Area Efficient Asynchronous SDM Routers Using 2-Stage Clos Switches

Wei Song, Doug Edwards, Jim Garside and William J. Bainbridge

School of Computer Science
The University of Manchester
Manchester M13 9PL U.K.
Motivation

• Asynchronous NoCs are low power and low latency.
  – No clock, no dynamic power, simple structure
• Virtual channels are good for QoS support but poor in area and speed, especially in BE traffic.
  – Large area in buffers, arbiters and switches
  – Deep buffers with complicated arbitration
• Spatial division multiplexing (SDM) provide good throughput with low area overhead
• Further reduce the area overhead using Clos switches.
Outline

• Asynchronous on-chip networks
• Virtual channel (VC) and its problems
• Spatial division multiplexing (SDM) router
  – Structure
  – Analyses
• 2-stage Clos switches for SDM routers
• Performance comparison
• Future work and conclusion
Asynchronous on-chip networks

Synchronous processing elements connected by asynchronous routers.

Basic building block: 4-phase multi-rail pipelines
Wormhole and virtual channel

Wormhole router

VC router
Issues of VC in asynchronous NoCs

Benefits:
Excellent QoS support with low latency.

Issues:
1. Extra and deeper input buffers. Longer latency and larger area.
2. Direct connection with the crossbar making it V times large. Larger area.
3. An extra VC allocator.
   QoS: P x P x V
   Best effort: PV x PV
   Longer latency and larger area.
Benefits:
High throughput performance. Also support QoS (in theory)

Compare with VC:
1. No extra buffer.
2. The switch is V times large, as the same as VC.
3. Large switch allocator: PV x PV, as the same as the VC allocator.
4. Arbitration once per frame.

Issue:
Smaller bandwidth leads to long latency.
Performance: Wormhole vs. VC vs. SDM

32-bit 5 ports 4 VC/virtual circuits, 8x8 network, uniform random traffic
SDMCS: SDM router using sliced pipelines

Switch area

\[ A(\text{switch}) = PV*PV*W/V \]
\[ = V*P^2W \]
2-Stage Clos switches and area saving

- No output modules.
- Trun model optimisation in central modules.
- More than 50% area reduction when $V \geq 4$. 

![Diagram of 2-Stage Clos switches](image)
Router structure

Distributed Clos scheduler also reduces area.

Area reduction and frame latency

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Area (μm²)</th>
<th>2 virtual circuits: no area reduction</th>
<th>3 virtual circuits: 21% reduction</th>
<th>4 virtual circuits: 50% reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline:W32</td>
<td>[Graph]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM:V2W32</td>
<td>[Graph]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM-Clos:V2W32</td>
<td>[Graph]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM:V3W48</td>
<td>[Graph]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM-Clos:V3W48</td>
<td>[Graph]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM:V4W32</td>
<td>[Graph]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM-Clos:V4W32</td>
<td>[Graph]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Throughput is slightly compromised due to the heuristic scheduling in the Clos switch.
Area efficiency and throughput

area efficiency = saturation throughput / router area

The efficiency boost proves the throughput drop is marginal.
Energy consumption

Faraday 130nm

Period:
Baseline: W32
  2.2 ns

SDM: V4W32
  2.8 ns

SDM-Clos: V4W32
  2.8 ns

Slower routers consume less energy.
Conclusions

• Using 2-stage Clos switches reduces the area overhead of asynchronous SDM routers significantly when V>2.
• It slightly compromises throughput but area efficiency is improved.
• No obvious benefits in energy but SDM-Clos may consume less energy when deep buffers are used.
• Source available from:
  – http://opencores.org/project,async_sdm_noc
Possible future work

- Buffer the central modules
Thank you!

Question?