the linear model can be neglected in the FGSTI procedure. In table 2, the leakage current of cells in the TSMC 0.18\( \mu \)m standard cell library under two different ST conditions: with ST or without ST are compared. As shown in table 1, the leakage current of a gate with ST become larger with a larger \( (W/L) \). Therefore, the largest leakage current of a gate with ST is derived by setting the \( (W/L) \) of a ST to 16 which is the maximum ratio of ST in our FGSTI technique.

Table 2. Leakage current comparison of standard cells (FA)

<table>
<thead>
<tr>
<th>Cell Name</th>
<th>( p^2 )</th>
<th>( p^3 )</th>
<th>( p^1 )</th>
<th>Cell Name</th>
<th>( p^2 )</th>
<th>( p^3 )</th>
<th>( p^1 )</th>
<th>Cell Name</th>
<th>( p^2 )</th>
<th>( p^3 )</th>
<th>( p^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAND2X4</td>
<td>149.06</td>
<td>45.03</td>
<td>313.0</td>
<td>AND3X4</td>
<td>3490.3</td>
<td>55.4</td>
<td>1008</td>
<td>INVX4</td>
<td>3490.3</td>
<td>55.4</td>
<td>1008</td>
</tr>
<tr>
<td>NAND3X4</td>
<td>89392.7</td>
<td>45.5</td>
<td>1854</td>
<td>HUX4</td>
<td>80876.7</td>
<td>55.4</td>
<td>1513</td>
<td>INVX4</td>
<td>80876.7</td>
<td>55.4</td>
<td>1513</td>
</tr>
<tr>
<td>INVX4</td>
<td>14233.2</td>
<td>35.9</td>
<td>368</td>
<td>NORDX4</td>
<td>11626.1</td>
<td>55.4</td>
<td>516</td>
<td>INVX4</td>
<td>11626.1</td>
<td>55.4</td>
<td>516</td>
</tr>
<tr>
<td>NORDX4</td>
<td>14906.8</td>
<td>37.6</td>
<td>328</td>
<td>CLKINVX4</td>
<td>38736.4</td>
<td>55.4</td>
<td>1043</td>
<td>NORDX4</td>
<td>38736.4</td>
<td>55.4</td>
<td>1043</td>
</tr>
<tr>
<td>XORX4</td>
<td>89358.9</td>
<td>55.4</td>
<td>1129</td>
<td>NAND4X4</td>
<td>72254.2</td>
<td>55.4</td>
<td>1441</td>
<td>XORX4</td>
<td>72254.2</td>
<td>55.4</td>
<td>1441</td>
</tr>
<tr>
<td>NAND2X1</td>
<td>12861.2</td>
<td>54.9</td>
<td>259</td>
<td>ANOX4</td>
<td>26975.7</td>
<td>55.4</td>
<td>909</td>
<td>NAND2X1</td>
<td>26975.7</td>
<td>55.4</td>
<td>909</td>
</tr>
<tr>
<td>NAND3X1</td>
<td>14186.1</td>
<td>49.7</td>
<td>298</td>
<td>ANOX4</td>
<td>13788.6</td>
<td>55.4</td>
<td>258</td>
<td>NAND3X1</td>
<td>13788.6</td>
<td>55.4</td>
<td>258</td>
</tr>
<tr>
<td>ANDX4</td>
<td>48999.1</td>
<td>54.8</td>
<td>915</td>
<td>CLKINVX8</td>
<td>26975.7</td>
<td>55.4</td>
<td>822</td>
<td>ANDX4</td>
<td>26975.7</td>
<td>55.4</td>
<td>822</td>
</tr>
</tbody>
</table>

As shown in table 2, the leakage current difference is at least 238X. Referring to equation (5) and (7), the delay difference is less than 20% of the original gate delay under the same condition. However, the delay difference of a gate with different ST size is much larger, for example, setting the \( (W/L) \) of a ST to 1 will lead to about 140% additional delay comparing with the original gate without ST. Also from table 1, the leakage current variation range due to the change of ST size can be neglected since it is much smaller compare to the leakage saving of changing a gate’s ST condition.

In a word, although leakage current is reduced by sizing the ST, ST placement is not affected by ST sizing due to the large gap between their effects on leakage saving.

With technology scaling down, the leakage current difference may be smaller under different ST condition, but it will still be very large due to high \( V_{th} \) ST and stacking effect. Hence we can draw a conclusion that the leakage reduction depends on where to insert ST and the leakage difference of each gate under different ST condition; while the area penalty is decided by the ST sizing procedure.

We further assume that ST placement and sizing are independent in a FGSTI design. Therefore, we develop a two-phase FGSTI technique: first, ST placement can be performed to decide where to put ST and achieve most of the leakage saving; and then ST sizing can be used to reduce the area overhead along with further leakage current reduction.

3. TWO-PHASE FGSTI TECHNIQUE

In this section two-phase FGSTI technique is modeled using linear programming. First ST placement phase shows how to place the ST as many as possible in order to reduce the total leakage, and then an optimal sizing method is given out for ST sizing phase to reduce the area overhead based on the ST placement information. At the end of this section, the simultaneous placement and sizing method [1] is briefly reviewed for comparison.

3.1 ST placement

A novel ST placement method is proposed that tries to maximize the leakage saving in the circuits through mixed integer linear programming (MILP). First, the object function for the total leakage current is constructed as below:

\[
I(G) = \sum_{i=1}^{N}(I_{\text{on}}(v)\times(1-\text{ST}(v))+I_{\text{on}}^\text{ST}(v)\times \text{ST}(v))
\]

(9)

where \( \text{ST}(v) \) is a binary variable to represent gate \( v \)’s ST condition, \( \text{ST}(v) = 1 \) means gate \( v \) has ST inserted and \( \text{ST}(v) = 0 \) means gate \( v \) is without ST. As ST size is not considered, we choose the largest ST size \( (W/L) \) in equation (2) to obtain the minimum delay overhead. The leakage current of gate \( v \) with ST is given by:

\[
I_{\text{on}}^\text{ST}(v) = A(v)\times(W/L)_{max}
\]

(10)

It can be derived that \( I_{\text{on}}^\text{ST}(v) \) is a constant for each gate to simplify the MILP model further.

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

\[
t_{m}(m) = 0 \quad m \in PI
\]

(11)

\[
t_{e}(n) + d(n) \leq T_{req}
\]

(12)

\[
t_{e}(i) + d(i) = t_{e}(j) \quad \forall (i,j) \in E \quad i, j \in V
\]

(13)

where \( PI \) and \( PO \) refer to the primary input and primary output gates of the circuit; \( t_{e}(v) \) 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 using equation (5) and (7):

\[
d(v) = d(v) + \Delta d(v) \times \text{ST}(v)
\]

(14)
where \( d^{st}(v) \) and \( \Gamma \) are constant for each gate \( v \). Similarly we choose the largest ST size \((W/L)_{\text{max}}\) to get the minimum delay overhead.

Minimize:

\[
I(G) = \sum_{v \in G} (I^+ - (1 - ST(v)) + (A(v) \times (W/L)_{\text{max}}) \times ST(v))
\]

Subject to:

(Timing constraints)

\[
l_p(m) = 0 \quad m \in PI
\]

\[
l_p(n) + d(n) \leq T_{eq} \quad n \in PO
\]

\[
l_p(i) + d(i) \leq l_p(j) \quad \forall (i, j) \in E, i, j \in V
\]

\[
d(v) = d^{st}(v) + \Gamma d^{st}(v) \times ST(v) \quad v \in V
\]

[Variable bounds]

\[
ST(v) \text{ are binary variables}
\]

**Figure 2.** MLP model for leakage minimization through ST placement

The general form of our MLP model for ST placement is shown in figure 2. ST placement is similar as dual \( V_d \) assignment with fixed high and low \( V_d \) values, thereby it can also be solved by sensitive based heuristic algorithms which are previously dealing with dual \( V_d \) assignment [9-11].

### 3.2 Optimal ST Sizing

After the ST condition for each gate \( v \) is decided, we use linear programming to get the optimal ST size. First the objective function for optimal ST sizing is given out as below:

\[
\text{Area}(ST) = \sum_{v \in G} (W/L) \times ST(v)
\]

where \( ST(v) \) is a binary value given out in the ST placement phase; \((W/L)\), is a continuous variable. Because the transistor length for ST is assumed to be a constant: the minimum length, \((W/L)_{\text{min}}\), is used instead of \( W \times L \) to represent the ST area. Moreover, the expression for \((W/L)\), can be derived from equation (7) and (8) as below:

\[
(W/L) = \frac{I_{dss}(v)}{\mu C_s(V_{dd} - V_{th})} \times \frac{1}{V_d}
\]

\[
= \frac{I_{dss}(v)}{\mu C_s(V_{dd} - V_{th})} \left( \frac{1}{2} \left[ 1 - \left( \frac{d^{st}(v)}{d^{st}(v)} \right)^\Gamma \right] \right) \left( \frac{V_{dd} - V_{th}}{V_{dd}} \right) \]  \quad (16)

The timing constraints can also be expressed as equation (11), (12) and (13). The propagation delay \( d^{st}(v) \) of gate \( v \) with ST can be rewrite using equation (7) and (8) as:

\[
d^{st}(v) = d^{st}(v) + \Delta d(v)
\]

\[
d^{st}(v) = d^{st}(v) + \left( 1 - \frac{2V_d}{V_{dd} - V_{th}} + \frac{1}{(W/L)_{\text{min}}} \right) \left( V_{dd} - V_{th} \right)
\]

\[
d^{st}(v) = d^{st}(v) \left( 1 - \frac{I_{dss}(v)}{\mu C_s(V_{dd} - V_{th})} \right) \left( \frac{V_{dd} - V_{th}}{V_{dd}} \right)
\]

With a given boundary of \((W/L)\), \( [(W/L)_{\text{min}} \ (W/L)_{\text{max}}] \), the boundary of \( d^{st}(v) \) can be easily gained: \([d^{\text{min}}(v), d^{\text{max}}(v)]\) using equation (17). Consequently, the general form of our LP model for ST sizing is show in figure 3.

### 3.3 Simultaneous ST placement and sizing

The object function of simultaneous ST placement and sizing is very similar to ST placement as shown in equation (9):

\[
I(G) = \sum_{v \in G} (I^+ - (1 - ST(v)) + (A(v) \times (W/L)_{\text{max}}) \times ST(v))
\]  \quad (18)

**Figure 3.** LP model for optimal ST sizing

where \( ST(v) \) and \((W/L)\), are variables which decide where to put ST and how to size ST respectively.

The timing constraints also follow equation (11), (12) and (13). Refer to equation (14), gate delay \( d(v) \) for gate \( v \) can be derived as:

\[
d(v) = d^{st}(v) + \left( 1 - \frac{2V_d}{V_{dd} - V_{th}} + \frac{1}{(W/L)_{\text{min}}} \right) \left( V_{dd} - V_{th} \right)
\]

\[
d(v) = d^{st}(v) \left( 1 - \frac{I_{dss}(v)}{\mu C_s(V_{dd} - V_{th})} \right) \left( \frac{V_{dd} - V_{th}}{V_{dd}} \right)
\]

\[
d(v) = d^{st}(v) + \frac{(W/L)_{\text{opt}}}{(W/L)_{\text{max}}} \times \Phi(W/L)_{\text{opt}}\times ST(v)
\]

As we can see from equation (18) and (19), this problem is actually a non-linear programming model. In [1], Taylor series expansion and piecewise linear approximation technique are used to get a mixed integer linear programming model. Some dummy variables are needed for linear approximation and more linearization constraints are added in the MLP model for each dummy variable. Unfortunately, the model size becomes extremely large with the increasing gate number in the circuit.

### 4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

#### 4.1 Implementation

All ISCAS85 benchmark circuit netlists are synthesized using Synopsys Design Compiler and a TSMC 0.18\( \mu m \) 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 equation (2) for all the standard cells is estimated using HSPICE simulation results under different \((W/L)\). The values of various transistor parameters are taken from the TSMC 0.18\( \mu m \) process library. \( V_{dd}=1.8V, V_{T\text{High}}=500mV, V_{T\text{Low}}=300mV, \) and \( I_{dss}=200\mu A \) for all the gates in the circuit. The timing constraints are set up with a specialized static timing analysis (STA) tool [10], and the MLP and LP models for ST placement and sizing are automatically generated. We use an LP solver named \texttt{lp_solve} [22] to solve the models.

We assume \( 1 \leq (W/L) \leq 16 \), corresponding to a least delay variance of 6% if ST is assigned to every gate in the circuit. We perform our two-phase \textit{FGSTI} technique by first using the MLP model to get \( ST(v) \) for all the gates in the circuit and then solve the LP model to get the optimal \((W/L)\), based on the results of \( ST(v) \). The MLP model to simultaneously determine ST placement and sizing are also solved using the same LP solver under a same set of parameters for comparison with the two-phase \textit{FGSTI} technique.

#### 4.2 Results for two-phase \textit{FGSTI} technique

For 0%, 3%, 5% circuit slowdown, we can not get a valid solution from conventional fixed slowdown method [12]. Thus the
leakage current saving for 0%, 3%, 5% circuit slowdown are compared between our two-phase FGSTI technique and MLP method [1]. As shown in table 3, our two-phase FGSTI technique can achieve 78.91% leakage saving even when the circuit slowdown is 0%. When the circuit slowdown is 3%, 5%, the leakage saving of our two-phase FGSTI technique is 92.55%, 97.97% respectively. Because of less approximation in ST placement phase, more ST’s can be assigned to different gates and additional leakage saving is achieved. The leakage saving is about on average 2% more than the MLP method [1].

In table 4, we show that our two-phase FGSTI technique can achieve a very impressive runtime saving. As the LP model for ST sizing only need seconds to solve, the runtime saving is largely due to two reasons: one is the two-phase procedure of FGSTI technique and the other is less variables and constraints used in MLP model for ST placement. For circuit C432, there are only 271 constraints and 338 variables in our MLP model for ST placement; however, in [1] there are 2975 constraints and 1183 variables. Although the MLP problem still need a long time to solve, as we can see from some of the benchmark, especially the small ones, our two-phase FGSTI technique can achieve at least 10X runtime saving. We only list the results of 4 benchmarks, because other benchmarks take hours to get the optimal results. The stopping time criteria is set to 4 hours for larger circuits. Heuristic algorithms can get near optimal results with a very fast speed, but as we all know the heuristic may lead to local optimal and can not guarantee the optimality of the result; thus we can use the results of LP models as a reference.

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 changed to one: ST sizing, while ignoring the ST placement. Our LP model for ST sizing leads to a same result with the fixed slowdown method and the MLP method in table 5. With 7% circuit slowdown, our ST sizing LP model causes 75.48% ST area saving compared to fixed slowdown method and the result is almost the same with MLP method. In table 5, ST area is calculated using equation (15), just summing up all the \((W/L)_i\), since the length of ST is a constant.

<table>
<thead>
<tr>
<th>ISCAS benchmark circuits</th>
<th>ST placement</th>
<th>ST sizing</th>
<th>Total</th>
<th>Two-phase FGSTI</th>
<th>Two-phase FGSTI</th>
<th>MLP</th>
<th>MLP</th>
<th>Fixed slowdown</th>
<th>Fixed slowdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>C432</td>
<td>0.491</td>
<td>1.234</td>
<td>1.725</td>
<td>32.047</td>
<td>1.502</td>
<td>1.593</td>
<td>3.095</td>
<td>1905.094</td>
<td>0.551</td>
</tr>
<tr>
<td>C499</td>
<td>0.856</td>
<td>1.454</td>
<td>2.310</td>
<td>2400.862</td>
<td>2.396</td>
<td>2.403</td>
<td>165825.17</td>
<td>22.016</td>
<td></td>
</tr>
<tr>
<td>C800</td>
<td>1.351</td>
<td>1.919</td>
<td>3.270</td>
<td>2931.374</td>
<td>3.221</td>
<td>3.498</td>
<td>6.665</td>
<td>3528.766</td>
<td>0.878</td>
</tr>
<tr>
<td>C578</td>
<td>0.315</td>
<td>0.707</td>
<td>1.022</td>
<td>211734.246</td>
<td>211734.246</td>
<td>211734.246</td>
<td>1.334</td>
<td>211734.246</td>
<td>211734.246</td>
</tr>
</tbody>
</table>

4.3 ST type Consideration

From table 3, when the circuit slowdown is below 6%, not all the gates in the circuit can be assigned with ST. FGSTI technique can cause a gate with ST to drive a gate without ST which leads to a floating state at the output of the gate with ST and large power dissipation in the gate without ST. As mentioned in [18], leakage feedback gate structure [23] shown in figure 4 is used in order to avoid the floating states. We assume that the leakage feedback structure can achieve the same delay as the normal ST insertion with a larger area and dynamic power consumption penalty. We examine all the gates with ST in the circuits and find out how many gates with ST should be changed into leakage feedback structure.

In table 6, when the circuit slowdown is 0%, about 82.75% of the total gates can change into gate with ST, and about 37.10% of the gates with ST should change into leakage feedback structure. When circuit slowdown is 3% and 5%, about 93.47% and 98.03% of the total gates can be changed into gates with ST, and only 19.78% and 9.90% of them should be changed into leakage feedback structure respectively. When the circuit slowdown is 0%, some of the benchmarks, such as C499, C1355, need to change 80.41% and 66.41% of original ST into leakage feedback struct-

### Table 3. Leakage current comparison between two-phase FGSTI and MLP method

| ISCAS85 benchmark circuits | Origin \(I_{leak}\) (pA) | Total gate Num. | ST gate Num. | Two-phase FGSTI | Two-phase FGSTI | MLP | MLP | Fixed slowdown |
|---------------------------|--------------------------|-----------------|--------------|-----------------|-----------------|-----|-----|----------------|----------------|
| C432                      | 4009.417                 | 169             | 1759.291     | 130             | 3643.911        | 127 | 1463.719 | 131             | 205.459         |
| C499                      | 21374.955                | 204             | 14479.83     | 97              | 14749.294       | 101 | 1451.785 | 102             | 257.367         |
| C800                      | 9261.315                 | 363             | 8194.241     | 152             | 8171.241        | 152 | 1024.245 | 154             | 126.149         |
| C1355                     | 1874.533                 | 546             | 474.856      | 308             | 872.238         | 307 | 320.238 | 307             | 320.238         |
| C1908                     | 5519.285                 | 1229            | 5156.976     | 1235            | 5192.020        | 1235 | 387.976 | 1235            | 387.976         |
| C3540                     | 4059.852                 | 1862            | 2040.401     | 1817            | 2351.211        | 1817 | 1201.211 | 1817            | 1201.211        |
| C3515                     | 6629.285                 | 2329            | 1669.947     | 2253            | 1841.916        | 2253 | 78.946 | 2253            | 78.946          |
| C2670                     | 4904.834                 | 2407            | 3277.753     | 1948            | 4065.315        | 1956 | 2543.315 | 1956            | 2543.315        |
| C7528                     | 9522.934                 | 3566            | 3012.342     | 3415            | 4190.606        | 3383 | 1230.606 | 3383            | 1230.606        |

### Table 4. Runtime comparison between two-phase FGSTI and MLP method (Time in s)

<table>
<thead>
<tr>
<th>ISCAS85 benchmark circuits</th>
<th>Two-phase FGSTI</th>
<th>MLP</th>
<th>Fixed slowdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST placement</td>
<td>ST sizing</td>
<td>Total</td>
<td>Two-phase FGSTI</td>
</tr>
<tr>
<td>C432</td>
<td>0.491</td>
<td>1.234</td>
<td>1.725</td>
</tr>
<tr>
<td>C499</td>
<td>0.856</td>
<td>1.454</td>
<td>2.310</td>
</tr>
<tr>
<td>C800</td>
<td>1.351</td>
<td>1.919</td>
<td>3.270</td>
</tr>
<tr>
<td>C578</td>
<td>0.315</td>
<td>0.707</td>
<td>1.022</td>
</tr>
<tr>
<td>C2670</td>
<td>0.431</td>
<td>1.077</td>
<td>1.508</td>
</tr>
</tbody>
</table>
Figure 4. Leakage feedback structure.

There are still some unsolved problems in FGSTI technique as our future work. Fast heuristic algorithms are needed for ST placement phase because the MLP model is very time consuming and can not handle large circuits. Furthermore, the detailed comparison between FGSTI and BBSTI techniques should be carefully examined in the physical level, such as place and routing penalty.

6. References


[22] http://groups.yahoo.com/group/lp_solve/