Center for Embedded Computer Systems
University of California, Irvine

Strength of Diversity: Exploiting Cheap Heterogeneous Noisy Sensors for Accurate Full-Chip Thermal Estimation and Prediction

Santanu Sarma, N. Dutt and P. Gupta†

Center for Embedded Computer Systems
University of California, Irvine
Irvine, CA 92697-2620, USA

†Department of Electrical Engineering
University of California Los Angeles
{santanus, dutt}@ics.uci.edu, †puneet@ee.ucla.edu

CECS Technical Report No: CECS TR-14-01

Jan 15, 2014

Strength of Diversity: Exploiting Cheap Heterogeneous Noisy Sensors for Accurate Full-Chip Thermal Estimation and Prediction

S.Sarma, N.Dutt, and P.Gupta†

***Abstract*—Thermal sensor characteristics and placement directly impacts the effectiveness and accuracy of full-chip thermal characterization necessary for dynamic thermal management (DTM) and reliable on-chip operation of multi/many-core chips. Temperature sensor characteristics widely vary in their area, power, and accuracy; the number of deployable sensors is constrained by the on-chip area/power constraints. However, recent approaches have considered only one type of sensor without leveraging the diversities among different sensor types. In this paper, we exploit the flexibility and trade-off in area/power and error characteristics of varied thermal sensors to perform a heterogeneous sensor allocation and placement (HSAP) to precisely recover the full-chip thermal map with high-fidelity. Unlike traditional sensor allocation and placement techniques that consider only one sensor type, HSAP finds the best combination of the heterogeneous sensors along with their placement for a given sensor area and power budget such that the full-chip thermal characterization error is minimized. Experimental results with multicore Alpha processor show significant improvements compared to the state-of-the-art in terms of reconstruction accuracy for the same sensor area and power budget. In particular, our HSAP approach achieves superior accuracy (around 10-100x error reduction with three types of sensors in comparison to a single type without any additional overhead) and execution speedup of over $20\times$ for full-chip thermal monitoring over a state-of-the-art technique.**

***Index Terms*—Thermal variation characterization, Variability, Thermal Sensors, Sensor allocation, Heterogeneous sensor placement.**

I. Introduction

As integrated circuit technology continues to scale to the nanoscale era, power and thermal issues become increasingly important and major concern for processor design [1], [2], [3]. Earlier works have shown that elevated temperatures directly impact all key circuit metrics including: lifetime and reliability, speed, power, and costs. Thermal hotspots reduce the mean time to failure as most failure mechanisms (e.g., electromigration, time dependent dielectric breakdown, and negative bias temperature instability) have strong temperature dependencies [4]. With more than 50% of all integrated circuit failures being accounted to thermal issues [4], a mere 10^oC 15^oC rise in the operating temperature could halve the life span of the chip [5]. Not only does the fault rate double for every 10^oC increase in temperature [6], but different thermal expansion coefficients of chip materials also cause mechanical stresses that can eventually crack the chip/package interface [3], resulting in increased packaging cost. Additionally, accuracy of thermal measurements directly affects the efficiency of thermal management as well as the performance of the CPU [7], [8]. Inaccuracies in thermal tracking decreases the processors performance and wastes power. In particular, it was shown that a 1 ^oC accuracy translates to 2W power savings while 1.5^{o}C accuracy in temperature measurement is equivalent to 1W of CPU power in mobile computers [7]. Due to localized heating, the temperature variation within a single chip can reach up to 10s of degrees. For instance, within-die temperature variation of up to 50 ^oC was reported in [1]. Due to lack of proximity, sensor measurements and hot spot temperatures could differ by up to 10 ^0C [7]. Such inaccuracies in thermal estimates can either trigger early or late activation of DTM resulting in unwanted performance loss [8] or severe reliability degradations [4].

S.Sarma and N.Dutt are with the Department of Computer Science, ICS, University of California, Irvine, USA, e-mail: {santsnu, dutt}@ics.uci.edu.

†P.Gupta is with Department of Electrical Engineering, UCLA, Los Angeles, email: puneet@ee.ucla.edu

As the effectiveness and efficiency of dynamic thermal management approaches heavily relies on the accuracy of on-chip temperature measurement and estimation, several types of thermal sensors varying in accuracy, resolution, area, and power consumption have been designed and proposed. The specifications and relative area and power characteristics of some of the recent smart sensors are listed in Table I and Fig. (1) respectively. For an area limited design, sensor S1 would be preferred. On the other hand, for a power limited design, sensor S3 would be preferred over S1 because of its lowest power characteristics. For design spaces between the two extremes of area or power limited design, neither S1 nor S3 but a combination of sensors (e.g. S1, S2, S3) would potentially provide a better solution by leveraging the diversity of each sensor type. Additionally, in order to accurately capture the wide within-die temperature variations, a large number of them must be deployed throughout the chip to collect thermal data in real time. For example, Intel's Dunnington Xeon processor with 6 cores contains 12 thermal sensors [14], AMD's quad-core opteron deploys 38 sensors [15], and IBM's POWER7 microprocessors includes 40 temperature sensors [16] at different locations.

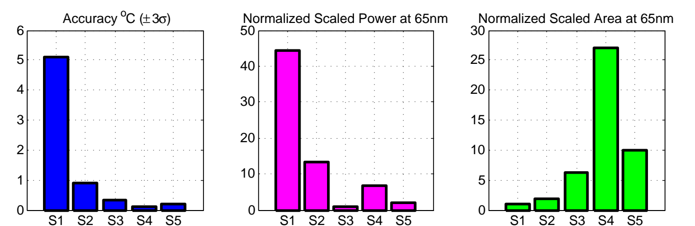


Figure 1. Sensor accuracy, normalized power and area of the digital temperature sensors of Table (I) scaled to 1.0V, 65 nm technology.

As temperature sensors along with their peripheral circuits introduce substantial overhead in silicon area and power consumption, it is extremely important to minimize the design overhead associated with temperature sensors and their placement–without surrendering the accuracy of thermal monitoring. On one-hand large number of temperature sensors are needed for accurate thermal monitoring, on

Table I
COMPARISON OF RECENT SMART TEMPERATURE SENSORS

Sensor	Resolution(^0C)	Accuracy(^0C)	Range(^0C)	Power	Area(mm^2)	Technology	Reference
S1	0.139	-5.1~+3.4	0~60	$150\mu W$@1.0V	0.01	$65nm$	Chung et al.[9]
S2	0.16	-0.7~+0.9	0~100	0.49mW@3.3V	0.175	$0.35\mu m$	Chen et al. [10]
S3	0.0918	-0.25~+0.35	0~90	$36.7\mu W$@3.3V	0.6	$0.35\mu m$	Chen el al. [11]
S4	0.01	±0.1	-55~125	$247\mu W$@3.3V	4.5	$0.7\mu m$	Pertijs et al. [12]
S5	0.03	±0.2	-70~125	$10\mu W$@1.2V	0.1	65 nm	Sebastiano et al. [13]

the other hand they incur substantial die area real estate and power. To effectively address this inherent trade-off in thermal monitoring, we propose heterogeneous sensor allocation and placement (HSAP) using heterogeneous mix of sensors. In this approach, we exploit the flexibility and trade-off in area, power, and accuracy characteristics of varied thermal sensors to perform a improved sensor allocation and placement while precisely recovering the full-chip thermal map for multicore architectures. Unlike state-of-the-art sensor allocation and placement techniques that use a single type of homogeneous sensor, our proposed HSAP approach finds the best combination or the mix of different sensors for given sensor area and power budget along with their placement such that the full-chip thermal characterization error is minimized. Contrary to the existing techniques that minimizes the number of sensors to reduce the area and power overhead, our proposed approach maximizes the number of samples used in the reconstruction without increasing the sensing overhead (considering the sensor area and power consumption). The contributions of our work are that we:

- redefine the sensor allocation and placement problem using combination of heterogeneous sensors,
- present an algorithm to find the best sensor combination and their location without either surrendering the accuracy of thermal monitoring or exceeding the sensor area power budget,
- show significant improvements in reconstruction error (around 10-100x reduction with three types of sensors) for same overheads and execution speedup of over $20\times$ in comparison to a state-of-the-art technique.

II. MOTIVATION AND RELATED WORK

Full chip thermal monitoring and reconstruction suffers from an inherent trade-off between accuracy of monitoring and implementation overhead (in terms of sensor die area and power) incurred in the monitoring infrastructure. Both the accuracy of the reconstruction and the overhead incurred in the monitoring infrastructure is directly impacted by the number of sensors, their locations, and accuracy [17]. Specifically, full chip thermal reconstruction accuracy predominantly depends on the number on-chip sensors [18], [17], their placement [18], [17], as well as sensor accuracy and error characteristics [8]. Although, improvement in reconstruction accuracy was studied by several computational means like interpolation [19], [20], [21], correlation of the sensor measurements, and transformed domain processing (e.g. FFT [22], DCT [18]), a tacit assumption in all these works [23], [19], [20], [18], [22], [17] is that sensors characteristics are homogeneous which is no longer valid for nanometer technologies in presence of variability [8], [1]. These works [23], [19], [20], [18], [22], [17] have neither considered the sensor error characteristics nor did they exploit the trade-offs in sensor characteristics to improve reconstruction accuracy. In this paper, we exploit these trade-off characteristics among different sensor type with regard to the reconstruction accuracy and show that by using mix of many weak and inaccurate sensors along with few robust and accurate ones, the full-chip reconstruction accuracy can be improved by an order of magnitude.

Our work is motivated by the recent work of Ranieri et al. [17] which addresses the problem of sensor allocation and placement for a given set of thermal traces. Ranieri et al. [17] considers the effect of the sensor accuracy in the thermal profile reconstruction but does not specifically account for the sensor area and power overheads. While [17] improves the reconstruction accuracy by exploiting the correlation among the thermal maps, our work jointly considers the sensor accuracy, area, and power overhead and exploits the trade-off in different sensor types along with the correlation information among the thermal maps. Our work differs from [17] in several ways. First, our approach is more generic and can be applied for both homogenous and heterogeneous sensor types where as the approach in [17] is only applicable for homogeneous sensors. As the thermal sensor accuracy can substantially vary due to process variations at different location of the die [8], the methods in [17] fails to account these effects. HSAP can easily be applied in such scenarios. Second, the reconstruction approach in [17] is based on ordinary least square (OLS), whereas our reconstruction approach is based on generalized least square (GLS) which is more accurate and efficient for general and heteroscedastic noise distribution. Third, our greedy sensor placement algorithm is relatively faster than that of [17] as we directly find the best M locations one after the other in stead of rejecting $N-M$ (where $M \ll N$) futile locations. Fourth, we provide efficient solutions that can provide improvement of the order of $10\times -100\times$ in reconstruction accuracy for the same sensing area and power overhead and execution speedup of over $20\times$.

III. THERMAL PROFILE RECONSTRUCTION AND NOISY SENSOR PLACEMENT

Let $t[i,j]$ be the two-dimensional thermal map where i and j represents the coordinate of the location within the two-dimensional floor plan. The two-dimensional thermal map $t[i,j]$ can be represented using a one-dimensional vector $x[k]$ where $1 \le k \le N$ by stacking the columns such that $x[k] = t\left[k\, mod\, H, floor\left[\frac{k}{W}\right]\right]$. Thus, $\mathbf{x} = \{x_1, x_2, .., x_N\}$ represents the full temperature field at each grid locations $N = WH$ of the die. Let $\hat{\mathbf{x}}$ be the reconstructed signal of $\mathbf{x}$ in presence of noise and truncation errors and $\mathbf{x_S}$ be the measurement with total sensors $M \ll N$ at locations $\mathbf{L} = \{i_1, i_2, ..., i_M\}$ such that $\mathbf{x_S}$ =$\mathbf{x}(\mathbf{L})$. In presence of sensor noise and errors $\boldsymbol{\epsilon}$, the sensor measurement is:

$$\mathbf{x_S} = \mathbf{x}(\mathbf{L}) + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{\epsilon}$ is suitably represented by a probability distribution such as multivariate Gaussian distribution ($\boldsymbol{\epsilon} \sim \mathscr{N}_M(\boldsymbol{\mu},\boldsymbol{\Sigma})$) with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$. Let $\widehat{\mathbf{x}}_\mathbf{S}$ be an estimate of $\mathbf{x_S}$ which is the measurement from R heterogeneous sensor types with accuracy σ_r, power p_r, area a_r for sensor type r. Let $\mathbf{m} = \{m_1, m_2, ..., m_R\}$ represent combination of the heterogenous sensors, i.e. m_r is the number of sensors of type r that is used in the measurement at locations $\mathbf{L_r}$ (representing the location of the sensors of type r) and $\mathbf{x_{S_r}}$ represent the measurement from sensor of type r located at $\mathbf{L_r}$ such that :

$$\mathbf{x_{S_r}} = \mathbf{x}(\mathbf{L_r}) + \boldsymbol{\epsilon_r}$$

where $\boldsymbol{\epsilon_r} \sim \mathcal{N}_{\mathbf{m_r}}(\boldsymbol{\mu_r}, \boldsymbol{\Sigma}_r)$ is Gaussian noise of dimension m_r, mean $\boldsymbol{\mu_r}$, and variance $\boldsymbol{\Sigma_r}$. If we represent the covariance of the sensor error $\boldsymbol{\epsilon_s}$ as $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$, where $\mathbf{V}$ is a matrix and σ is a positive bounding constant corresponding to max value in $\boldsymbol{\Sigma}$, there are two special cases (a) $\mathbf{V} = \mathbf{I}$ (b) $\mathbf{V} = diag(V_1, V_2, ..., V_R)$. In the first case, where the variance are same for each component (sensor), is called **homoscedasticity** and in the second where variance of each components are **not** same is called **heteroscedasticity.** As we have different types of sensors with varied error characteristics,

the sensor error ϵ_s has heteroscedastic distribution. Thus, combined sensor measurements is heteroscedastic and is given by:

$$\mathbf{x_S} = \mathbf{x_{S_1}} \bigcup \mathbf{x_{S_2}} ... \bigcup \mathbf{x_{S_R}} = \left\{ \bigcup_{r=1}^{R} \mathbf{x_{S_r}} \right\} \quad (2)$$

in presence of the sensor noise $\epsilon_s = \bigcup_{\mathbf{r=1}}^{\mathbf{R}} \epsilon_r$ at the locations :

$$\mathbf{L} = \mathbf{L_1} \bigcup \mathbf{L_2} .... \bigcup \mathbf{L_R} = \left\{ \bigcup_{\mathbf{r=1}}^{\mathbf{R}} \mathbf{L_r} \right\} \mid \mathbf{L_i} \bigcap \mathbf{L_j} = \emptyset, \forall \mathbf{i} \neq \mathbf{j}. \quad (3)$$

As the number of sensor of a particular type is equal to the size of the location vectors, i.e. $m_r = |\mathbf{L_r}|$, the total number of sensors is $M = \sum_{r=1}^{R} m_r = |\mathbf{L}|$. Similarly, the total sensor area and power consumed by the sensors is given by:

$$\begin{aligned} \textstyle\sum_{r=1}^{R} a_r * m_r \leq AB \\ \textstyle\sum_{r=1}^{R} p_r * m_r \leq PB \end{aligned} \quad (4)$$

where AB and PB are the sensor area and power budget respectively.

IV. Signal Estimation and Recovery

The sensor measurement signal in (1) can be represented using a generalized linear regression model:

$$\mathbf{x_S} = \mathbf{\Phi}_s \boldsymbol{\alpha}_s + \boldsymbol{\epsilon}_s \quad (5)$$

where $\mathbf{\Phi}_s$ is the basis or kernel matrix, $\boldsymbol{\alpha}_S$ are the coefficients of the expansion over the basis $\mathbf{\Phi}_s$. With the assumption of zero mean and deterministic basis, following theorems are well known results in linear regression analysis and estimation theory [24].

Theorem 1. *Let* $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ *be the linear regression model with zero mean homoscedastic noise distribution* ($E(\boldsymbol{\epsilon}) = \mathbf{0}$, $Cov(\boldsymbol{\epsilon}) = \boldsymbol{\sigma^2 I}$) *and deterministic basis* $\mathbf{\Phi}$ *such that* $rank(\mathbf{\Phi}^{n \times p}) = p$ *then* $rank(\mathbf{\Phi}'\mathbf{\Phi}) = p$ *and* $\left(\mathbf{\Phi}'\mathbf{\Phi}\right)^{-1}$ *exists. In that case, the ordinary least square (OLS) solution is unique and is given by:*

$$\widehat{\boldsymbol{\alpha}}_{OLS} = \mathbf{\Phi}^{\dagger}\mathbf{x} = (\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{x} \quad (6)$$

with the orthogonal projection as $\widehat{\mathbf{x}} = \mathbf{\Phi}\widehat{\boldsymbol{\alpha}}_{OLS} = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{x} = \mathbf{P}\mathbf{x}$ *where* $\mathbf{P} = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'$ *is the projection matrix and* $\mathbf{P}\mathbf{\Phi} = \mathbf{\Phi}$. *The OLS estimate has the following properties:*

- $E(\widehat{\boldsymbol{\alpha}}_{OLS}) = \boldsymbol{\alpha}$,
- $Cov(\widehat{\boldsymbol{\alpha}}_{OLS}) = \sigma^2(\mathbf{\Phi}'\mathbf{\Phi})^{-1}$,
- $\widehat{\boldsymbol{\alpha}}_{OLS}$ is the best linear unbiased estimator (BLUE), i.e. the minimum variance unbiased estimator.

Theorem 2. *Let* $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ *be the generalized linear regression model where* $rank(\mathbf{\Phi}^{n \times p}) = p$, *with zero mean heteroscedastic noise distribution such that* $E(\boldsymbol{\epsilon})$=0,$Cov(\boldsymbol{\epsilon}) = \mathbf{\Sigma} = \sigma^2\mathbf{V}$, *with known* $\mathbf{V}$, *then there exists a transformation of* $\mathbf{x}$ *to a new response vector which has a covariance matrix* $\sigma^2\mathbf{I}$. *Ordinarily least square applied to the transformed* $\mathbf{x}$ *yields a solution:*

$$\widehat{\boldsymbol{\alpha}}_{GLS} = (\mathbf{\Phi}'\mathbf{V}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{V}^{-1}\mathbf{x} \quad (7)$$

with the orthogonal projection as $\widehat{\mathbf{x}} = \mathbf{\Phi}\widehat{\boldsymbol{\alpha}}_{GLS} = (\mathbf{\Phi}'\mathbf{V}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{V}^{-1}\mathbf{x} = \widetilde{\mathbf{P}}\mathbf{x}$ *where* $\widetilde{\mathbf{P}} = (\mathbf{\Phi}'\mathbf{V}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{V}^{-1}$ *is the generalized projection matrix and* $\widetilde{\mathbf{P}}\mathbf{\Phi} = \mathbf{\Phi}$. *The generalized least square (GLS) estimate is the best linear unbiased estimator (BLUE) with the following properties:*

- $E(\widehat{\boldsymbol{\alpha}}_{GLS}) = \boldsymbol{\alpha}$,
- $E(\widehat{\boldsymbol{\alpha}}_{GLS}) = \boldsymbol{E}(\widehat{\boldsymbol{\alpha}}_{OLS})$,
- $Cov(\widehat{\boldsymbol{\alpha}}_{GLS}) = \sigma^2(\mathbf{\Phi}'\mathbf{V}^{-1}\mathbf{\Phi})^{-1}$,
- $Cov(\widehat{\boldsymbol{\alpha}}_{OLS}) = \sigma^2(\mathbf{\Phi}'\mathbf{\Phi})(\mathbf{\Phi}'\mathbf{V}\mathbf{\Phi})(\mathbf{\Phi}'\mathbf{\Phi})^{-1}$.

From theorems (1) and (2) it can be inferred that OLS is the best linear unbiased estimator (BLUE) in presence of homoscedasticity and GLS is the BLUE in presence of heteroscedesticity**.** Under heteroscedesticity, OLS is still unbiased linear estimator, but not the best estimator (i.e. not efficient). In other words, OLS is not the minimum variance estimate in presence of heteroscedesticity. The usual variance of the OLS estimator is biased and thus inefficient. Consequently, we use the GLS estimate in (7) to find the coefficients from the sensor measurement as

$$\widehat{\boldsymbol{\alpha}}_s = (\mathbf{\Phi}_s'\mathbf{V}^{-1}\mathbf{\Phi}_s)^{-1}\mathbf{\Phi}_s'\mathbf{V}^{-1}\mathbf{x_S} \quad (8)$$

where we call $\mathbf{\Phi}_s$ as the sensing matrix.

A. Full Signal Reconstruction and Recovery in Presence of Noise

As the number of sensors are very few compared to the number of grids / monitoring points in the die ($M \ll N$), the spatial thermal profile has to be recovered from the few measurements $\mathbf{x_S}$. If we apply the generalized linear regression model for the full thermal map, we have $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha} + \boldsymbol{\varepsilon_p}$, where $\mathbf{\Phi}$ is a deterministic basis of size $N \times N$, and the coefficients $\boldsymbol{\alpha}$ are estimated as $\widehat{\boldsymbol{\alpha}}$ by GLS to get $\widehat{\mathbf{x}}$, an estimate of the thermal map due to the true process noise $\boldsymbol{\varepsilon_p}$. As thermal maps are often sparse, we can approximate the thermal map with a linear combination of K columns of $\mathbf{\Phi}$ and K elements of $\boldsymbol{\alpha}$ out of N such that $\widetilde{\mathbf{x}} = \mathbf{\Phi_K}\boldsymbol{\alpha_K}$. Let $\mathbf{K} = \{j_1, j_2, ..., j_K\}$ be the vector of the locations of the coefficients in $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha_K} = \boldsymbol{\alpha}(\mathbf{K})$. Note that this approximation makes $\mathbf{\Phi_K}$ of dimension $N \times K$ and $\boldsymbol{\alpha_K}$ of dimensions $K \times 1$. This approximation or truncation introduces an addition error or noise term $\boldsymbol{\varepsilon_t}$ in the regression models such that $\mathbf{x} = \mathbf{\Phi_K}\boldsymbol{\alpha_K} + \boldsymbol{\varepsilon_p} + \boldsymbol{\varepsilon_t}$. When the number of sensors M is equal to (or greater than) the number of basis vectors or columns K , $\boldsymbol{\alpha_K}$ can be represented by $\boldsymbol{\alpha_s}$ in addition to a sensor noise term $\boldsymbol{\varepsilon_s}$ in the thermal map such that $\mathbf{x} = \mathbf{\Phi_K}\boldsymbol{\alpha_s} + \boldsymbol{\varepsilon_p} + \boldsymbol{\varepsilon_t} + \boldsymbol{\varepsilon_s}$. Representing the total noise as sum of all the noise components as $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon_p} + \boldsymbol{\varepsilon_t} + \boldsymbol{\varepsilon_s}$, we can make an estimate of the thermal map as:

$$\widehat{\mathbf{x}} = \mathbf{\Phi_K}\widehat{\boldsymbol{\alpha}}_s = \mathbf{\Phi_K}\left[(\mathbf{\Phi}_s'\mathbf{V}^{-1}\mathbf{\Phi}_s)^{-1}\mathbf{\Phi}_s'\mathbf{V}^{-1}\right]\mathbf{x_S}. \quad (9)$$

Note that the sensing matrix $\mathbf{\Phi}_s$ is formed from the basis matrix $\mathbf{\Phi}$ corresponding to the sensor locations $\mathbf{L}$ and coefficient locations $\mathbf{K}$ such that $\mathbf{\Phi}_s = \mathbf{\Phi}(\boldsymbol{L}, \boldsymbol{K})$. The reconstruction matrix $\mathbf{\Phi_K}$ is formed from K columns corresponding to the coefficient locations such that $\mathbf{\Phi_K} = \mathbf{\Phi}(:, \boldsymbol{K})$ where the operator ':' represents all rows of $\mathbf{\Phi}$. For valid reconstruction in (9) , $\mathbf{V}^{-1}$ and $(\mathbf{\Phi}_s'\mathbf{V}^{-1}\mathbf{\Phi}_s)^{-1}$ must exist and results in $M \geq K$ as a requirement.

To evaluate the reconstruction / recovery accuracy over a set of T thermal traces $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, .., \mathbf{x_T}\}$ of size $T \times N$ with each row indicating a trace $\mathbf{x}$, we define the residual vector $\widehat{\boldsymbol{\varepsilon}}_i = \mathbf{x_i} - \widehat{\mathbf{x}}_i$ for the thermal trace $\mathbf{x_i}$ and the residual sum of square $RSS_i = \widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i = \|\mathbf{x_i} - \widehat{\mathbf{x}}_i\|^2$ such that the mean square error of the trace $\mathbf{x_i}$ is $MSE_i = E(\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i) = \frac{1}{N}\sum_{j=1}^{N}(x_i[j] - \widehat{x}_i[j])^2$. The total average mean square error over all the traces is :

$$MSE = \frac{1}{TN}\sum_{i=1}^{T}\sum_{j=1}^{N}(x_i[j] - \widehat{x}_i[j])^2 . \quad (10)$$

V. Noisy Heterogeneous Sensor Allocation And Placement (HSAP) Problem

A. Problem Statement

The HSAP problem is defined as an optimization problem which selects the best combination of heterogeneous sensors, their locations along with the basis vector combinations (basis selection) such that

the full chip thermal reconstruction error is minimum subject to the constraints discussed in section (III). The formal statement of the problem is as follows:

Given R heterogeneous sensor types with each sensor type r having power consumption p_r, and area a_r, accuracy σ_r, the problem is to select the combination of sensors $\mathbf{m} = \{m_1, m_2, ..., m_R\}$ (where m_r is the number of sensors of type r with the total number of sensor $M = \sum_{r=1}^{R} m_r$), their placement $\mathbf{L} = \{i_1, i_2, ..., i_M\}$, and the coefficient locations $\mathbf{K} = \{j_1, j_2, ..., j_K\}$ to select the basis vectors such that the full-chip reconstructed residual error is minimized subject to the constraints as in (11).

$$
\begin{aligned}
& \underset{\mathbf{L}, \mathbf{m}, \mathbf{K}, \widehat{\boldsymbol{\alpha}}_{\boldsymbol{s}}}{Minimize} && \| \mathbf{x} - \boldsymbol{\Phi}_{\boldsymbol{K}} \widehat{\boldsymbol{\alpha}}_{\boldsymbol{s}} \|_2^2 \\
& Subject\ to \quad : && \| \widehat{\boldsymbol{\alpha}}_{\boldsymbol{s}} \|_0 \leq K \leq M = \textstyle\sum_{r=1}^{R} m_r \\
& && \textstyle\sum_{r=1}^{R} a_r * m_r \leq AB \\
& && \textstyle\sum_{r=1}^{R} p_r * m_r \leq PB \\
& && \mathbf{L} = \bigcup_{r=1}^{R} L_r ; \mathbf{L_i} \bigcap \mathbf{L_j} = \emptyset, \forall i \neq j
\end{aligned}
\tag{11}
$$

VI. Greedy Solution for HSAP Problem

The HSAP problem is a generalization of the sparse regression problem [25] with additional constraints and variables. The sparse regression problem [25] is NP-hard and consequently the HSAP problem is NP-hard. With NP-hardness established, an optimal polynomial time algorithm is unreachable. Consequently, we propose heuristics and simplification strategies in developing a greedy solution for the HSAP problem.

A. Stage-wise Greedy Solution (gHSAP)

The stage-wise greedy solution approach is motivated by the work by Ranieri et al.[17] and Reda el al.[21]. gHSAP attempts to decouple the HSAP into heterogeneous sensor selection as the first stage, the basis vector and coefficient location selection as second, and the sensor placement and reconstruction as the third stage. By using the observation in section IV, we decouple the HSAP in to two optimization problems such that solution to both directly improves the reconstruction accuracy. By doing so, we can easily deploy effective existing solutions at much reduced complexity. We discuss all the stages and the optimization algorithms in the following subsections.

1) Heterogeneous Sensor Selection: It has been shown in [25] that the probability of reconstructing a signal exponentially increases with increasing number of samples. If the number of samples is chosen as $M_S \geq cKln\left(\frac{N}{K}\right)$, it is possible to reconstruct every K sparse coefficients with a probability exceeding $1 - e^{-NM_s}$. As the number of samples M_S exponentially impacts the reconstruction accuracy, it is logical to maximize them while choosing them from good sensors. However, as the number of samples M_s is directly related to the number of sensors M as well as their area, power, and accuracy trade-offs, we define an optimization to maximize the number of samples as a weighted sum of the sample obtained from a sensor of particular type within the area and power budget. The weights w_r can be chosen based on accuracy to give preference to or penalize one type over the other in selecting the samples. We cast the problem as an integer linear program (ILP) by defining a vector $\mathbf{m} = [m_1, m_2, ..., m_R]$ such that $M_S = \sum_{r=1}^{R} w_r m_r = \mathbf{w}^{'} \mathbf{m}$. The ILP thus is stated as follows:

$$
\begin{aligned}
& \underset{\mathbf{m}}{Maximize} \quad M_S = \textstyle\sum_{r=1}^{R} w_r m_r = \mathbf{w}^{'} \mathbf{m} \\
& subject\ to \textstyle\sum_{r=1}^{R} p_r * m_r \leq PB \\
& \qquad\qquad \textstyle\sum_{r=1}^{R} a_r * m_r \leq AB
\end{aligned}
\tag{12}
$$

The ILP in (12) can easily be solved using any standard solver. The solution of the ILP in (12) determines the sensor combinations and total sensor used in the sensor placement and reconstruction.

2) Basis Vector and Coefficient Location Selection: The purpose of this step is to find the K best basis vectors in $\boldsymbol{\Phi}$ and their locations $\mathbf{K} = \{j_1, j_2, ..., j_K\}$ to form the orthonormal basis matrix $\boldsymbol{\Phi_K}$ such that the approximation $\widetilde{\mathbf{x}} = \boldsymbol{\Phi_K} \boldsymbol{\alpha_K}$ is optimal. One approach is to use the greedy based orthogonal matching pursuit (OMP) [25] algorithm to find the best basis vectors. Another approach is based on the dimensionality reduction technique as in [17] which states that the approximation error in the reconstruction can be represented as the sum of the eigen values of $\boldsymbol{\Phi}$, corresponding eigen vectors of which are not included in $\boldsymbol{\Phi_K}$. If we define a covariance matrix $\mathbf{C_x}$ formed from the set of thermal traces $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_T}\}$ with eigen values $\{\lambda_n\}_{n=1}^{N}$, than the orthonormal basis $\boldsymbol{\Phi_K}$ that introduces the least error in the approximation is formed from the K eigen vectors of $\mathbf{C_x}$ with the largest eigen values $\{\lambda_n\}_{n=1}^{K}$. The minimum approximation error $\xi = \mathbf{E}\left[|\mathbf{x} - \widetilde{\mathbf{x}}|^{\mathbf{2}}\right]$ can be represented as [17]

$$
\xi = \sum_{n=K}^{N} \lambda_n. \tag{13}
$$

Note that as the number of basis vectors K increases, approximation error ξ decreases in (13). In other words, the reconstruction error can be reduced by forming the basis $\boldsymbol{\Phi_K}$ with the eigen vectors corresponding to the dominant eigen values of $\mathbf{C_x}$, as well as by increasing $K \leq M$.

3) Sensor Allocation and Placement : In this stage, sensor allocation and placement is carried out using total number of sensors computed in the previous stage. We allocate the best and most accurate sensors to the most crucial locations. The sensor placement algorithm provides the sensor location in order of their importance and we map the sensor types to these location according to their accuracy. The placement algorithm iteratively finds the sensor locations using a greedy approach [21] in polynomial time. The algorithm picks the location based on highest temperatures iteratively by computing the orthogonal components at available location and then picking the location with highest orthogonal norm. Once the locations are found, the thermal profile is reconstructed using (9). The algorithm is summarized in Fig. 2.

Algorithm: Greedy Heterogeneous Sensor Allocation and Placement (gHSAP)

Input: Thermal traces $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_T}\}$, Sensor types R, Sensor Specifications, $\mathbf{V}$, Sensor Area Budget AB, Sensor Power Budget PB
Output: Total no of sensors M, Sensor combination $\mathbf{m}$, Sensor location $\mathbf{L}$, Coefficient Location $\mathbf{K}$,Sensing matrix $\boldsymbol{\Phi_s}$,$\hat{\mathbf{x}}$

---

1) Solve the heterogeneous sensor selection ILP in (12) and compute M for the given sensor area power budget
2) Estimate $\mathbf{C_x} = \mathbf{E}\left[\mathbf{x}[\mathbf{i}]\mathbf{x}[\mathbf{j}]\right]$ from the set of thermal traces $\mathbf{X}$. $\mathbf{X}$ is normalized to reflect a zero mean process.
3) Compute the eigen values of $\mathbf{C_x}$ and the corresponding eigenvectors to form the basis $\boldsymbol{\Phi}$
4) Construct $\boldsymbol{\Phi_K}$ from $\boldsymbol{\Phi}$ corresponding to the largest eigen values given by coefficient location $\mathbf{K}$ such that $\boldsymbol{\Phi_K} = \boldsymbol{\Phi}(:, \mathbf{K})$
5) Greedy Sensor Allocation & Placement
   a) Let $\mathbf{L} = \emptyset$ and $\mathbf{S} = \{1, .., N\}$
   b) $s_1 = \arg\max_{s_1 \in S} \| \mathbf{X}_{\{\mathbf{s}\}} \|_\mathbf{2}$
   c) Let $\mathbf{L} = \mathbf{L} \cup \{s_1\}$ and $\mathbf{S} = \mathbf{S} - \{s_1\}$
   d) For $k = 2..M$ do
      i) Project $\mathbf{X_S}$ into the column space of $\mathbf{X_L}$: $\overline{\mathbf{P}} = \mathbf{X_L} \mathbf{X_L^\dagger} \mathbf{X_S}$
      ii) Find the orthogonal components $\mathbf{N} = \mathbf{X_S} - \overline{\mathbf{P}}$
      iii) Let $s_i = \arg\max_{s_i \in S} \| \mathbf{N}_{\{\mathbf{s}\}} \|_\mathbf{2}$
      iv) Let $\mathbf{L} = \mathbf{L} \cup \{s_i\}$ and $\mathbf{S} = \mathbf{S} - \{s_i\}$

      end
6) Construct the Sensing Matrix $\boldsymbol{\Phi_s} = \boldsymbol{\Phi}(\boldsymbol{L}, \boldsymbol{K})$ and sensor variance scaling matrix $\mathbf{V}$
7) Reconstruct the thermal map $\widehat{\mathbf{x}} = \boldsymbol{\Phi_K}\left[(\boldsymbol{\Phi}_{\boldsymbol{s}}^{'} \mathbf{V}^{-1} \boldsymbol{\Phi_s})^{-1} \boldsymbol{\Phi}_{\boldsymbol{s}}^{'} \mathbf{V}^{-1}\right] \mathbf{x_S}$

Figure 2. Heterogeneous Sensor Allocation and Placement for Full-chip Thermal Reconstruction.

VII. Experimental Setup

Simulation Setup: We evaluate the effectiveness of our methodology, by setting up a tool chain that simulates the temperatures for single and multi-core architectures of up to 4 cores at 65nm technology node as shown in Fig. 6. We utilize HotSpot [26] for thermal simulation and McPAT [27] to estimate the power for each block of the processor. We use the Alpha 21264 processor as our baseline core. The Alpha 21264 is an out-of-order speculative execution core that is commonly used as a test-bench core in thermal management research [26], [28]. HotSpot takes in the floor-plan of the processor and the workload that will run on each core to produce the steady state and transient temperatures at each location of the grid. Using workload instruction traces, dynamic power traces for each micro-architectural unit are calculated and then fed together with the floor-plan into the thermal simulator to compute both the transient and the steady-state temperature. We use total of 25 benchmarks from SPEC 2000 suites to randomly allocate these workloads to a core in the multi-core architecture. The combination of these workload assigned to different cores provide a rich set of thermal traces to characterize the thermal profile of the multi-core processor. We discretize the thermal profile by using a grid of $W = 64$ and $H = 64$ and used 25 benchmarks and their combinations to generate total of $T = 3194$ traces in our analysis.

VIII. Experimental Results

We present and discuss results for the set of thermal maps for single and multi-core architectures using the diverse set of heterogeneous temperature sensors as tabulated in Table I. Although there is no limitation, we only consider three types of sensors $\{S_1, S_2, S_3\}$ from Table I in the following example. Fig. 3 shows the effect of sensor heterogeneity on the total number of samples that can be collected for given sensor power and area overhead. The x-axis represents the product of the area and power. Recall that, sensor S_1 is used for an area limited design and sensor S_2 for a power limited design. For typical case of area and power budgets in between the two extremes of area or power limited design, neither S_1 nor S_2 is the optimal sensor type. In fact, to show the impact of choosing a particular sensor type, we plot the number of samples that can be collected for each type and their combination for a given overhead in Fig.3. As more number of samples directly improves the measurement and reconstruction accuracy, we observe from Fig.3 that sensor S_1 is better than S_3 and sensor S_2 is better than S_1 as well as S_3 for this area power design overhead. However, the combination of three sensors types $\{S_1, S_2, S_3\}$ accommodates more samples than any individual sensor type. Consequently, the heterogeneous combination of the sensors outperforms any individual sensor type in accommodating more sensors to improve the thermal monitoring process. Note that sensor S_2 would provide a better reconstruction compared to S_1 or S_3 if a single sensor were to be selected and hence HSAP can be used to select a suitable sensor type for given area power overhead.

The impact of increased number of samples and their combination for the given area-power overhead is reflected in the MSE and is shown in Fig. 3. A significant improvement in accuracy is observed for different combinations of the sensors for a given sensor area and power budget. The number of samples that can be collected for different design corners and trade-offs is shown in Fig. 4. As seen from Fig. 4 the heterogeneous mix of sensors $\{S_1, S_2, S_3\}$ can collect more samples by accommodate more number of sensors than the best area-efficient sensor type S_1 at all area power design corners. Fig. 5 shows the saving in overhead for a given accuracy requirement as well as improvement in accuracy for a given overhead in comparison with two state-of-the-art techniques, namely k-LSE[18] and EigenMaps[17]. As they do not specify the type of the sensor to be used, we used the best area efficient sensor S_1 for the given overhead for these two methods, and the sensor combination $\{S_1, S_2, S_3\}$ for HSAP. It is clear from Fig. 5 that for a specified reconstruction accuracy, HSAP can save die area or power for the sensors by appropriately choosing the right mix of sensors. As observed from Fig. 5, HSAP provides accuracy improvement of around 10-100x for same overheads compared to the k-LSE and and EigenMaps. This improvement in turn can provide better detection of hotspots and worst case temperature gradient. Besides, the total execution time for HSAP is 358.9 sec and that of EigenMaps is 8099.7 sec, an improvement of $22.56\times$, when implemented in Matlab running on Intel i7, 2.4GHz machine.

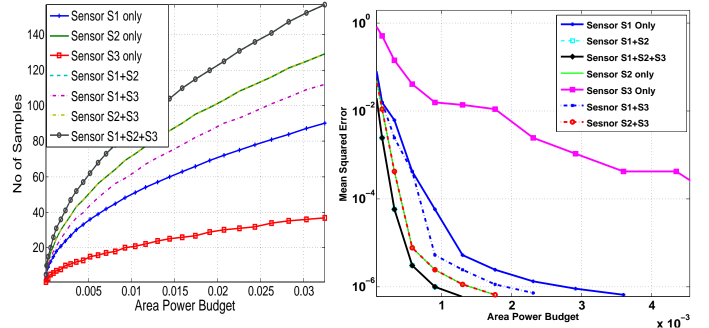


Figure 3. (a) Number of samples that can be used in reconstruction for given Area Power Budget for various sensor combinations (equal weights for all types) (b) MSE with different combination of sensors for given area power overhead.

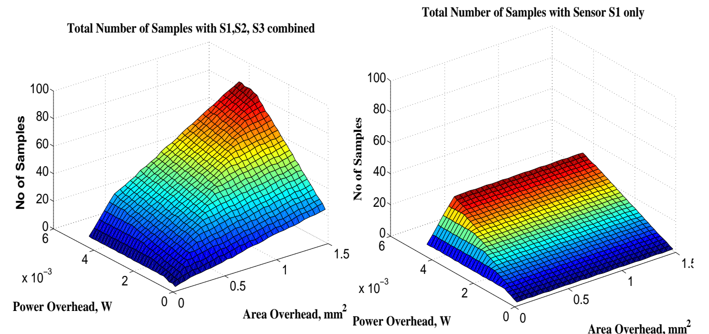


Figure 4. Number of samples that can be collected for different design corners and trade-offs (a) mix of sensors type S1,S2, S3 (b) S1 only. More the number of samples used in reconstruction, lesser would be the reconstruction error.

IX. Conclusions

Full chip thermal monitoring and reconstruction suffers from an inherent trade-off between accuracy of monitoring and implementation overhead (in terms of sensor die area and power). Both the accuracy of the reconstruction and the overhead incurred in the monitoring infrastructure is directly impacted by the number of samples, sensor locations, and accuracy of the sensors. To effectively address this inherent trade-off in thermal monitoring, we propose a new method called heterogeneous sensor allocation and placement (HSAP) that exploits the flexibility and trade-off in area and power characteristics of varied thermal sensors to perform a heterogeneous sensor allocation and placement for precisely recovering the full-chip thermal map. Unlike state-of-the-art sensor allocation and placement techniques that use a single type of homogeneous sensor, our HSAP approach finds the best combination or the mix of heterogeneous sensors for given sensor area and power budget along with their placement such that the full-chip thermal characterization error is minimized. Instead of quantifying the overheads in terms of total number of sensors, HSAP directly considers the sensor accuracy, area, power overhead of each sensor types in its formulation and thus provides a more flexible and accurate approach. HSAP exploits the trade-off in area and power of several sensor types to directly improve the accuracy of the full-chip thermal reconstruction. Simulations and experimental results

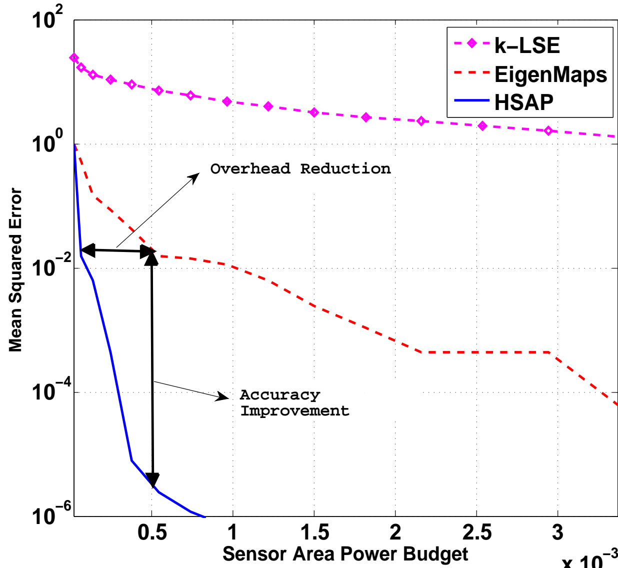


Figure 5. Comparison of reconstruction error with state-of-the-art methods. Both k-LSE[18] and EigenMaps[17] use sensor type S1 while proposed HSAP uses sensor combination S1,S2, and S3.

shows an improvement of 10-100x in reconstruction accuracy with three sensor types and execution speedup of over $20\times$ in comparison to a state-of-the-art technique using the most area efficient sensor type.

References

[1] S. Borkar *et al.*, “Parameter variations and impact on circuits and microarchitecture,” in *Design Auto. Conf., 2003. Proc.*, 2003, pp. 338–342.
[2] K. Skadron *et al.*, “Temperature-aware microarchitecture,” *SIGARCH Comput. Archit. News*, vol. 31, no. 2, pp. 2–13, May 2003.
[3] D. Brooks *et al.*, “Power, thermal, and reliability modeling in nanometer-scale microprocessors,” *Micro, IEEE*, vol. 27, no. 3, pp. 49–62, 2007.
[4] M. Pedram *et al.*, “Thermal modeling, analysis, and management in VLSI circuits: Principles and methods,” *Proc. of the IEEE*, vol. 94, no. 8, pp. 1487–1501, 2006.
[5] R. Viswanath *et al.*, “Thermal performance challenges from silicon to systems,” 2000.
[6] C. J. Lasance, “Thermally driven reliability issues in microelectronic systems: status-quo and challenges,” *Microelectronics Reliability*, vol. 43, no. 12, pp. 1969 – 1974, 2003.
[7] E. Rotem *et al.*, “Temperature measurement in the Intel Core Duo Processor,” in *Proc. Intl Workshop Thermal Investigations of ICs*, 2006, pp. 23–27.
[8] Y. Zhang *et al.*, “Accurate temperature estimation using noisy thermal sensors for gaussian and non-gaussian cases,” *Very Large Scale Integration (VLSI) Systems, IEEE Trans. on*, vol. 19, no. 9, pp. 1617–1626, 2011.
[9] C.-C. Chung *et al.*, “An autocalibrated all-digital temperature sensor for on-chip thermal monitoring,” *Circuits and Systems II: Express Briefs, IEEE Trans. on*, vol. 58, no. 2, pp. 105–109, 2011.
[10] P. Chen *et al.*, “A time-to-digital-converter-based cmos smart temperature sensor,” *Solid-State Circuits, IEEE Jr. of*, vol. 40, no. 8, pp. 1642–1648, 2005.
[11] ——, “A time-domain sar smart temperature sensor with curvature compensation and a 3 σ inaccuracy of -0.4 ^oc +0.6 ^oc over a 0 ^oc to 90 ^oc range,” *Solid-State Circuits, IEEE Jr. of*, vol. 45, no. 3, pp. 600–609, 2010.
[12] M. A. P. Pertijs *et al.*, “A cmos smart temperature sensor with a 3 sigma; inaccuracy of plusmn;0.1 deg;c from -55 deg;c to 125 deg;c,” *Solid-State Circuits, IEEE Jr. of*, vol. 40, no. 12, pp. 2805–2815, 2005.
[13] F. Sebastiano *et al.*, “A 1.2v 10 μw NPN-based temperature sensor in 65nm CMOS with an inaccuracy of 0.2 ^oc (3σ) from -70^oc to 125 ^oc,” in *Solid-State Circuits Conf. Digest of Tech. Papers (ISSCC), 2010 IEEE Int.*, 2010, pp. 312–313.
[14] R. Kuppuswamy *et al.*, “Over one million tpcc with a 45nm 6-core Xeon ® CPU,” in *Solid-State Circuits Conf. - Digest of Tech. Papers, 2009. ISSCC 2009. IEEE Int.*, 2009, pp. 70–71,71a.
[15] J. Dorsey *et al.*, “An integrated quad-core opteron processor,” in *Solid-State Circuits Conf., 2007. ISSCC 2007. Digest of Tech. Papers. IEEE Int.*, 2007, pp. 102–103.
[16] M. Ware *et al.*, “Architecting for power management: The IBM® POWER7 TM approach,” in *High Performance Comp. arch. (HPCA), 2010 IEEE 16th Int. Sym. on*, 2010, pp. 1–11.
[17] J. Ranieri *et al.*, “Eigenmaps: Algorithms for optimal thermal maps extraction and sensor placement on multicore processors,” in *Design Auto. Conf. (DAC), 2012 49th ACM/EDAC/IEEE*, 2012, pp. 636–641.
[18] A. Nowroz *et al.*, “Thermal monitoring of real processors: Techniques for sensor allocation and full characterization,” in *Design Auto. Conf. (DAC), 2010 47th ACM/IEEE*, 2010, pp. 56–61.
[19] S. Memik *et al.*, “Optimizing thermal sensor allocation for microprocessors,” *Comp.-Aided Design of IC and Sys., IEEE Trans. on*, vol. 27, no. 3, pp. 516–527, 2008.
[20] R. Mukherjee *et al.*, “Systematic temperature sensor allocation and placement for microprocessors,” in *Design Auto. Conf., 2006 43rd ACM/IEEE*, 2006, pp. 542–547.
[21] S. Reda *et al.*, “Improved thermal tracking for processors using hard and soft sensor allocation techniques,” *Comp.s, IEEE Trans. on*, vol. 60, no. 6, pp. 841–851, 2011.
[22] R. Cochran *et al.*, “Spectral techniques for high-resolution thermal characterization with limited sensor data,” in *Design Auto. Conf., 2009. DAC ’09. 46th ACM/IEEE*, 2009, pp. 478–483.
[23] H. Jung *et al.*, “A stochastic local hot spot alerting technique,” in *Design Auto. Conf., 2008. ASPDAC 2008. Asia and South Pacific*, 2008, pp. 468–473.
[24] C. R. Rao *et al.*, *Linear Models and Generalizations: Least Squares and Alternatives*. Springer, 2008.
[25] J. Tropp *et al.*, “Signal recovery from random measurements via orthogonal matching pursuit,” *Information Theory, IEEE Trans. on*, vol. 53, no. 12, pp. 4655–4666, 2007.
[26] W. Huang *et al.*, “Hotspot: a compact thermal modeling methodology for early-stage VLSI design,” *Very Large Scale Integration (VLSI) Systems, IEEE Trans. on*, vol. 14, no. 5, pp. 501–513, 2006.
[27] S. Li *et al.*, “McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures,” in *Microarchitecture, 2009. MICRO-42. 42nd Annual IEEE/ACM Int. Sym. on*, 2009, pp. 469–480.
[28] J. Long *et al.*, “Thermal monitoring mechanisms for chip multiprocessors,” *ACM Trans. Archit. Code Optim.*, vol. 5, no. 2, pp. 9:1–9:33, Sep. 2008.
[29] D. Donoho, “Compressed sensing,” *Information Theory, IEEE Trans. on*, vol. 52, no. 4, pp. 1289–1306, 2006.
[30] A. Çivril and M. Magdon-Ismail, “On selecting a maximum volume sub-matrix of a matrix and related problems,” *Theoretical Computer Science*, vol. 410, no. 47, pp. 4801–4811, 2009.

Appendix A
Direct Greedy Solution (gHSAP)

The direct greedy approach in Fig. VI-A consist of three stages that aims at decoupling the problem into subproblems for easy solution. We discuss the subproblems and some of the implementation issues in the following sections.

A. Selection of Weights in Sensor Selection ILP

The first stage of the HSAP performs the sensor selection based on the area, power and sensor accuracy. An optimization problem to maximize the number of samples used in the reconstruction is formulated as ILP. The intuition is based on the results derived in compressive sensing [29] and dimensionally reduction techniques, which states that the probability of reconstruction can be exponentially increased with increasing samples for a sparse signal. We establish a relationship between total samples used in the reconstruction with that of the sensor accuracy, area, and power overhead so as to maximize number of samples. The number of samples M_s is defined as a weighted sum of the samples collected from a particular

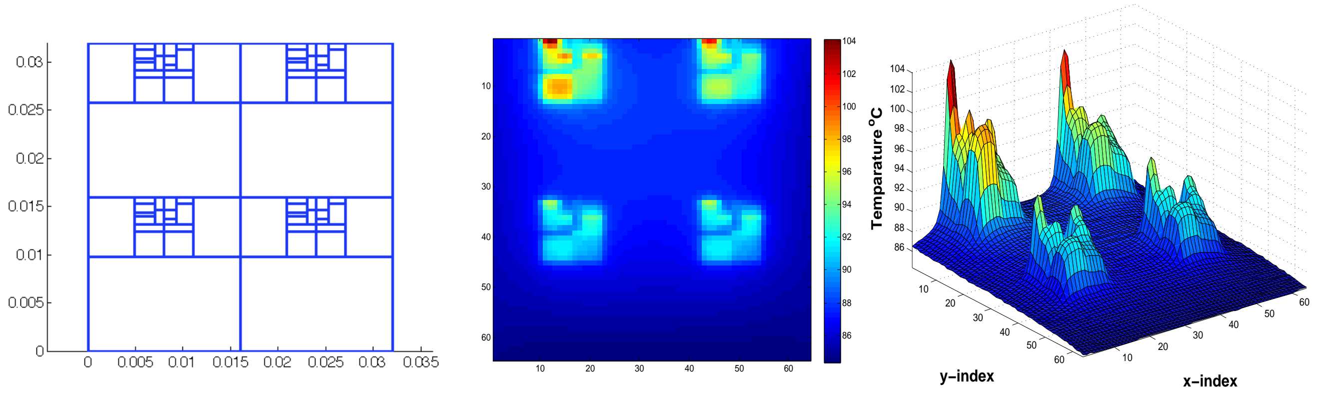


Figure 6. (a) Multicore Alpha processor floorplan (b) original thermal profile with SPEC 2K benchmarks (c) Recovered thermal profile from noisy sensor measurements.

type of sensor. The weights will decide which sensor type will have maximum contributions to the reconstruction samples. Thus by suitably selecting the weights we can either penalize a particular type or sensor or give preference to another. Selecting the weights can determine the number of samples collected by a particular sensor type for given area-power budget of the sensors. In selecting the sensors, one approach would be to give equal preference to each sensor types. In such a scenario, irrespective of the accuracy of the sensor, the number of samples are predominantly determined by the area and power budget. If the design is area constraint i.e., the area budget AB is small, the sensors with the least area overhead would be selected. On the other hand, if the design is power constraint such that the sensor power budget PB is small, the solution to the ILP would be guided toward the sensors with least power budgets. As both the constraints are simultaneously achieved, the solution meets design corners in between the two extremes.

Furthermore, as the overall reconstruction accuracy is effected by the sensor accuracy affects, higher the accuracy of the sensors, better would be the overall reconstruction. To consider the sensor accuracy in the selection of the sensors, we make the weights proportional to the accuracy (or inversely proportional to the the sensor noise variance). This way, we can account for the sensor accuracy in the sensor selection while fine tuning the results to a better solution.

B. Sensor Allocation and Mapping

Once the sensor combination and total number of sensors are selected, it is imperative to look for the mapping of the sensor of various types to the the specified locations. As some sensor location contribute more information to the reconstruction process than the others, it is intuitably logical to place the higher accuracy sensors to more information rich locations. The sensor placement algorithm returns placement of the sensors in order of their importance. We rank the sensors according to their accuracy and map the most accurate sensors to the most dominant locations. Another opportunity (or concern) that arises because of varied operating range of different sensors, need to be addressed during the sensor mapping process. A resoanble approch is to divide the full processor die into regions of hotter to colder thermal zones. For example, the integer register units (IRU) region of most processor (specifically Alpha processor) are having higher temperature than that of the caches memories. Sensor with high dynamic ranges can be mapped to these elevated regions while the lower range sensors can be placed in the relatively colder blocks.

C. Sensor Placement Algorithm

The sensor placement approach in Fig. 2 is inspired by the concept of volume sampling [30]. The algorithm maintains two set $\mathbf{L}$ and $\mathbf{S}$ where $\mathbf{L}$ is the set of chosen sensor location and $\mathbf{S}$ is the set of available sensor locations. Initially, the chosen sensor location set $\mathbf{L}$ is empty, and the $\mathbf{L}$ is the set to the set of all possible sensor locations. In the next step, the algorithm picks the location with the highest temperature based on the Euclidian norm. The location $\mathbf{L}$ and $\mathbf{S}$ are updated accordingly. The algorithm then iteratively computes the orthogonal components of the column vectors at the available locations and finds the highest Euclidian norm of the orthogonal components. The location corresponding to the highest Euclidian norm of the orthogonal components is included to the set of sensor locations $\mathbf{L}$ and removed from $\mathbf{S}$. The process is iterated for M locations.

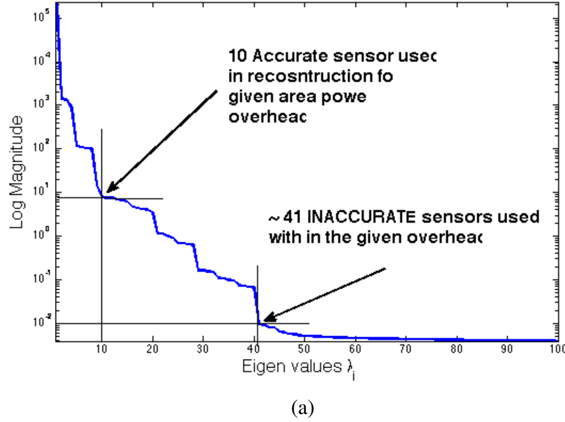
D. Effect of Approximation and Sensor Accuracy on MSE

The effect of both approximation and sensor noise on the reconstruction mean square error is bounded by the condition number κ of $\mathbf{\Phi_s}$ and the noise energy of $\| \boldsymbol{\epsilon} \|_2$ as [17]:

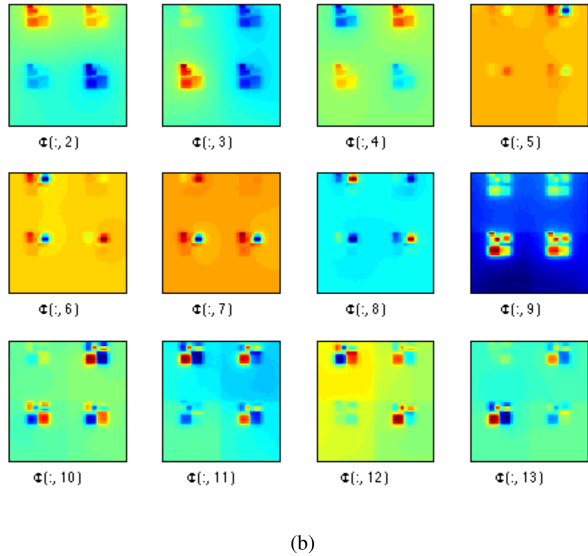
$$\frac{\| \hat{x} - x \|_2}{\| x \|_2} = \mathcal{O}(\kappa^2(\mathbf{\Phi_s})) \| \boldsymbol{\epsilon} \|_2 \,. \tag{14}$$

If the condition number is close to one, the matrix is well conditioned which means its inverse can be computed with good accuracy. If the condition number is large, then the matrix is said to be ill-conditioned. Practically, such a matrix is almost singular, and the computation of its inverse, or solution of a linear system of equations in is prone to large numerical errors. The numerical error that is introduced is often much larger than that introduced by ill-conditioning of $\mathbf{\Phi_s}$. The error $\boldsymbol{\varepsilon_t}$ introduced due to either truncation or inadvertent ill-conditioning in section IV-A is much higher than that of the contribution $\boldsymbol{\varepsilon_s}$ due to sensor accuracy. Bounding the sensor noise energy in 14 by the weakest sensor accuracy, we can bound the maximum error contribution due to sensor noise in MSE 14. In other words, the error contribution due to sensor inaccuracy is not as significant as basis truncation or ill-conditioning, under certain conditions. Thus we can effectively gain in accuracy by slightly trading sensor accuracy by allowing many weak, inaccurate sensors but eventually gaining more by increasing the number of basis vectors in the reconstruction. More number of sensors results in inclusion of the dominant eigen vectors in the reconstruction and remove the repercussions significantly in lieu of moderate increase in error due to sensor inaccuracy. Exploiting this tradeoff is cardinal to the accuracy improvement in HSAP. This trade-off is also depicted in the eigen values of the basis matrix $\mathbf{\Phi}$ as shown in Fig. 7a where the magnitude is of the order of 10^7 within 50 eigen values. If we accommodate 10 accurate sensors in a given area-power budget, we will be able to use 10 eigen vectors as in Fig. 7b in the reconstruction and thus the approximation error introduced

would be equal to the sum of rest of the eigen values – roughly $> 10^1$ as marked in Fig. 7a. On the other hand, if 41 relatively inaccurate sensors are used in the same are power budget, they will be able to cover 41 dominant eigen vectors and thus reduce the error to roughly $> 10^{-2}$. So, there is factor of 10^3 error reduction due to increasing the dominant basis vectors in the approximation even though the sensor inaccuracy has increase to say by an order of 10^1. Overall, there is a improvement of 10^2 in the reconstruction with heterogenous mix of relatively inaccurate sensors.



(a)



(b)

Figure 7. (a)Eigen values and their magnitude (b) Eigen Vectors for first 12 dominant eigen values corresponding to the basis matrix $\mathbf{\Phi}$ formed from the covariance matrix $\boldsymbol{C_x}$ of the thermal traces.

Appendix B
Future Works

The computational complexity of choosing M sensors from N locations is combinatorial of C_M^N and that of choosing K coefficients (or basis vectors) among N possible combination is C_K^N. The combined combinatorial complexity is $C_M^N * C_K^N$ which is proportional to $(N!)^2$. This combinatorial complexity makes HSAP a very difficult problem to solve. We are exploring efficient solution using global optimization techniques, specifically randomized and pattern search based optimization approach. A hybrid optimization approach, where greedy algorithm gHSAP provides the initial solution to a direct pattern search with mesh adaptive search (MADS) algorithm shows promising improvements in accuracy and speed.