System-on-Chip monitoring networks targeting nanometer technologies

CMOS Emerging Technologies
Marisa López Vallejo, Pablo Ituero
ETSIT-UPM
marisa@die.upm.es, pituero@die.upm.es
Outline

• Introduction
• Monitors taxonomy
• Other approaches
• Proposal
• Conclusions
Introduction

- Millions of transistors in a single die allow the implementation of very complex architectures:
  - SoC, MPSoC, NoC, Multi-core processors

Source: http://www.intel.com/research/silicon/mooreslaw.htm
However...

Aggressive technology scaling comes at the cost of significant yield reduction

- Increased process variations
  - Variability-aware design becomes a must
- Tackling with higher power densities
  - Hot-spots mitigation through DTM
- Worse aging behavior
  - DFR
Process variations

• Within-die variations make the design of reliable complex systems more complicated
  ▫ Line edge roughness
  ▫ Interface roughness
  ▫ Random discrete dopants
  ▫ Gate tunneling...

• Effects
  ▫ Critical path variations
  ▫ Static power
  ▫ VT variations
  ▫ Yield,
  ▫ Noise robustness

• Solutions
  ▫ Variations-aware design
  ▫ Design For Manufacturability
  ▫ Critical path monitoring
  ▫ Calibration

Funded by FP7-STREP REALITY: Reliable and Variability tolerant System-on-a-chip Design in More-Moore Technologies

Local Versus Global Statistical Variability
Increasing Proportion of Overall Variability

Past

Present

Source [6]
Aging

- Aging causes
  - NBTI: negative-bias temperature instability
  - TDDB: time-dependent dielectric breakdown
  - Electromigration

- Aging effects
  - Critical path delay
  - Leakage power increase
  - Failures (reduced reliability)

- Aging-aware design
  - Autonomous computing
  - Dynamic Reliability Management, DRM

Temperature issues

- Spatial and time distributed hotspots
- Adverse effects
  - NBTI, TDDB, increased leakage, reliability, mechanical strains
- Thermal-aware design
  - Cooling
  - Temperature sensing and monitoring
  - Dynamic Thermal Management (DTM)
    - DVFS schemes
    - Workload management

On-chip monitoring and calibrating network

- After fabrication, all techniques conceived to mitigate these adverse effects require
  - In-situ monitors
    - Sensors are more reliable
  - Ultra light-weight interconnection infrastructure
    - Reduced area, power overhead
    - Transparent to the designer
    - Multi-purpose
      - Monitoring and calibrating functions
      - Targeting all previous needs
Required monitors

- Very reduced area and power
- Digital output
- Linear response
- Simple interface
- Transparent
- Easy to integrate in CMOS
Temperature monitors

- Efficiently allocated in the silicon to detect hotspots
- Low self-heating especially important
- Range: depends on the application.
  - For example, intel i5 20-111ºC
- Accuracy: 1 ºC
- Resolves in the order of milliseconds
- Spatial constant: 1 mm
Example of Temperature Sensor

- Based on the thermal dependence of the leakage currents
- Time to digital
- Digital interface with a logarithmic counter
  - 0.35 um AMS Technology
  - 10250 nm²
  - 1.05-65.5 nW at 5 samples/s
  - Resolution 0.28°C
  - Maximum error (3σ) = ±1.97°C

![Temperature Sensor Schematic](image)

Critical path monitors

- Provide real-time performance information based upon the variation of the delay in a path
- Time to digital conversion
- Output provided to DVFS techniques
- Range-Accuracy:
  - F.i. 25 mV AC and 10 mV DC (65nm)
- Time period: depends on the application.
  - From 1/clk to 1/Mclk
  - Slow monitors record worst/average delay
- Quantization: few bits
- Spatial distribution: up to hundreds
Supply voltage monitors

- Power supply behind safety threshold can be due to
  - Resistive path
  - High instant current
- Monitors avoid unreliable operation of the chip
- Time constant: short
  - Quick actions are demanded
- Range-accuracy: 10-20 mV
- Spatial distribution: only close to the circuits that demand higher currents
Aging monitors

- Digitally quantify reliability/performance/power degradation due to aging
  - TDDB, NBTI, hot-carrier effect...
- Most of them based on critical path monitors
- Quantization: few bits
- Time period: large
- Spatial distribution: up to thousands
## Monitors Taxonomy Summary

<table>
<thead>
<tr>
<th>Magnitude</th>
<th>Time Period</th>
<th>Density</th>
<th>Quantization needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Milliseconds</td>
<td>Hundreds/chip</td>
<td>≈ 8 bits</td>
</tr>
<tr>
<td>Critical path</td>
<td>Microseconds</td>
<td>Hundreds/chip</td>
<td>≈ 7 bits</td>
</tr>
<tr>
<td>Supply voltage</td>
<td>Nanoseconds</td>
<td>Few/chip</td>
<td>≈ 6 bits</td>
</tr>
<tr>
<td>Aging</td>
<td>Long</td>
<td>Thousands/chip</td>
<td>≈ 4 bits</td>
</tr>
</tbody>
</table>
Network justification

- Current complex designs may need up to thousand sensors for monitoring
- Flexibility to deal with
  - Calibration and monitoring
  - Different monitors
- No previous design of a common infrastructure
  - Only for thermal monitoring
- But a clear need is shown in the literature
System using temperature monitors

Previous works study the number and allocation of temperature sensors for DTM

No information on the way they are connected is provided

System using temperature monitors: previous approaches

• Evolution:
  ▫ From a single sensor per die to hundreds

• Previous approaches
  ▫ Boundary-scan like

Maestro: reliability-aware system

Previous approaches

- Only targeting temperature sensors for DTM
- Boundary scan-like
- Replicated digitalization stage
Our proposal

- Ultra light-weight network
  - Simplicity is the key to success
- Constrained to:
  - Monitors with time-to-digital conversion
  - Common interface
- Shared digitalization for all monitors
- Time-multiplexing scheme to improve performance
Monitoring network

- **Monitor_1**: Sensor, Network Interface, Control
- **Monitor_2**: Sensor, Network Interface
- **Monitor_n**: Sensor, Network Interface
- **Digitalization**
- **Control Logic**

![Diagram of Sensor Network Interface Monitor](image)
Comparison

- 32 monitor network
- 90 nm UMC Library
- Area savings: 25% each monitor + 16% net
- Power savings: 16%

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Area</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single monitor</td>
<td>Control</td>
</tr>
<tr>
<td>Boundary-ScanLike</td>
<td>504</td>
<td>9140</td>
</tr>
<tr>
<td>Digital Resource Sharing</td>
<td>374</td>
<td>9294</td>
</tr>
</tbody>
</table>
Conclusions

• Current nanometer technologies allow the integration of million of transistors at the cost of
  ▫ Increased power densities
  ▫ Lower reliability
• Designers should take care of all adverse effects
  ▫ DFM, DTM, DPM
• Need of a common infrastructure for monitoring and callibrating the increased amount of sensors required by nowadays circuits
• Proposed network
  ▫ Digitalization sharing, time multiplexing based
  ▫ Common interface
  ▫ Low area and power overhead
  ▫ Scalability, flexibility
References


