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

Yufu Zhang, Bing Shi and Ankur Srivastava

Introduction

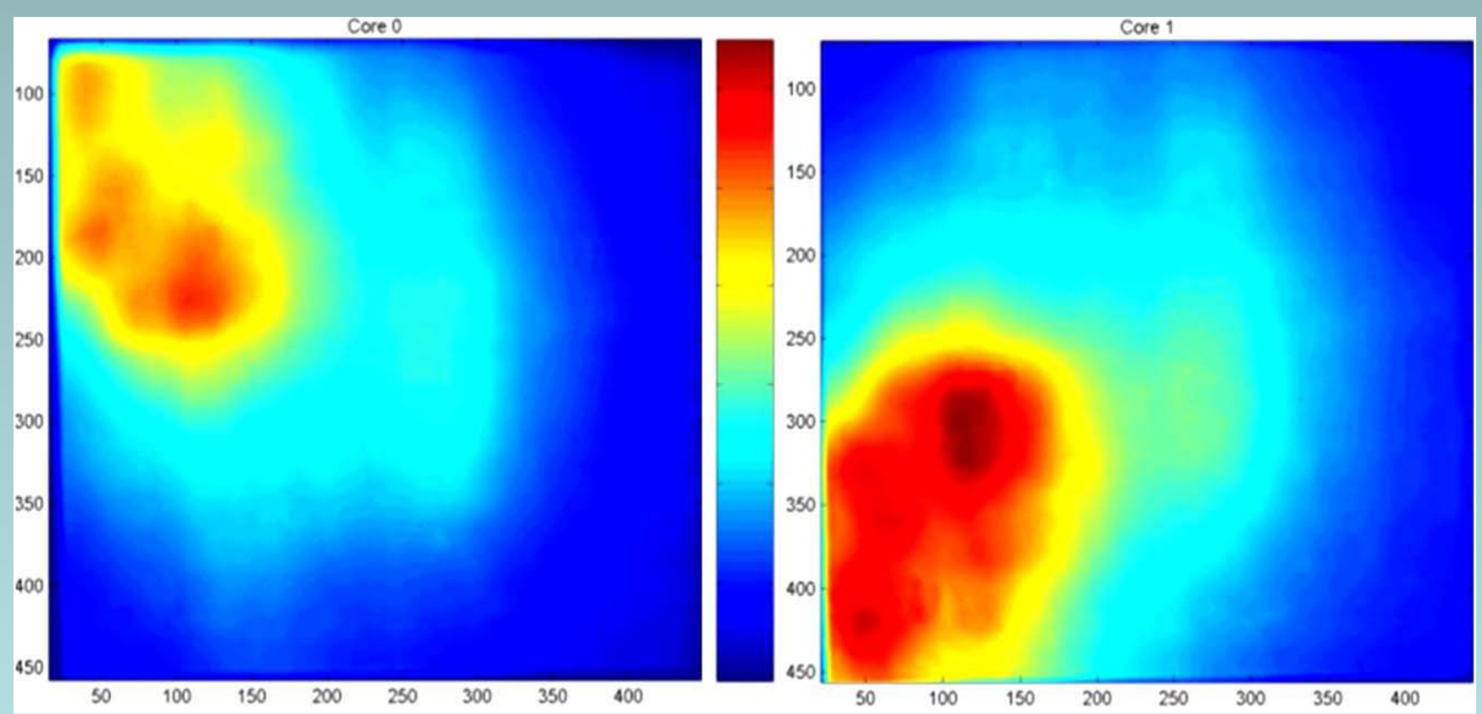
The Heating Problem:

• As technology continues to scale down, the leakage power continues to increase.

• The chip temperature can easily rise up to 150 degrees Celsius.

• Highly unreliable and error-prone chip behavior (even break-down sometimes), degraded performance.

• The random chip workload and the variability inherent in the fabrication process made the situation even worse.



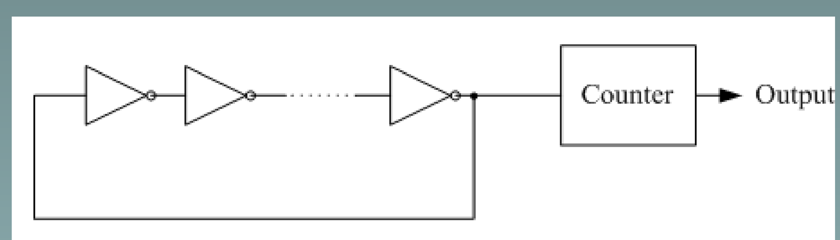
[1] Eren Kursun and Chen-Yong Cher, "Variation-awar Thermal Characterization and Management of Multi-core Architectures", Proc. IEEE International Conference on Computer Design, October 2008

How to Cure the Chip Fever?

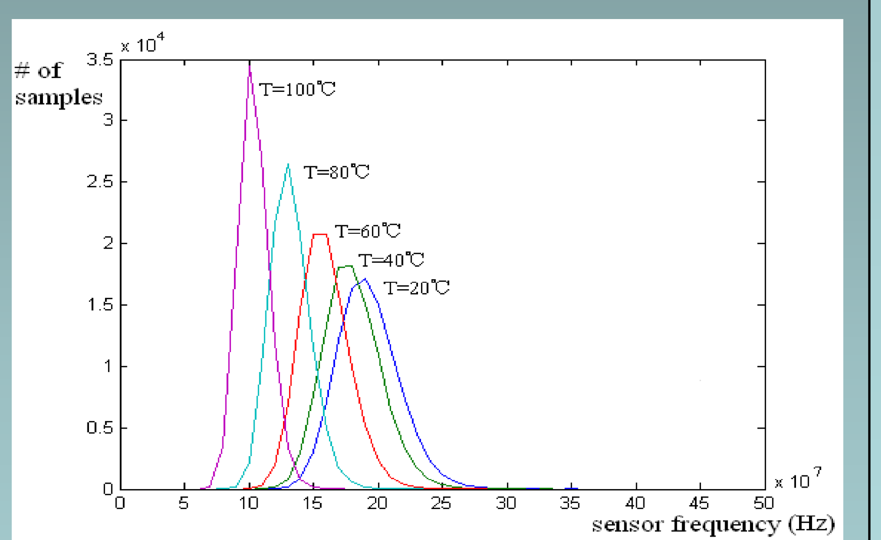
Due to the above problems, an integrated solution is highly desired to address the thermal problem.

1. The Sensing (Testing)

• On-chip thermal sensors can be implemented by Ring Oscillators



• The sensors themselves can be affected by noise and various fabrication randomness.



• Design and place sensors smartly to minimize overhead

2. The Diagnosis

• Reconstruct the entire thermal profile given limited and noise-corrupted sensor observations

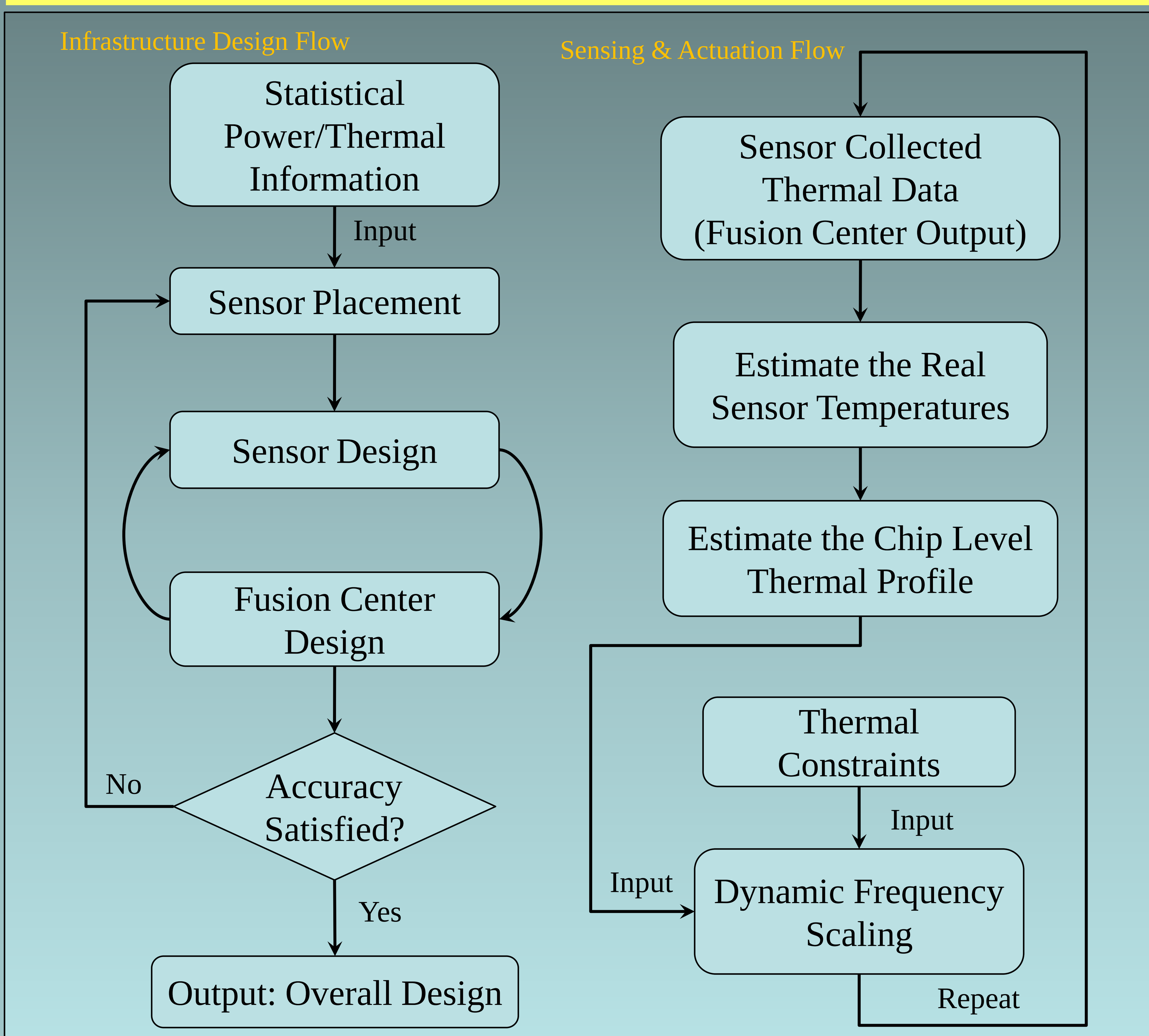
3. The Treatment

• Dynamic frequency scaling under thermal constraints

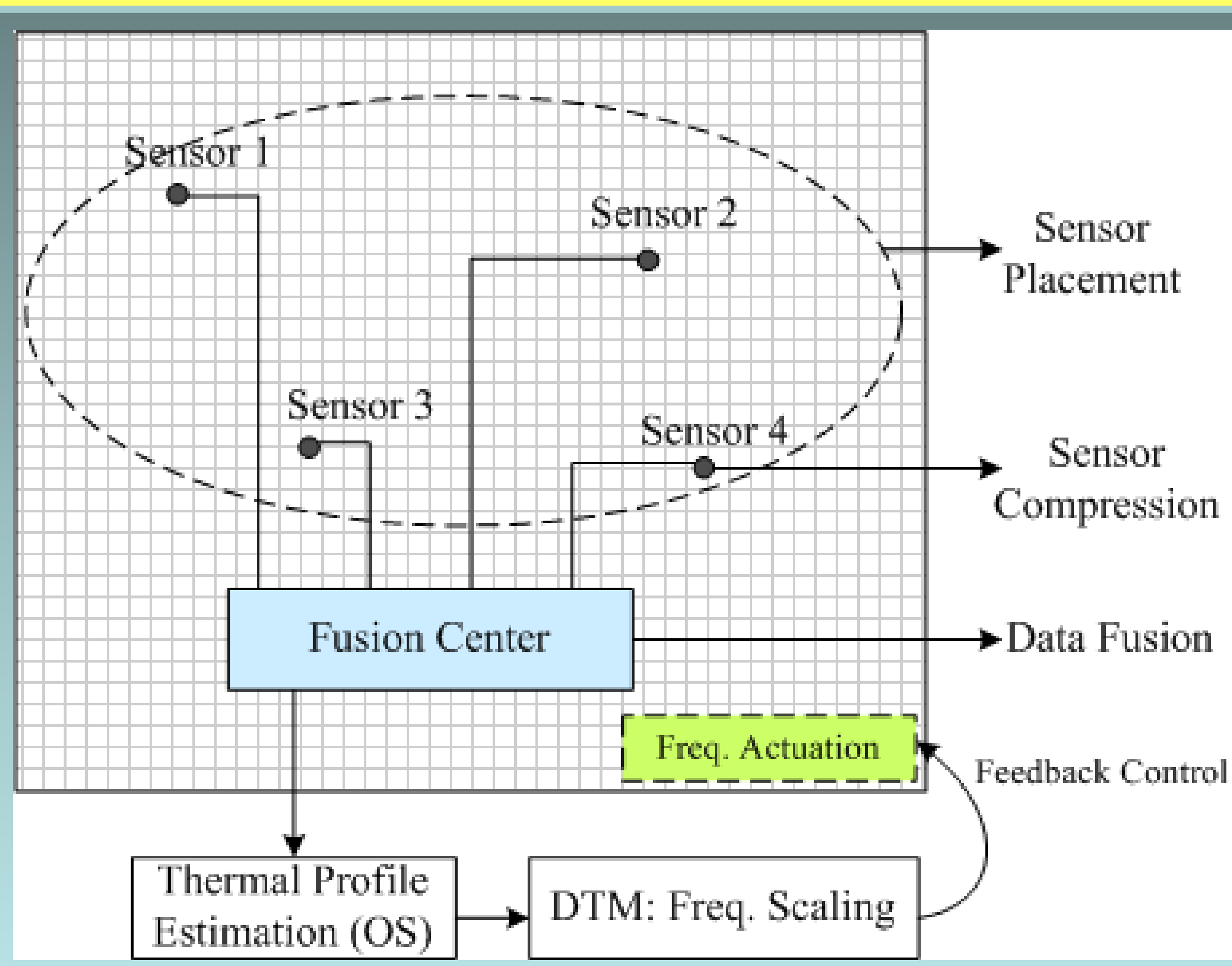
--- Estimated thermal profile as the input to guide decision making

--- Use more flexible constraints to improve performance

The Methodology Flow



Infrastructure



Sensing Infrastructure Design

Sensor Design:

--- Make the sensors more robust to noise, compress the sensors for minimal area/power overhead

Sensor Placement:

--- Exploit the thermal correlations among different chip modules

--- Better accuracy can be achieved with less sensors.

Fusion Center Design:

--- Use hypothesis testing to reconstruct the accurate sensor temperatures from the compressed and noise-corrupted sensor readings

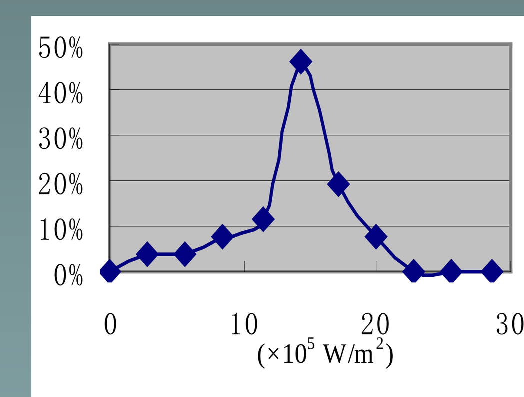
--- Combine all sensor readings and send to OS

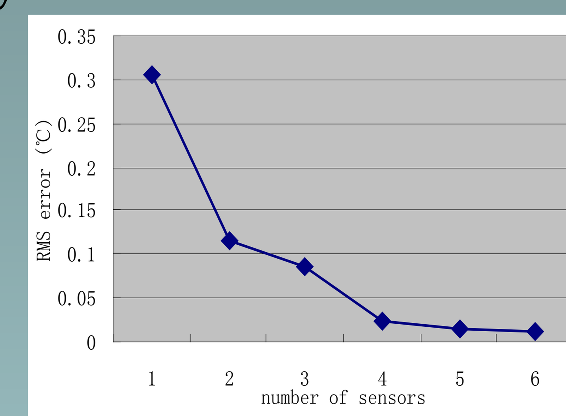
Feedback Control Design

Thermal Profile Estimation:

1. Gaussian Case

When the underlying thermal/power randomness are jointly Gaussian, the optimal estimation for all chip locations (in the MMSE sense) is simply the expected temperature conditioned on the sensor observations



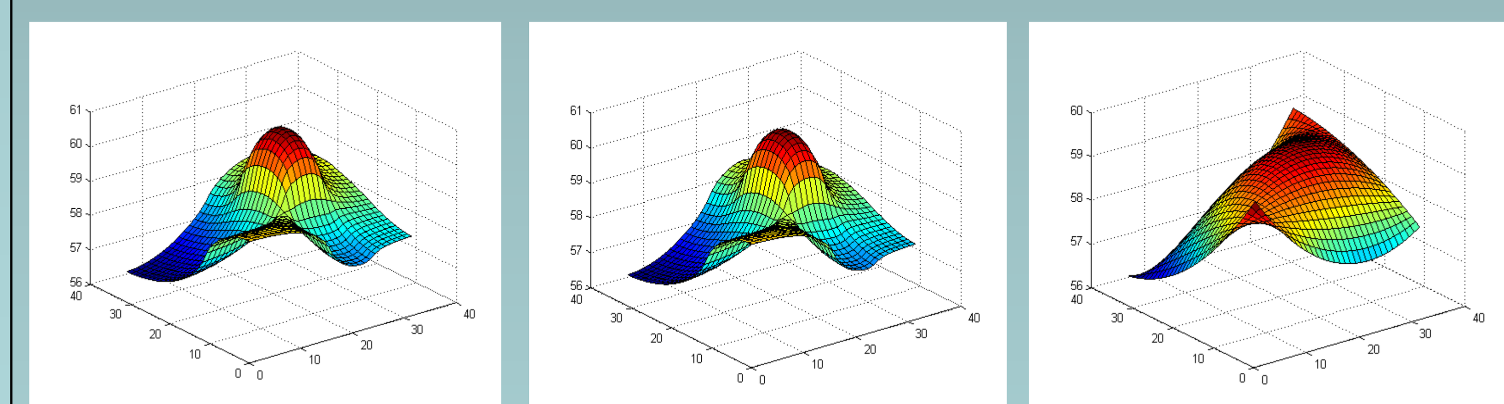


$$E(\vec{P}|\vec{T}_s) = \overline{\mu_P} + \Sigma_{pp} A_s^T (A_s \Sigma_{pp} A_s^T)^{-1}(\vec{T}_s - A_s \overline{\mu_p})$$

2. Non-Gaussian Case

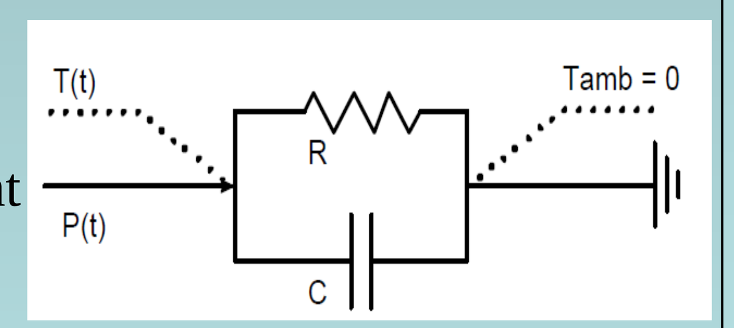
Fit a Gaussian PDF using moment matching

Use Hypothesis Testing – details omitted



3.Dynamic Frequency Scaling

Model the thermal behavior by an RC circuit, in which voltage/current represents temperature/power.



• Soft constraint: allow the temperature to violate the constraint, as long as the total duration of violation is within a certain threshold.

• Optimal solution: always run the processor at the maximum frequency first, and then shut it down so as to let T go back to T_m.