Low Power Gated Bus Synthesis using Shortest-Path Steiner Graph for System-on-Chip Communications

Renshen Wang\textsuperscript{1}  
Nan-Chi Chou\textsuperscript{2}  
Bill Salefski\textsuperscript{2}  
Chung-Kuan Cheng\textsuperscript{1}

\textsuperscript{1}University of California San Diego  
\textsuperscript{2}Mentor Graphics Corporation
Goal of this talk

- Power saving techniques on bus communications
  - AMBA bus architecture
  - Steiner tree $\rightarrow$ Steiner graph
- Minimal power for on-chip communications under practical constraints
  - Fixed placement of components
  - Limited routing resource
  - Compatible with existing protocols
Introduction

- Low power design
- For battery life of embedded systems
- Lower CO$_2$ emission and environmental impact
On-chip Bus

- Connection bridge of System-on-Chips
  - Global connection, inter-module
  - Scaling up (relatively, to Moore’s law)

- Consuming about 15% of system power
  - “Power analysis of system-level on-chip communication architectures” by Lahiri & Raghunathan. *CODES* 2004
Bus Power

- Microprocessor 1 sending message to slave device 1
- What is necessary
- What is being taken by current bus architectures
- Low power efficiency on
  - Wires (14% of power on communication)
  - Component interfaces (35% of power)
Power Saving

- Clock gating
  - Masking off unused components
  - Dynamic power

- Power gating
  - Turning off unused components
  - Leakage power

- Bus gating
  - Dynamic power
On-chip Bus Architectures

- AMBA Ahb protocol
  - Also in IBM CoreConnect, Altera Avalon, etc

- Pros: Simplicity

- Cons: Low efficiency on power and wires
Objective of Improvement

- Greatly reduce bus power consumption
- Not greatly consume other resources
- Bus gating: may result in increased complexity in design flow
Step 1: Steiner Tree

- Use distributed multiplexers
- Add distributed de-multiplexers
  - Minimum rectilinear Steiner arborescence (MRSA)
  - Arborescence = Shortest-path Steiner tree
MRSA Construction

- RMST, NP-complete
- MRSA, NP-complete (correction to the paper)
- Basic RSA/G heuristic for MRSA

  - Merge subtrees such that the merging point is as far from source as possible

Given a source $s$ and $n$ terminals $t_1, \ldots, t_n$, $v_1, \ldots, v_N$ are the Hanan grid nodes sorted by $\Delta_s(v_1) > \cdots > \Delta_s(v_N)$;

\[ P \leftarrow \emptyset; \]

for $i = 1$ to $N$ do

  if there is $t_j$ at $v_i$, then \hspace{1cm} (TMO)

  \[ P \leftarrow P \cup \{v_i\}; \]

  \[ X \leftarrow P \cap \{v_j | \Delta_s(v_j) = \Delta_s(v_i) + \Delta(v_i, v_j)\}; \]

  if $|X| \geq 2$ then \hspace{1cm} (SMO)

    merge the nodes in $X$ rooted at $v_i$

    \[ P \leftarrow (P \cap \overline{X}) \cup \{v_i\}; \]

return the arborescence rooted at $s$;
MRSA Construction (cont.)

- Heuristics are explored in
  - “Efficient algorithms for the minimum shortest path Steiner arborescence problem…” by Cong, Khang, & Leung. IEEE TCAD 1998

- k-IDeA (iterated $k$-deletion for arborescence)
  - Remove up to $k$ nodes when running RSA/G
  - The best set of skipped nodes are marked as permanently deleted
  - Repeat iterations until no further improvement
  - 2-IDeA has the best overall performance
Step 2: Steiner Graph

- Single arborescence
  - Not good enough for all
- Shortest-path Steiner graph
  - Shortest path between each master-slave pair
  - Wire length may increase (but not necessarily)
Shortest-path Steiner Graph

- Multiple arborescences
  - By multiple MRSA constructions

- Wires can be shared
  - Starting from the second arborescence

- Example with 2 sources
  - 8 terminals need to be connected for $s_2$
SPSG Construction

- Revised RSA/G for SPSG

Given existing Steiner graph $G$, source $s_k$, terminals $t_1, \ldots, t_n$, and $v_1, \ldots, v_N$ are same as in RSA/G;

Routine Necessitate(vertex $v$);

$U \leftarrow \{u \in G \text{ and exists a wire path from } v \text{ to } u \text{ of length } \Delta_{s_k}(v) - \Delta_{s_k}(u)\};$

$T' \leftarrow T' \cup \{u_m \in U \text{ with minimum } \Delta_{s_k}(u)\};$

$T' \leftarrow \phi;$

for $i = 1$ to $n$ do Necessitate($t_i$);

$P \leftarrow \phi;$

for $i = 1$ to $N$ do

if $v_i \in T'$ then $P \leftarrow P \cup \{v_i\}$; \hspace{1cm} (TMO)

$X \leftarrow P \cap \{v_j | \Delta_{s_k}(v_j) = \Delta_{s_k}(v_i) + \Delta(v_i, v_j)\};$

if ($|X| \geq 1$ and $v_i \in G$) then

for each $(u \in X)$ connect($v_i, u$);

$P \leftarrow P \cap X;$

Necessitate($v_i$);

else if ($|X| \geq 2$) then \hspace{1cm} (SMO)

merge the nodes in $X$ rooted at $v_i$

$P \leftarrow (P \cap \overline{X}) \cup \{v_i\};$

return; \hspace{1cm} (the MRSA rooted at $s_k$ is added to $G$)

$v_i = t_3$

$P = \{t_2\}$

$X = \{t_2\}$

$P = \{t_3\}$
SPSG Construction (cont.)

- k-IDeA (iterated $k$-deletion for aborescence)
  - Same iterations as in MRSA construction
  - To construct each aborescence
- Run k-IDeA iterations on each source $s_i$
Switch Control on Graph Nodes

- Each node has degree 3 or 4
  - Needs a switch to guide signals
- A switch of node degree 4
- A path through a 4-way switch
  - $C(4,2) = 6$ possible cases
- May need 3 control signals
Reducing Control Signals

- Path guidance requirement
  - Paths only exist between master-slave pairs
  - Actual cases can be less than $C(4,2)$ or $C(3,2)$

- Decoder duplication
  - Reduce wire at the cost of silicon resource
  - Not adopted in our experiments

- More techniques…
Experimental Results

- Average wire capacitance in transmission

Table 3: Average data transaction wire capacitance

<table>
<thead>
<tr>
<th>Case ((m, n))</th>
<th>AHB</th>
<th>MRSA</th>
<th>SPSG</th>
<th>(P_{\text{save}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0 (1,16)</td>
<td>5934</td>
<td>1638</td>
<td>469</td>
<td>72.4% - 92.1%</td>
</tr>
<tr>
<td>T0 (2,16)</td>
<td>7584</td>
<td>2138</td>
<td>613</td>
<td>71.8% - 91.9%</td>
</tr>
<tr>
<td>T0 (3,16)</td>
<td>8418</td>
<td>2038</td>
<td>635</td>
<td>75.8% - 92.5%</td>
</tr>
<tr>
<td>T1 (3,16)</td>
<td>14550</td>
<td>4067</td>
<td>1754</td>
<td>72.0% - 87.9%</td>
</tr>
<tr>
<td>T2 (2,30)</td>
<td>12305</td>
<td>2074</td>
<td>669</td>
<td>83.1% - 94.6%</td>
</tr>
<tr>
<td>T3 (3,16)</td>
<td>8230</td>
<td>2019</td>
<td>691</td>
<td>75.5% - 91.6%</td>
</tr>
<tr>
<td>T4 (5,15)</td>
<td>9674</td>
<td>1862</td>
<td>689</td>
<td>80.8% - 92.9%</td>
</tr>
<tr>
<td>T5 (6,16)</td>
<td>11210</td>
<td>2012</td>
<td>694</td>
<td>82.1% - 93.9%</td>
</tr>
<tr>
<td>T6 (8,8)</td>
<td>8952</td>
<td>1945</td>
<td>689</td>
<td>78.3% - 92.3%</td>
</tr>
<tr>
<td>T7 (12,6)</td>
<td>12107</td>
<td>1933</td>
<td>657</td>
<td>84.0% - 94.6%</td>
</tr>
<tr>
<td>T8 (16,10)</td>
<td>14377</td>
<td>1905</td>
<td>683</td>
<td>86.7% - 95.2%</td>
</tr>
<tr>
<td>T9 (8,16)</td>
<td>11652</td>
<td>1883</td>
<td>618</td>
<td>83.8% - 94.7%</td>
</tr>
<tr>
<td>T10 (8,16)</td>
<td>12899</td>
<td>2103</td>
<td>749</td>
<td>83.7% - 94.2%</td>
</tr>
<tr>
<td>T11 (6,12)</td>
<td>8970</td>
<td>1897</td>
<td>668</td>
<td>78.9% - 92.6%</td>
</tr>
<tr>
<td>T12 (12,12)</td>
<td>13242</td>
<td>1936</td>
<td>669</td>
<td>85.4% - 94.9%</td>
</tr>
</tbody>
</table>

\[
\frac{C_{\text{SPSG}} - C_{\text{AHB}}}{C_{\text{AHB}}}
\]

MRSA

SPSG
Experimental Results (cont.)

- Impact on Total Wire Length

<table>
<thead>
<tr>
<th>Case ((m, n))</th>
<th>AHB</th>
<th>MRSA</th>
<th>SPSG</th>
<th>(L_{\text{increment}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0 (1,16)</td>
<td>11200</td>
<td>10316</td>
<td>9656</td>
<td>-7.9%~ -13.8%</td>
</tr>
<tr>
<td>T0 (2,16)</td>
<td>14000</td>
<td>11850</td>
<td>14078</td>
<td>-15.4%~ 0.6%</td>
</tr>
<tr>
<td>T0 (3,16)</td>
<td>16000</td>
<td>13350</td>
<td>21697</td>
<td>-16.6%~35.6%</td>
</tr>
<tr>
<td>T1 (3,16)</td>
<td>27000</td>
<td>15131</td>
<td>18238</td>
<td>-44.0%~32.5%</td>
</tr>
<tr>
<td>T2 (2,30)</td>
<td>23257</td>
<td>16386</td>
<td>24535</td>
<td>-29.5%~5.5%</td>
</tr>
<tr>
<td>T3 (3,16)</td>
<td>15248</td>
<td>11554</td>
<td>14274</td>
<td>-24.2%~6.4%</td>
</tr>
<tr>
<td>T4 (5,15)</td>
<td>17840</td>
<td>13729</td>
<td>22792</td>
<td>-23.0%~27.8%</td>
</tr>
<tr>
<td>T5 (6,16)</td>
<td>20988</td>
<td>15778</td>
<td>35287</td>
<td>-24.8%~68.1%</td>
</tr>
<tr>
<td>T6 (8,8)</td>
<td>16623</td>
<td>15800</td>
<td>25297</td>
<td>-5.0%~52.2%</td>
</tr>
<tr>
<td>T7 (12,6)</td>
<td>22640</td>
<td>13227</td>
<td>27113</td>
<td>-41.6%~19.8%</td>
</tr>
<tr>
<td>T8 (16,10)</td>
<td>27316</td>
<td>17018</td>
<td>40844</td>
<td>-37.7%~49.5%</td>
</tr>
<tr>
<td>T9 (8,16)</td>
<td>22183</td>
<td>16055</td>
<td>35267</td>
<td>-27.6%~59.0%</td>
</tr>
<tr>
<td>T10 (8,16)</td>
<td>24173</td>
<td>15906</td>
<td>34337</td>
<td>-34.2%~42.0%</td>
</tr>
<tr>
<td>T11 (6,12)</td>
<td>16920</td>
<td>12566</td>
<td>22070</td>
<td>-25.7%~30.4%</td>
</tr>
<tr>
<td>T12(12,12)</td>
<td>25369</td>
<td>17256</td>
<td>38305</td>
<td>-32.0%~51.0%</td>
</tr>
</tbody>
</table>
Tradeoff between power & wire

- Power reduction curve
  - Depending on routing resource
  - Always large improvement
Future Works

- Wirelength reduction
  - Compress narrow rectangles into lines

- Bandwidth extension (bus matrix)
  - Allow multiple access
  - Power-bandwidth co-optimization
Some Insights

- What does this gated bus look like?
  - A map
  - Streets created for efficient commuting
Questions?

- Thank you for your attention!

- This research is under the support of NSF CCF-0811794 and California MICRO program