Thermal Monitoring of Real Processors: Techniques for Sensor Allocation and Full Characterization

Abdullah Nazma Nowroz
Division of Engineering
Brown University
Providence, RI 02912
Email: abdullah_nowroz@brown.edu

Ryan Cochran
Division of Engineering
Brown University
Providence, RI 02912
Email: ryan_cochran@brown.edu

Sherief Reda
Division of Engineering
Brown University
Providence, RI 02912
Email: sherief_reda@brown.edu

ABSTRACT

The increased power densities of multi-core processors and the variations within and across workloads lead to runtime thermal hot spots locations of which change across time and space. Thermal hot spots increase leakage, deteriorate timing, and reduce the mean time to failure. To manage runtime thermal variations, circuit designers embed within-die thermal sensors that acquire temperatures at few selected locations. The acquired temperatures are then used to guide runtime thermal management techniques. The capabilities of these techniques are essentially bounded by the spatial thermal resolution of the sensor measurements. In this paper we characterize temperature signals of real processors and demonstrate that on-chip thermal gradients lead to sparse signals in the frequency domain. We exploit this observation to (1) devise thermal sensor allocation techniques, and (2) devise signal reconstruction techniques that fully characterize the thermal status of the processor using the limited number of measurements from the thermal sensors. To verify the accuracy of our methods, we compare our temperature characterization results against thermal measurements acquired from a state-of-the-art infrared camera that captures the mid-band infrared emissions from the back of the die of a 45 nm dual-core processor. Our results show that our techniques are capable of accurately characterizing the temperatures of real processors.

ACM Categories & Subject Descriptors
B.7.1 [Integrated Circuits]: Types and Design Styles.
General Terms: Design, Performance, Algorithms.
Keywords: Thermal characterization, spectral methods, k-LSE, compressive sensing, sensors allocation.

1. INTRODUCTION

While a chip's mean temperature is determined by its total power consumption, junction *hot spot* temperature is determined by the temporal and spatial distribution of power [6]. Depending on their intended deployment, current and future processors will fall within a range between being *power limited* (e.g., battery-operated devices) and being *hot spot limited* (e.g., high-end workstations) [6]. Hot spots impact directly all key circuit metrics, including: lifetime and reliability, speed, power, and costs. Hot spots 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 [14]. Furthermore, different thermal expansion coefficients of chip materials cause mechanical stresses that can eventually crack the chip/package interface [2]. Elevated temperatures also slow down devices and increase interconnect delays which might lead to timing failures [2]. High temperature also increase leakage power, which could lead to thermal runaway [8]. Many-core processors with 100s$-$1000s of cores will localize power consumption which further increases thermal problems [7].

Given the measurements of a few thermal sensors, the general objective of this paper is to devise a *physical layer* that takes as inputs the measurements from the sensors, and then provides to the thermal management unit the locations of the hot spots and the thermal slack that is available at every location on the die. The contributions of this paper are as follows.

- We elaborate the frequency domain characterizations of thermal signals directly acquired from a real processor using a thermal infrared camera. The sparsity of these signals is analyzed in the frequency domain. We also explore different choices of frequency-domain bases.
- Based on the concept of signal energy in the frequency domain, we propose a new method for thermal sensor allocation that allocates thermal sensors in a manner that allows accurate hot spot detection and full thermal characterization.
- Using the few measurements of thermal sensors, we propose a number of full signal reconstruction techniques that are capable of locating the sparse components of the temperature signal in the frequency domain and determining their magnitudes.
- In contrast to previous works which relied on simulators, we directly verify our method on an operational 45-nm dual-core processors using an infrared camera. Our realistic setup enables us to monitor the spatial and temporal thermal variations during runtime at a spatial resolution of 30 μm and 100 frames per second.
- We elucidate the trade-off between the number of thermal sensors and the accuracy of full thermal characterization and thermal hot spot estimation. We demonstrate that our proposed techniques are capable of outperforming existing techniques in the literature.

The organization of this paper is as follows. Section 2 overviews some of the recent relevant methods in the literature. Section 3 introduces the frequency-domain as a method to analyze temperature signals, and in Section 4 we propose new techniques for thermal sensor allocation based on the frequency domain. In Section 5 we propose a number of new, accurate methods for full thermal characterization. We verify our methods through an extensive set of experimental results in Section 6, and finally, Section 7 summarizes the main results of this work.

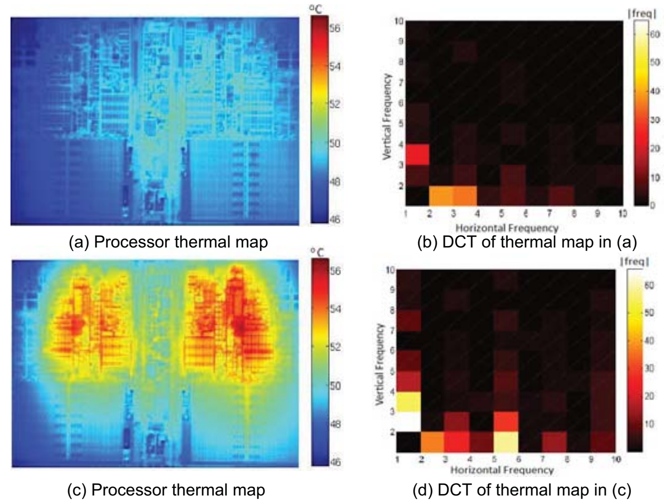
2. BACKGROUND AND PREVIOUS WORK

Most runtime thermal management methods rely on thermal sensors for thermal measurements [17, 10]. Inaccurate thermal monitoring arises from: limited number of sensors, remoteness of sensors from hot spots locations, manufacturing variability, and analog-to-digital conversion accuracy [17]. These inaccuracies decrease the processor's performance and waste power. Rotem *et al.* [17] report that 1 °C accuracy translates to 2 W power savings, and that due to lack of proximity, sensor measurements and hot spot temperatures could differ by up to 10 °C. While increasing the number of thermal sensors can reduce these errors, thermal sensors and their support circuitry (e.g., A/D converters) and wiring occupy valuable silicon area, and thus designers tend to limit their numbers.

A number of recent papers discuss *design methods* that allocate sensors near potential hot spot locations, and *runtime methods* that estimate the *hot spot* temperature and *full chip* temperatures during runtime using the measurements of the thermal sensors [9, 11, 13]. There are two reasons for characterizing the full thermal status of the processor in addition to hot spot detection:

- While hot spot temperature estimation enables overall chip throttling and shut down decisions, full chip temperature estimation provides the necessary information required for fine-grain thermal management. This information is essential for thermal-driven spatial thread-migration or dynamic frequency and voltage scaling techniques for many-core processors [16, 18]. For example, Intel's Core i7 utilizes the *thermal slack* available at one core to boost the performance of another core.
- Full thermal mapping can be used to compute the runtime power dissipation for each functional unit of the processor. While design tools can provide detailed power consumption estimates when input test vectors are provided, they provide inaccurate estimates when power consumption due to software execution (e.g. O.S. and applications) is desired. Hence computing the detailed power consumptions from thermal data can give more accurate results. Existing techniques utilize infrared camera to acquire the thermal data [12]. Full thermal characterization of processors using only the thermal sensor measurements can eliminate the need for infrared cameras greatly simplifying thermal characterization setup and costs.

For thermal sensor allocation, Mukherjee and Memik [13, 11] describe a clustering algorithm that computes the thermal sensor positions that best serve clusters of potential hot spot locations. The locations of these hot spots are identified via extensive workload thermal simulation. For runtime hot spot and full thermal characterization, two techniques have been proposed. In the first technique, Long *et al.* [9] advocate using a grid-based interpolation scheme that identifies the hot spot around each sensor by interpolating the measurements at its immediate neighbors. In the second technique, Cochran *et al.* [15] advocated spectral techniques that are capable of fully reconstructing the temperature at all locations of the processor die. The main idea of [15] is to regard the spatial temperature as a space-varying signal and to utilize the Nyquist-Shannon sampling theory to devise methods that can reconstruct the full thermal status from the measurements of the thermal sensors.



(a) Processor thermal map (b) DCT of thermal map in (a)
(c) Processor thermal map (d) DCT of thermal map in (c)

Figure 1: Thermal maps of Athlon dual II processor and their corresponding DCT frequency-domain representations.

3. FREQUENCY DOMAIN TECHNIQUES

Exact thermal estimation requires solving the partial differential heat equation, or a lumped first-order approximation for it, using as input the detailed power consumption of various processor units together with a model of the chip-package structure. This detailed information is not available during runtime, and furthermore, there is no sufficient computational resources to estimate the temperature in real time. Thus, we focus on devising computationally-efficient *frequency-domain* techniques that only use the measurements of the thermal sensors to fully characterize the processor's temperature. If t is the spatial temperature signal which is expressed as a $N \times 1$ vector, then the frequency domain representation of t can be expressed as

$$t = \Phi C, \text{or } C = \Phi^\dagger t \tag{1}$$

where Φ is an $N \times N$ matrix with columns that form a orthonormal basis, $\Phi^\dagger$ is the conjugate transpose of Φ, and $C = \{C_1, \dots, C_N\}$ is $N \times 1$ vector that has the frequency-domain coefficients of t. There are many choices for orthonormal bases. One choice for Φ is the 2-D Discrete Fourier Transform (DFT) matrix, where an element at row u and column v of Φ is computed as,

$$\Phi_{uv} = \frac{1}{N \times N} e^{-2\pi i q_u q_v / N} e^{-2\pi i r_u r_v / N}, \tag{2}$$

where r_u and q_u are the remainder and quotient of dividing u by N respectively, and r_v and q_v are the remainder and quotient of dividing v by N respectively. Another choice is the 2-D Discrete Cosine Transform (DCT) matrix , where each element in the matrix is computed as,

$$\Phi_{uv} = \alpha_u \alpha_v \cos \frac{\pi(2q_v + 1)q_u}{2N} \cos \frac{\pi(2r_v + 1)r_u}{2N}, \tag{3}$$

where α_u and α_v are normalization factors. The biggest advantage of using frequency-domain techniques is that they transform a non-sparse signal, t in this case, to a sparse signal C with mostly zero coefficients. The sparsity of the spatial thermal signal in the frequency domain is confirmed as follows. Using our infrared camera, we capture the thermal maps of an AMD Athlon II dual-core processor during operation as shown in Figures 1.a and 1.c. We plot the the DCT transform of the two thermal maps in Figure 1.b and 1.d. Note that we only plot few frequency coefficients at low

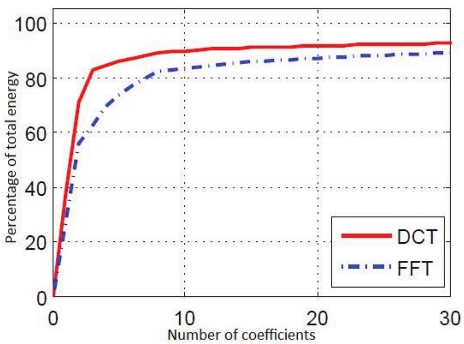


Figure 2: Fraction of signal energy captured as a function of the number of frequency-domain coefficients.

frequencies as the rest of the coefficients are very close to zero. Furthermore, Figure 1.d displays coefficients that are larger in magnitude and at a larger frequency ranges than the coefficients of Figure 1.b. These differences in the magnitudes and locations of the frequency domain coefficients are a quantitive metric for the visually apparent strong spatial gradients of the temperature signal of Figure 1.c in comparison to Figure 1.b.

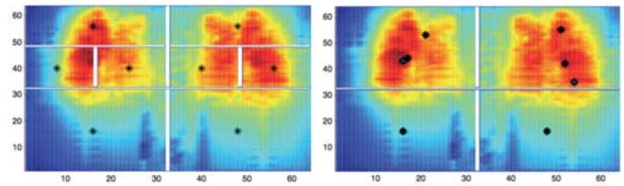
Signal *energy* is an important concept in frequency-domain analysis, where it is defined as the sum of squares of the magnitudes of its coefficients in the frequency domain. The concept of energy gives a quantitive metric to compare various frequency-domain representations. We use this concept to confirm that DCT is a better basis than DFT for thermal signals. Using the thermal maps of Figure 1, we plot in Figure 2 the percentage of energy captured as a function of the number of frequency domain coefficients, for both the FFT and DCT bases. From the plot, it is evident that it is possible to capture most of the energy content of the thermal signal with only few coefficients. Furthermore if we compare DFT and DCT, we can see from the plot that the DCT captures the same energy as the DFT using a fewer number of coefficients. This result shows that DCT has more energy concentrated in fewer number of coefficients making it a sparser representation. Thus, we choose DCT as our orthogonal basis for further analyses.

4. PROPOSED THERMAL SENSOR ALLOCATION TECHNIQUES

In this section we propose a new thermal sensor allocation method that allows both accurate thermal hot spot estimation and full thermal characterization of the processor. In general, techniques that focus on placing sensors only near potential hot spot locations will produce poor results for full thermal characterization as they will have no information on the thermal slack available at the other locations on the die. Hot spot driven methods "see the trees in the forest" but lacks a picture of how the "forest" looks like. This full picture deficiency is undesirable for fine-grain thermal management techniques. Our approach attempts to address this deficiency by proposing methods that are simultaneously capable of accurate hot spots estimation and full thermal characterization.

Our proposed thermal sensor allocation method is based on the observation that signals with large spatial thermal variations lead to frequency-domain representations that have significant energy at higher frequencies as demonstrated earlier in Section 3. Thus, the general idea of our method is to allocate more thermal sensors to regions of the chip with higher frequency-domain energy. In particular, our sensor allocation method – which is inspired by min-cut placement techniques [5] – recursively bisects the die area alternating between vertical and horizontal cuts. At each cut level, the algorithm calculates the spectral energy of the signal (after subtracting its DC value) of each of the two regions of the bisection and then allocates the number of sensors available at the cut-level in proportion to the spectral energy of both regions. During the recursive bisection, when the number of sensors for a region is less than or equal to some *threshold limit*, then the thermal sensor allocation algorithm can use one of two possible methods to allocate the sensors to its region:

- **Energy-Center Sensor Allocation.** In the first method which is used when the threshold limit is equal to one, a sensor is placed at the geometric center of the region. The method is aimed at maximizing the accuracy of full thermal characterization of the whole region as will be verified in the experimental results section. However, this method might lead to wrong predictions for the region's hot spot temperatures if the locations of the hot spots are not at the center of the region.
- **Energy-Cluster Sensor Allocation.** In the second method, the sensor(s) available for a region are placed at the centroids of the potential thermal hot spots located at the region. The centroid locations are computed using the k-means clustering algorithm. Essentially this method attempts to strike a balance between hot spot estimation and full thermal characterization. It lets the recursive energy-based bisection guides the *global* allocation of the available sensors over the chip, and then it lets the hot spot locations determine the *detailed* or exact location of the thermal sensors. If the region is expected not to have any hot spots and is allocated one sensor, then the sensor is allocated to the center of the region; otherwise, recursive bisection continues till only one sensor is allocated per region.



(a) energy-center sensor allocation (b) energy-cluster sensor allocation

Figure 3: Proposed thermal sensor allocation based on spectral energy density.

For example, Figure 3.a and 3.b show the bisection of the thermal map of Figure 1.c using eight sensors. The difference between the two figures is that Figure 3.a gives the bisection and sensor allocation for the energy-center sensor allocation method, and Figure 3.b shows the bisection and allocation for the energy-cluster allocation method when the limit is equal to three sensors per region.

5. PROPOSED FULL RUNTIME THERMAL CHARACTERIZATION TECHNIQUES

During regular processor operation, the full temperature field $t = \{t_1, t_2, \ldots, t_N\}$ of the die at all N locations is not available; instead, only n samples obtained from the n thermal sensors are available. Let $y = \{i_1, i_2, \ldots, i_n\}$ denote the indices or locations of the n sensors in the temperature field, and $t(y)$ denote the temperature measurements at these locations. Then the objective of full thermal characterization is to estimate the temperature at each of the N points given the measurements of the n sensors. If the thermal signal is k-sparse in the frequency domain, and if the locations of the non-zero signal coefficients in the frequency domain are known, then let $s = \{j_1, j_2, \ldots, j_k\}$ denote the locations or indices of these non-zero coefficients. Given the sensor samples $t(y)$ and the orthonormal basis Φ, the signal samples $t(y)$ can be expressed

$$\begin{aligned} t(i_1) &= \Phi_{i_1,j_1} C_{j_1} + \Phi_{i_1,j_2} C_{j_2} + \cdots + \Phi_{i_1,j_k} C_{j_k} \\ \vdots &= \vdots \\ t(i_n) &= \Phi_{i_n,j_1} C_{j_1} + \Phi_{i_n,j_2} C_{j_2} + \cdots + \Phi_{i_n,j_k} C_{j_k} \end{aligned}$$

These set of equations can be written succinctly using matrix notation as

$$t(y) = \Phi(y, s)C(s), \tag{4}$$

where $\Phi(y, s)$ denote the matrix formed using Φ's rows with indices y and Φ's columns with indices s, and $C(s)$ is the vector formed of the C elements at indices s. In this case the best C that gives the total least square errors (LSE) of Equation (4) is given by

$$C_{LSE} = (\Phi(y, s)^{\dagger}\Phi(y, s))^{-1}\Phi(y, s)^{\dagger}t(y). \tag{5}$$

Full temperature characterization is achieved by multiplying the original basis set Φ with C_{LSE}; i.e., $t = \Phi C_{LSE}$. While Equation (5) gives a convenient closed-form to find the best frequency-domain representation, it is also possible to find the LSE solution of Equation (4) iteratively using gradient descend methods [1]. The advantage of gradient descend methods is that they allow a relatively smooth trade-off between computational runtime and solution accuracy. In all cases, solving Equation (4) hinges on the ability to determine the locations of the non-zero coefficients of C; that is, s must be determined. We propose two approaches to tackle this problem:

A. k-LSE using Pre-determined Thermal Characterization. In the first approach, we utilize our observation of Section 3 that the energy of the temperature signals acquired from real processors are mostly concentrated in the low frequency range. Consequently, the locations of the k non-zero coefficients can be picked from the low-frequency range. The number of coefficients picked depend on the number of available thermal sensors where $k < n$ to make sure that the solution to Equation (5) is stable. To prioritize low-frequency coefficients over high-frequency coefficients, we propose the order given in Figure 4. Essentially our proposed method picks coefficient locations to match the frequency-domain representations of thermal signals acquired from real processors.

B. Compressive Sensing. Compressive sensing techniques attempts to simultaneously find the locations and magnitudes of the non-zero components of the sparse signal [3]. We investigate two compressive sensing techniques. In the first technique (`CS-L1_MIN`), the following Second-Order Cone Programming (SOCP) formulation

$$\min \quad ||C||_1 \tag{6}$$

$$\text{subject to} \quad t(y) = \Phi(y, :) \times C, \tag{7}$$

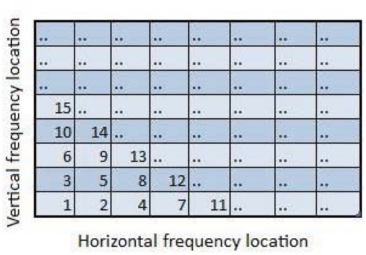


Figure 4: Order of coefficients for the `k-LSE` method.

Procedure: STOMP($\cdot$)
Input: Initialize $r = t(y)$, C as a zero $N \times 1$ vector, and $s = \emptyset$.
Output: Full characterization of temperature signal.

While $|r|$ < noise threshold

1. $C_r = \Phi(y, :)^{\dagger} r$
2. Find $p = \{j | C_r(j) > \max(C_r) - \text{noise threshold}\}$
3. Let $s = s \cup p$
4. Compute $C(s) = (\Phi(y, s)^{\dagger}\Phi(y, s))^{-1}\Phi(y, s)^{\dagger}t(y)$
5. Let $r = t(y) - \Phi(y, :)C$

Return $t = \Phi C$

Figure 5: Procedure STOMP ($\cdot$) for full thermal characterization using compressive sensing.

is solved to minimize the L_1 norm of C, where $\Phi(y, :)$ is the matrix formed from the full rows of the Φ matrix. Minimizing the L_1 norm forces the SOCP solver to choose the solutions that are sparsest. For a real-time setting, solving a SOCP formulation can be unacceptable from an computational perspective. Thus, we explore a second technique (`CS-STOMP`) based on the greedy iterative procedure Stagewise Orthogonal Matching Pursuit (StOMP) [4]. The StOMP algorithm transforms the sampled signal $t(y)$ into a negligible residual by identifying the significant non-zero components from the signal in the frequency domain one at a time in a greedy fashion as given in Figure 5. In Step 1, the algorithm computes the frequency domain representation C_r of the residual signal (the residual signal r is initialized to $t(y)$ at the beginning of the algorithm), and then in Step 2, it determines the locations of non-zero coefficients from C_r that are above a certain noise threshold. Using the selected coefficient locations, the algorithm then solves the least squares estimation formulation in Step 4 to determine the magnitude of these selected coefficients. Finally, it reconstructs the signal in Step 5 and subtracts it from the original sampled signal to produce a new residual. The algorithm is iterated until the residual goes close to zero and all the significant non-zero components in the signal are recovered.

A fundamental assumption for both the compressive sensing techniques is that the sampling is performed randomly. In real processors thermal sensors are placed at strategic locations near hot spots, and consequently, the theoretical attractiveness of compressive sensing is no longer guaranteed.

6. EXPERIMENTAL RESULTS

To test the effectiveness of our methods, we put together the following sophisticated experimental setup:

- For thermal imaging, we utilize a FLIR SC5600 infrared camera with a mid-infrared spectral range of 2.5 μm – 5.1 μm. The camera is capable of operating at 100 Hz with a spatial resolution of 30 μm with a 0.5× microscopy kit.
- We use a real dual-core 45 nm AMD Athlon II X2 240 processor running at 2.1 GHz. We remove the processor's heat spreader to expose the back of its die to the infrared camera, and then use a thin film (about 1 mm of thickness) of infrared transparent mineral oil as the heat spreader as shown in Figure 6. The oil is pumped continuously through a heat exchange system. Our oil-based cooling system is capable of dissipating about 40 W.
- To collect the thermal traces involved in the experiments we use the CPU SPEC 2006 benchmark set which has 29 applications. We collect 100 temperatures traces obtained from the infrared

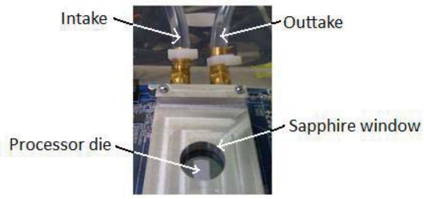


Figure 6: Setup for oil-based infrared transparent heat spreader.

camera after executing each benchmark individually and after executing two benchmarks at a time.

In our experiments we report the following two metrics:

- *Full thermal characterization error.* For each temperature trace, we compute the average absolute error between the true temperatures as measured by the infrared camera and as estimated by our signal reconstruction methods. We normalize the average error by the difference between the maximum and minimum temperatures in the trace. This normalization helps put the characterization error in perspective: a 0.5 °C error is more significant if the difference between maximum and minimum temperatures is 5 °C rather than 20 °C. We report the average absolute error computed for all 100 temperature traces.
- *Hot spot estimation error.* The computation of the hot spot error is similar to the computation of the full thermal characterization error except that the error is equal to the difference between the maximum hot spot temperature as reported by the thermal infrared camera and the maximum of the temperatures at the locations of the thermal sensors.

The objective of the first experiment is to demonstrate the effectiveness of our frequency-domain techniques in fully reconstructing the thermal signal. In this experiment we compute the full thermal characterization error as a function of the number of sensors which are placed randomly (in subsequent experiments we explore different methods for thermal sensor placement). We compare here four full thermal characterization methods: the `spectral` method [15], the proposed `k-LSE` method, the `CS-L1_MIN` method, and the `CS-STOMP` method. The performance of these methods as a function of the number of sensors is given in Figure 7. The results show that the proposed `k-LSE` method gives superior results compared to other methods. Note that `CS-L1_MIN` and `CS-STOMP` fail to produce meaningful results when the number of sensors is relatively low. The main reason for the poor performance of generic compressive sensing techniques is that they attempt to compute both the locations and magnitudes of the frequency-domain coefficients, and hence there is a chance they end up picking wrong high-frequency coefficients. In contrast our proposed `k-LSE` method is devised with the nature of thermal characterizations encountered in real processors in mind, and hence it is more powerful than a generic technique. Our method `k-LSE` also outperforms the `spectral` method ([15]) as it uses a sparser basis (DCT instead of FFT) and a more accurate direct algebraic approach that guarantees minimizing the total square error.

In the second experiment we demonstrate the accuracy of thermal characterization of the `k-LSE` method as a function of sensor placement method. We vary the number of thermal sensors from 4 to 32, and for each number of thermal sensors, we report the mean of the absolute error between the full characterization results as computed by the `k-LSE` method and the true thermal map as given by the thermal camera. We compare here four different methods for thermal sensor placement: (1) `random` gives the

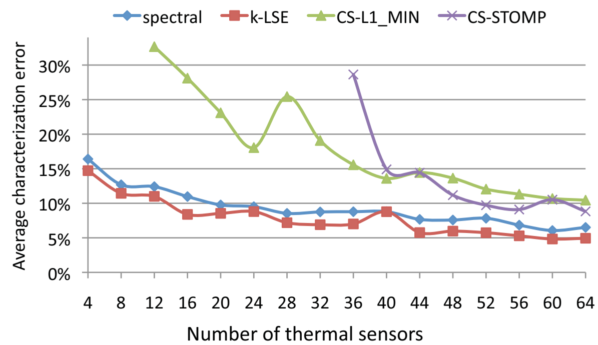


Figure 7: Average error in full thermal characterization using various temperature reconstruction methods.

characterization with a random sensor placement; (2) `k-cluster` gives the characterization error when k-means clustering is used to determine the sensor placement [13, 11]; (3) `energy-cluster` gives the characterization error when our proposed energy-driven sensor allocation method is used for global sensor allocation and k-means clustering is used for detailed sensor placement; and (4) `energy-center` gives the characterization error when our proposed energy-driven sensor allocation method is used and the sensors are placed at the centers of their respective regions. The plot in Figure 8 gives the full thermal characterization error under the aforementioned sensor allocation strategies. In comparing the results, we find that proposed `energy-center` gives the least full characterization error. Except for `k-cluster` all methods display the expected trend of decreasing error as the number of sensors increases. The reason for `k-cluster` "poor" performance is that `k-cluster` attempts to get the sensors closely to the potential hot spots locations at the expense of fairly estimating the temperature at other die locations. Ultimately when the number of sensors is equal to the number of potential hot spot locations, `k-cluster` will place a sensor at exactly each hot spot location further increasing the full thermal characterization error.

The objective of the third experiment is to demonstrate the effectiveness of our frequency-domain based thermal sensor allocation technique in detecting hot spots. In this experiment we compute the error in hot spot estimation using only the temperatures observed at the thermal sensor locations in comparison to the true maximum temperature on the die as recorded by the infrared camera. We compare here four sensor allocations methods: a plain random allocation (`random`), a method based on clustering technique proposed in [11, 13] (`k-cluster`), the proposed frequency-domain based techniques (`energy-center`) and (`energy-cluster`). We repeat this experiment using different number of sensors and report

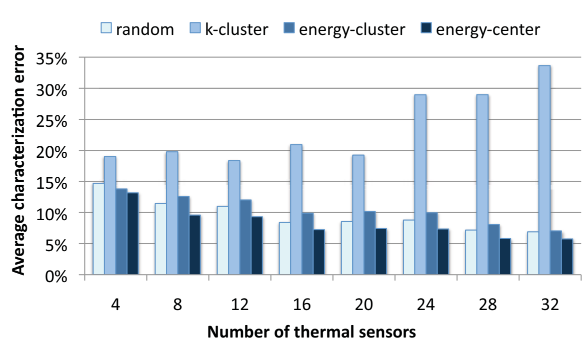


Figure 8: Average error in full thermal characterization using the `k-LSE` for various sensor allocation strategies.

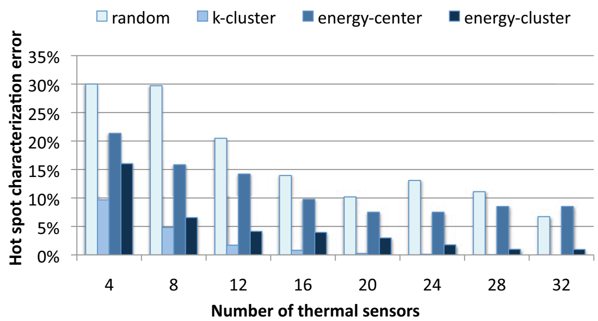


Figure 9: Average error in hot spot estimation using various number of sensors and sensor allocation strategies.

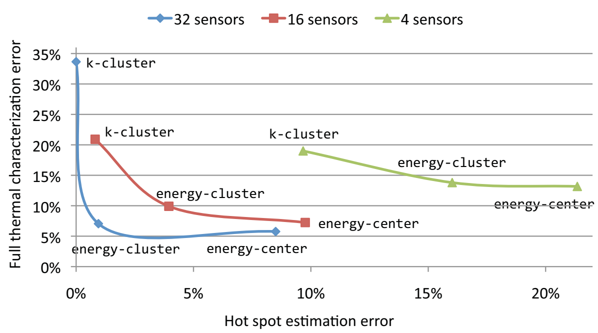


Figure 10: Hot spot error and full thermal characterization error as a function of thermal sensor placement methodology. `k-LSE` method is used for full thermal reconstruction.

the errors in hotspot estimation in Figure 9. The results show that sensor allocation `energy-cluster` method gives close results within $1-2\%$ to `k-cluster`, while attaining a large accuracy advantage in full thermal characterization as demonstrated by Experiment 2.

To elucidate the trade-off between hot spot estimation and full thermal characterization, we summarize the performance of the proposed `k-LSE` method as a function of the thermal sensor allocation method in Figure 10. It is clear that `energy-cluster` provides an excellent compromise: it comes very closely to the accuracy of `energy-center` in full thermal characterization and `k-cluster` in hot spot estimation.

7. CONCLUSIONS

In this paper we presented a new methodology for thermal monitoring techniques for real processors. We proposed static design-time thermal sensor allocation technique, and runtime full thermal characterization techniques. Our work utilizes frequency-domain signal representations to drive both static and runtime thermal monitoring techniques. We proposed frequency-domain representations based on the DCT basis which achieves better results than the traditional FFT basis. We characterized the DCT representations of spatial thermal signals of real processor, and utilized this characterization to devise a number of effective full thermal estimation methods such as `k-LSE`, `CS-L1_MIN`, and the `CS-STOMP`, which are designed for fine-grain thermal management techniques. We also proposed thermal sensing allocation techniques that recursively allocates the sensors to the different die regions depending on their spectral energy. Our `energy-cluster` sensor allocation method gives a good balance between reducing full thermal characterization error and hot spot estimation error. We also explored compressive thermal sensing techniques and traditional signal analysis techniques as means for thermal characterization. Using a sophisticated experimental setup, we demonstrated the superiority of our techniques using a real 45-nm dual-core processor and a state-of-the-art thermal infrared camera.

Acknowledgments: This work is partially supported by a DoD ARO grant number 55544-CS-RIP and a NSF CAREER grant number 0952866.

8. REFERENCES

[1] E. Alpaydin. *Introduction to Machine Learning*. The MIT Press, first edition, 2004.

[2] D. Brooks, R. Dick, R. Joseph, and L. Shang. Power, Thermal, and Reliability Modeling in Nanometer-Scale Microprocessors. *IEEE Micro*, 27(3):49 – 62, 2007.

[3] E Candès. Compressive Sampling. *Proceedings of the International Congress of Mathematicians*, pages 1–20, 2006.

[4] D Donoho, Y Tsaig, I Drori, and J Starck. Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit. *Stanford University Technical Report*, Jan 2006.

[5] A. E. Dunlop and B. W. Kernighan. A Procedure for Placement of Standard-Cell VLSI Circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 4(1):92–98, 1985.

[6] H. Hamann, A. Weger, J. Lacey, Z. Hu, and P. Bose. Hotspot-Limited Microprocessors: Direct Temperature and Power Distribution Measurements. *IEEE Journal of Solid-State Circuits*, 42(1):56–65, 2007.

[7] W. Huan, M. R. Stan, K. Sankaranarayanan, R. J. Ribando, and K. Skadron. Many-Core Design from a Thermal Perspective. In *Design Automation Conference*, pages 746–749, 2008.

[8] S.-C. Lin and K. Banerjee. Cool Chips: Opportunities and Implications for Power and Thermal Management. *IEEE Transactions on Electron Devices*, 55(1):245 – 255, 2008.

[9] J. Long, S. Memik, G. Memik, and R. Mukherjee. Thermal Monitoring Mechanisms for Chip Multiprocessors. In *ACM Transactions on Architecture and Code Optimization*, volume 5(2), pages 9:1–9:23, 2008.

[10] R. McGowen, C. Poirier, C. Bostak, and J. Ignowski. Power and Temperature Control on a 90-nm Itanium Family Processor. *IEEE Transactions on Solid-State Circuits*, 41(1):229–237, 2006.

[11] S. Memik, R. Mukherjee, M. Ni, and J. Long. Optimizing Thermal Sensor Allocation for Microprocessors. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 27(3):516–527, 2008.

[12] F. J. Mesa-Martinez, M. Brown, J. Nayfach-Battilana, and J. Renau. Measuring Performance, Power, and Temperature from Real Processors. In *International Symposium on Computer Architecture*, pages 1–10, 2007.

[13] R. Mukherjee and S. Memik. Systematic Temperature Sensor Allocation and Placement for Microprocessors. In *Design Automation Conference*, pages 542 – 547, 2006.

[14] M. Pedram and S. Nazarin. Thermal Modeling, Analysis, and Management in VLSI circuits: Principles and Methods. *Proceedings of the IEEE*, 94(8):1487–1501, 2006.

[15] R. Cochran and S. Reda. Spectral Techniques for High-Resolution Thermal Characterization with Limited Sensor Data. In *Design Automation Conference*, pages 478–483, 2009.

[16] R. Rao, S. Vrudhula, and C. Chakrabarti. Throughput of Multi-Core Processors Under Thermal Constraints. In *International Symposium on Low Power Electronics and Design*, pages 201–206, 2007.

[17] E. Rotem, J. Hermerding, C. Aviad, and C. Harel. Temperature Measurement in the Intel Core Duo Processor. In *Proceedings of the International Workshop on Thermal Investigations of ICs*, pages 23–27, 2006.

[18] Y. Wang, K. Ma, and X. Wang. Temperature-Constrained Power Control for Chip Multiprocessors with Online Model Estimation. In *International Symposium on Computer Architecture*, pages 314–324, 2009.