Practical and Theoretical Considerations on Low-Power Probability-Codes for NoC

Alberto Garcia-Ortiz
Leandro Soares Indrusiak

9th of September 2010
Patmos’10, Grenoble
Networks-on-Chip

Network on Chip: layered solution which reuse global interconnects

Advantages:
- Efficient use of global wires
- Generic interconnect architecture (IP reuse, predictable back-end)
- Scales with number of elements (in contrast to bus or shared bus)
- Layered approach (abstraction of timing, reliability, etc)

Problem: power efficiency

Energy efficiency of NoCs

Power consumption is a major obstacle for the incorporation of NoCs in embedded systems

Power problem increases when Virtual Channels are used
- Virtual channels permit QoS in the NoC by using time multiplexing
- Problem of multiplexing: increases in the transition activity (and thus, power)
## Low-power coding

### Coding opportunity

1. Reduce power dissipation in communication arch.
2. Applicable to other aspects (reliability, speed...)
3. Adapt existing theory on information/transmission to the requirements of on-chip interconnects

### Coding overhead

1. Power dissipation of coder/coder
2. Additional delay because of coding
3. Overhead of redundant lines
Agenda

Goal

Analyze the possibility of using **simple** low-power codes in NoCs based on virtual channels

- Intro: NoCs with virtual channels
- Effect of virtual channels on activity
  - Probability-Multiplex-Decorr code
- Probability-Multiplex code
- Theoretical analysis
- Experimental results
NoC structure
Virtual channels

- Flits are multiplexed in time
- Needs separate channel buffers for each virtual channel
**Virtual channels**

- Flits are multiplexed in time
- Needs separate channel buffers for each virtual channel
Idea: insert coder (and decoder) in Network interface to reduce switching activity

Problem: Virtual channel destroy correlation, thus Low-power code has no effect!!
Low-power switch with PMD coding

Idea: add XOR correlator and decorrelator
Low-power switch with PM coding

Idea: use only the Probability coder (instead of low-power coder)
Why standard techniques do not work?

Time multiplexing in NoC destroys correlation

Coding

Multiplex

T=12/6

T=7/6

T=38/11

A. Garcia-Ortiz / ITEM – U. Bremen
Why standard techniques do not work?

Time multiplexing in NoC destroys correlation

- **Coding**
  - 11111
  - 11111
  - 11111
  - 00010
  - 00001
  - 11110

- **Decorr**
  - 00100
  - 00100
  - 00100
  - 00011
  - 00011
  - 00110

- **Multiplex**
  - 00100
  - 11111
  - 00000
  - 11110
  - 00100

- **T**
  - 12/6
  - 7/6
  - 7/6
  - 38/11
Time multiplexing maintains low probability of inputs

PMD code

Coding

Multiplex

Decorr

T = 12/6

P = 7/6

P = 15/11

T = 15/11
PM code

Time multiplexing maintains low probability of inputs

\[
\begin{align*}
T &= 12/6 \\
P &= 7/6 \\
T &= 24/11
\end{align*}
\]
PM versus PMD

$tm_i = 2p_i(1 - p_i)$

PM less efficient reducing the activity than PMD
Exact coding

- Find the optimal $E_p$ (given the prob. characteristics of the signal $x[n]$)
- For a given $x[n-1]$  $f(a) = E(a, x[n-1])$ must be injective and minimizing power
Exact coding: graphical description
Exact coding: mathematical description

- **Original Gaussian**
  \[
  f_{XY}(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[ -\frac{x^2 + y^2 - 2\rho xy}{2\sigma^2(1-\rho^2)} \right]
  \]

- **Sort**
  \[
  \text{sort}(f_{XY}(x, k)) = \frac{2}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[ -\frac{x^2}{2\sigma^2(1-\rho^2)} \right] \exp\left[ -\frac{k^2}{2\sigma^2} \right]
  \]

- **Addition**
  \[
  p_W(w) = \begin{cases} 
    \frac{2}{\sqrt{2\pi}\sigma\sqrt{1-\rho^2}} \exp\left[ -\frac{w^2}{2\sigma^2(1-\rho^2)} \right] & \text{if } w \geq 0 \\
    0 & \text{if } w < 0 
  \end{cases}
  \]

- **Cost with vdm**
  \[
  E_{\text{Link}}(X) = \sum_w \frac{2}{\sqrt{2\pi}\sigma\sqrt{1-\rho^2}} \exp\left[ -\frac{w^2}{2\sigma^2(1-\rho^2)} \right] vdmCost_B(w)
  \]

  \[
  H_G(\sigma, \rho) = \frac{1}{2} \log_2(2\pi e) + \log_2\left(\sigma\sqrt{1-\rho^2}\right)
  \]

  \[
  E_{\text{Link}}(X) = \phi_G(H_G, B)
  \]
Experimental results

- Analysis of activity in buses and buffers
- Simulation performed in an idealized NoC

- Unlimited buffers
- Idealized producer and consumer

- Simulation with different data signals

  - Music
    - Bach
  - Male voice
  - Data OFDM
  - PATMOS image
  - Data gzip
## Experimental results

### Comparison of PMD and PM with different strategies

<table>
<thead>
<tr>
<th>Code</th>
<th>Raw Image PMD</th>
<th>Male voice PMD</th>
<th>Music PMD</th>
<th>GZIP exe PMD</th>
<th>OFDM data PMD</th>
<th>Mean PMD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PM</td>
<td>PMD</td>
<td>PM</td>
<td>PMD</td>
<td>PM</td>
<td>PM</td>
</tr>
<tr>
<td>K1</td>
<td>3.84</td>
<td>3.10</td>
<td>2.48</td>
<td>3.20</td>
<td>2.90</td>
<td>3.23</td>
</tr>
<tr>
<td>K0</td>
<td>4.03</td>
<td>3.05</td>
<td>2.65</td>
<td>3.33</td>
<td>3.16</td>
<td>3.29</td>
</tr>
<tr>
<td>vbm</td>
<td>5.69</td>
<td>3.15</td>
<td>3.60</td>
<td>3.96</td>
<td>4.06</td>
<td>4.00</td>
</tr>
<tr>
<td>corr+ none</td>
<td>1.21</td>
<td>1.93</td>
<td>2.41</td>
<td>3.10</td>
<td>2.13</td>
<td>2.49</td>
</tr>
<tr>
<td>corr+ K1</td>
<td>1.20</td>
<td>1.92</td>
<td>1.94</td>
<td>2.60</td>
<td>1.93</td>
<td>2.24</td>
</tr>
<tr>
<td>corr+ K0</td>
<td>1.10</td>
<td>1.78</td>
<td>1.78</td>
<td>2.42</td>
<td>1.63</td>
<td>1.98</td>
</tr>
<tr>
<td>corr+ vbm</td>
<td>0.87</td>
<td>1.53</td>
<td>1.92</td>
<td>2.89</td>
<td>1.45</td>
<td>2.32</td>
</tr>
<tr>
<td>dbm+ none</td>
<td>1.02</td>
<td>1.67</td>
<td>2.17</td>
<td>2.90</td>
<td>1.66</td>
<td>2.13</td>
</tr>
<tr>
<td>dbm+ K1</td>
<td>1.03</td>
<td>1.70</td>
<td>1.63</td>
<td>2.28</td>
<td>1.43</td>
<td>1.81</td>
</tr>
<tr>
<td>dbm+ K0</td>
<td>1.15</td>
<td>1.84</td>
<td>1.81</td>
<td>2.41</td>
<td>1.72</td>
<td>1.87</td>
</tr>
<tr>
<td>dbm+ vbm</td>
<td>0.80</td>
<td>1.42</td>
<td>1.82</td>
<td>2.78</td>
<td>1.27</td>
<td>2.09</td>
</tr>
</tbody>
</table>
Comparison of PMD and PM

![Bar chart comparing PMD and PM across different signals: Raw, Male, Music, GZIP, OFDM, Mean.](chart.png)

<table>
<thead>
<tr>
<th></th>
<th>Network Interface</th>
<th>Data Link</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMD</td>
<td>141</td>
<td>210</td>
<td>576</td>
</tr>
<tr>
<td>PM</td>
<td>141</td>
<td>210</td>
<td>0</td>
</tr>
</tbody>
</table>

A. Garcia-Ortiz / ITEM – U. Bremen
Leakage reduction

Additional reduction in leakage with leakage skewed flip-flops

![Graph showing leakage improvement vs. probability with data points for un-coded and PMD-coded data.]
Conclusions

- QoS and virtual channels in NoCs are getting an important issue.
- Existing low-power coding techniques were not able to decrease power consumption in NoCs based on virtual channels.
- Proposed PM and PMD coding approach provides a framework to apply coding in NoCs based on virtual channels.
- Achieved reduction of 34% (45% for DSP signals).
- Leakage in buffers reduced by 30%.

To do: Analysis in a non-idealized NoC implementation.
Thanks for your attention
Classification of coding strategies

Knowledge of signal
- General purpose
- Dedicated
- Adaptative

Coder memory requirements
- Memoryless
- With memory

Additional lines for link
- Redundant
- Non-redundant

Trade-off complexity vs. power minimization

General purpose memoryless non-redundant code

General purpose 2-tap memory non-redundant code
Example of coding strategies

K0, K1: Memoryless code
BI: Bus Invert (redundant code)
VBM: Value Based Mapping
DBM: Data Based Mapping
Why standard techniques do not work?

- Standard low-power coding is based in reducing transition activity
- But effect of coding is destroyed by virtual channels !!

<table>
<thead>
<tr>
<th>Coding</th>
<th>Signal transition (raw image)</th>
<th></th>
<th>Signal transition (average)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Relative</td>
<td>Total</td>
<td>Relative</td>
</tr>
<tr>
<td>un-encoded</td>
<td>1.21</td>
<td>100.0%</td>
<td>3.16</td>
<td>256.2%</td>
</tr>
<tr>
<td>k1</td>
<td>1.20</td>
<td>99.7%</td>
<td>3.10</td>
<td>252.1%</td>
</tr>
<tr>
<td>k0</td>
<td>1.10</td>
<td>90.7%</td>
<td>3.05</td>
<td>252.2%</td>
</tr>
<tr>
<td>bi</td>
<td>3.43</td>
<td>283.4%</td>
<td>4.50</td>
<td>371.7%</td>
</tr>
<tr>
<td>corr+k1+decorr</td>
<td>1.20</td>
<td>99.4%</td>
<td>3.10</td>
<td>256.1%</td>
</tr>
<tr>
<td>corr+k0+decorr</td>
<td>1.10</td>
<td>90.7%</td>
<td>3.05</td>
<td>252.2%</td>
</tr>
<tr>
<td>corr+vbm+decorr</td>
<td>0.87</td>
<td>71.6%</td>
<td>3.99</td>
<td>330.1%</td>
</tr>
<tr>
<td>corr+bi</td>
<td>1.14</td>
<td>94.0%</td>
<td>4.50</td>
<td>371.7%</td>
</tr>
<tr>
<td>dbm+ none+decorr</td>
<td>1.02</td>
<td>84.0%</td>
<td>3.99</td>
<td>329.6%</td>
</tr>
<tr>
<td>dbm+k1+decorr</td>
<td>1.03</td>
<td>85.3%</td>
<td>3.99</td>
<td>330.0%</td>
</tr>
<tr>
<td>dbm+k0+decorr</td>
<td>1.15</td>
<td>94.0%</td>
<td>3.99</td>
<td>330.0%</td>
</tr>
<tr>
<td>dbm+vbm+decorr</td>
<td>0.80</td>
<td>65.9%</td>
<td>4.00</td>
<td>330.4%</td>
</tr>
<tr>
<td>dbm+ bi</td>
<td>1.00</td>
<td>82.3%</td>
<td>4.49</td>
<td>371.2%</td>
</tr>
</tbody>
</table>
Low-power switch

- Idea: insert coder (and decoder) in Network interface to reduce switching activity.
- Problem: Virtual channel destroy correlation, thus Low-power code has no effect!!