Test Infrastructure Design for Core-Based System-on-Chip Under Cycle-Accurate Thermal Constraints

Thomas Edison Yu†, Tomokazu Yoneda†, Krishnendu Chakrabarty‡ and Hideo Fujiwara†

† Nara Institute of Science and Technology
‡ Duke University

Nara Institute of Science & Technology
Outline

• Background
  – Core-based testing
  – Power / Heat-related problems during test
  – Limits of power-constrained SoC testing
• Related Works
• Objectives
• Proposed Method:
  – Fixed-TAM architecture
  – Scheduling techniques
  – Thermal model & cost function
• Experimental Results
• Summary
Core-Based System-on-Chip

- SoC (System-on-Chip):
  - One-chip system
  - Use pre-designed functional blocks called IP cores
  - Reduced design and manufacturing cost
    - IP design and test re-use

- Test challenges
  - High test data volume
  - Limited access to internal circuitry
  - Long test application time (TAT)
  - High test power and temperature

Courtesy of WindowsForDevices.com
Design for Test (DFT) for SoCs

- **TAM (Test Access Mechanism)**
  - Dedicated test bus connecting test source/sink and core-under-test (CUT)

- **Wrapper**
  - Isolates CUT from other cores during test
  - Interfaces TAM & core
  - Enables test reuse

Limited # of I/O pins

Most cores not directly controllable or observable
Test Scheduling

- Determine test sequence for each core
- Minimize test time under certain constraints
  - Power, TAM width, temperature, etc.
  - Parallel testing results in high test power and temperature
Power & Heat-related Issues During Test

• High test power dissipation
  – can cause chip damage or random errors
  – result in yield loss

• High power dissipation can cause overheating
  – every 20°C rise in temperature = approx. 5-6% timing delay
  – chip packaging are designed for worst case typical application
Limits of Power-constrained SoC Testing

- Ignore non-uniform spatial power distribution across chip
  - layout, core proximity affects temperature
  - can't ensure thermal-safety
- Ignore effect of time on temperature

*Results using HotSpot temp. simulator (Skadron et al., ISCA'03)
Objectives

• Given an SoC, test data and maximum allowable temperature,

Propose a thermal-safe test architecture design and test scheduling methodology
  – the given thermal constraint is satisfied at any time during test
  – the test application time is minimized
Related Works

• Thermal-safe test scheduling (*Rosinger et al.*, *DATE’05*)
  – group cores with fixed wrappers into test sessions
  – minimize temperature per session

• Uniform heat distribution (*Liu et al.*, *DFT’05*)
  – use layout information to determine wrapper configuration and test schedule

• Test set partitioning and interleaving (*He et al.*, *DFT’06*)
  – partition test set when temperature constraint is exceeded
  – interleave test partitions sequentially, allow other cores to cool down
Characteristics of Related Works

• Use constant power per core
• Ignore effects of active cores on non-active cores
• Do not optimize both the TAM and Wrapper configuration

• Thermal-safe TAM / Wrapper Co-optimization (Yu et al., ATS’07)
  – use cycle-accurate power profiles per core wrapper configuration
  – consider heat exchange across all cores and test sequence via heat dissipation paths
  – first work to optimize TAM and Wrapper under a thermal constraint
Research Contributions

• Improvements over work done for ATS`07
  – Allow schedule reshaping
  – Allow dynamic test partitioning and interleaving
  – Allow insertion of bandwidth matching circuitry
  – Find solutions under much tighter temperature constraints

• Preserve advantages of ATS`07 work
  – Cycle-accurate power and temperature profiles
  – Consider inter-core temperature effects
  – Optimize TAM & Wrapper architecture
Target Test Architecture

- Characteristics:
  1. TAM has fixed partitioning
  2. Cores are assigned to only one TAM partition
  3. Cores in the same TAM scheduled sequentially, order is variable
  4. Cores in different TAMs can be scheduled in parallel

Schedule A
Schedule B

SoC d695

TAM 1 = 16bits
TAM 2 = 8bits
TAM 3 = 8bits

Time t
Test Schedule Reshaping, Partitioning & Interleaving

• Test schedule reshaping
  – Reconfigure schedule to minimize temperature of target core
  – Ex. Avoid testing hot cores in parallel and in sequence

• Test partitioning and interleaving
  – Partition test before temperature exceeds constraint

Max. Temp = 110°C

Max. Temp = 100°C

Max. Temp = 95°C
Bandwidth Matching

- Frequency throttling
  - lower scan frequency => lower power
- Add bandwidth matching circuitry to TAM partition
  - 2*TAM width, ½ freq. => temperature reduction ideally w/o TAT increase
Bandwidth Matching

- Frequency throttling
  - lower scan frequency => lower power
- Add bandwidth matching circuitry to TAM partition
  - 2*TAM width, ½ freq. => temperature reduction ideally w/o TAT increase
Proposed Method Flow

- Scheduling is NP-Hard
  - need heuristic scheduling algorithm

- Thermal simulation is time consuming
  - need simplified thermal model
  - need simple thermal cost function

Diagram:

- Initial TAM design & schedule
- Thermal simulation
  - $\max T \leq \text{Temp}_{\text{max}}$
  - NO
- Reshape possible
  - YES
  - Reshape schedule
  - NO
- Can partition hot core?
  - YES
  - Do partitioning
  - NO
- Virtual TAM limit reached?
  - YES
  - END
  - NO
- Do bandwidth matching
Thermal Model

• Model SoC as a network of thermal resistances
  – first proposed by P. Rosinger, K. Chakrabarty, et.al (DATE’05)
  – takes advantage of thermal and electrical duality
  – only consider lateral thermal resistances

- Models lateral heat flow
- More heat flow from cores = higher temperature
Thermal Cost Function

- \( T_{cont_j}(c_i) \): thermal contribution of Core \( j \) to Core \( i \)

\[
T_{cont_j}(c_i) = \frac{R_{ji}}{R_{TOT,j}} \times P_{avg_j} \times \frac{T_{rel_{ji}}}{T_{AT,i}}
\]

where:
- \( R_{ji} \): Lateral thermal resistance from core \( c_j \) to \( c_i \) \((R_{ii} = 0)\)
- \( R_{TOT,j} \): Total lateral resistance of core \( c_j \)
- \( P_{avg_j} \): Average power dissipation of \( c_j \)
- \( T_{rel_{ji}} \): Relative test time of \( c_j \) and \( c_i \)

Minimize thermal contribution to hotspot core

Reset cost values and revert to initial schedule
Experimental Setup

• Benchmark *ITC’02 SoCs*
  – d695, p22810 with hand-crafted layouts
  – cycle-accurate power profiles from Samii et al.

• Scheduling parameters
  – $TAM = 16, 24, 32, 64$
  – $Tmpmax$: initially set to max temperature of schedule under no power & thermal constraint
    • decreased by $5^\circ$C intervals
Max. Temperature & Power vs. Temp. Constraint

- d695 with TAM = 16bits

Max. Temp, Max. Power vs. Temp. Constraint

Relatively constant max. power but temperature was lowered further
Max. Temperature & TAT vs. Temp. Constraint

- d695 with TAM = 16bits

Relatively minimal TAT change but temperature was lowered further
Max. Temperature & Power vs. Temp. Constraint

- p22810 with TAM = 32bits

lower max. power doesn’t ensure lower temperature
Max. Temperature & TAT vs. Temp. Constraint

- p22810 with TAM = 32bits
## Minimum Temperature Comparison

Max. 40% reduction in minimum temperature

<table>
<thead>
<tr>
<th>SoC</th>
<th>TAM=16</th>
<th>TAM=24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATS’07</td>
<td>Proposed</td>
</tr>
<tr>
<td>d695</td>
<td>92.79C</td>
<td>55.8C</td>
</tr>
<tr>
<td>p22810</td>
<td>133.02C</td>
<td>77.71C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SoC</th>
<th>TAM=32</th>
<th>TAM=64</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATS’07</td>
<td>Proposed</td>
</tr>
<tr>
<td>d695</td>
<td>77.15C</td>
<td>69.91C</td>
</tr>
<tr>
<td>p22810</td>
<td>109.36C</td>
<td>71.17C</td>
</tr>
</tbody>
</table>
Summary

• Studied the impact of test-set partitioning, bandwidth matching on thermal-aware TAM / Wrapper optimization and test scheduling
• Algorithm based on computationally tractable thermal-cost model makes thermal simulation more useable; ensures thermal safety
• The results show that:
  – method allows more flexibility to trade-off temperature and TAT while minimizing TAT increase
  – method provides solutions even under tight temperature constraints, including situations where previous work fails to find a solution