CLOCK DISTRIBUTION NETWORKS WITH GRADUAL SIGNAL TRANSITION TIME RELAXATION FOR REDUCED POWER CONSUMPTION

SHERIF A. TAWFIK
Department of Electrical and Computer Engineering,
University of Wisconsin — Madison,
Madison, Wisconsin 53706-1691
tawfik@wisc.edu

VOLKAN KURSUN
Department of Electronic and Computer Engineering,
The Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong

Revised 5 May 2008

Clock distribution network consumes a significant portion of the total chip power since the clock signal has the highest activity factor and drives the largest capacitive load in a synchronous integrated circuit. A new methodology is proposed in this paper for buffer insertion and sizing in an H-tree clock distribution network. The objective of the algorithm is to minimize the total power consumption while satisfying the maximum acceptable clock transition time constraints at the leaves of the clock distribution network for maintaining high performance. The new methodology employs nonuniform buffer insertion and progressive relaxation of the transition time requirements from the leaves to the root of the clock distribution network. The proposed algorithm provides up to 30% savings in the total power consumption without sacrificing clock skew as compared to a standard algorithm with uniform buffer insertion aimed at maintaining uniform transition time constraints at all the nodes of a clock tree in a 180 nm CMOS technology.

Keywords: Clock-tree; slew rate; low-power clocking; buffer insertion; short-circuit power; dynamic switching power.

1. Introduction

Clock distribution network consumes a significant portion of the power, area, and metal resources of an integrated circuit (IC). The enhancement of clock frequency and the growth of die size cause the power consumption of the clock distribution subsystem to increase significantly with each new technology generation. The power consumption of the clock distribution network is estimated to contribute 30–40% to the total power in today’s state-of-the-art high-performance microprocessors.1,6
Power consumption in a clock distribution network can be reduced by the optimal placement and sizing of clock buffers for a given skew budget,\textsuperscript{7} clock gating,\textsuperscript{8} and/or distributing the clock signal with a reduced voltage swing.\textsuperscript{5-12} Few previous works on clock distribution network synthesis are aimed at reducing the power consumption of a clock distribution network under clock transition time constraints. Signal transition time is defined as the duration for a signal to change from 10\% to 90\% (or 90\% to 10\%) of the full voltage swing. The signal slew rate is the reciprocal of the signal transition time. Maintaining sharp clock signal transition times is critical for high-speed operation and robustness of the clocked elements. Shorter clock transition time, however, typically requires an increased number of larger buffers, thereby leading to higher power consumption.

In Ref. 2, buffers with a fixed (predetermined) size are inserted on a given routed clock tree with the objective of minimizing the total number of utilized buffers subject to a given upper boundary on the transition time of the clock signal at the input of all the buffers. Short-circuit power is ignored and buffer sizing is not considered in Ref. 2. The constraint on the transition time of the clock signal at the input of the buffers is assumed to be supplied by the user. The criteria for the choice of the transition time constraints at the internal nodes of the clock distribution network are not presented in Ref. 2.

In Ref. 3, buffer insertion, buffer sizing, and wire sizing are performed on a routed H-tree clock distribution network with the objective of minimizing the total power consumption including the short-circuit power. The transition times of the clock signal at the inputs of the inserted buffers are constrained to be below a set of given thresholds. Buffers are inserted only at the beginning of the branches in the clock tree. Buffer and wire sizes are selected from a given discrete set. Restricting the buffer insertion to be performed only at the beginning of each branch produces a sub-optimal design. Furthermore, the criteria for choosing the transition time constraints at the internal nodes of the tree are not presented.

A nonuniform buffer insertion technique is proposed in Ref. 4 in order to achieve sharper signal transition times at the receiver end of long wires at the expense of higher power consumption. All the buffers except the buffer closest to the receiver end are uniformly spaced. The buffer closest to the receiver end drives a shorter wire. The power consumption is increased due to the higher short-circuit current produced by the last stage buffer with this technique. The nonuniform buffer insertion technique proposed in Ref. 4 can potentially reduce the power consumption provided that the signal transition time at the receiver end can be maintained similar to a wire with uniformly spaced buffers.

In this paper, a new buffer insertion and sizing algorithm is proposed for an H-tree clock distribution network. The objective of the proposed algorithm is to minimize the total power consumption (including the short-circuit power component) while maintaining sharp clock edges at the leaves of the clock distribution network. Nonuniform buffer insertion is performed with the algorithm proposed in this paper, thereby lowering the total power consumption. The power consumption
is further reduced by gradually relaxing the transition time constraints from the leaves to the root of the clock distribution network.

The paper is organized as follows. The power consumption and signal transition time trade-offs in buffered wires are discussed in Sec. 2. The proposed buffer insertion and sizing algorithm is described in Sec. 3. Experimental results are provided in Sec. 4. A branch and bound formulation is presented in Sec. 5. Finally, conclusions are offered in Sec. 6.

2. Buffered Wire Segment

In this section, the power consumption and signal transition time trade-offs in buffered wires are explored. A buffered wire segment is shown in Fig. 1. An inverter \((I_D)\) drives the wire segment of length \((L)\) and width \((W)\). The wire is loaded with an inverter \((I_L)\) at the receiver end. Both inverters are sized \(S\) times a unit size inverter (unit inverter: \(W_n = W_{\text{min}}\) and \(W_p = 2.5 \times W_{\text{min}}\)). An input signal having a transition time \(T_{\text{r-in}}\) is applied to the driver. The total power consumption and the transition time \(T_{\text{r-out}}\) at the receiver end (Node 1) are measured for different \({T_{\text{r-in}}, S, \text{and } L}\) combinations. The wire width is fixed at 0.25 \(\mu\)m. The simulations are carried out in a 0.18 \(\mu\)m CMOS technology.

The dependence of the output transition time and the average power consumption on the uniform size of the buffers for different wire lengths are illustrated in Figs. 2 and 3, respectively. As shown in Fig. 2, the output transition time decreases with the increased size of the inverters until a minimum output transition time is reached. Further increase in the uniform buffer size degrades the output slew rate due to the higher load imposed by \(I_L\). The target output transition time cannot be achieved with sizing alone when the minimum transition time is higher than the target value for certain wire lengths. For these specific cases, reducing the wire length between the inverters is necessary. This is achieved by inserting more inverters along the wire. This is analogous to moving from the upper to the lower curves in Fig. 2. The average power consumption of the buffered wire segment increases monotonically with the buffer size and the wire length as shown in Fig. 3.
Fig. 2. The signal transition time at Node 1 for different buffer sizes and wire lengths. The input transition time is 100 ps.

Fig. 3. The average power consumption for different buffer sizes and wire lengths. The input transition time is 100 ps.

The variation of the output transition time with the input transition time and the uniform buffer size is illustrated in Fig. 4. The spacing between the different curves is narrow in Fig. 4, indicating a relatively weak dependence of the receiver end transition time on the input slew rate of $I_D$. When the input transition time is relaxed by 200 ps (from 50 ps to 250 ps), the maximum increase in the receiver end transition time is 31.8 ps. This result indicates that it is possible to significantly relax the transition time constraints at the internal nodes all the way up to the
final leaves in a clock tree, while successfully providing sharp clock edges to the
clocked elements beyond the leaves.

As shown in Fig. 5, the power consumption increases with longer input transition
time due to the higher short-circuit current. In Figs. 2 and 4, the maximum buffer
size (beyond which the output transition time starts to increase) does not vary
considerably with the wire length and the input transition time. This observation is
useful to determine the maximum buffer size while constructing a clock distribution network.

Advantages of relaxing the transition time constraints at the internal nodes of a multi-stage buffered tree for reducing the overall power consumption are explored next. Two cascaded buffered wire segments are shown in Fig. 6. Initially, the buffer insertion and sizing for the two segments are performed such that the transition times at Node\(_1\) and Node\(_2\) are equal. If the transition time at Node\(_1\) is allowed to increase, the sizes or the number of the buffers in buffer-wire\(_1\) can be smaller leading to a reduction in the power consumption of buffer-wire\(_1\). The degradation of the slew rate at Node\(_1\), however, increases the short-circuit power consumption in buffer-wire\(_2\). Furthermore, to maintain the same transition time at Node\(_2\), the sizes or the number of the buffers of buffer-wire\(_2\) need to be larger thereby further increasing the power consumption in buffer-wire\(_2\). The total power consumption can be reduced by imposing unequal transition time constraints at different nodes provided that the reduction in the power consumption of buffer-wire\(_1\) dominates the increase in the power consumption of buffer-wire\(_2\) by appropriately adjusting (relaxing) the targeted slew rate at the internal nodes.

As a numerical example, assume that the length (width) of each wire is 0.125 mm (0.5 \(\mu\)m) in Fig. 6. A single buffer is inserted at the beginning of each wire. The transition time of the input signal is 100 ps and the transition time at Node\(_2\) is restricted to be equal to 100 ps. When the transition time at Node\(_1\) is also restricted to be equal to 100 ps, the power consumed in buffer-wire\(_1\) is 41.5 \(\mu\)W (41.5 \(\mu\)W). Alternatively, when the transition time at Node\(_1\) is relaxed to 140 ps while maintaining 100 ps transition time on Node\(_2\), the power consumed in buffer-wire\(_1\) is 31 \(\mu\)W (45 \(\mu\)W). The total power consumption is thereby reduced by 9.2\% with the relaxation of the transition time constraint at Node\(_1\).

### 3. Buffer Insertion and Sizing Algorithm

In this section, the new buffer insertion and sizing algorithm is presented. The proposed algorithm is evaluated using time-domain SPICE simulations automated with PERL scripts. The proposed algorithm employs nonuniform buffer insertion and progressive transition time relaxation at the internal nodes of an H-tree clock distribution network for reducing the total power consumption.
The implementation framework is as follows. Two PERL subroutines are employed at the lowest level. The first subroutine takes a set of circuit parameters and generates a SPICE input file that includes the circuit netlist and the measurement cards. The second subroutine parses the SPICE output file and returns the values of the measured quantities (e.g., total power consumption and transition times at specific nodes). SPICE simulations are run in between these two subroutines. At the next higher level, buffer insertion and sizing for a single wire are performed with a new algorithm denoted by BISW (Buffer Insertion and Sizing for a Wire). At the top level, buffer insertion and sizing for an H-tree clock distribution network are executed through a new algorithm denoted by BIST (Buffer Insertion and Sizing for a Tree).

The terminology employed in describing the algorithms is as follows. The given H-tree clock distribution network is composed of $m$ levels. For example, the H-tree clock distribution network shown in Fig. 7 is composed of four levels. Each level is labeled with a number, starting with 1 at the leaves. The length of a wire is $L$. The buffer size is $S$. The buffer size is assumed to be continuous with a lower (upper) boundary of $S_{\text{min}}$ ($S_{\text{max}}$). The minimum spacing between the buffers is $x$. All the buffers are uniformly spaced along a wire (uniform distance = $y$) except the last buffer that drives a wire of length $[L \mod y]$.

The flowchart of the buffer insertion and sizing algorithm for a single wire (BISW) is shown in Fig. 8. The input parameters are the wire length ($L$), input transition time ($T_{\text{r-in}}$), target output transition time (Target), minimum buffer size ($S_{\text{min}}$), maximum buffer size ($S_{\text{max}}$), minimum spacing between the buffers ($x$), and the size of the load inverter ($S_L$). The algorithm is iterative. For the first iteration $S$ is set to the minimum value $S_{\text{min}}$. The buffer spacing ($y$) is set to the maximum

![Fig. 7. An H-tree clock distribution network. The index of each level is indicated with the circled numbers. The squares represent the fixed loads at the leaves.](image-url)
possible value which is the length of the wire $L$. The netlist of the circuit is then created such that a buffer of size $S$ is inserted with periodic distances of $y$ along the wire. At the end of the wire a load inverter of size $S_L$ is inserted. The file is simulated in SPICE. The output transition time ($T_{r-out}$) and the power consumption are measured. The buffer size is doubled in an inner loop until $T_{r-out}$ is below the
Target or the buffer size exceeds $S_{\text{max}}$. If $S_{\text{max}}$ is reached with $T_{\text{r-out}} > \text{Target}$, it is concluded that sizing alone is not sufficient to reach the target output transition time and the current iteration is exited. If $T_{\text{r-out}}$ is below Target, a binary search for $S$ is performed such that $T_{\text{r-out}}$ is within 1% of Target. For the binary search to be valid, $T_{\text{r-out}}$ must be a monotonically decreasing function of $S$. This condition is satisfied by the proper selection of $S_{\text{max}}$ as presented in Sec. 2. $S$, $y$, and average power consumption are recorded at the end of the binary search. If $y$ is larger than $2 \cdot x$, the value of $y$ is decreased by $x$ and the next iteration is performed. If $y$ is less than $2 \cdot x$, then all the iterations are performed. The values of $y$ and $S$ that provide the minimum power consumption are reported as the optimum solution with respect to the generated solutions.

For buffer insertion and sizing in an H-tree clock distribution network (BIST), the BISW algorithm is repeatedly employed with a bottom-up tree traversal starting from level 1. The symmetry of the H-tree clock distribution network is exploited to reduce the complexity of the algorithm. The branches in each level are identical. Therefore, it is sufficient to solve the buffering problem for a single branch in each level. The BIST algorithm is listed in Table 1. Nonequal signal transition times

### Table 1. Pseudo-code of the buffer insertion and sizing algorithm for an H-tree clock distribution network (BIST).

<table>
<thead>
<tr>
<th>Algorithm BIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
</tr>
<tr>
<td>- H-tree network with $m$ levels</td>
</tr>
<tr>
<td>- Signal transition time at the root ($T_{\text{r-m}}$)</td>
</tr>
<tr>
<td>- Signal transition times at the leaves ($T_{\text{r-0}}$)</td>
</tr>
<tr>
<td>- Maximum signal transition time ($T_{\text{max}}$)</td>
</tr>
<tr>
<td>- Size of the load inverters at the leaves ($S_0$)</td>
</tr>
<tr>
<td>- Minimum spacing between the buffers ($x$)</td>
</tr>
<tr>
<td>- Minimum buffer size ($S_{\text{min}}$)</td>
</tr>
<tr>
<td>- Maximum buffer size ($S_{\text{max}}$)</td>
</tr>
<tr>
<td>Step 1 Compute the transition time at the inputs of each level using the formula</td>
</tr>
<tr>
<td>$T_{\text{r-i}} = T_{\text{r-0}} + i \cdot \frac{(T_{\text{max}} - T_{\text{r-0}})}{(m-1)}$, $i = 1, 2, \ldots, m-1$</td>
</tr>
<tr>
<td>Step 2</td>
</tr>
<tr>
<td>- Select a wire at level 1</td>
</tr>
<tr>
<td>- Set the size of the load inverter to $S_0$</td>
</tr>
<tr>
<td>- Call BISW</td>
</tr>
<tr>
<td>- Record the optimum buffer size $S$ and spacing $y_1$</td>
</tr>
<tr>
<td>- For $i = 2 \rightarrow m$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>End For</td>
</tr>
<tr>
<td>- Simulate the entire clock distribution tree and measure the power consumption</td>
</tr>
<tr>
<td>Outputs</td>
</tr>
</tbody>
</table>
at the internal nodes are achieved by progressively increasing the transition time requirement from the target value at the leaves to a maximum transition time at the output of the last level (closest to the root). The maximum acceptable transition time is determined by the signal integrity requirements (the clock signal becomes triangular in shape for transition times larger than approximately 30% of the clock period assuming a 50% duty cycle). The algorithm invokes the BISW algorithm with the parameters of the branch at level 1. The optimum solution for that level is recorded. The load for the next level is twice the load represented by the buffer size of level 1 since the tree is binary. The algorithm then invokes BISW for a branch in level 2 with the new load. The process is repeated until all the levels are visited. The final design is then simulated and the total power consumption is reported.

The complexity of the proposed algorithm is estimated by the number of times SPICE is invoked as a function of the number of clocked elements. If the number of clocked elements is \( n \), then the number of levels in the H-tree is

\[
m = \lceil \log_2(n) \rceil.
\]

(1)

The number of times SPICE is invoked for each call to BISW is

\[
C_{\text{BISW}} = \left\lfloor \frac{L}{x} \right\rfloor \times \left\lceil \log_2 \left( \frac{S_{\text{max}}}{S_{\text{min}}} \right) \right\rceil \times k,
\]

(2)

where \( L \) is the length of the wire, \( x \) is the minimum spacing between the buffers, \( S_{\text{min}} \) is the minimum buffer size, \( S_{\text{max}} \) is the maximum buffer size, and \( k \) is the number of times SPICE is invoked in the binary search. In order to find an upper boundary for \( k \), an extra exit condition for the binary search is needed when the difference between the upper \( (S_u) \) and the lower \( (S_L) \) boundaries used in the binary search (see Fig. 8) is below a given threshold \( S_d \). At the beginning of the binary search, \( S_L \) is one half of \( S_U \). The worst case occurs when \( S_u = S_{\text{max}} \) and \( S_L = S_{\text{max}}/2 \). In this case, the upper boundary for \( k \) is

\[
k = \left\lfloor \log_2 \left( \frac{S_{\text{max}}}{2S_d} \right) \right\rceil.
\]

(3)

By combining Eqs. 1–3, the upper boundary for the complexity of the BIST algorithm is

\[
C_{\text{BIST}} = \left\lfloor \log_2(n) \right\rfloor \times \left\lfloor \frac{L_{\text{max}}}{x} \right\rfloor \times \left\lceil \log_2 \left( \frac{S_{\text{max}}}{S_{\text{min}}} \right) \right\rceil \times \left\lfloor \log_2 \left( \frac{S_{\text{max}}}{2S_d} \right) \right\rceil,
\]

(4)

where \( L_{\text{max}} \) is the length of the longest branch in the clock distribution network. The complexity of the proposed algorithm is therefore a logarithmic function of the number of clocked elements.

4. Experimental Results

The experimental results are given in this section. Comparisons are provided between the proposed algorithm based on nonuniform buffer spacing and progressively relaxed transition times from the leaves to the root of a clock tree versus
the standard approach of uniform buffer spacing and maintaining equal transition times at all the nodes. Comparisons at the nominal process corner with uniform die temperature are given in Sec. 4.1. The clock distribution networks are characterized under temperature and process parameter variations in Sec. 4.2.

4.1. Comparisons at the nominal process corner with uniform die temperature

Three test circuits are employed. The first test circuit is a 4-level H-tree network as shown in Fig. 7. The second and the third test circuits are 5- and 6-level H-tree networks, respectively. Each circuit spans a 20 mm x 20 mm die. The circuits are designed in a 180nm CMOS technology. The clock frequency is 1 GHz. A uniform die temperature of 125°C is assumed in this section. The input transition time at the root is 50 ps. The target transition time at the leaves is 100 ps. The size of the load inverters at the leaves is 2X the minimum size inverter. The largest inverter is 32X the minimum size inverter. For the minimum size inverter, the width of the NMOS (PMOS) transistor is 220 nm (550 nm). The minimum buffer spacing is taken as one tenth of the shortest branch. For the first circuit, the length (width) of the branches in levels 1, 2, 3, and 4 is 2.5 mm (0.5 µm), 2.5 mm (0.5 µm), 5 mm (1 µm), and 5 mm (1 µm), respectively. The parasitic resistance and capacitance are extracted for each wire segment. Each wire segment is modeled as a Π network.

As listed in Table 2, the proposed algorithm based on nonuniform buffer spacing and progressive transition time relaxation is effective for reducing the total power consumption. With the proposed techniques, the total power consumption

<table>
<thead>
<tr>
<th>Number of levels of H-tree</th>
<th>Average power consumption (mW)</th>
<th>Normalized average power consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Uniform buffer insertion Equal transition times</td>
<td>15.30</td>
</tr>
<tr>
<td></td>
<td>Nonuniform buffer insertion Gradually relaxed transition times</td>
<td>10.66</td>
</tr>
<tr>
<td>5</td>
<td>Uniform buffer insertion Equal transition times</td>
<td>21.89</td>
</tr>
<tr>
<td></td>
<td>Nonuniform buffer insertion Gradually relaxed transition times</td>
<td>15.24</td>
</tr>
<tr>
<td>6</td>
<td>Uniform buffer insertion Equal transition times</td>
<td>29.00</td>
</tr>
<tr>
<td></td>
<td>Nonuniform buffer insertion Gradually relaxed transition times</td>
<td>20.89</td>
</tr>
</tbody>
</table>
is reduced by up to 30% due to the utilization of smaller and fewer buffers as compared to a conventional network with uniform buffer insertion and equal transition time constraints.

4.2. Impact of process and temperature variations on clock skew

The impact of temperature and process parameter fluctuations on the clock skew is evaluated in this section for the clock distribution networks designed with equal transition times and uniform buffer insertion approach and the clock distribution networks designed with the proposed gradually relaxed transition times and nonuniform buffer insertion techniques. The impact of temperature and process parameter fluctuations is considered both independently and concurrently.

The effect of temperature fluctuations on circuit delay is presented for both CMOS gates and wires. The propagation delay of a CMOS gate driving a load capacitance $C_L$ considering the nonzero transition times of the inputs is:

$$t_p = \left(\frac{1}{2} - \frac{V_{DD} - V_{th}}{V_{DD} (1 + \alpha)}\right) T_{r-in} + \frac{C_L V_{DD}}{2 I_{D0}}, \quad (5)$$

where $V_{th}$ is the transistor threshold voltage, $V_{DD}$ is the supply voltage, $\alpha$ is the velocity saturation coefficient ($1 \leq \alpha \leq 2$), $T_{r-in}$ is the input transition time, and $I_{D0}$ is a function of the carrier mobility $\mu$ and the gate overdrive voltage ($V_{GS} - V_{th}$). The primary parameters in the delay equation that depend on temperature are the carrier mobility $\mu$ and the threshold voltage $V_{th}$. Both $\mu$ and $|V_{th}|$ decrease as the temperature increases. The degradation of mobility tends to increase the delay. Alternatively, the decrease in the absolute value of the threshold voltage tends to lower the delay. The combined effect of the mobility and threshold voltage variations on delay depends on other parameters such as the supply voltage and the input signal slope. Variations in gate overdrive are smaller as compared to the carrier mobility variations when the temperature fluctuates in circuits operating at the nominal supply voltage ($V_{DD}-nominal = 1.8$ V) in a 180 nm CMOS technology. The transistor saturation current and the circuit speed, therefore, degrade with the increased temperature.

The first term in Eq. (5) is proportional to the input signal transition time. This term is also a function of the gate overdrive ($V_{DD} - V_{th}$). As the input transition time increases, the significance of this term on the propagation delay is enhanced. Hence, the temperature fluctuations-induced variation of the gate overdrive becomes more effective on the delay variations at a higher input transition time. The increased significance of the gate overdrive on the delay leads to enhanced counterbalancing of the mobility degradation. Therefore, the temperature fluctuations-induced delay variations are suppressed with a higher input transition time.

The effect of the input transition time on the delay variations with temperature fluctuations is evaluated with the circuit shown in Fig. 9. The supply voltage is swept from 0.6 V to 1.8 V. The average propagation delay at 25°C and 125°C and the percent variation in the average propagation delay with temperature are
measured for each value of the supply voltage. The input signal transition time is assumed to be 50 ps and 160 ps in the first and second set of simulations, respectively. The temperature fluctuations-induced delay variation is reduced by 33% with a 160 ps input transition time as compared to a 50 ps input transition time at the nominal supply voltage ($V_{DD} = 1.8$ V), as shown in Fig. 9.

In clock distribution networks, long wires with significant resistance are employed. The effect of temperature on wire resistance and delay is therefore more pronounced in clock trees. Furthermore, the widths of the wires are scaled in order to enhance the integration density with each new technology generation. The increasing resistance of the wires plays an important role in determining the performance of scaled CMOS integrated circuits in the nanometer regime. Resistance of a metal wire increases approximately linearly with the temperature according to the formula\(^1\)

\[
R(T) = R_0(1 + k(T - T_0)) \, ,
\]

where $R_0$ is the resistance at temperature $T_0$, $T$ is the temperature of the wire, and $k$ is the temperature coefficient. The signal propagation delay across a wire, therefore, increases at a higher temperature.

Each clock distribution network is characterized for clock skew caused by a nonuniform die temperature profile. The temperature profile of an IC varies over time. Four on-chip temperature profiles are considered in this study for characterizing the clock skew as shown in Fig. 10. Both vertical and horizontal temperature gradients are considered with a different location for the hottest spot in each
temperature profile. The temperature range of each profile is 25°C (room temperature) to 125°C. The temperature of 125°C is reported as a typical hot spot temperature for the state-of-the-art microprocessors fabricated in a 180nm CMOS technology.\textsuperscript{15} For each temperature profile, the maximum clock skew between the leaves of the clock distribution networks is measured as listed in Table 3 and as depicted in Fig. 11. The temperature fluctuations-induced clock skew is reduced by up to 22% with the proposed gradually relaxed transition time and nonuniform buffer insertion technique as compared to the uniform buffer insertion and equal transition times approach.

The impact of process parameter variations on the clock skew with a uniform die temperature is evaluated next. The gate oxide thickness, the channel length, the channel doping, the transistor width, and the wires’ widths are assumed to have independent Gaussian distributions. The $3\sigma$ variation of the gate oxide is assumed to be 5%. Alternatively, the $3\sigma$ variations of the remaining parameters are assumed

---

Table 3. Clock skew of the different clock distribution networks with nonuniform temperature profiles.

<table>
<thead>
<tr>
<th>Temperature profile</th>
<th>Clock skew (ps)</th>
<th>Uniform buffer insertion equal transition times</th>
<th>Nonuniform buffer insertion gradually relaxed transition times</th>
<th>% reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four-level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>profile 1</td>
<td>37.0</td>
<td>29.6</td>
<td>20.0</td>
<td></td>
</tr>
<tr>
<td>profile 2</td>
<td>51.2</td>
<td>39.7</td>
<td>22.4</td>
<td></td>
</tr>
<tr>
<td>profile 3</td>
<td>38.0</td>
<td>29.6</td>
<td>22.1</td>
<td></td>
</tr>
<tr>
<td>profile 4</td>
<td>51.3</td>
<td>40.0</td>
<td>22.0</td>
<td></td>
</tr>
<tr>
<td>H-tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>profile 1</td>
<td>24.4</td>
<td>23.7</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>profile 2</td>
<td>37.0</td>
<td>33.5</td>
<td>9.4</td>
<td></td>
</tr>
<tr>
<td>profile 3</td>
<td>27.6</td>
<td>25.3</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>profile 4</td>
<td>37.2</td>
<td>33.8</td>
<td>9.1</td>
<td></td>
</tr>
<tr>
<td>Five-level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>profile 1</td>
<td>33.0</td>
<td>30.6</td>
<td>7.2</td>
<td></td>
</tr>
<tr>
<td>profile 2</td>
<td>44.5</td>
<td>39.7</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>profile 3</td>
<td>33.0</td>
<td>29.2</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>profile 4</td>
<td>44.8</td>
<td>39.9</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>H-tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Fig. 10. Four different on-chip temperature profiles considered in this study.
Clock Distribution Networks with Gradual Signal Transition Time Relaxation

Fig. 11. Clock skew for the nominal process corner and different nonuniform die temperature profiles.

Fig. 12. Clock skew distribution under process parameter variations (uniform die temperature).

...Monte Carlo simulations with 250 samples are run to evaluate the clock skew of the clock distribution networks. The skew distributions of different H-tree networks are shown in Fig. 12. The mean clock skew is increased by up to 36% with the clock distribution networks designed with the proposed gradual relaxation of transition time and unequal buffer insertion approach as compared to the clock distribution networks designed with standard equal transition time and uniform buffer insertion approach. This increase in clock skew is due to the increased dependence of the gate delays on the threshold voltage with a higher input transition time as explained earlier.
The impact of both process parameter variations and nonuniform die temperature are considered together next. The Monte Carlo simulations are rerun in this case with the nonuniform die temperature profiles shown in Fig. 10. The clock skew distributions for a four-level H-tree clock distribution network is shown in Fig. 13. The mean clock skew of the clock network designed with the proposed gradual relaxation of transition time and nonuniform buffer insertion approach is reduced by up to 14.4% as compared to a clock distribution network designed with equal transition time and uniform buffer insertion as shown in Fig. 13.

5. Branch and Bound Formulation

In this section, a branch and bound formulation of the buffer insertion and sizing problem is presented. The branch and bound formulation allows an exhaustive search to be performed on a reduced solution space. The solution space is reduced using pruning techniques. Two pruning techniques are presented in this section. Unlike the heuristic algorithm proposed in Sec. 3, the configuration obtained using the branch and bound formulation is guaranteed to be the optimum. The runtime
of the branch and bound algorithm, however, grows exponentially with the number of levels in the H-tree. The branch and bound algorithm is therefore practical only for small-sized problems.

The BIST heuristic algorithm presented in Sec. 3 is a greedy algorithm. The local optimum solution for the current level is selected regardless of the effect on the remaining levels. The size of the load inverter for the next higher level is twice the buffer size of the current level (binary tree). If a sub-optimal solution with a smaller buffer size is selected for the current level, the power consumption of the higher level will be reduced due to the smaller load buffers. A better global solution can be obtained if the reduction in the power consumption of the higher level is less than the increase in the power consumption of the current level. To reach a global optimum, a single solution cannot be selected at each level but rather all the solutions must be selected for the current level and propagated to the next level. This creates a tree of solutions and is called branching.

The branch and bound algorithm is illustrated in Fig. 14. At each level every possible buffer spacing \((y)\) is considered and the buffer sizing is performed using a binary search approach like the one employed in BISW. The tree of solutions shown in Fig. 14 is traversed depth first. At the last level the buffer sizes are determined for all the levels and the total power consumption for the clock distribution network is determined. The global optimum solution is determined at the end of the tree.

![Fig. 14. Illustration of the branch and bound algorithm.](image-url)
traversal by visiting all the nodes in the tree of solutions and by keeping a record of the minimum total power consumption.

Two pruning techniques are applied in order to avoid visiting all the nodes in the solution tree. With the first pruning technique, if a solution for the current level gives higher power consumption at a larger buffer size, this solution is not propagated to the next level to avoid incurring higher power consumption in both the current and the higher levels. With the second pruning technique, a branch in each level is initially and individually optimized with a minimum load represented by an inverter sized 2X the minimum inverter. The power consumption representing the minimum power consumption for each level is recorded. The minimum power consumption for the whole clock distribution network is initialized to infinity. During the solution tree traversal, if at a node the sum of the power consumption for the current level, previous levels, and the minimum power consumption of the subsequent levels is higher than the best solution obtained so far, then this new solution is rejected and not propagated to the following levels. To implement the second pruning technique a record of the minimum total power consumption is kept.

The minimum power consumption achievable, determined by the branch and bound algorithm for a 4-level H-tree clock distribution network, is 10 mW. The power consumption with the proposed BIST heuristic algorithm is within 6.6% of the absolute minimum achievable power consumption determined by the branch and bound formulation.

6. Conclusions

In this paper, a new heuristic algorithm is proposed for buffer sizing and insertion in an H-tree clock distribution network for minimizing the total power consumption while satisfying the transition time constraints at the leaves of the clock distribution network. The algorithm employs nonuniform buffer insertion and progressive relaxation of the transition time requirements from the leaves to the root of a clock tree in order to reduce the total power consumption. Up to 30% reduction in the power consumption is achieved without increasing the clock skew by applying these two techniques simultaneously to a clock distribution network. The algorithm is based on time-domain SPICE simulation instead of relying on simple and inaccurate circuit models. The proposed algorithm exploits the symmetry in the H-tree clock distribution network in order to reduce the runtime complexity. The runtime complexity of the algorithm grows logarithmically with the number of clocked elements.

References