On-chip Bus Thermal Analysis and Optimization

Feng Wang, Yuan Xie, N. Vijaykrishnan and M. J. Irwin
Penn State University
March, 2006
Outline

- Motivation
- Bus Energy & Thermal model
- Thermal impact of system bus with technology scaling
- Theoretical analysis of bus thermal optimization
- Thermal spreading encoding method
- Experimental results
- Conclusions
Motivation

- On-chip bus power consumption: an important part of the overall system power consumption
- The interconnect temperature can be as high as 90°C [Shang]
- High temperature has a negative impact on interconnect performance and reliability
- The on-chip bus has not gained enough attention
Energy Model

- Heat generated on the interconnects regardless the current direction
- The energy drawn from power [Sotiadis]
  - Energy stored in the capacitance
  - Energy dissipated as heat ($E_{heat0}$)
- Energy generated when bus line discharged
  - Energy stored in the capacitance
  - Energy dissipated as heat ($E_{heat1}$)
Energy Model

- Heat dissipated in the interconnects and drivers or receivers ($E_{heat0} + E_{heat1}$)
- Optimal length of the wire segment ($S_{opt}$)
- Optimal size of the gate ($L_{opt}$)
- Heat dissipated in the interconnects

$$\frac{l_{opt}r(E_{heat1} + E_{heat0})}{(l_{opt}r + \frac{r_0}{s_{opt}})T}$$
Chiang’s Thermal Model

- Assumptions:
  - Heat dissipated through the underlying layer
  - Ignore thermal conductivity variation along the length
  - Via effects diminished beyond critical length
Chiang’s Thermal Model

- For edge wires

\[ P_1 = C_1 \times \frac{\partial T_1}{\partial t} + \frac{T_1 - T_0}{R_1} + \frac{T_1 - T_2}{R_{\text{inter}}} \]

\[ P_{32} = C_{32} \times \frac{\partial T_{32}}{\partial t} + \frac{T_{32} - T_0}{R_{32}} + \frac{T_{32} - T_{31}}{R_{\text{inter}}} \]

- For the wires in the middle

\[ P_i = C_i \times \frac{\partial T_i}{\partial t} + \frac{T_i - T_0}{R_i} + \frac{(2T_i - T_{i-1} - T_{i+1})}{R_{\text{inter}}} \]
Technology Scaling

- Not worst case analysis
- Technology Parameters [ITRS 2004 edition]
- Thermal conductivity of low k materials, clock, interconnect pitch, Number of interconnects layer

![Graph showing peak temperature increases over substrate with technology scaling](image-url)

- Peak Temperature Increases over Substrate with Technology Scaling
- Peak Temperature Rise (unit °C)
- Technology Nodes: 90nm, 65nm, 45nm, 32nm
- Graphs for gzip and mesa
Problem formulation:
- Given that the total power consumption across the system bus $P(t)$
- Find an optimal distribution of the power consumption among the bus wires

Optimal power consumption distribution:

$$P_i = \frac{P_{total}}{Buswidth}$$
Thermal Spreading Encoding

- Distribute the switching activities among the bus line evenly
- Rotating the bus line position at a certain period
- More effective combined with the simple low power coding scheme
Spreading coding

Advantages:
- Low area overhead
- Low power overhead
- Low complexity

Diagram:
- Raw Address
- 32 x 32 Cross Bar
- Shift Register
- Coded Address

Flow:
- Original bus
- Low power bus encoder
- Spreading encoder
- Coded bus
- Data
Experiment results

- Instruction Address Bus:
  - large power consumption
  - Regular switching patterns
- Switching activities distribution
- Power distribution
- Temperature reduction
Instruction Address Bus

Switching Activities of Address Bus before Encoding

Switching Activities of Address Bus after Encoding
Instruction Address Bus

Average Power of Address Bus before Encoding (Unit Length m)

Average Power of Address Bus after Encoding (Unit Length m)
Instruction Address Bus

Peak Temperature Reduction on Address Bus

Temperature Reduction (°C)

- T0cac
- spreading
- T0+spreading

Bars for gzip, vpr, parser, mesa, ammp, vortex, mcf, equake
Instruction Address Bus

- Compared to T0CAC -- a very effective low power encoding
- Spreading + $T0$ coding better than T0CAC
  - Simple
  - Lower power overhead
  - More effective in temperature reduction
Instruction Bus

- Instruction data bus
  - Large power consumption
  - Irregular switching patterns
- T0CAC can not be used
- Much better than Bus Inverting
Switching Activities of Instruction Bus before Encoding

Normalized Value vs. Instruction Lines

Switching Activities of Instruction Bus after Encoding

Normalized Value vs. Instruction Lines
Instruction Bus

Peak Temperature Reduction on Instruction Bus

-1 0 1 2 3 4 5 6 7 8

Temperature Reduction (°C)

<table>
<thead>
<tr>
<th>gzip</th>
<th>vpr</th>
<th>parser</th>
<th>mesa</th>
<th>ammp</th>
<th>vortex</th>
<th>mcf</th>
<th>equake</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus Inverting</td>
<td>Spreading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

-1 0 1 2 3 4 5 6 7 8

Bus Inverting

Spreading
Conclusion

- Characterize the thermal impact on on-chip bus
- Improved bus energy and thermal models
- Optimal power distribution for temperature reduction
- Spreading+T0 encoding better than T0CAC
- Effective for both instruction address and instruction bus