Technology Mapping for Low Leakage Power and High Speed with Hot-Carrier Effect Consideration

Chang Woo Kang and Massoud Pedram
University of Southern California/EE-systems
ckang@usc.edu
2D-3

Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion
Introduction

Deep sub-micron CMOS Technology

Power dissipation (active + standby)  Reliability

Leakage Power Dissipation  Hot-Carrier Effect

Short battery life  Function failure

Technology independent logic synthesis

Input pattern estimation  Technology Mapping  Predict aging process

Physical Design

Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion
Subthreshold Current

- Becomes quite large as the threshold voltage and the channel length are reduced
- Increases exponentially as the threshold voltage is decreased linearly

\[ I_{\text{subthreshold}} = A \cdot \exp \left( \frac{1}{kT/q} (V_{gs} - V_{th} - \gamma \cdot V_{th}) \right) \]

Threshold Voltage Scaling

\[ \tau = \frac{CV_{dd}}{(V_{dd} - V_{TH})^\chi} \]

- \( V_{dd}/V_{th} = 3\sim4 \)
  - Satisfy the DC noise margin
  - Maintain circuit performance
- Supply voltage is scaled down with each technology generation
The amount of subthreshold current is strongly dependent on the applied binary input pattern.

Techniques for leakage power reduction:
- Transistor stacking
- Input vector assignment
- Dual-threshold voltage devices

Active Mode Leakage Power

<table>
<thead>
<tr>
<th>Technology</th>
<th>Active power</th>
<th>Active mode leakage power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25µm</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>0.18µm</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>0.13µm</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>0.10µm</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>0.07µm</td>
<td>250</td>
<td></td>
</tr>
</tbody>
</table>

[Intel]
Hot-Carrier Effect

- Injection of high-energy electrons into the gate oxide near the drain region
- Can result in a substantial device parameter degradation
  - Threshold voltage
  - Transconductance
  - Linear and saturation drain current

Hot-Carrier Effects

- Change of critical timing paths due to transistor degradation
- Critical from the viewpoint of long term system reliability
- Long saturation region during transition causes device aging
- Therefore, devices are worn out quickly by
  - Slow slew rate of inputs
  - High load output capacitance
  - High switching activity
Motivation

- Leakage power dissipation must be minimized during not only the standby mode but also the active mode
- Leakage power dissipation can be reduced during the logic synthesis phase
- Aging delay due to hot-carrier effect ought to be considered during the technology mapping step to improve the circuit reliability

Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion
**Gate Modeling**

**Dynamic Power**

\[
P_{dyn} = \frac{1}{2} \times C_{load} \times \frac{V_{dd}^2}{T_{cycle}} \times sw
\]

- **C\textsubscript{load}** capacitive load
- **V\textsubscript{dd}** supply voltage
- **T\textsubscript{cycle}** clock cycle time
- **sw** switching activity

**Gate Modeling**

**Leakage Power**

- Input vector probability
- Lookup subthreshold current for input vectors

\[I_{\text{subthreshold}} = \text{Subthreshold current when } U \text{ is applied}\]

\[Pr(U) = \text{Signal probability of having this input vector}\]
Gate Modeling

Leakage Power

- Average leakage power dissipation
- 3-INPUT NAND gate

<table>
<thead>
<tr>
<th>input</th>
<th>subthreshold current</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>[000]</td>
<td>2.30E-08</td>
<td>0.1</td>
</tr>
<tr>
<td>[001]</td>
<td>2.96E-08</td>
<td>0.2</td>
</tr>
<tr>
<td>[010]</td>
<td>2.96E-08</td>
<td>0.4</td>
</tr>
<tr>
<td>[011]</td>
<td>6.66E-08</td>
<td>0.1</td>
</tr>
<tr>
<td>[100]</td>
<td>2.91E-08</td>
<td>0</td>
</tr>
<tr>
<td>[101]</td>
<td>6.09E-08</td>
<td>0.1</td>
</tr>
<tr>
<td>[110]</td>
<td>6.03E-08</td>
<td>0.05</td>
</tr>
<tr>
<td>[111]</td>
<td>7.92E-08</td>
<td>0.05</td>
</tr>
</tbody>
</table>

\[ P_{\text{leak}} = V_{dd} \times \sum U [I_{\text{subthreshold}}(U) \times pr(U)] \]

Gate Modeling

Hot-carrier effect

\[ T_{\text{fresh}} = \tau_{\text{rig}} + R_{\text{rig}} C_{\text{load}} \]

\[ T_{\text{aged}} = T_{\text{fresh}} \times \left(1 + SW \times \alpha C_{\text{load}} \times \frac{\beta}{T_{\text{avg.slew}}} \right) \]

- \( T_{\text{aged}} \): aged input-to-output delay
- \( T_{\text{fresh}} \): fresh input-to-output delay
- \( T_{\text{avg.slew}} \): average input slew rate
- \( \alpha \): degradation factor due to the output load
- \( \beta \): degradation factor due to the input slew rate
- \( SW \): switching probability
Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion

Review

Technology Mapping

1. Adding two curves to generate one parent curve

2. Merging of power-delay curves

- Post-order traversal
  - Adding and lower-bound merging
- Pre-order traversal
  - Selecting the best point satisfying required time
Leakage Power Aware Mapping

\[ P(n, g) = \frac{1}{2} C_{\text{diff}}(n, g) \frac{V_{\text{dd}}^2}{T_{\text{cycle}}} + V_{\text{dd}} \sum_{i} I_{\text{subth}}(U) pr(U) \]

\[ + \sum_{n_{\text{in}}(n, g)} \left( \frac{1}{2} C_{\text{load}}(n) \frac{V_{\text{dd}}^2}{T_{\text{cycle}}} + \frac{P(n_{i}, g_{i})}{\text{fanout}(n)} \right) \]

- Total power dissipation in a node \( n \) mapped by gate \( g = \)
  
  1. Dynamic power dissipation on diffusion capacitance
  2. Leakage power dissipation
  3. Dynamic power dissipation on gate capacitance
  4. Dynamic power dissipation in transitive fanin/fanout

Dual-Threshold Cell Mapping

- Two types of library cells
  - High-threshold cells on non-critical timing paths
    - Low leakage but slow
  - Low-threshold cells on critical timing path
    - High leakage but fast

- Current practice is to initially map to high-threshold gates and then selectively replace some gate with low-threshold counterparts from the library
  - Less powerful

- In our approach, we do the selection of high versus low threshold devices during the mapping step
Dual-Threshold Cell Mapping

- Mapper naturally selects low-threshold cells for critical paths from the power-delay curve

![Graph showing power-delay curve with low-threshold and high-threshold regions]

Hot-carrier Effect Aware Mapping

- Predict the aging effect and apply it for arrival time computation

![Diagram illustrating aging effect and arrival time with violation and prediction markers]
Hot-carrier Effect Aware Mapping

arrival \((n, g, C_n)\) = \(\max_{\text{net inputs}(n, g)} (T_{aged} + \text{arrival}(n_i, g_i, C_i))\)

\[ T_{aged} = T_{fresh} \times \left( 1 + s \times \alpha \frac{C_{load} \beta}{T_{avg\_slew}} \right) \]

- Arrival time is the maximum delay from primary inputs to the gate output
- Aged delay is considered as a gate delay

Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion
High-Speed Heuristics

Primary Output Ordering

- Different order of mapping the logic cones generate different area, power dissipation, and speed
- Delay of a mapped cone is roughly proportional to the logic depth of the corresponding primitive cone
- Sort in descending order of their logic depths
- Simple heuristic, but effective

High-Speed Heuristics

Pin Permutation

- Careful pin assignment for signals can result in reduced propagation delay through a CMOS gate
- Equivalent pins belong to the same set
- Pin permutation is performed during technology mapping
Outline

- Introduction
- Background
- Gate Modeling
- Technology Mapping
- High-Speed Heuristics
- Simulation Results
- Conclusion

Simulation Setup

- LP-SIS
- 0.18µm CMOS technology commercial library
- Dual-threshold library set
  - High-threshold library : 0.4V
  - Low-threshold library : 0.2V
  - Typically, ~10X leakage difference
- Calculation of signal probabilities using ordered binary decision diagram (OBDD)
Results
Leakage Power Aware Mapping

- Power dissipation reduction due to leakage power aware mapping
- Low-threshold library cells are used
- Dynamic power increases but leakage power decreases significantly

Results
Power Dissipation

- Power dissipation reduction due to dual threshold gates
- High-threshold voltage cells on non-critical paths
- Low-threshold voltage cells on critical paths
- Low total power dissipation
Results

Delay

Circuit speedup by mapping low-threshold voltage gates on critical paths

Area

Gate area reduction by mapping low-threshold voltage gates on the critical paths
## Results

### Aging-aware Mapping

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Aging-unaware mapping</th>
<th>Aging-aware mapping</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>C432</td>
<td>2.21</td>
<td>2.54</td>
<td>2.08</td>
</tr>
<tr>
<td>C499</td>
<td>1.52</td>
<td>1.74</td>
<td>1.5</td>
</tr>
<tr>
<td>C880</td>
<td>1.39</td>
<td>1.41</td>
<td>1.33</td>
</tr>
<tr>
<td>C1355</td>
<td>1.91</td>
<td>2.09</td>
<td>1.72</td>
</tr>
<tr>
<td>C1908</td>
<td>2.65</td>
<td>2.82</td>
<td>2.45</td>
</tr>
<tr>
<td>C2670</td>
<td>1.89</td>
<td>1.93</td>
<td>1.81</td>
</tr>
<tr>
<td>count</td>
<td>2.18</td>
<td>2.13</td>
<td>2.06</td>
</tr>
<tr>
<td>sqrt18ml</td>
<td>1.93</td>
<td>2.06</td>
<td>1.97</td>
</tr>
<tr>
<td>f51m</td>
<td>2.01</td>
<td>2.04</td>
<td>1.89</td>
</tr>
<tr>
<td>ahu2</td>
<td>2.1</td>
<td>2.14</td>
<td>2.03</td>
</tr>
<tr>
<td>i2</td>
<td>0.49</td>
<td>0.5</td>
<td>0.45</td>
</tr>
<tr>
<td>i7</td>
<td>1.13</td>
<td>1.14</td>
<td>1.1</td>
</tr>
</tbody>
</table>

- Aging-aware mapping enhances system reliability by limiting degraded delay during technology mapping.

### Primary Output Ordering

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Without PO ordering</th>
<th>With PO ordering</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area (µm²)</td>
<td>Power (µW)</td>
<td>Delay (ns)</td>
</tr>
<tr>
<td>C432</td>
<td>2219</td>
<td>82</td>
<td>2.48</td>
</tr>
<tr>
<td>C499</td>
<td>3881</td>
<td>136.9</td>
<td>1.65</td>
</tr>
<tr>
<td>C880</td>
<td>4232</td>
<td>170.4</td>
<td>1.84</td>
</tr>
<tr>
<td>C1355</td>
<td>6390</td>
<td>348.5</td>
<td>2.04</td>
</tr>
<tr>
<td>sqrt18ml</td>
<td>385</td>
<td>14.5</td>
<td>0.84</td>
</tr>
<tr>
<td>f51m</td>
<td>1269</td>
<td>56</td>
<td>2.09</td>
</tr>
<tr>
<td>ahu2</td>
<td>4283</td>
<td>131.1</td>
<td>2.93</td>
</tr>
<tr>
<td>i7</td>
<td>5271</td>
<td>142.4</td>
<td>1.51</td>
</tr>
</tbody>
</table>

- In general, large deep cone can be mapped first to provide maximum mapping flexibility.
## Results

### Pin Permutation

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Without Pin Permutation</th>
<th>With Pin Permutation</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area (µm²)</td>
<td>Power (µW)</td>
<td>Delay (ns)</td>
</tr>
<tr>
<td>C432</td>
<td>3207</td>
<td>101.5</td>
<td>2.27</td>
</tr>
<tr>
<td>C499</td>
<td>3395</td>
<td>125.8</td>
<td>1.53</td>
</tr>
<tr>
<td>C880</td>
<td>2637</td>
<td>155.1</td>
<td>1.9</td>
</tr>
<tr>
<td>CT355</td>
<td>3428</td>
<td>370.8</td>
<td>1.61</td>
</tr>
<tr>
<td>r4ml</td>
<td>402</td>
<td>45.9</td>
<td>0.57</td>
</tr>
<tr>
<td>f51m</td>
<td>1026</td>
<td>79.82</td>
<td>1.81</td>
</tr>
<tr>
<td>i2n2</td>
<td>2950</td>
<td>148.3</td>
<td>2.58</td>
</tr>
<tr>
<td>i2</td>
<td>1712</td>
<td>136.3</td>
<td>0.48</td>
</tr>
<tr>
<td>i7</td>
<td>4128</td>
<td>109.1</td>
<td>1.64</td>
</tr>
</tbody>
</table>

- Delay is always reduced, but in some cases, area and power go up

**ASP-DAC 2003** 35/36

## Conclusion

- A technology mapping technique considering leakage power dissipation was presented
- 52% reduction in leakage power dissipation and 27% decreases in total power dissipation was achieved
- An aging delay model capturing the hot-carrier effect was proposed and used during technology mapping
- 10.6% reduction in aged delay of the circuits was achieved
- Primary output ordering and pin permutation were effective in improving the circuit speed

**ASP-DAC 2003** 36/36