An Integer Programming Placement Approach to FPGA Clock Power Reduction

Alireza Rakhshanfar and Jason Anderson
Dept. of Electrical and Computer Engineering
University of Toronto

IEEE/ACM ASP-DAC
Yokohama, Japan

January 2011
Mobile Devices
Motivation

- FPGA dynamic power is 7-14x higher than ASICs [Kuon TCAD’07].
  - Long metal wire segments.
  - Overhead of programmability.
- Power consumption excludes them from low-power applications such as mobile devices.
- Significant reductions are needed to close gap with ASICs.
FPGA Power Consumption

• Dynamic power is a function of toggle rates and capacitances of signals.

• Clock power is a major source of power dissipation due to their high fan-out and toggle frequency.
  – Accounts for 20-39% of dynamic power consumption [Degalahal ASP-DAC’05].

• This work: CAD technique to reduce FPGA clock power.
FPGA Clock Network

Xilinx Virtex-5 family

Root Spines

Clock Region

Vertical Half-Spines

Horizontal Spines

Logic Block
Spine Reduction

4 Clocks: A, B, C and D

Spine Count: 9

Vertical Half-Spines

Horizontal Spines
Spine Reduction
Spine Reduction

Spine Count: 4

Reduction: 56%
Related Work

• Prior work at Actel Corp. [Hsu IET-CDT’10] has used this technique in annealing-based placement framework.
  – Annealing works with any cost function.

• However, many placement algorithms use analytical techniques.

• Our technique is a post-placement task:
  – Can be incorporated in any placement algorithm.
Integer Linear Programming

- Formal optimization technique with broad applications.
- Linear objective function.
- Linear equality and inequality constraints.
- NP-hard to solve.
- ILP Solver: lp_solve.
Our Approach

Binary Variables: clkA_1, clkA_2, clkB_1, clkB_2

clkA_1 is true if clock A is used in the first column

Block M: Clock A
Block N: Clock B
Our Approach

Binary Variables:
- clkA_1, clkA_2
- clkB_1, clkB_2
- blkM_11, blkM_12, blkM_21, blkM_22
- blkN_11, blkN_12, blkN_21, blkN_22

blkM_11 is true if block M is placed in location 1,1.

Block M: Clock A
Block N: Clock B
Our Approach

Binary Variables:  
clkA_1, clkA_2, clkB_1, clkB_2  
blkM_11, blkM_12, blkM_21, blkM_22  
blkN_11, blkN_12, blkN_21, blkN_22

Objective Function:  
\[ \min \ clkA_1 + clkA_2 + clkB_1 + clkB_2 \]

Block M: Clock A  
Block N: Clock B
Our Approach

Objective Function: \( \text{min} \clkA_1 + \clkA_2 + \clkB_1 + \clkB_2 \)

Binary Variables: \( \clkA_1, \clkA_2, \clkB_1, \clkB_2 \)
\( \blkM_{11}, \blkM_{12}, \blkM_{21}, \blkM_{22} \)
\( \blkN_{11}, \blkN_{12}, \blkN_{21}, \blkN_{22} \)

Placement:

Block M: Clock A
Block N: Clock B

\( \blkM_{11} + \blkM_{12} + \blkM_{21} + \blkM_{22} = 1 \)
\( \blkN_{11} + \blkN_{12} + \blkN_{21} + \blkN_{22} = 1 \)
Our Approach

Exclusivity:

\[
\begin{align*}
\text{blkM}_{12} + \text{blkN}_{12} & \leq 1 \\
\text{blkM}_{21} + \text{blkN}_{21} & \leq 1 \\
\text{blkM}_{22} + \text{blkN}_{22} & \leq 1 \\
\text{blkM}_{11} + \text{blkN}_{11} & \leq 1
\end{align*}
\]

Placement:

\[
\begin{align*}
\text{blkM}_{11} + \text{blkM}_{12} + \text{blkM}_{21} + \text{blkM}_{22} = 1 \\
\text{blkN}_{11} + \text{blkN}_{12} + \text{blkN}_{21} + \text{blkN}_{22} = 1
\end{align*}
\]

Objective Function:

\[
\min \text{ clkA}_{1} + \text{clkA}_{2} + \text{clkB}_{1} + \text{clkB}_{2}
\]

Binary Variables:

\[
\text{clkA}_{1}, \text{clkA}_{2}, \text{clkB}_{1}, \text{clkB}_{2} \\
\text{blkM}_{11}, \text{blkM}_{12}, \text{blkM}_{21}, \text{blkM}_{22} \\
\text{blkN}_{11}, \text{blkN}_{12}, \text{blkN}_{21}, \text{blkN}_{22}
\]

Exclusivity:

\[
\begin{align*}
\text{blkM}_{11} + \text{blkN}_{11} & \leq 1 \\
\text{blkM}_{12} + \text{blkN}_{12} & \leq 1 \\
\text{blkM}_{21} + \text{blkN}_{21} & \leq 1 \\
\text{blkM}_{22} + \text{blkN}_{22} & \leq 1
\end{align*}
\]

Block M: Clock A
Block N: Clock B
Our Approach

**Spine Count:**

\[
\text{blkM}_{11} + \text{blkM}_{12} \leq T \cdot \text{clkA}_1 \\
\text{blkM}_{21} + \text{blkM}_{22} \leq T \cdot \text{clkA}_2 \\
\text{blkN}_{11} + \text{blkN}_{12} \leq T \cdot \text{clkB}_1 \\
\text{blkN}_{21} + \text{blkN}_{22} \leq T \cdot \text{clkB}_2
\]

**Exclusivity:**

\[
\text{blkM}_{11} + \text{blkN}_{11} \leq 1 \\
\text{blkM}_{12} + \text{blkN}_{12} \leq 1 \\
\text{blkM}_{21} + \text{blkN}_{21} \leq 1 \\
\text{blkM}_{22} + \text{blkN}_{22} \leq 1
\]

**Placement:**

\[
\text{blkM}_{11} + \text{blkM}_{12} + \text{blkM}_{21} + \text{blkM}_{21} = 1 \\
\text{blkN}_{11} + \text{blkN}_{12} + \text{blkN}_{21} + \text{blkN}_{21} = 1
\]

**Objective Function:**

\[
\text{min } \text{clkA}_1 + \text{clkA}_2 + \text{clkB}_1 + \text{clkB}_2
\]

**Binary Variables:**

\[
\text{clkA}_1, \text{clkA}_2, \text{clkB}_1, \text{clkB}_2, \\
\text{blkM}_{11}, \text{blkM}_{12}, \text{blkM}_{21}, \text{blkM}_{22}, \\
\text{blkN}_{11}, \text{blkN}_{12}, \text{blkN}_{21}, \text{blkN}_{22}
\]

**T:** large constant
Our Approach

- Multiple solutions to an ILP problem.
- Original placement is optimized for wirelength and timing
- Anchoring term: encourages blocks to remain in their original position.

Original Objective Function: \( \min \ clkA_1 + clkA_2 + clkB_1 + clkB_2 \)

Augmented Objective Function: \( \min \ clkA_1 + clkA_2 + clkB_1 + clkB_2 + 0.1 \times (blkM_{12} + blkM_{21} + blkM_{22} + blkN_{11} + blkN_{12} + blkN_{21}) \)
Managing Runtime

Optimization window
Permissible Move Distance

Manhattan distance of 1

Manhattan distance of 2
Methodology

Benchmark Circuits

Placement

VPR Placer [Rose FPGA’09]

Power Optimization

ILP (Our approach)

Routing

VPR Router

Power, wirelength, timing
Measurement Metrics

- Clock signals are routed from horizontal spines to vertical half-spines on an as-needed basis.
- Clock signals usually span many columns and require many vertical half-spines.
- Clock network capacitance mainly determined by the # of vertical half-spines used.
- **Our metrics:** number of vertical half-spines (spine count), wirelength, timing
Comparative Baseline

- Aims to reduce spine count in an annealing-based placer (Actel Corp.) [Hsu IET-CDT’10].
- Cost function:
  \[
  \text{Cost} = a \cdot \text{WireLength} + b \cdot \text{Timing} + c \cdot \text{ClockPowerCost}
  \]
  \text{ClockPowerCost} = \text{sum of spine costs}
- Spine cost:
  - when no logic blocks use a spine, spine cost=0.
  - when blocks are initially added to a spine, spine cost ramps quickly.
  - when spine is ½ full, spine cost increases slowly up to a maximum when the spine is full.
Experimental Methodology

- FPGA architecture: 10 4-LUT/flip-flop pairs per logic block and length-4 wire segments.

- Altered 20 combinational and sequential benchmark circuits to contain multiple clock signals.

- For each circuit, the logic blocks were arbitrarily assigned one of the four clock domains.
Spine Count Reduction

% reduction

Permissible Move Distance

0% 10% 20% 30% 40% 50% 60% 70%

Window Size 2
Window Size 4

1 2 3 4
Post Routing Wirelength Increase
Annealing Based Approach

Cost = $a \cdot \text{WireLength} + b \cdot \text{Timing} + c \cdot \text{ClockPowerCost}$

- Spine reduction: 60%
- Wirelength increase: 3-4%

Increasing $c$
## Comparison

<table>
<thead>
<tr>
<th>Optimization Approach</th>
<th>Spine Reduction</th>
<th>Wirelength Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our approach-Permissible move distance 2</td>
<td>52-54%</td>
<td>6-7%</td>
</tr>
<tr>
<td>Our approach-Permissible move distance 3</td>
<td>58-61%</td>
<td>10-11%</td>
</tr>
<tr>
<td>Annealing-based [Hsu IET-CDT'10]</td>
<td>60%</td>
<td>3-4%</td>
</tr>
</tbody>
</table>
## Projected Power Benefit

<table>
<thead>
<tr>
<th></th>
<th>% Total Dynamic Power</th>
<th>% Reduction</th>
<th>Projected power reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock power</td>
<td>25%</td>
<td>50%</td>
<td>12.5%</td>
</tr>
<tr>
<td>Logic-signal power</td>
<td>50%</td>
<td>-6%</td>
<td>-3%</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td>~9%</td>
</tr>
</tbody>
</table>
Conclusions

• Post-placement ILP based clock power optimization technique.
• Can be used in conjunction with placement algorithm.
• Over 50% reduction of clock spine resources with minimal damage to traditional placement metrics.
Future Work

• Evaluate proposed ILP-based technique within comprehensive power-aware FPGA CAD system that optimizes power throughout flow, including synthesis, packing, place and route.
Questions
<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Clk Spine Count</th>
<th># Wire Segs</th>
<th>Clk Spine Count</th>
<th># Wire Segs</th>
</tr>
</thead>
<tbody>
<tr>
<td>alu4</td>
<td>73</td>
<td>8148</td>
<td>34</td>
<td>8664</td>
</tr>
<tr>
<td>apex2</td>
<td>92</td>
<td>13021</td>
<td>39</td>
<td>13758</td>
</tr>
<tr>
<td>apex4</td>
<td>74</td>
<td>9740</td>
<td>32</td>
<td>10410</td>
</tr>
<tr>
<td>bigkey</td>
<td>66</td>
<td>13870</td>
<td>32</td>
<td>14306</td>
</tr>
<tr>
<td>clma</td>
<td>217</td>
<td>64527</td>
<td>109</td>
<td>68509</td>
</tr>
<tr>
<td>des</td>
<td>78</td>
<td>17913</td>
<td>42</td>
<td>18927</td>
</tr>
<tr>
<td>difeq</td>
<td>65</td>
<td>5578</td>
<td>31</td>
<td>6340</td>
</tr>
<tr>
<td>dsip</td>
<td>67</td>
<td>13056</td>
<td>30</td>
<td>13733</td>
</tr>
<tr>
<td>elliptic</td>
<td>113</td>
<td>18857</td>
<td>46</td>
<td>20433</td>
</tr>
<tr>
<td>ex1010</td>
<td>154</td>
<td>36901</td>
<td>71</td>
<td>39430</td>
</tr>
<tr>
<td>ex5-p</td>
<td>60</td>
<td>9404</td>
<td>27</td>
<td>10185</td>
</tr>
<tr>
<td>frisc</td>
<td>118</td>
<td>22549</td>
<td>48</td>
<td>24023</td>
</tr>
<tr>
<td>misex3</td>
<td>73</td>
<td>8537</td>
<td>34</td>
<td>8848</td>
</tr>
<tr>
<td>pdc</td>
<td>164</td>
<td>42780</td>
<td>70</td>
<td>44606</td>
</tr>
<tr>
<td>s298</td>
<td>97</td>
<td>7716</td>
<td>45</td>
<td>8123</td>
</tr>
<tr>
<td>s38417</td>
<td>93</td>
<td>26605</td>
<td>88</td>
<td>29296</td>
</tr>
<tr>
<td>s38584.1</td>
<td>119</td>
<td>30842</td>
<td>92</td>
<td>33685</td>
</tr>
<tr>
<td>seq</td>
<td>86</td>
<td>11830</td>
<td>40</td>
<td>12884</td>
</tr>
<tr>
<td>spla</td>
<td>147</td>
<td>28673</td>
<td>65</td>
<td>30117</td>
</tr>
<tr>
<td>tseng</td>
<td>49</td>
<td>4981</td>
<td>22</td>
<td>5599</td>
</tr>
<tr>
<td>Geomean</td>
<td>99.0</td>
<td>15563.9</td>
<td>45.1</td>
<td>16660.7</td>
</tr>
</tbody>
</table>

Window size 4, move distance 2