Minimal Sparse Observability of Complex Networks: Application to MPSoC Sensor Placement and Run-time Thermal Estimation & Tracking

Santanu Sarma and Nikil Dutt

Department of Computer Science, University of California Irvine, CA, USA.
Email:{santanus,dutt}@ics.uci.edu

Abstract—**This paper addresses the fundamental and practically useful question of identifying a minimum set of sensors and their locations through which a large complex dynamical network system and its time-dependent states can be observed. The paper defines the minimal sparse observability problem (MSOP) and provides analytical tools with necessary and sufficient conditions to make an arbitrary complex dynamic network system completely observable. The mathematical tools are then used to develop effective algorithms to find the sparsest measurement vector that provides the ability to estimate the internal states of a complex dynamic network system from experimentally accessible outputs. The developed algorithms are further used in the design of a sparse Kalman filter (SKF) to estimate the time-dependent internal states of a linear time-invariant (LTI) dynamical network system. The approach is applied to illustrate the minimum sensor in-situ run-time thermal estimation and robust hotspot tracking for dynamic thermal management (DTM) of high performance processors and MPSoCs.**

Index Terms—**Complex Networks, Sparsity, Observability, Controllability, Control Theory, Compressive Sensing, Thermal modeling, Temperature sensor placement, Estimation, Prediction, CyberPhysical Systems.**

I. INTRODUCTION

OBSERVING and controlling complex networks is of paramount importance in science and engineering. The ability to experimentally access and accurately observe the internal states of a system offers means to quantitatively describe dynamic behaviors of any complex network system. Such networks exist in a wide range of systems including network of chemicals linked by chemical reactions [10], [11], the Internet, network of routers and computers connected by physical links in multi/many-core processor, as well as many/multi processor thermal and heat-flow networks [8]. A necessary step towards observing a complex network system is to fully understand the observability of complex networks with linear dynamics [11]. Consider a network of n nodes described by the following set of ordinary differential equations [10]:

$$\begin{aligned} \dot{\mathbf{x}} &= \widetilde{\mathbf{A}}\mathbf{x} + \widetilde{\mathbf{B}}\mathbf{u} \\ \mathbf{y} &= \widetilde{\mathbf{C}}\mathbf{x} \end{aligned} \tag{1}$$

where the vector $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ stands for the states of nodes, $\widetilde{\mathbf{A}} \in \mathbb{R}^{n \times n}$ represents the coupling matrix of the system, in which a_{ij} represents the weight of a directed link from node j to i (for undirected networks $a_{ij} = a_{ji}$), $\mathbf{u_k} = [u_1, u_2, ..., u_p]^T$ are the set of controllers or control inputs, $\widetilde{\mathbf{B}} \in \mathbb{R}^{n \times p}$ is the control matrix, $\mathbf{y}$ are the measurement, and $\widetilde{\mathbf{C}} \in \mathbb{R}^{\mathbf{m} \times \mathbf{n}}$ is the measurement matrix. A system is called observable if we can reconstruct its complete internal states $\mathbf{x}$ from the measured outputs $\mathbf{y}$. Although simultaneous measurement and sensing of all the internal variables offers a complete description of the system's state, in practice experimental access is limited to only a subset of variables due to cost of the sensing infrastructure, placement restrictions, as well as unavailability of suitable and effective sensing mechanisms. Identifying a set of such crucial points that can provide complete insight into the internal dynamics of a complex network system is fundamental to effective and high performance system design.

In this paper, we explore this fundamental question of observing the internal dynamics of a linear dynamical system using a minimal set of observation points by using the notion of sparsity as developed in the emerging field of compressive sensing [4]. We define the minimal sparse observability problem to find the sparsest measurement vector which makes a linear dynamical complex network system completely observable. We formulate and develop analytical tools to find the minimum number of nodes (sensors) for any arbitrary type of networks. The mathematical tools are then used to develop effective algorithms to find the sparsest measurement vector that enables the ability to estimate the internal states of a complex dynamic system from experimentally accessible outputs. The developed algorithms are further used in the design of a sparse Kalman filter (SKF) to estimate the time-dependent internal states of a high performance processor system and its dynamic thermal management (DTM) and control with the minimum number of on-chip sensors.

II. MOTIVATION AND RELATED WORK

The conditions of observability and controllability of LTI systems were initially introduced by Kalman [9] and have been used extensively in control theory. Although the classical rank condition [9], [10] proposed by Kalman provides a test for checking the controllability and observability for given system matrices, but the process of systematically finding these system matrices (measurement and control matrices) have not been addressed. The very recent groundbreaking work of Liu et al. [10] addressed the process of making a complex system completely controllable using few controlling or driving nodes. The paper finds the minimal number of driving nodes (or controllers) that would be necessary for driving a complex network to a specific state by using a graphical approach. Specifically, they developed a minimum input theory to efficiently characterize the structural controllability of directed networks, allowing a minimum set of driver nodes to be identified to achieve full control. In particular, the structural controllability of a directed network can be mapped into the problem of maximum matching [7], [17], where external control is necessary for every unmatched node. Liu et al. [11] extended their graphical approach to observability of complex networks in their very recent work [11] .

This work was partially supported by the NSF Variability Expedition award CCF-1029783.

978-3-9815370-2-4/DATE14/©2014 EDAA

Although the graphical approach based on structural controllability theory offers a general tool for *directed* networks, the approach fails if the assumption of independence of free link parameters and non-symmetry of the structural matrices is violated [10]. In other words, the graphical approach can not be applied to any arbitrary complex network with structure and configurations of the link weights where the parameters (-i.e., the elements of the systems matrices) are not independently varying. Specifically, for *undirected* networks, the symmetric characteristic of the network matrix accounts for the violation of the assumption of the structural matrix, even with random weights [10]. To overcome this limitation, recently a more generic algorithmic approach was proposed in [13]. Our work is motivated by the work in [13] but differs in its objective and problem formulation. The work in [13] finds the sparsest control matrix $\mathbf{B}$ for which a complex network is fully controllable whereas we find the sparsest measurement matrix $\mathbf{C}$ and minimum number of sensors as well as their locations for which the complex dynamical network system is completely observable. Our work is closest to the very recent work of Lie et al. [11] but differs in two respect. First, [11] considers a graphical approach (GA) based on structural properties of system matrices to make the system structurally observable, whereas we pursue a generic algorithmic approach to make the system completely observable. Second, the approach in [11] can not be applied to any arbitrary system with structural symmetries, whereas no such limitations hampers our approach and thus our approach can be applied to any arbitrary system. In our paper, we also outline the design of a sparse Kalman filter to illustrate full-state observability of complex thermal networks of real high performance processors and multiprocessor system-on-chips (MPSoC).

The remainder of the paper is organized in the following sections. Section II provides a brief overview of the related works followed by the preliminaries in Section III. The sparse observability problem is defined in Section IV with the description of the properties and complexity of the problem. Algorithms to solve the MSOP are described in the subsequent sub-sections. A specific example use case of thermal sensor placement and run-time in-situ thermal profile estimation and robust hotspot tracking is described in Section V supported by simulation and experimental results.

III. Preliminaries & System Model

A. Dynamic System Model

We consider the discrete-time equivalent of a linear time-invariant system in (1) as follows :

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k \end{aligned} \tag{2}$$

where $\mathbf{x_k} \in \mathbb{R}^{\mathbf{n}}$ are the system states at k^{th} time instant, $\mathbf{u_k} \in \mathbb{R}^{\mathbf{p}}$ are the system inputs , $\mathbf{y_k} \in \mathbb{R}^{\mathbf{m}}$ are the measurements, $\mathbf{A} \in \mathbb{R}^{n\times n}$,$\mathbf{B} \in \mathbb{R}^{n\times p}$, $\mathbf{C} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}$, are system matrices and $\mathbf{x_0}$ is the initial state of the system. Note that the discrete-time system matrices are obtained for the sampling time t_s as [12], [5]:

$$\begin{aligned} \mathbf{A} &= e^{\widetilde{\mathbf{A}}*t_s} \\ \mathbf{B} &= \int_0^{t_s} e^{\widetilde{\mathbf{A}}(t_s-\tau)}\widetilde{\mathbf{B}}d\tau. \end{aligned} \tag{3}$$

B. Observability & Controllability of a System

The system described by equation (2) is said to be controllable if it can be driven from any initial state to any desired final state in finite time, which is possible if and only if the $n \times np$ controllability matrix

$$\mathbf{Q_c} = \left[\mathbf{B}, \mathbf{AB}, \mathbf{A^2B}, ..., \mathbf{A^{n-1}B}\right] = \mathtt{C}(\mathbf{A}, \mathbf{B}) \tag{4}$$

has full rank, that is

$$rank(\mathbf{Q_c}) = n. \tag{5}$$

This represents the mathematical condition for controllability, and is called Kalman's controllability rank condition [9], [12].

Observability, on the other hand, requires us to establish a relationship between the outputs $\mathbf{y_k}$, the state vector $\mathbf{x_k}$, and the inputs $\mathbf{u_k}$ in a manner that we can uniquely infer the system's complete initial state $\mathbf{x_0}$. The linear dynamic system described by equation (2) is said to be observable if we can reconstruct the system's complete internal state from its outputs, which is possible if and only if the $nm \times n$ observability matrix

$$\mathbf{Q_o} = [\mathbf{C^T}, (\mathbf{CA})^{\mathbf{T}}, ..., (\mathbf{CA^{n-1}})^{\mathbf{T}}]^{\mathbf{T}} = \mathcal{O}(\mathbf{A}, \mathbf{C}) \tag{6}$$

has full rank, that is

$$rank(\mathbf{Q_o}) = n. \tag{7}$$

The controllability and observability of a LTI dynamic system is related by the following duality property as described in theorem (1).

Theorem 1. *[9]A linear dynamical system described in* (2) *is observable (controllable) if and only if the dual system*

$$\begin{aligned} \mathbf{x_{k+1}} &= -\mathbf{A^T}\mathbf{x_k} + \mathbf{C^T}\mathbf{u_k} \\ \mathbf{y_k} &= \mathbf{B^T}\mathbf{x_k} \end{aligned} \tag{8}$$

is controllable (observable).

Proof: Substituting the system matrices in (4) produces the observability matrix (6) of the dual system and vice versa. See [9], [12], [3] for the detailed proof.

■

IV. Problem Formulation

We define the minimal sparse observability problem using the above definitions of observability and controllability as follows:

A. Minimal Sparse Observability Problem (MSOP)

For the dynamic system defined by equation (2), the minimal sparse observability problem is defined as the sparsest measurement matrix $\mathbf{C}$, i.e., with smallest number of nonzero entries in $\mathbf{C}$, for which the system described by equation (2) is completely observable. We use the following theorems to show that MSOP is a NP-hard problem.

Theorem 2. *[13]For any $p \geq 1$, finding matrix $\mathbf{B} \in \mathbb{R}^{n\times p}$,with smallest number of nonzero entries in $\mathbf{B}$ such that the system $\mathbf{x_{k+1}} = \mathbf{Ax_k} + \mathbf{Bu_k}$ is controllable is NP-hard.*

Proof: See [13]

■

Theorem 3. *The minimal sparse observability problem defined in* (IV-A) *is NP-hard. In other words, for any $m \geq 1$, finding the matrix $\mathbf{C} \in \mathbb{R}^{m\times n}$ with the smallest number of nonzero entries that will make the system in* (2) *observable is NP-hard.*

Proof: We use the duality theorem in (1) to construct a dual system $(-\mathbf{A^T}, \mathbf{C^T}, \mathbf{B^T})$ as in (8). We then use theorem (2) to prove the NP-hardness of finding the sparsest $\mathbf{C^T}$ matrix that will make the dual system controllable. Since the minimal controllability of the dual system is NP-hard, hence the sparsest $\mathbf{C}$ that will make the original system observable is NP-hard. Hence the minimal sparse observability problem defined in (IV-A) is NP-hard.

■

Theorem 4. *[13]For any $p \geq 1, m \geq 1$ finding matrix $\mathbf{B} \in \mathbb{R}^{n\times p}$,$\mathbf{C} \in \mathbb{R}^{m\times n}$ with smallest number of nonzero entries (in $\mathbf{B}$ and $\mathbf{C}$) such that the system $\mathbf{x_{k+1}} = \mathbf{Ax_k} + \mathbf{Bu_k}$ and $\mathbf{y_k} = \mathbf{Cx_k}$ is both controllable and observable is NP-hard.*

Proof: The proof follows from theorem (2) and (3). For details see [13].

■

B. Greedy Solution to MSOP

Since the minimal controllability problem as described in [13] and the minimal sparse observability problems are NP-hard, polynomial time optimal solutions are unreachable. A randomized and deterministic algorithm was proposed for the minimal controllability problem in [13]. We extend the algorithm in [13] and propose a Greedy algorithm for the minimal sparse observability problem. The algorithm for the minimal sparse observability is presented in Fig. (1) and (2).

Algorithm #1: Minimal Sparse Controllability

Input: System matrix $\mathbf{A}$
Output: Sparse $\mathbf{B}$ such that System (2) is Controllable

---

1) Initialize $\mathbf{B}$ to zero vector and rank difference $e_r^* = 1$
2) While $e_r^* > 1$,
   a) For $i = 1..n$
      i) If $\mathbf{B}[i]$=0, then for $j = 1..2n + 1$, set $\widetilde{\mathbf{B}}[\mathbf{i},\mathbf{j}] = \mathbf{B} + \mathbf{j} * \mathbf{v_i^{\perp}}$ where $\mathbf{v}^{\perp}{}_i$ is the i^{th} basis vector
      ii) $\widetilde{\mathbf{Q}}_\mathbf{c} = \mathtt{C}(\mathbf{A}, \widetilde{\mathbf{B}}[i,j])$; $\mathbf{Q_c} = \mathtt{C}(\mathbf{A}, \mathbf{B})$
      iii) Set $e_r(i,j) = rank(\widetilde{\mathbf{Q}}_\mathbf{c})$-$rank(\mathbf{Q_c})$
      end
   b) Let $(i^*, j^*) \in arg\ max_{(i,j)}\,\{e_r(i,j)\}$ and let $e_r^* = e_r(i^*, j^*)$
   c) if $e_r^* > 0$,set $\mathbf{B} \leftarrow \mathbf{B}$+$j^* * \mathbf{v}_{i*}^{\perp}$
   end
3) Output $\mathbf{B}$

Figure 1. Greedy Minimal Controllability Algorithm.

Algorithm #2: Minimal Sparse Observability

Input: System matrix $\mathbf{A}$
Output: Sparse $\mathbf{C}$ such that system (2) is Observable , Minimum number of Sensor n_s, Sensor Locations

---

1) Compute the system matrices for the dual system
   $\overline{\mathbf{A}} = -\mathbf{A^T}$,
2) Find the sparse control matrix for the dual system using the algorithm (IV-B) in Fig.(1)
   $\overline{\mathbf{B}} =$ minimal_sparse_controllability($\overline{\mathbf{A}}$)
3) Compute the controllability matrix of original system as
   $\mathbf{C} = \overline{\mathbf{B}}^T$
4) Output
   a) Minimum No of Sensor n_s= $rank(\mathbf{C^T})$
   b) Sensor Locations are independent rows of $\mathbf{C^T}$
   c) Measurement matrix $\mathbf{C}$

Figure 2. Greedy Algorithm for Minimal Sparse Observability Problem.

V. Application to Run-time Thermal Estimation and Hotspot Tracking for DTM

A. Thermal Network Model of Multi-core Processor Systems

The thermal behavior of the multi/many core processor is modeled using heat-flow dynamics [16]. The heat-flow dynamics describe the temperature values at different locations on the die depending on various factors such as power consumptions of functional units, layout of the chip and the package characteristics. The differential equations describing the heat flow have a form dual to that of electrical current, represented using lumped values of thermal R and Cs network, and forms the basis for commonly used micro-architectural thermal models [16], [8]. This complex dynamical thermal network is represented in state space form [8], [5] with the grid cell or subsystem block temperatures as states and the power consumption of each blocks as inputs to this system. The outputs of this state space model are the temperatures at the sensor locations which can be observed by the temperature sensor readings. The system matrices $\mathbf{A}$ and $\mathbf{B}$ are constant and are computed based on the floor-plan of the processor and the process parameters [5]. We consider the Alpha 31386 processor and its multi-core architectures that have been extensively used in previous research works [16], [8], [5]. A quad-core Alpha 31386 processor floorplan is shown in Fig. 3 where each processor core is having 18 functional blocks/ subsystem units. Fig. 4 shows the blocks in different layers of the chip and package. Note that the block in the top most layer i.e., the die only consumes power and the blocks in the thermal interface layer, heat spreader, and the heat sinks help in dissipating the heat generated in the die. The equivalent RC network representation of the processor's thermal dynamic system is shown in Fig. 4(b).

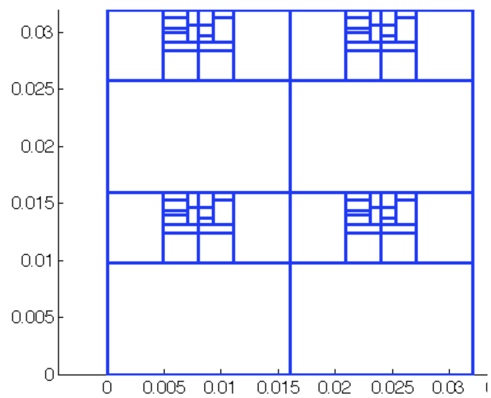
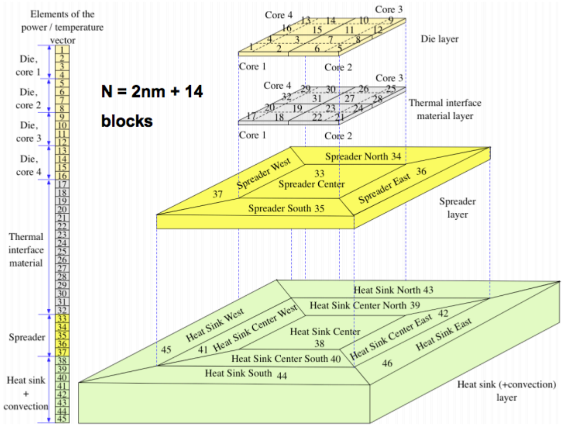


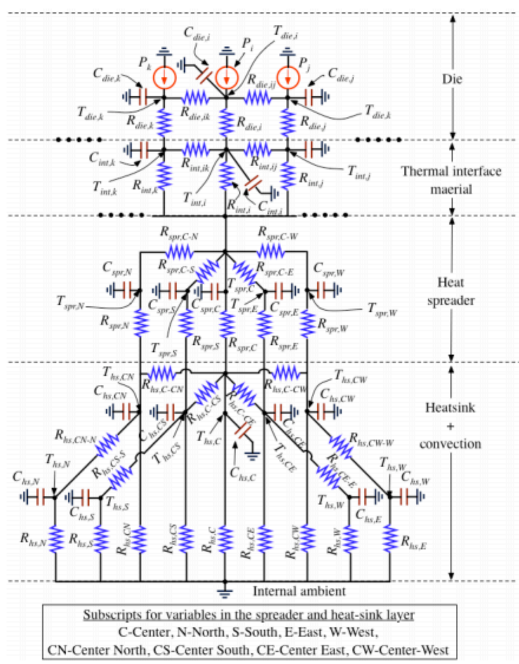
Figure 3. Floor plan of quad-core processor based on Alpha 31386.

B. Minimum Sensor Set and Their Optimal Placement

With increasing number of cores in the processor and projection of hundreds and even thousands of cores [1], the thermal dynamics of such a processor can be extremely complex with more than thousands of blocks or functional units. Consequently, it is prudent to consider them as large complex dynamic networks, requiring systematic analysis. We use our mathematical tools and algorithms developed in section IV to perform sparse observability analysis on these networks. We consider the Alpha 31386 processor and its multi-core architectures as discussed in the previous section. Our objective is to determine if the thermal network of the Alpha processor is completely observable and find the sparsest measurement matrix $\mathbf{C}$ such that the processor network is completely observable. We use the greedy algorithm in Fig. 2 to find the sparsest measurement matrix $\mathbf{C}$. The algorithm returns that the single-core Alpha 31386 processor thermal network is completely observable with a single sensor. Consequently, we would be able to estimate the temperature of the 18 subsystem units of the processor (resulting in 2x18+14=50 nodes on the network, see Fig. 4(a)) and autonomously track them from a single sensor measurement. The placement of the sensor returned by the algorithm is the first block in the processor floorplan with a sensor gain of 0.272. Note that the complexity of the algorithm in Fig. 2 is determined by the rank computation of the controllability



(a)



(b)

Figure 4. Thermal network representation of high performance processors. (a) blocks in the different layers of the chip and their corresponding nodes in the thermal networks (b) RC equivalent circuit representation of the thermal dynamic network. [Figures taken from presentation of [16] and [6]]

matrix $\mathbf{Q_c}$ which is $\mathcal{O}(n^3)$. The rank computation uses singular value decomposition (SVD) which has the computational complexity of $\mathcal{O}(n^3)$. To further validate and verify the the completely observability of the system, we construct a Kalman filter using this single sensor observation as well as placement and track the peak temperature (hotspot) of the processor. We compare the results of the thermal hotspot tracking of the alpha processor with that of one of the state-of-the-art thermal and hotspot tracking method [14].

C. Sparse Kalman Filter (SKF)

To construct a full-state observer, we use the well known Kalman filtering approach. The thermal dynamics of the processor is modeled using the discrete linear state-space system [8], [5] in presence of variability induced process noise [2] as:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k \tag{9}$$

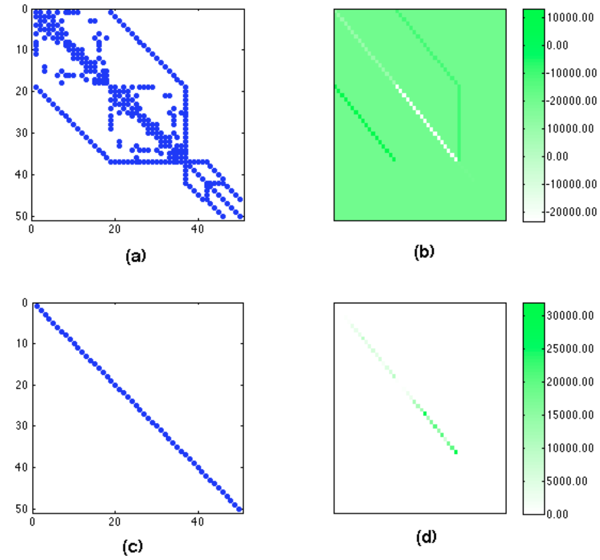


Figure 5. System matrices of the micro-architectural thermal model of Alpha 31386 processor. (a) sparsity pattern of the the system coupling matrix $\mathbf{A}$ (size 50×50) in continuous domain (b) representation of the coupling matrix $\mathbf{A}$ (size 50×50) in the discrete domain (c) sparsity pattern of the control matrix $\mathbf{B}$ (size 50×50) in continuous domain (d) the $\mathbf{B}$ matrix (size 50×50) in discrete domain. The matrices are obtained from the Hotspot thermal simulator [8] for the Alpha 31386 processor floorplan.

where $\mathbf{x_k} \in \mathbb{R}^\mathbf{n}$ are the system states (i.e, the temperatures of each block), $\mathbf{u_k} \in \mathbb{R}^\mathbf{P}$ are the system inputs (i.e, power consumption at each block), $\mathbf{A}$,$\mathbf{B}$ are system matrices as discussed earlier. The process noise $\mathbf{w}_k$ is zero-mean, white random signals with known covariance matrices, $\mathbf{Q_k} = \mathbf{E}\left[\mathbf{w_k}\mathbf{w_k^T}\right]$. Our objective is to select minimal number of sensors and their placement such that we have minimum number of sensors in the measurement equation:

$$\breve{\mathbf{y}}_k = \breve{\mathbf{C}}\mathbf{x}_k + \breve{\eta}_k \tag{10}$$

where $\breve{\mathbf{y}}_\mathbf{k} \in \mathbb{R}^\mathbf{m}$ are the measurement (i.e, the temperature sensor measurement), $\breve{\mathbf{C}}$ is the minimum sensor measurement matrix, and measurement noise $\breve{\eta}_\mathbf{k}$ is zero-mean, white random signals with known covariance matrices $\breve{\mathbf{R}}_\mathbf{k} = \mathbf{E}\left[\breve{\eta}_\mathbf{k}\breve{\eta}_\mathbf{k}^\mathbf{T}\right]$. The process noise $\mathbf{w}_k$ and measurement noise $\breve{\eta}_\mathbf{k}$ are assumed to be mutually uncorrelated.

The Kalman filter maintains the states $\hat{\mathbf{x}}_{\mathbf{k}|\mathbf{k}}$ which means the estimate of $\mathbf{x_k}$ given the measurement $\mathbf{y_k}, \mathbf{y_{k-1}}, \dots$, and the error covariance of the states, $\mathbf{P_{k|k}}$, is the covariance of the states $\mathbf{x_k}$ given the measurement $\mathbf{y_k}, \mathbf{y_{k-1}}, \dots \mathbf{y_0}$. Kalman filter performs the following recursive processing as in Fig. 6 to estimate the states from the measurement for the given input.

D. Run-time Thermal Profile Reconstruction & Hotspot Tracking

Temperature adversely affects power and reliability of processor systems. For safe and reliable operation, the peak temperature of the processor has to be be always maintained or controlled below a safe threshold. However, with the change in workloads and phasic behavior of workloads, the power consumption of each block vastly vary. As temperature sensors along with their peripheral circuits introduce substantial overhead in silicon area and power consumption, it is extremely important to minimize the number of temperature sensors without surrendering the accuracy of thermal monitoring. On one-hand large number of temperature sensors are needed for accurate thermal monitoring in presence of large dynamic power variations, on the other hand they incur substantial die area real

Algorithm #3: Sparse Kalman Filter

Input: Measurements $\breve{\mathbf{y}}_\mathbf{k}$, input $\mathbf{u}_\mathbf{k}$, state space model $\mathbf{A}, \mathbf{B}, \breve{\mathbf{C}}, \mathbf{Q}_\mathbf{k}, \breve{\mathbf{R}}_\mathbf{k}$, Initial Values $\mathbf{x}_\mathbf{0}, \mathbf{P}_\mathbf{0}$
Output: State estimates $\hat{\mathbf{x}}_\mathbf{k}$, measurement estimate $\hat{\mathbf{y}}_\mathbf{k}$

---

1) Initialize the values of $\hat{\mathbf{x}}_{\mathbf{0}|-\mathbf{1}} = \mathbf{x}_\mathbf{0}$,$\hat{\mathbf{P}}_{\mathbf{0}|-\mathbf{1}} = \mathbf{P}_\mathbf{0}$,
2) Perform the following every sampling step :
   a) Predict states and states covariance:
      i) $\hat{\mathbf{x}}_{\mathbf{k}|\mathbf{k}-\mathbf{1}} = \mathbf{A}\hat{\mathbf{x}}_{\mathbf{k}|\mathbf{k}-\mathbf{1}} + \mathbf{B}\mathbf{u}_\mathbf{k}$
      ii) $\mathbf{P}_{\mathbf{k}|\mathbf{k}-\mathbf{1}} = \mathbf{A}\mathbf{P}_{\mathbf{k}-\mathbf{1}|\mathbf{k}-\mathbf{1}}\mathbf{A}^\mathbf{T} + \mathbf{Q}_{\mathbf{k}-\mathbf{1}}$
   b) Calculate the Kalman gain, update the states and states covariance:
      i) $\mathbf{K}_\mathbf{k} = \mathbf{P}_{\mathbf{k}|\mathbf{k}-\mathbf{1}}\breve{\mathbf{C}}^\mathbf{T}\left(\breve{\mathbf{C}}\mathbf{P}_{\mathbf{k}|\mathbf{k}-\mathbf{1}}\breve{\mathbf{C}}^\mathbf{T} + \mathbf{R}\right)^{-1}$
      ii) $\hat{\mathbf{x}}_{\mathbf{k}|\mathbf{k}} = \hat{\mathbf{x}}_{\mathbf{k}|\mathbf{k}-\mathbf{1}} + \mathbf{K}_\mathbf{k}\left(\breve{\mathbf{y}}_\mathbf{k} - \breve{\mathbf{C}}\mathbf{x}_{\mathbf{k}|\mathbf{k}-\mathbf{1}}\right)$
      iii) $\mathbf{P}_{\mathbf{k}|\mathbf{k}} = (\mathbf{I} - \mathbf{K}_\mathbf{k}\breve{\mathbf{C}})\mathbf{P}_{\mathbf{k}|\mathbf{k}-\mathbf{1}}$
3) Output the results:
   a) State estimate: $\hat{\mathbf{x}}_\mathbf{k} = \hat{\mathbf{x}}_{\mathbf{k}|\mathbf{k}}$
   b) Measurement estimate: $\hat{\mathbf{y}}_\mathbf{k} = \breve{\mathbf{C}}\hat{\mathbf{x}}_{\mathbf{k}|\mathbf{k}}$

Figure 6. Minimum sensor state estimation using Kalman filter.

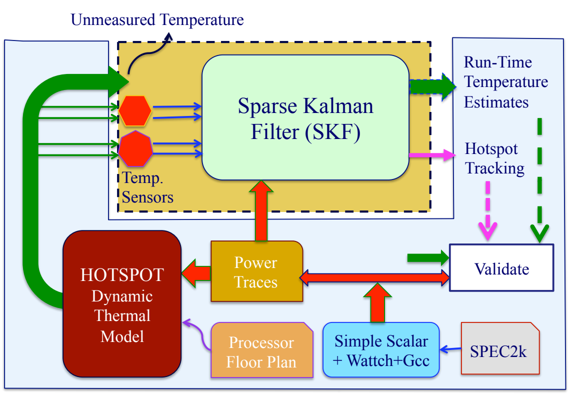


Figure 7. Simulation and validation framework for run-time thermal estimation using Sparse Kalman Filter (SKF). The SKF estimates the full chip thermal profile using minimum number of sensors while filtering the effect of sensor, measurement, and process noise.

estate and power consumptions overhead. Given this scenario, the minimal sparse observability problem directly addresses this trade-off by providing the minimum number of sensors to accurately observe the thermal dynamics.

In order to experimentally verify the results, we have created a simulation framework as shown in Fig. 7. Fig.8 shows the thermal profile estimation of the complete processor using a sparse Kalman filter presented in Fig. 6 with just one sensor. The thermal network for the single-core Alpha processor is observable using a single sensor and the thermal profile estimated by the SKF is very accurate. Even in presence of sensor noise, the SKF provide the best statistically possible estimation and tracking of the thermal profile. The approach presented can easily be applied to multi-core configuration; the number of sensors and their location are obtained using the algorithm in Fig.2. Fig.9 shows improved tracking of the hotspot in the Alpha processor in comparison to one of the state-of-the-art method [14] for all the SPEC 2000 benchmarks. As should be evident from Fig. 9, the minimum sensor SKF provides superior hotspot tracking that is paramount for reliable operations of processors.

In addition, the SKF can also alleviate the variability induced process noise in deep sub-micron technologies as well as suppress sensor noise by suitably filtering the noise during the estimation steps. Fig. 10 shows the filtering of the sensor noise from the measurements during the estimation procedure. Such an approach could also be beneficial in the closed loop dynamic thermal management and performance improvement of the processor as illustrated by in Fig. 11. During the dynamic thermal management operation, the processor throttles its frequency (i.e., reduces the frequency of operation and hence power by certain percentage, e.g., 20 %) whenever the peak temperature crosses a specified threshold. In the absence of a accurate hotspot tracker, the peak thresholds are violated; resulting in deep reliability issues and imminent damage to the processor. On the other hand, the noise in the sensor can spuriously early trigger DTM operation causing unwanted performance loss. Fig. 11 illustrates the scenario where the SKF enabled DTM filter-out the noise and spurious triggers, resulting in performance improvement of approx. 28 % in this specific example scenario.

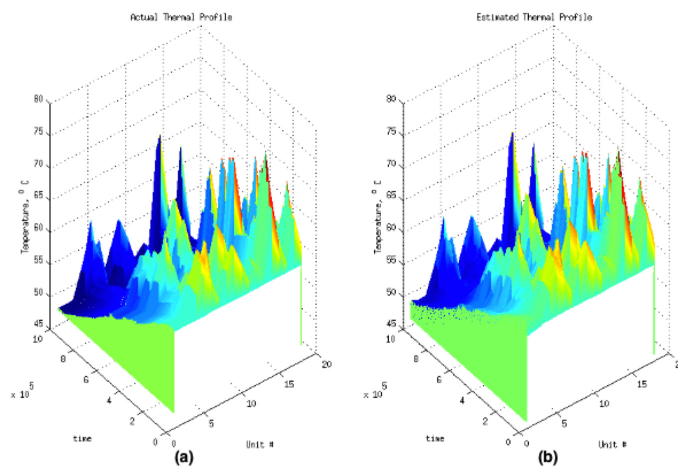


Figure 8. Thermal profile estimation (a) actual thermal profile (b) estimated by SKF using a single sensor for all the SPEC 2000 benchmarks. The estimated error is with in 0.3% for all the blocks.

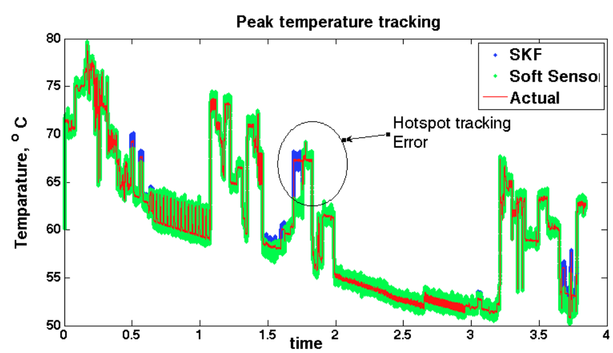


Figure 9. Robust hotspot tracking of the Alpha processor using SKF in comparison to the state-of-the-art hotspot tracking approach in [14].

E. Overheads and Complexity

The algorithmic solution to the MSOP has the computational complexity determined by the SVD computations. The complexity of finding the sparsest measurement matrix through the SVD based rank computations result in $\mathcal{O}(n^3)$. On the other hand, the complexity of the SKF that uses the sparse measurement matrix, is also $\mathcal{O}(n^3)$. This is determined by the matrix inverse involved in the computation of the Kalman gain $\mathbf{K}_\mathbf{k}$. In a scenario, where the time scale at which

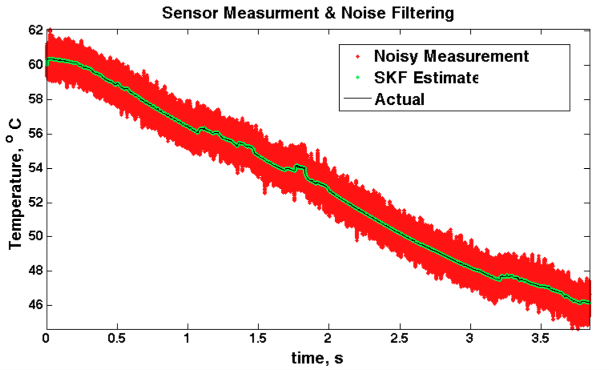
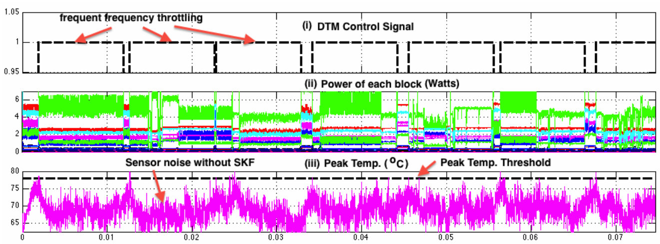
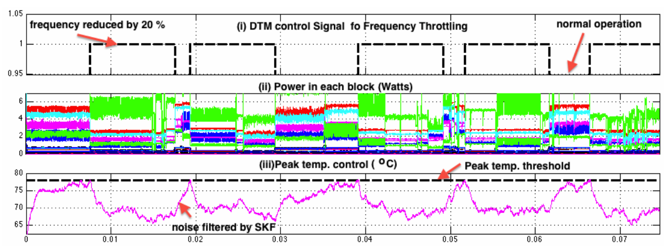


Figure 10. Sensor measurement (with noise variance of $\pm 1^o C$) and noise filtering with SKF. Effect of both measurement noise and variability induced process noise can be mitigated by the SKF to achieve statistically superior estimates.



(a)



(b)

Figure 11. DTM control of the Alpha processor with minimal sensor placement (a) DTM with direct noisy sensor reading (b) with SKF. Because of the noise filtering by the SKF, less frequent frequency throttling is initiated in the DTM, which improves performance of the processor system (approx. by 28%).

the noise characteristics change is much larger than the time scale at the thermal networks are studies (e.g., month or even years), the system and the noise covariance of the Kalman filter can be assumed to be time-invariant [15]. As a result, we can use a steady state sparse Kalman filter for which it is not necessary to compute the estimation error covariance or the Kalman gains in real-time [15], rather the gains can be replaced by constant gains. This reduces the computational overhead of the real-time thermal estimation using the SKF from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ while still providing good accuracy. Although, a calibration step may be needed prior to running the SKF with steady state gains, this can be achieved by computing the Kalman gains initially from the SKF and then switching to steady state SKF with constant gains.

VI. Conclusions

One of the most challenging problem in modern network science and engineering is the controlling and observing of complex dynamical networks. Observability is fundamental to having deeper insight in any complex dynamic networks. It is paramount in the understanding of the interplay between the complex network topology and the underlying dynamic behavior. In this paper, we explore this fundamental question of observing the internal dynamics of a complex dynamic network systems using minimal set of observation points by using the notion of sparsity. We define the minimal sparse observability problem (MSOP) to find the sparsest measurement vector and prove that the problem is NP-hard. Our main result is the development of an algorithm to find the sparsest measurement matrix that will make any arbitrary linear dynamical network completely observable. We develop effective greedy algorithms to find the sparsest measurement and use it in the design of a sparse Kalman filter (SKF) to estimate the time-dependent internal states of complex dynamical networks. We apply the approach to complex thermal networks of real processor systems and demonstrate the applicability in run-time thermal profile estimation and hotspot tracking using minimal number of on-chip sensors for effective dynamic thermal management of such processors.

Acknowledgment

This work was partially supported by the NSF Variability Expedition award CCF-1029783.

References

[1] S. Borkar. Thousand core chips-a technology perspective. In *Design Automation Conference, 2007. DAC '07. 44th ACM/IEEE*, pages 746 –749, june 2007.

[2] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De. Parameter variations and impact on circuits and microarchitecture. In *Design Automation Conference, 2003. Proceedings*, pages 338 – 342, june 2003.

[3] Stephen L Campbell, Nancy K Nichols, and William J Terrell. Duality, observability, and controllability for linear time-varying descriptor systems. *Circuits, Systems and Signal Processing*, 10(4):455–470, 1991.

[4] D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[5] Yongkui Han, Israel Koren, and C Mani Krishna. TILTS: A fast architectural-level transient thermal simulation method. *Journal of Low Power Electronics*, 3(1):13–21, 2007.

[6] Vinay Hanumaiah, Sarma Vrudhula, and Karam S Chatha. Performance optimal online dvfs and task migration techniques for thermally constrained multi-core processors. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 30(11):1677–1690, 2011.

[7] John E. Hopcroft and Richard M. Karp. A n5/2 algorithm for maximum matchings in bipartite. In *Switching and Automata Theory, 1971., 12th Annual Symposium on*, pages 122–125, 1971.

[8] Wei Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M.R. Stan. Hotspot: a compact thermal modeling methodology for early-stage vlsi design. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 14(5):501 –513, may 2006.

[9] Rudolf Emil Kalman. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial & Applied Mathematics, Series A: Control*, 1(2):152–192, 1963.

[10] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.

[11] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Observability of complex systems. *Proceedings of the National Academy of Sciences*, 110(7):2460–2465, 2013.

[12] D. G. Luenberger. *Introduction to Dynamic Systems: Theory, Models, and Applications*. Wiley, 1979.

[13] A. Olshevsky. The Minimal Controllability Problem. *ArXiv e-prints*, April 2013.

[14] Sherief Reda, Ryan J Cochran, and Abdullah Nazma Nowroz. Improved thermal tracking for processors using hard and soft sensor allocation techniques. *Computers, IEEE Transactions on*, 60(6):841–851, 2011.

[15] Dan Simon. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. Wiley. com, 2006.

[16] Kevin Skadron, Mircea R. Stan, Karthik Sankaranarayanan, Wei Huang, Sivakumar Velusamy, and David Tarjan. Temperature-aware microarchitecture: Modeling and implementation. *ACM Trans. Archit. Code Optim.*, 1(1):94–125, March 2004.

[17] Lenka Zdeborová and Marc Mézard. The number of matchings in random graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(05):P05003, 2006.