Design of Energy-Efficient On-Chip Networks

Vladimir Stojanović
Integrated Systems Group
MIT
Manycore System Roadmap

- 64-tile system (64-256 cores)
  - 4-way SIMD FMACs @ 2.5 – 5 GHz
  - 5-10 TFlops on one chip
  - Need 5-10 TB/s of off-chip I/O
  - Even larger bisection bandwidth
The rise of manycore machines

Only way to meet future system feature set, design cost, power, and performance requirements is by programming a processor array

- Multiple parallel general-purpose processors (GPPs)
- Multiple application-specific processors (ASPs)

 Intel Network Processor
  1 GPP Core
  16 ASPs (128 threads)

 IBM Cell
  1 GPP (2 threads)
  8 ASPs

 Sun Niagara
  8 GPP cores (32 threads)

 Picochip DSP
  1 GPP core
  248 ASPs

 Cisco CRS-1
  192 Tensilica GPPs

“The Processor is the new Transistor” [Rowen]

Intel 4004 (1971):
  4-bit processor,
  2312 transistors,
  ~100 KIPS,
  10 micron PMOS,
  11 mm² chip

ISSCC 2010 Tutorial
Interconnect bottlenecks

Manycore system

- CPU
- Cache
- DRAM DIMM

Interconnect Network

- cores

Interconnect Network

Bottlenecks due to energy and bandwidth density limitations

Need to jointly optimize on-chip and off-chip interconnect network

ISSCC 2010 Tutorial
Scaling to many cores

- Today’s approaches
- Many meshes
  - Slow, latency varies greatly
  - Easy to implement
- Large crossbars
  - Fast, predictable latency
  - Hard to build and scale

TILE64
[Bell08]

ISSCC 2010 Tutorial
Rainbow-Falls 2-stage Crossbar

Bisection Bandwidth 461GB/s

[Patel09]
On-chip network topology spectrum

Mesh
CMesh
Clos
Crossbar

Increasing diameter

Easy to design
Hard to program

Increasing radix

Hard to design
Easy to program

In power constrained systems – Need to look at networks in a cross-cut approach
Connect physical implementation (channels, routers, power) with
network topology, routing and flow-control

Radix – Number of inputs and outputs of each switching node
Diameter – largest minimal hop count over all node pairs
NOCs Tutorial Roadmap

• Networking Basics
• Building Blocks
• Evaluation
NOCs Tutorial Roadmap

• Networking Basics
  – Topologies
  – Routing
  – Flow-Control

• Building Blocks

• Evaluation
Message definitions

- Basic trade-off
  - Minimize overheads (large size)
  - Efficient use of resources (small size)

RI - Routing Info
SN - Sequence #

Basic unit of bw and storage allocation (flow-control)
Sent across channel in a clock cycle
Latency Components

- Zero-load latency
  - Average latency w/o contention

\[ T_0 = H_{\text{min}} t_r + \frac{D_{\text{min}}}{v} + \frac{L}{b} \]

- Router delays
- Channel delays
- Serialization delay

- \( H_{\text{min}} \) – average minimum number of hops
- \( t_r \) – Router delay
- \( D_{\text{min}} \) – average minimum distance
- \( v \) – signal velocity
- \( L \) – packet length in bits
- \( b \) – router-to-router channel bandwidth

\[ T_0 = 2t_r + (t_{xy} + t_{yz}) + \frac{L}{b} \]
Ideal network throughput (capacity)

- Maximum traffic that can be sustained by all cores
- Mesh throughput
  - 50% of data crosses the bisection assuming uniform random traffic
- Bisection bandwidth $= 2\sqrt{N}b$
- Data crossing the bisection $= \frac{1}{2}Nb_{\text{core}}$
- Maximum on-chip throughput
  \[ \Theta_{\text{ideal}} = Nb_{\text{core}} = 4\sqrt{N}b \]

$N =$ number of cores
$b =$ router-to-router link bandwidth
$b_{\text{core}} =$ rate at which each core generates traffic
Network performance plots

Zero-load latency includes effects of routing and flow-control

Latency (s)

$T_0$

$H_{\text{avg}} t_r + L/b$

$H_{\text{min}} t_r + L/b$

Offered Traffic (bits/s)

$\lambda_S$

$\Theta_R$

$\Theta_{\text{ideal}}$

Topology

Routing

Flow-control
Tori

- Low-radix, large diameter networks
- N-ary, K-cube (mesh)
  - N nodes per dimension
  - K dimensions

- Cubes have 2x larger bisection bandwidth

[Dally04]
TILE64

- 64 cores at 750 MHz
- Memory BW 25 GB/s
- 240 GB/s bis. Bw

[Bell08]

ISSCC 2010 Tutorial
TILE64 Networks

STN – Static network
TDN – Tile Dynamic network
UDN – User Dynamic network
MDN – Memory Dynamic network
IDN – I/O Dynamic network

32 bit channels on all networks
Wormhole, dimension-order routed
5-port routers with credit-based flow-control

STN – Scalar operand network
TDN and MDN implement the memory sub-system
UDN/IDN – Directly accessible by processor ALU (message-based, variable length)
Improving Tori - Express cubes

• Increase bisection bandwidth, reduce latency
  – Add expressways - long “express” channels

One dimension of 16-ary express cube with 4-hop express channels

Add extra channels to diversify and/or increase bisection
Buterflies

• N-ary, K-fly
  – N nodes per switch
  – K stages
• Example
  – 2-ary 4 fly

[Dally04]
Path diversity problem

- Buterflies have no path diversity
- Bad performance for some traffic patterns
  - e.g. shuffle permutation
- Wide spread in BW
- Inherently blocking
- Fixed in Clos topologies

[Dally04]
Clos networks

- Redundant paths – more uniform throughput
Logical to Physical Mapping

Router group

8-ary 3-stage Clos

Three 8 x 8 Routers
(I-VIII, a-h, A-H)

Two 8 x 8 Routers
(l-VIII,a-h)

Eight 8 x 8 Routers
(middle stage A-H)

• Same topology – different physical mapping
Topology comparison

Table 1: Example Network Configurations – Networks sized to support 128 b/cyc per tile under uniform random traffic. $N_C =$ number of channels, $b_C =$ bits/channel, $N_{BC} =$ number of bisection channels, $N_R =$ number of routers, $H =$ number of routers along data paths, $T_R =$ router latency, $T_C =$ channel latency, $T_{TC} =$ latency from tile to first router, $T_S =$ serialization latency, $T_0 =$ zero load latency.
Routing Algorithms

• Deterministic routing algorithms
  – Always same path between x and y
    • Poor load balancing (ignore inherent path diversity)
    • Quite common in practice
      – Easy to implement and make deadlock-free.

• Oblivious algorithms
  – Choose a route w/o network’s present state
    • E.g. random middle-node in Clos

• Adaptive algorithms
  – Use network’s state information in routing
    • Length of queues, historical channel load, etc
Deterministic Routing

2-ary 3-fly

6-ary 2-cube

Destination-tag Butterflies

Dimension-order Tori

[Dally04]
Oblivious Routing

- Valiant’s algorithm (Randomized Routing)

Folded Clos (Fat Tree)

Randomly select nearest common ancestor switch

Randomly select middle switch

Randomly select middle node

Dimension-order to/from node

6-ary 2-cube

[Valiant04]

ISSCC 2010 Tutorial
Flow Control

- Bufferless flow-control (*Circuit Switching*)
- Buffered flow-control (*Packet Switching*)
  - Packet-based (store&forward, cut-through)
  - Flit-based (wormhole, virtual channels)
- Buffer Management
  - Credit-based, on-off, flit-reservation
Circuit switching

Pros
- Simple to implement (simple routers, small buffers)

Cons
- High latency (R+A) and low throughput

Dally04

R - Request
A - Acknowledgment
Acquires channel state at each hop

D - Data packets
e.g. Two, four-flit packets

T - Tail flit
Deallocate channels

Blocked request
held at switch

ISSCC 2010 Tutorial
Example - Pipelined Circuit Switching

64 core 2D mesh, 125 mW/router
Network efficiency 3 pJ/bit

[Anderson08]
Packet-buffered Flow Control

Buffer and channel allocated to whole packet

[Dally04]

• Store-and-forward

  Start next hop after whole packet received

  5-flit packet

\[ T_0 = H \left( t_r + \frac{L}{b} \right) \]

• Cut-through

  Start next hop after head flit received

  5-flit packet

\[ T_0 = H t_r + \frac{L}{b} \]

Contention for channel 2

Both ineffective in use of buffer storage

Contention latency increased in channels

ISSCC 2010 Tutorial
Flit-buffered Flow Control

• Wormhole

  I – idle, W – waiting, A - allocated

More efficient buffer usage than cut-through
But, may block a channel mid-packet

[Dally04]
Flit-buffered Flow Control

- Wormhole vs. Virtual-Channel  [Dally92]

[Dally04]
Virtual-channels – Bandwidth Allocation

Inputs compete for bandwidth
Flit-by-Flit

# flits in VC buffer (cap 3)

Fair Arbitration

Winner-take-all
Arbitration

Reduced latency
No throughput penalty

[Dally04]

ISSCC 2010 Tutorial
Virtual-channel Router

Each channel only as deep as round-trip credit latency
More buffering, more virtual channels

[Dally04]

(a) One 16-flit buffer
(b) Two 8-flit buffers
(c) Four 4-flit buffers
Credit-based buffer management

\[ F \geq \frac{t_{crt} b}{L f} \]
NOCs Tutorial Roadmap

• Networking Basics
• Building Blocks
  – Channels
  – Routers
• Evaluation
Building block costs

Router vs. channel energy

- Simple routers and channels roughly balanced
- Narrower networks scale better

Router Area Breakdowns

90nm technology

ISSCC 2010 Tutorial
Channels: Electrical technology

**Design constraints**
- 22 nm technology
- 500 nm pitch
- 5 GHz clock

**Design parameters**
- Wire width
- Repeater size
- Repeater spacing
Channels: Equalized interconnects

- FFE shapes transmitted pulse
- DFE cancels first trailing ISI tap
- Lower energy cost due to output voltage swing attenuation

[Menisk07, Kim08, Kim09]
Repeated interconnects vs Equalized interconnects

Data-dependent energy (DDE) is 4-10x lower for equalized interconnects, while fixed energy (FE) is comparable.
Channels: Silicon photonic technology

Energy spent in O-E conversion = 25 - 60 fJ/bt (independent of link length)
Photodetector loss = 0.1 dB
Filter drop loss = 1.5 dB
Receiver sensitivity = -20 dBm

[Gunn06, Orcutt08]
Silicon photonic link – WDM

Through ring loss = $1 \times 10^{-4} - 1 \times 10^{-2}$ dB/ring

- Dense WDM improves bandwidth density
  - E.g. 128 $\lambda$/wg, 10 Gbps/$\lambda$
Silicon photonic link – Energy cost

- E-O-E conversion cost – 50-150 fJ/bt (independent of length)
- Thermal tuning energy – 2-20µW/K/heater
  - Increases with ring count
- External laser power
  - Dependent on losses in photonic devices
Electrical vs Optical links – Energy cost

- Elec: Electrical
- Opt-A: Optical-Aggressive
- Opt-C: Optical-Conservative

- Optical laser power not shown (dependent on the physical layout)
- Thermal tuning energy
- Transmitter - Receiver energy

Approximately:
- 6x
- 2x

ISSCC 2010 Tutorial
## Channel Technologies

### On-chip links

<table>
<thead>
<tr>
<th></th>
<th>Latency (cyc)</th>
<th>Energy (fJ/b)</th>
<th>Density (Gb/s/µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimally repeated wire (2.5 mm)</td>
<td>1</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>Equalized link (2.5 mm)</td>
<td>2</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>Photonic link (2.5 mm)</td>
<td>2</td>
<td>100-200</td>
<td>320</td>
</tr>
<tr>
<td>Optimally repeated wire (10 mm)</td>
<td>2</td>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>Equalized link (10 mm)</td>
<td>2</td>
<td>120</td>
<td>10</td>
</tr>
<tr>
<td>Photonic link (10 mm)</td>
<td>2</td>
<td>100-200</td>
<td>320</td>
</tr>
</tbody>
</table>
Routers

per packet

Input VC state

<table>
<thead>
<tr>
<th>Field</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>Global state</td>
</tr>
<tr>
<td>R</td>
<td>Route</td>
</tr>
<tr>
<td>O</td>
<td>Output VC</td>
</tr>
<tr>
<td>P</td>
<td>Pointers</td>
</tr>
<tr>
<td>C</td>
<td>Credit count</td>
</tr>
</tbody>
</table>

Output VC state

<table>
<thead>
<tr>
<th>Field</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>Global state</td>
</tr>
<tr>
<td>I</td>
<td>Input VC</td>
</tr>
<tr>
<td>C</td>
<td>Credit count</td>
</tr>
</tbody>
</table>
Router pipeline

- Pipelined routing of a packet

- Pipeline stalls (virtual allocation stall)

RC – route computation
VA – virtual channel allocation
SA – switch allocation
ST – switch traversal
Speculation and Lookahead

Speculative allocation

Lookahead routing
(pass routing for next hop in head flit)
Crossbar switches

No Speedup – 68% capacity

2x Input Speedup – 90% capacity

2x Output Speedup – 87% capacity

2x Input & Output Speedup – 137% capacity

\[ \Theta = s_o \left(1 - \left(\frac{k-1}{k}\right)\frac{s}{s_o}\right) \]
Router design space exploration - Setup

\[ w = \text{Flit size (bits)} \]

\[ p = \text{Ports} = 5 \]

6-bit Destination Address for 64-core system

[Shamim09]
Matrix Crossbar

input 1
input 2
... input k
output 1
output 2
... output k
Mux Crossbar

Multiplexer Crossbar

in1[w-1:0] → out1[w-1:0]

in2[w-1:0] → out2[w-1:0]

in3[w-1:0] → out3[w-1:0]

in4[w-1:0] → out4[w-1:0]

in5[w-1:0] → out5[w-1:0]

ISSCC 2010 Tutorial
Example System

- 64 tiles.
- 1GHz frequency
- 1 Message = 512-bits
- 4 Messages per input port (2048-bits)
- Router Aspect Ratio 1
  - \( p = 5, 8, 12 \)
  - \( w = 32, 64, 128 \) (bits)
- Matrix xbar
- Mux xbar

Design space
18 Routers
5x5 Router Floorplan (128bit)
### 8x8 Routers Floorplan (128bit)

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>48</td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
</tr>
<tr>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
</tr>
</tbody>
</table>

- **in1**: Inlet 1
- **out1**: Outlet 1
- **in2**: Inlet 2
- **out2**: Outlet 2
- **in3**: Inlet 3
- **out3**: Outlet 3
- **in4**: Inlet 4
- **out4**: Outlet 4
- **in5**: Inlet 5
- **out5**: Outlet 5
- **in6**: Inlet 6
- **out6**: Outlet 6
- **in7**: Inlet 7
- **out7**: Outlet 7
- **in8**: Inlet 8
- **out8**: Outlet 8

**Legend:**
- 16word 128bits SRAM

---

ISSCC 2010 Tutorial
12x12 Routers Floorplan (128bit)
Area vs Port Width and Radix

- Mux crossbar always better
- 5-12 port routers scale well (sub $p^2$, $b^2$)
Power vs Port Width and Radix

- Mux crossbar always better
- 5-12 port routers scale well \((\text{sub } p^2, b^2)\)
Router Power Breakdown

Xbar and Buffer power roughly even

Improve Xbar with Ckt/channel design (equalized, low-swing)

Use less buffers (circuit switching, token flow control) [Anders08, Kumar08]
Router Area per core vs. # Ports

Area Fraction VS Ports

- 32bit matrix
- 64bit matrix
- 128bit matrix
- 32bit mux
- 64bit mux
- 128bit mux

ISSCC 2010 Tutorial
## Effects of Concentration

- **Mesh to Cmesh**
  - **5p routers to 8p routers**

### Table: Matrix Design vs. Mux Design

<table>
<thead>
<tr>
<th>Matrix Design</th>
<th>Area (mm²)</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 x 5p32b-mat</td>
<td>1.1664</td>
<td>332.304</td>
</tr>
<tr>
<td>1 x 8p64b-mat</td>
<td>0.4356</td>
<td>246.3924</td>
</tr>
<tr>
<td>4 x 5p64b-mat</td>
<td>1.2996</td>
<td>484.4544</td>
</tr>
<tr>
<td>1 x 8p128b-mat</td>
<td>0.8836</td>
<td>568.2672</td>
</tr>
<tr>
<td>2 x 8p32b-mat</td>
<td>0.5832</td>
<td>264.6312</td>
</tr>
<tr>
<td>1 x 12p64b-mat</td>
<td>0.6889</td>
<td>546.8928</td>
</tr>
<tr>
<td>2 x 8p64b-mat</td>
<td>0.8712</td>
<td>492.7848</td>
</tr>
<tr>
<td>1 x 12p128b-mat</td>
<td>1.7424</td>
<td>1584.54</td>
</tr>
<tr>
<td>8 x 5p32b-mat</td>
<td>2.3328</td>
<td>664.608</td>
</tr>
<tr>
<td>1 x 12p128b-mux</td>
<td>1.7424</td>
<td>1584.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mux Design</th>
<th>Area (mm²)</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 x 5p32b-mux</td>
<td>1.1664</td>
<td>268.3056</td>
</tr>
<tr>
<td>1 x 8p64b-mux</td>
<td>0.3721</td>
<td>203.268</td>
</tr>
<tr>
<td>4 x 5p64b-mux</td>
<td>1.2544</td>
<td>410.5872</td>
</tr>
<tr>
<td>1 x 8p128b-mux</td>
<td>0.7225</td>
<td>391.0116</td>
</tr>
<tr>
<td>2 x 8p32b-mux</td>
<td>0.5832</td>
<td>215.8464</td>
</tr>
<tr>
<td>1 x 12p64b-mux</td>
<td>0.5625</td>
<td>389.5896</td>
</tr>
<tr>
<td>2 x 8p64b-mux</td>
<td>0.7442</td>
<td>406.536</td>
</tr>
<tr>
<td>1 x 12p128b-mux</td>
<td>1.2769</td>
<td>926.2188</td>
</tr>
<tr>
<td>8 x 5p32b-mux</td>
<td>2.3328</td>
<td>536.6112</td>
</tr>
<tr>
<td>1 x 12p128b-mux</td>
<td>1.2769</td>
<td>926.2188</td>
</tr>
</tbody>
</table>

- Works well for small flits and number of ports

[Balfour06]
Orion 1.0 vs P & R design

Ratio (Power of Synthesized designs / Dynamic (no leakage) Power of Analytical Models)

ISSCC 2010 Tutorial
Orion 2.0 vs P & R design

[Kahng09]  [Shamim09]

Ratio (Power of Synthesized designs / Dynamic (no leakage) Power of Analytical Models)

Buffer
Xbar
Total

<table>
<thead>
<tr>
<th>Ports</th>
<th>32 bits</th>
<th>64 bits</th>
<th>128 bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 ports</td>
<td>0.72</td>
<td>0.75</td>
<td>0.78</td>
</tr>
<tr>
<td>8 ports</td>
<td>0.82</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>12 ports</td>
<td>0.92</td>
<td>0.95</td>
<td>0.98</td>
</tr>
</tbody>
</table>

ISSCC 2010 Tutorial

62
NOCs Tutorial Roadmap

• Networking Basics
• Building Blocks
• Evaluation
Landscape of on-chip photonic networks

Mesh

CMesh

Clos

Crossbar

[Shacham’07]
[Petracca’08]

[Joshi’09a]
[Pan’09]

[Vantrease’08]
[Psota’07]
[Kirman’06]
Clos with electrical interconnects

- 8-ary 3-stage Clos
  - 10-15 mm channels
  - Equalized
  - Pipelined Repeaters

Two 8 x 8 Routers
Eight 8 x 8 Routers
Centralized Multiplexer Crossbar

Electrical design

Photonic design

ISSCC 2010 Tutorial
Clos network using point-to-point channels

Electrical design

Photonic design

ISSCC 2010 Tutorial
Photonic Clos for a 64-tile system

- Tile
- .. 8 tiles per cluster
- 56 Waveguides (64λ/direction)
Photonic Clos for a 64-tile system

Photonic Transmitter Block

ISSCC 2010 Tutorial
Photonic Clos for a 64-tile system

Router Group & Photonic Transmitter-Receiver Block

ISSCC 2010 Tutorial
Photonic Clos for a 64-tile system
Photonic Clos for a 64-tile system

- 64 tiles
- 56 waveguides (for tile throughput = 128 b/cyc)
- 128 modulators per cluster
- 128 ring filters per cluster
- Total rings $\approx 28K \rightarrow 0.56$W (Thermal tuning)

ISSCC 2010 Tutorial
Photonic device requirements in a Clos

Optical laser power (W) contour

Percent area of photonic devices contour

Waveguide loss and Through loss limits for 2 W optical laser power (30% laser efficiency) constraint
Photonic Crossbar vs Photonic Clos

Crossbar

- 10 W power for thermal tuning circuits (1 μW/ring/K)
- For 2 W optical laser power
  - Waveguide loss < 1 dB/cm
  - Through loss < 0.002 dB/ring

Clos

- 0.56 W power for thermal tuning circuits (1 μW/ring/K)
- For 2 W optical laser power
  - Waveguide loss < 2 dB/cm
  - Through loss < 0.05 dB/ring
Simulation setup

- Cycle-accurate microarchitectural simulator
- Traffic patterns based on partition application model
  - Global traffic – UR, P2D, P8D
  - Local traffic – P8C
- 64-tile system, 512-bit messages
- Events captured during simulations to calculate power
Partition application model

- Tiles divided into logical partitions and communication is within partition
- Logical partitions mapped to physical tiles
  - Co-located tiles → Local traffic
  - Distributed tiles → Global traffic

Uniform random (UR)

2 tiles per partition that are distributed across the chip (P2D)

8 tiles per partition that are distributed across the chip (P8D)

8 tiles per partition that are co-located (P8C)

[Joshi’09]
Latency vs BW

- **flatFlyX2 vs mesh/cmeshX2**
  - **Saturation BW** → comparable (UR, P8D, P2D)
  - **Latency** → flatFlyX2 has lower latency

- **clos vs mesh/cmeshX2/flatFlyX2**
  - **Saturation BW** → uniform for all traffic, comparable to UR of mesh
  - **Latency** → uniform for all traffic, comparable to UR of mesh

---

**Ideal Throughput** $\theta_T = 8$ kb/cyc for UR

- flatFlyX2 vs mesh/cmeshX2
  - Saturation BW → comparable (UR, P8D, P2D)
  - Latency → flatFlyX2 has lower latency

- clos vs mesh/cmeshX2/flatFlyX2
  - Saturation BW → uniform for all traffic, comparable to UR of mesh
  - Latency → uniform for all traffic, comparable to UR of mesh
Mesh vs CMeshX2

- Repeater-inserted interconnects
  - cmeshX2 lower power than mesh at comparable throughput
- Equalized interconnects
  - cmeshX2 has further 1.5x reduction in power
  - Channel gains masked by router power
Power vs BW plots – repeater inserted pipelined vs equalized

1.5-2x lower power with equalized channels at comparable throughput
• Channel DDE reduces by 4-10x using equalized links
• Channel fixed power and router power need to be tackled
Latency vs BW – no VC vs 4 VCs

Saturation throughput improves using VCs
Small change in power at comparable throughput
Latency vs BW – no VC vs 4 VCs

8 kb/cycle

4 kb/cycle

Ideal UR throughput
Power vs BW – no VC vs 4 VCs, repeater inserted pipelined

25-50% lower power using VCs at comparable throughput

ISSCC 2010 Tutorial
Power vs BW—no VC case, repeater inserted

pipelined vs 4 VCs, equalized

2-3x lower power obtained using equalized interconnects and VCs at comparable throughput
• VCs an indirect way to increase impact of channel power
  – Narrower networks, lower power for same throughput, keep utilization high
Power-Bandwidth tradeoff

Channel width $b_c$
Ideal Throughput $\theta_T$

CMeshX2
128b
4kb/cycle

Clos
64b
4kb/cycle

2-3x on-chip power savings for global traffic (off-chip laser)
Power-Bandwidth tradeoff

Comparable on-chip power for local traffic (off-chip laser)

\[ b_c \]

\[ \theta_T \]

ISSCC 2010 Tutorial
Cross-cut approach for NOC design needed

- Application mapping
- Topology, Routing, Flow-control
- Improving Routers and Channels equally important
  - Opportunities for new technologies
  - New circuit design (low-swing, equalized)
  - System – DVFS, bus-encoding
To probe further (tools and sites)

- Orion Router Design Exploration Tool
  - http://www.princeton.edu/~peh/orion.html

- Router RTLs
  - Bob Mullins’ Netmaker
    (http://www-dyn.cl.cam.ac.uk/~rdm34/wiki)

- Network simulators
  - Garnet (http://www.princeton.edu/~niketa/garnet.html)
  - Booksim (http://nocs.stanford.edu/booksim.html)

Integrated Systems Group at MIT (vlada@mit.edu)
http://www.rle.mit.edu/isg/
Bibliography


[Kahng09] A. Kahng, B. Li, L-S. Peh and K. Samadi “ORION 2.0: A Fast and Accurate NoC Power and Area Model for Early-Stage Design Space Exploration” in Proceedings of Design Automation and Test in Europe (DATE), Nice, France, April 2009
Bibliography


Bibliography


