Impact of Process and Temperature Variations on Network-on-Chip Design Exploration

Bin Li, Li-Shiuan Peh, Priyadarsan Patra*

Princeton University
{binl, peh}@princeton.edu

*Intel Corporation
priyadarsan.patra@intel.com

April 9, 2008
Process and Temperature Variations

- Process variation has higher impact on sub-100 nm CMOS
  - Transistor dimension variation: Sub-wavelength lithography
  - Transistor characteristic variation: Dopant density fluctuation etc
- Temperature variations: caused by workload variation

[IBM research, Austin]
Impact on Power Consumption

- Process Variation
- Threshold Voltage Variation
- Leakage Power
- Workload
- Temperature Fluctuation

Need to consider process variation for leakage

Need to consider within-die temperature variation for leakage
Current Approaches

**Circuit level:**
Eg: Adaptive Body Biasing, Adaptive Voltage Scaling [Tschanz, JSSC’02]

Expensive to implement
Often too late for changes

**Microarchitecture level:**
Analytical modeling [eg. Bowman, JSSE’02]

Time consuming
Small scale

Need tools that predict the impact of variations on design metrics at early design stage

**System**
Current Approaches

- Workload variation not considered [e.g., Humenay, ASGI’06]
- Limitation in architecture variety, small scale [e.g., Chandra, ISLPED’06]
- Most work focuses on performance, not power

System level:
- Simulation with statistical model
- Develop tools that predict the impact of variations on power at early design stage
Network-on-Chip (NoC) for CMP

NoC power is large

Develop tools that predict the impact of variations on power for NoC
Outline

- Introduction and Motivation
- Methodology and Tool Development
- Case Study
- Conclusions and Future Work
Leverage Polaris Toolchain for Variations

Polaris

Step 1
Trident
Synthetic traffic generation

Design-space exploration tool

Step 2
LUNA
High-level on-chip network analysis

Microarchitecture parameters

Step 3
ORION
power and area models

power consumption
CMOS area
Performance (latency)

NoC designs projections

V. Soteriou, N. Eisley, H. Wang, B. Li, L.S. Peh, TVLSI’07
Polaris: ORION Power Model

Uniform temperature and no process variation

Modify ORION to account for process and temperature variations

Network Resource Utilization from LUNA

Leakage Power (Chen, ISLPED'03)

Dynamic Power

Trident
Synthetic traffic generation

Design-space exploration tool

LUNA
High-level on-chip network analysis

ORION
power and area models

Step 1
Step 2
Step 3

power consumption
CMOS area
Performance (latency)

NoC designs projections
New ORION Model (Process and Temperature Aware)

Yellow parts are the extended parts we added to ORION

Thermal Profile

ISAC Temperature profile

Power Profile

Power Profile

Total Power Profile at each process point

Set simulation number achieved?

Power Distributions

Network Resource Utilization from LUNA

D2D Process Variation Models

Utilization

Leakage Power HotLeakage

Dynamic Power

Step 1 Trident Synthetic traffic generation

Step 2 LUNA High-level on-chip network analysis

Step 3 ORION power and area models

Design-space exploration tool

Microarchitecture parameters

Power consumption

CMOS area

Performance (latency)

NoC designs projections
Outline

- Introduction and Motivation
- Methodology and Tool Development
- Case Study
- Conclusions and Future Work
Experimental setup

- Technology
  - 64-node chip
  - 65nm technology
  - Supply voltage: 1.2V
  - Frequency: 3.8GHz
  - Threshold voltage:
    - mean = 0.25 V
    - standard deviation = 6%
  - Die size: 14.4mm x 14.4mm x 0.6mm
### Design Space Explored

<table>
<thead>
<tr>
<th>Topology</th>
<th>2D mesh plain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2D mesh with express cube interval 2, 3</td>
</tr>
<tr>
<td></td>
<td>2D mesh with hierarchical link interval 2, 3, 4</td>
</tr>
<tr>
<td></td>
<td>2D torus plain</td>
</tr>
<tr>
<td></td>
<td>2D torus with express cube interval 2, 4</td>
</tr>
<tr>
<td></td>
<td>2D torus with hierarchical link interval 2, 3, 4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Buffer size (64-bit flits)</th>
<th>4, 8, 16, 32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual channels per link</td>
<td>1, 2, 4, 8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Routing</th>
<th>Deterministic routing (X-Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic</td>
<td>Long distance, moderately bursty, hot-spot traffic</td>
</tr>
</tbody>
</table>

What is the optimal on-chip network architecture for my application?
Effects of Process and Temperature Variations on Power

Distribution rather than deterministic

Average power increases

Process and temperature variations affect power consumption
Effects of Process and Temperature Variations on EDPPF

- Average
- Distribution
- Change

Early design exploration quantify effects of variations

Normalized EDPPF

Network architecture

x axis format: topology variant (buffer size, number of virtual channels)

EDPPF: Energy-Delay Product Per Flit
Other Aspects not Covered

☐ Sensitivity analysis
  ■ Only consider temperature variation
  ■ Only consider process variation
☐ Mean x standard-deviation metrics
Conclusions

- Early design stage tool that accounts for process and temperature variations

- Process and temperature variation strongly impacts power
  - Influence design choices
  - Need to be considered together
Future Work

☐ Study how within die process variations affect the network power consumption

☐ Studying process and temperature variation effects on network operating frequency
Thank you!