A Statistical Framework for Designing On-chip Thermal Sensing Infrastructure

Yufu Zhang, Bing Shi, Ankur Srivastava

University of Maryland, College Park

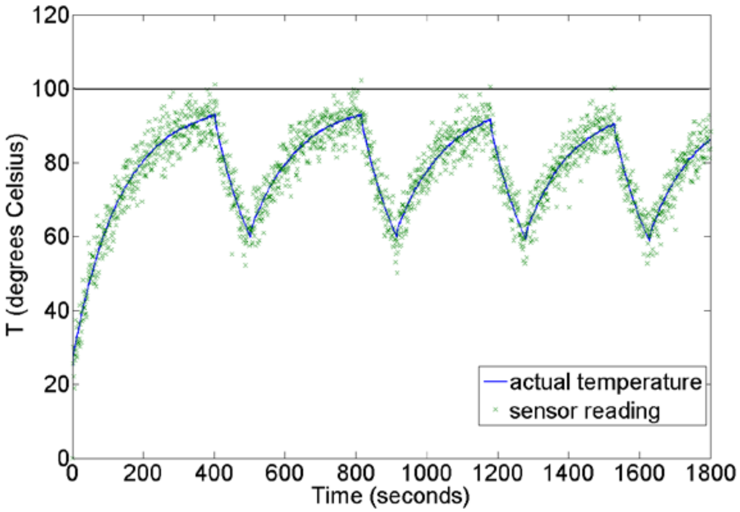
{yufuzh, bingshi, ankurs}@umd.edu

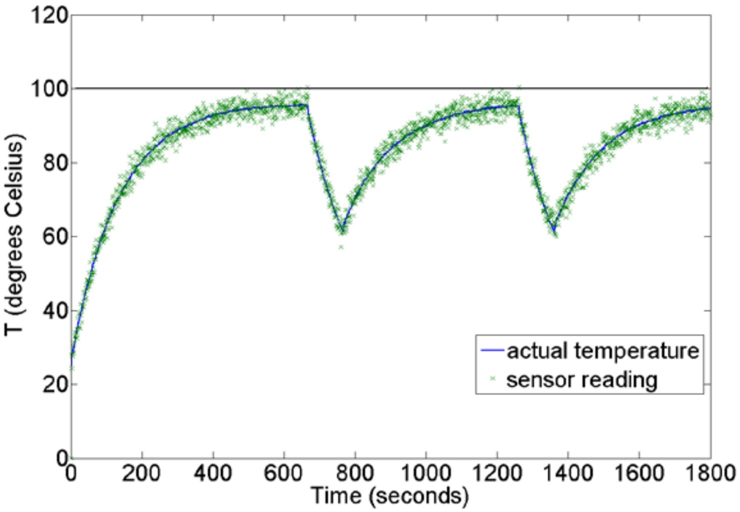
Outline

- Motivation/overview
- Fusion center design
- Sensor design/compression
  - Noisy sensor behavior
  - Exploiting the correlation
- Sensor placement
- Overall flow and interplay
- Results and conclusion

Motivation

- Thermal/power stress
  - Heavy task execution
  - Increasing chip density
  - Leakage power
- Dynamic thermal management (DTM)
  - Essentially sacrificing performance for lower temperature
  - Need accurate runtime thermal information

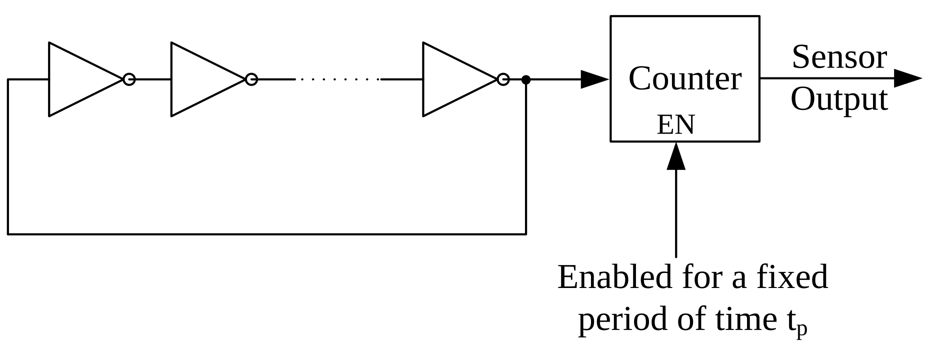




Motivation

- Need sensors to provide accurate runtime thermal input
- On-chip thermal sensors
  - On-chip sensors can sample the thermal state of the chip during runtime

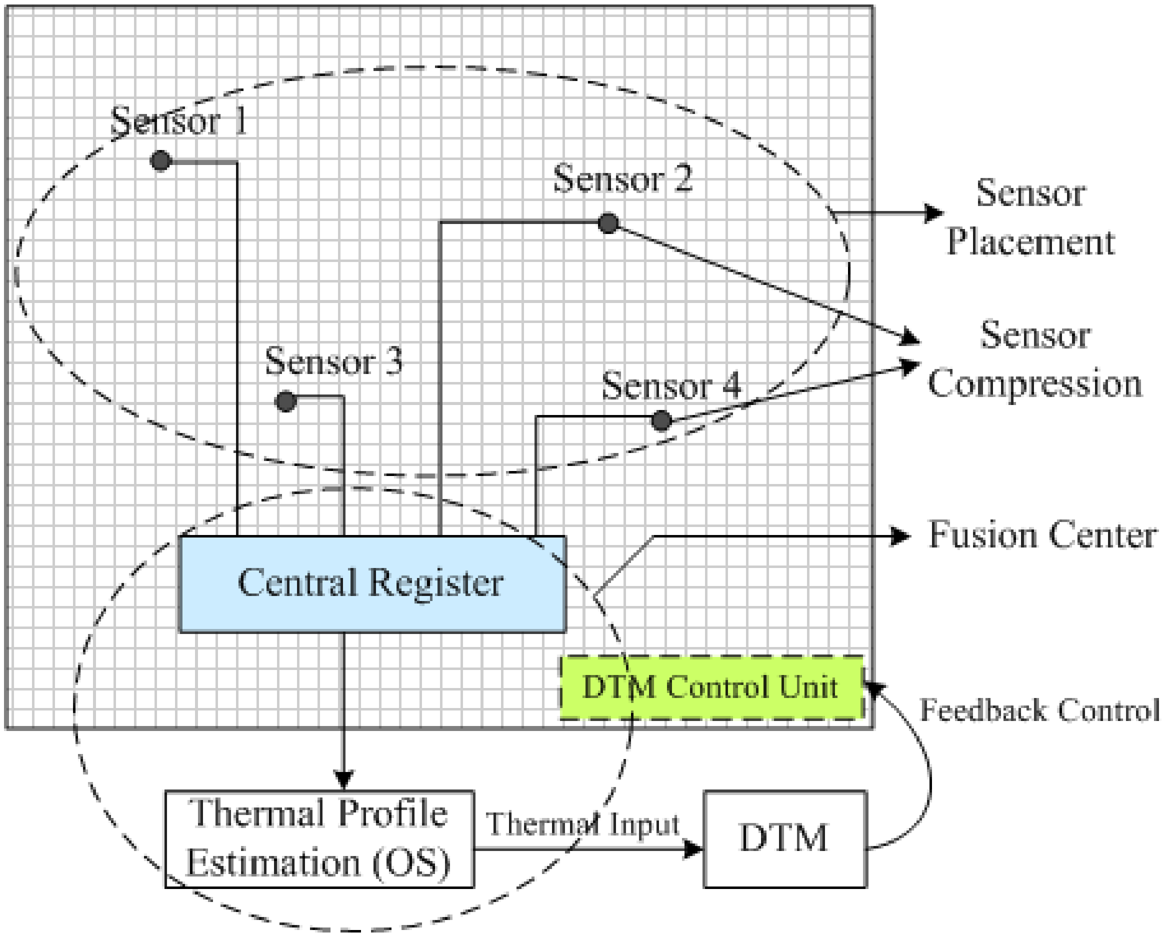
A simple ring oscillator-based thermal sensor



Motivation

- Several problems for a naïve thermal sensing scheme.
  - Sensors cannot go everywhere
  - Sensors are subject to noise
  - Resource is limited
- Our goal --- a complete thermal sensing infrastructure that includes:
  - Sensor design/compression
  - Sensor placement
  - Data fusion

Overall structure

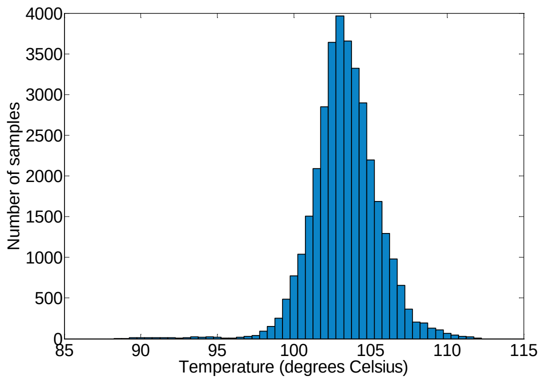


Fusion Center Design

- Central register (finite size ***M***)
  - Could be a single or multiple actual registers
- Fusion algorithm
  - Model the thermal profile as a random vector ***T***
  - Predict (***T***) given the sensor obs vector (***T***$_S$)
  - Exploit statistical information (mean, var, correlation etc.)
- Bayesian Estimation Philosophy

Scalar case: $$E(T \mid T_S) = \mu_T + \frac{\rho_{TS}\sigma_T}{\sigma_S}(T_S - \mu_S)$$

Vector case: $$E(\vec{T} \mid \vec{T}_S) = \vec{\mu}_T + \Sigma_{TS}\Sigma_{SS}^{-1}(\vec{T}_S - \vec{\mu}_S)$$



Fusion Center Design

- Given sensor input, the variance of ***T*** is reduced to:

$$\hat{\Sigma}_{TT} = E\left((T - E(T \mid T_s)) \cdot (T - E(T \mid T_s))^T \right)$$
$$= \Sigma_{TT} - \Sigma_{TS} \Sigma_{SS}^{-1} \Sigma_{ST}$$

  - Diagonal elements – variance of the thermal estimates.
  - Reflects the fundamental uncertainty of our estimation. (how far away our estimates are from the real temperature)
  - Used to drive sensor placement.
- A better metric to drive sensor placement?
  - Sensors are not like cameras
  - Generate the probability of capturing all hotspots

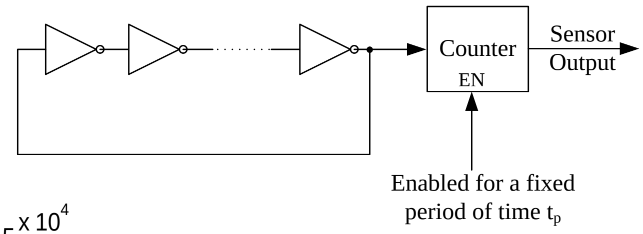
Sensor Design

- Noisy sensor behavior (Monte Carlo Simulation)

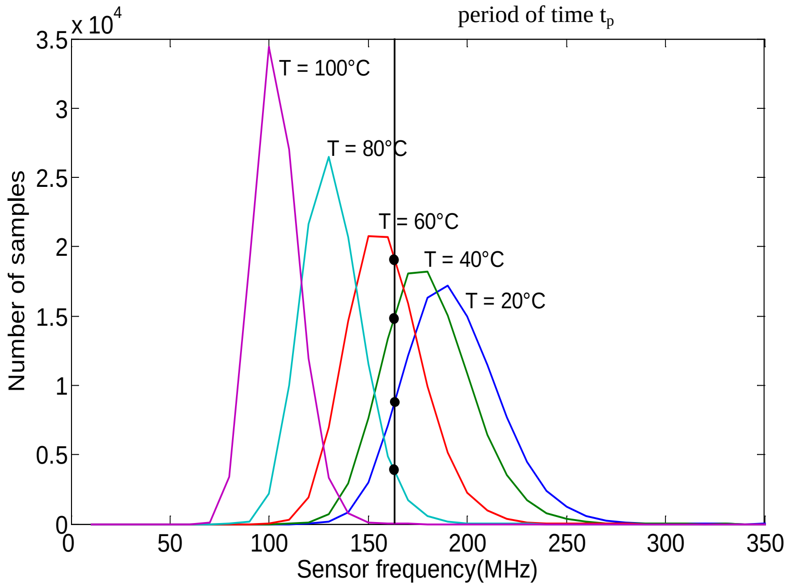
$$f = \frac{1}{P} = \frac{1}{N(t_{PHL} + t_{PLH})} \qquad t_{PHL} = \frac{2C}{\mu_n C_{ox}(W/L)_n (V_{DD} - V_t)}\left[\frac{V_t}{V_{DD} - V_t} + \frac{1}{2}\ln\left(\frac{3V_{DD} - 4V_t}{V_{DD}}\right)\right]$$

$$\mu_{n/p} = \mu_0 (T/T_0)^{-1.5}$$

$$V_t = V_{t0} + 0.002(T - T_0)$$



- Sensor readings are compressed as well due to center register size constraint
- Hypothesis testing



Sensor Design

- Target: minimize the expected prediction error:

$$Cost = E(|T_{pred} - T_{real}| \,\big|\, T_{obs})$$

$$= \sum_{i=1}^{n} |T_{pred} - H_i| \cdot \underline{prob(T_{real} = H_i | T_{obs})}$$

Bayes rule

$$= \frac{prob(T_{obs} | T_{real} = H_i) \cdot P_i}{prob(T_{obs})}$$

$$= \frac{prob(T_{obs} | T_{real} = H_i) \cdot P_i}{\sum_{j=1}^{n} prob(T_{obs} | T_{real} = H_j) \cdot P_j}$$

- Optimal decision rule:

$$T_{pred} = \delta(T_{obs}) = \arg\min_{T_{pred} = H_1 \ldots H_n} E(|T_{pred} - T_{real}| \,\big|\, T_{obs})$$

- Implement as an encoder at the sensor output

Sensor and fusion center co-design

- How do we compress sensors so that…
  - They fit into the central register
  - Collectively they provide better accuracy

    ---- more compressed sensors vs fewer non-compressed ones
- Bit allocation problem:
  - Decide how to allocate a total of ***M*** bits to ***n*** sensors so that the overall expected estimation error is minimum:

    (suppose s_i is the number of bits allocated to sensor ***i***)

Minimize $$E(error(s_1, s_2, ..., s_n))$$

Subject to $$\begin{cases} 0 \le s_i \le b_i \\ \sum_i s_i = M \end{cases}$$

Sensor Compression

- Target: to reduce the overall expected error caused by sensor compression.

$$TotalCost = E\left(error(s_1, s_2, ..., s_n)\right)$$

$$E(T_i \mid \vec{T}_S) = \mu_{T_i} + \Sigma_{TS}\Sigma_{SS}^{-1}(\vec{T}_S - \vec{\mu}_S)$$

$$= E\left(\sum_{\forall grids:i} \left| E(T_i \mid \vec{T}_s^c) - E(T_i \mid \vec{T}_s^a) \right|\right)$$

$$= E\left(\sum_{\forall rows} \left| \Sigma_{TS}\Sigma_{SS}^{-1}(\vec{T}_s^c - \vec{T}_s^a) \right|\right)$$

- Different compression scheme leads to different overall error.
- Can be formulated as a optimization problem (see details in our paper).

Sensor Placement

- Let ***"S"*** and ***"T"*** represent the set of sensor locations and all chip locations, respectively.
- Problem formulation:

$$choose \;\; S \subset T \;\; with \;\; |S| = n$$

$$such \; that \; trace(\hat{\Sigma}_{TT}) \; is \; minimized$$

- As mentioned earlier $\hat{\Sigma}_{TT}$ represents the fundamental uncertainty/variance associated with our thermal estimates

$$\hat{\Sigma}_{TT} = \Sigma_{TT} - \Sigma_{TS}\Sigma_{SS}^{-1}\Sigma_{ST}$$

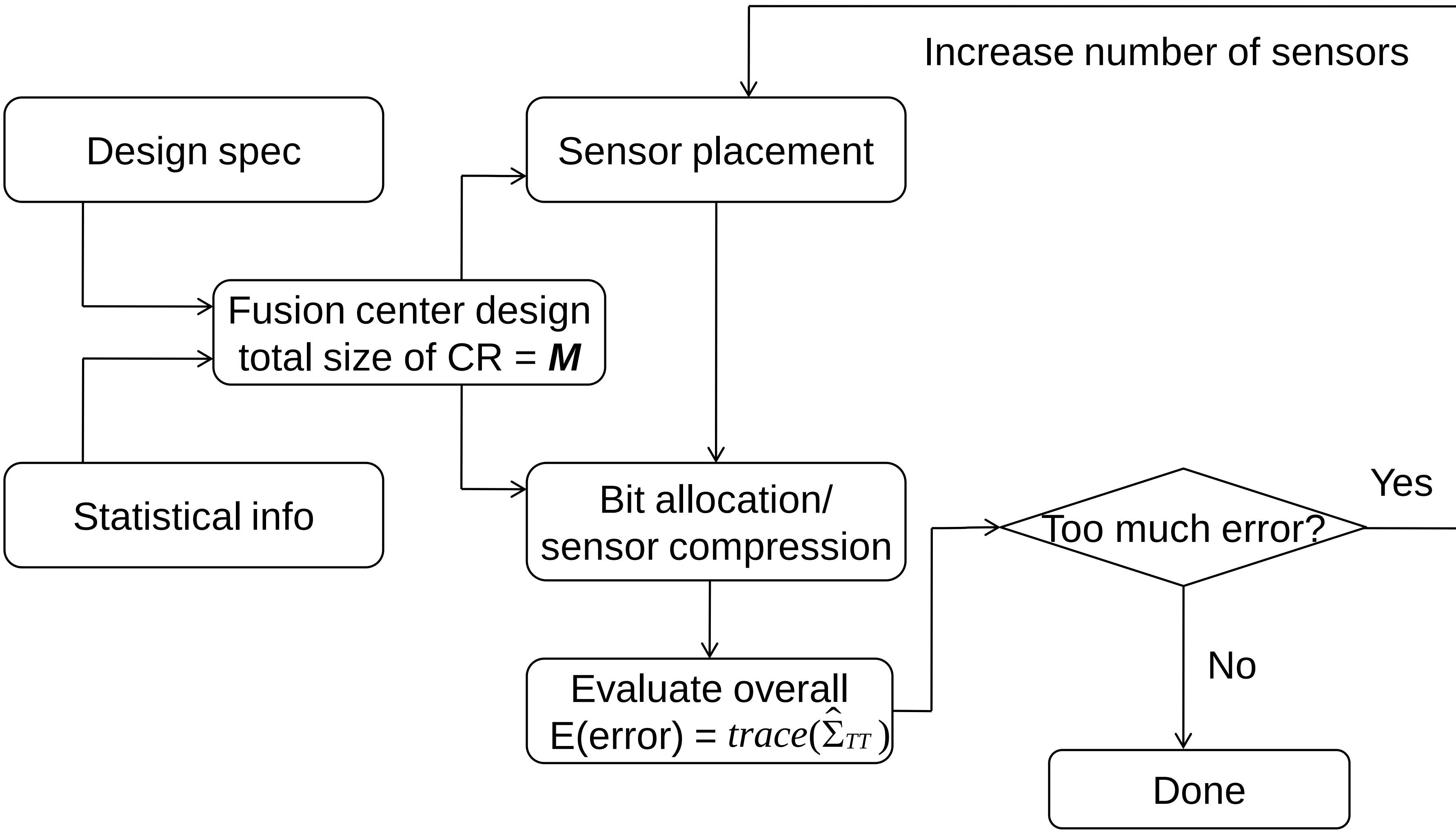
Sensor Placement Algorithm

Algorithm <Sensor Placement>

Input: Desired number of sensors n, all grid location set T
Output: Sensor location set S with $|S| = n$

$S \leftarrow \emptyset$
while $|S| < n$ **do**
  for all $i \in T \setminus S$ **do**
    $S_{new} \leftarrow S \cup i$
    Calculate the new $\hat{\Sigma}_{TT}$ using S_{new}
  end for
  Select i_{min} which results in the minimum trace$(\hat{\Sigma}_{TT})$ among all i
  $S \leftarrow S \cup i_{min}$
end while

Overall flow and interplay



Experimental Results

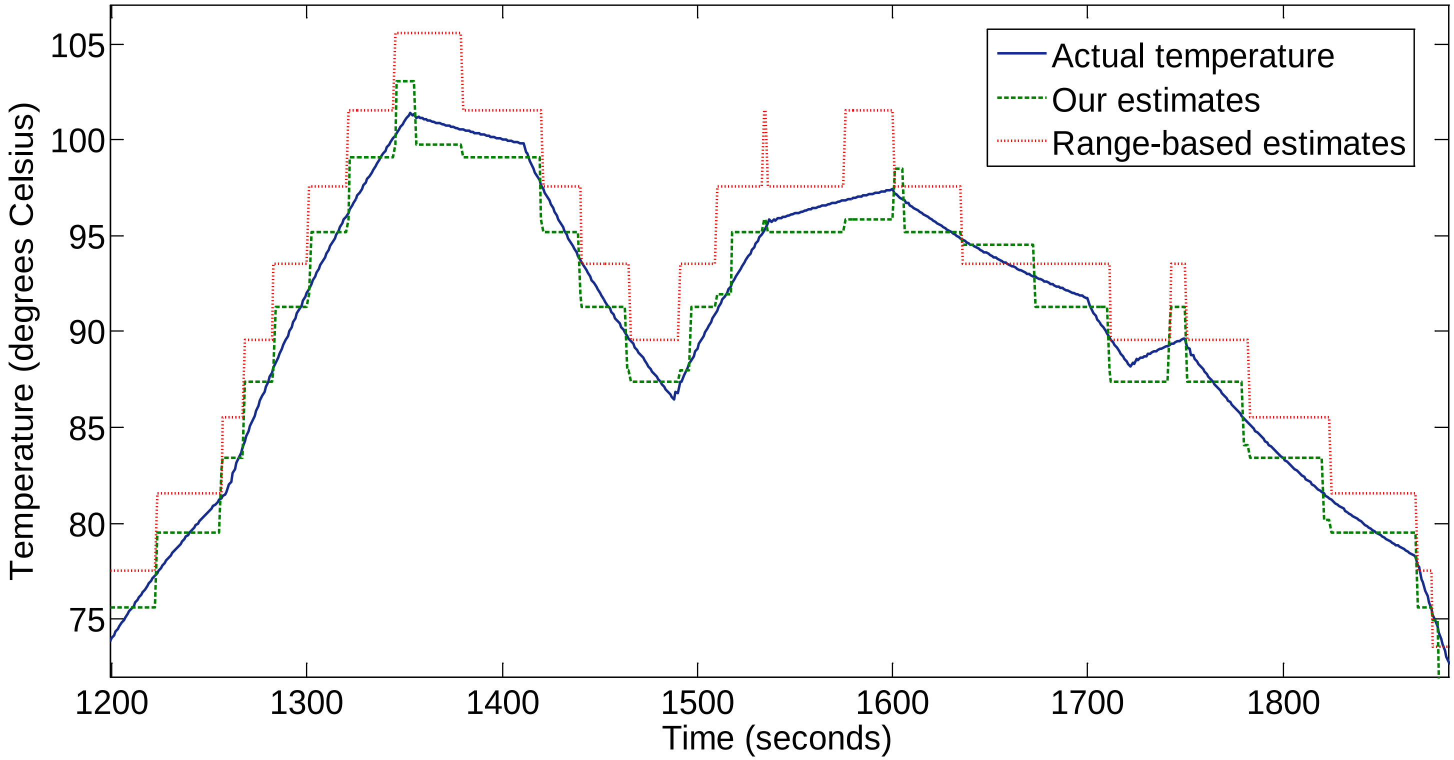


Fig. 1 Dynamic temperature tracking curves

Experimental Results

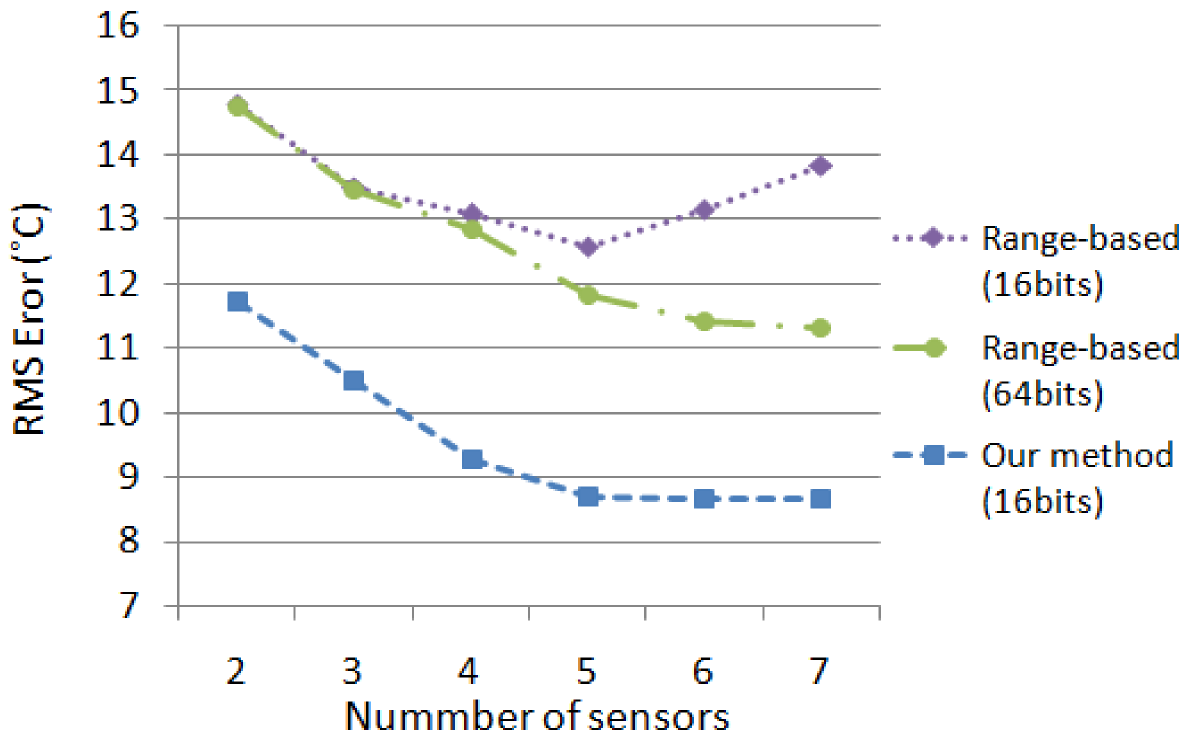


Fig. 2 RMS error comparison when increasing the number of sensors

Conclusion

- We presented a unified statistical framework for designing a complete thermal sensing infrastructure.
- Significant improvement in thermal sensing accuracy can be achieved with very small overhead
- Our methodology has the capability of trading off complexity for accuracy at will. It also takes into account various design considerations such as sensor noise and area constraints.

Thank you!