Curing the Chip Fever: Thermal Sensing and Actuation in Nano-scale Systems

Yufu Zhang, Bing Shi, Domenic Forte and Ankur Srivastava

Introduction

The Heating Problem:
- As VLSI technology continues to scale down, the leakage power increased dramatically.
- The chip temperature can easily rise up to 120 degrees Celsius.
- This leads to highly unreliable and error-prone chip behavior and performance degradation.
- Random chip workload and the variability inherent in the fabrication process made the situation even worse (see fig below).

Curing the Chip Fever

Due to the above problems, a systematic approach is highly desired to address the thermal stress issue. We propose the following:

1. Testing (Sensing)
   - Runtime monitoring: thermal sensors implemented on chip as Ring Oscillators
   - Problem: sensors are affected by noise and fabrication randomness;
   - How to design and place sensors to minimize overhead?

2. Diagnosis
   - Reconstruct the entire chip-level thermal profile given limited and noise-corrupted sensor observations.

3. Treatment
   - Dynamic frequency scaling under thermal constraints
     - Estimated thermal profile as the input to guide decision making
     - Use flexible constraints to improve performance

The Methodology Flow

Infrastructure Design Flow

- Statistical Power/Thermal Information Input
- Sensor Placement
- Sensor Design
- Fusion Center Design

Sensing & Actuation Flow

- Combine Sensor Data (Fusion Center Output)
- Filter Out Sensor Noise
- Estimate the Chip Level Thermal Profile
- Dynamic Frequency Scaling

Infrastructure Design

Sensor Placement:
- Exploit the thermal correlations among different chip modules;
- Target better accuracy with fewer sensors.

Sensor Design:
- Make the sensors more robust to noise;
- Compress the sensors according to their relative importance;
- Minimize area/power overhead.

Fusion Center Design:
- Filter out noise, combine all sensor readings and send to OS;
- Estimate chip-level thermal profile and feed it back to actuation unit.

Feedback Control Design

Thermal Profile Estimation:
1. Gaussian Case
   When the underlying thermal/power randomness exhibits Gaussian nature, we have the optimal linear estimator:
   \[ E(\hat{T}_{F}) = \mu_{F} + \sum_{k} A_{k} (\sum_{k} A_{k}^{\top})^{-1}(\bar{T}_{k} - \mu_{k}) \]

2. Non-Gaussian Case
   Use either moment matching heuristic or hypothesis Testing
   - details omitted
   Dynamic Tracking

3. Dynamic Frequency Scaling
   Once we know the runtime thermal information, we can take corrective actions such as frequency scaling
   - Soft constraint: allow the temperature to violate the constraint as long as the duration of violation is within a certain limit.
   - Optimal solution: run the processor at the maximum frequency first, and then shut it down so as to let T go back to T_{thermal}. 