2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Errors-In-Variables Identification of Thermal Models for Many-core Computing Systems

Roberto Diversi, Andrea Tilli, Andrea Bartolini, Luca Benini

Abstract— **A crucial issue in next generation many-core computing systems is the dynamic thermal control for run-time performance optimization. To this end, scalable, compact and effective thermal models are essential. In this paper, we present an algorithm based on the Frisch scheme to obtain identification of distributed core-centric interacting models, dealing with very noisy temperature measurements and high process noises. We apply the proposed solution to an Intel's Single-chip-Cloud-Computer (SCC), a many-core prototype with 48 cores.**

I. INTRODUCTION

Operating temperature has always been a critical aspect of the electronic devices. Absolute temperature not only impacts the performance of analog devices (i.e. noise and biasing), but it reduces the reliability of digital devices as it accelerates silicon aging, it reduces the internal timing margins and it increases the leakage current and power. In the recent past the thermal design of microprocessors was tackled statically. For a given architecture it was always possible to design a standard heat dissipation system (heatsink and fan) without limiting the potential computing performance. Indeed, the speed of the transistors were constrained by the minimum channel length and not by their power density. In contrast, today processors speeds are heavily constrained by the power density of the transistors and induced thermal issues [1].

According to the famous Moore's law [2], transistor density was expected to double every two years, along with a 40% increasing of the clock frequency, while keeping an almost constant silicon power density (i.e. without facing power and thermal problem). This rule-of-thumb for predicting microprocessor evolution has worked fine for more than 40 years. Unfortunately, in the last years, while the transistor scaling is still in line with the Moore's law, voltage supply cannot be scaled down any more stopping the related quadratic power reduction. As a consequence, clock frequency increase has been stopped too to bound the power consumption [3]. This has led microprocessor manufacturers to multicore and, recently, many-core architectures to obtain a larger throughput without increasing the frequency[1]. Nevertheless, even with such solutions, it is not possible leaving the overall power density almost unchanged as before, because the transistor density is anyway increasing quadratically, while their specific power consumption is reducing linearly [4]. As a matter of fact, if standard heatsinks and fans were used for heat dissipation, next generation many-core processors could not have all of the cores running at maximum frequency constantly without burning the chip. This effect is called *dark silicon* and is already visible in current general purpose multicore, where a significant silicon area is used to implement less power hungry units (i.e. cache memory) instead of processing units [5], [6].

In order to deal with such "thermal crisis", advanced active cooling systems, as micro Peltier cells, are under investigation [7]. But this kind of units seems far away to be reliable and sustainable from an energy consumption viewpoint (in particular for mobile applications) [8]. Therefore, nowadays and in the next future, the only practical solution is staying on standard heatsink and fan. But this would lead to a relevant computing performance downgrade, if just "old-style" passive and static thermal management is adopted. Indeed, with such kind of approach, computing performance should be a priori and statically capped, according to heatsink and fan limits, in order to guarantee thermal safety in every scenario, i.e. the maximum performance should be sized on the worst-case and then adopted in any working condition.

Taking the cue from these considerations, Dynamic Thermal Management (DTM) solutions have been recently adopted [9]. In [10] a strategy to speed-up the clock of the CPUs when only few of them are active is implemented. More advanced solutions [11], [12] ensure, at run-time, a safe working temperature by limiting the processor performance only under critical workload phases and environmental conditions. Knobs, available for performance modulation, are Dynamic Voltage Scaling (DVFS) [6] and core shutdown. The latter should be used carefully since it is rather abrupt, while the former enables an almost continuous modulation. In last years, Model-Predictive Controllers (MPC) have been proposed to exploit DVFS [11]–[13] for effective computing performance optimization, under thermal-capping constraints. This is a very promising approach, but it has been noted that MPC run-time overhead rapidly increases with the complexity of the chip thermal model and its control performance strongly depends on the model accuracy.

Unfortunately, chip manufacturers do not release dynamic thermal models of their chips and, most important, the thermal behavior of a chip strongly depends on its environment, namely the package, the heatsink, the fan, and the ambient temperature. Techniques for extracting and/or calibrating the thermal model of a die in its deployment environment are therefore required for MPC to be applicable in a real-life

The authors are with the Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy {roberto.diversi, andrea.tilli, andrea.bartolini, luca.benini}@unibo.it

[1]Actually, this has required also the development of parallel programming paradigms and tools, to obtain significant throughput increase. This is not a trivial issue, since many applications are not easy to be split and parallelized on different cores.

setting. In addition, particular attention has to be paid to the complexity of the identified model. Indeed, low-complexity models are preferred to reduce the computational burden of MPC policies.

The Single-Chip Cloud Computer (SCC) experimental processor [14] is a 48-core "concept vehicle" created by Intel Labs as a platform for many-core software research. It integrates hardware monitors and thermal sensors to track the chip workload phases and thermal behavior. Unfortunately, the built-in thermal sensors output is affected by significant noise. As a consequence, SCC is a challenging testbench for thermal model learning strategies. Moreover, its high number of cores set requirements on the complexity and scalability of the proposed algorithm.

Recently, a number of strategies for extracting compact thermal models directly from the core thermal response to a given power/workload stress input have been proposed [11], [12], [15]–[17]. The simplest ones are centralized and do not account for the multimodal nature of the thermal transients caused by the different building materials and their relative time-constants (i.e. die, heat-spreader and heatsink) [12]. In [11], [15], a first order dynamic thermal model is identified by solving a linear least-squares optimization problem. Moreover, the inter-core thermal interaction for the multi-core devices is accounted by solving a unique global least squares problem for all the cores at once. Sharifi and Rosing [18] show that when this model is available it can be used effectively to filter out measurement noise using a Kalman filter. These centralized model learning methodologies have been shown to poorly scale with the increase of the number of cores, leading to complex control solutions [12].

To address these limitations, distributed model learning approaches based on autoregressive structures have been recently proposed. Coskun et al. [16] use an autoregressive moving average (ARMA) technique for predicting the future thermal evolution for each core. This model is capable of predicting future temperature only based on its previous values. Since they do not account directly for workload-to-power dependency a SPRT (Sequential Probability Ratio Test) technique is then used to early detect changes in the statistical residual distribution (average, variance) and than to re-train the model when it is no longer accurate. Juan et al. [17] use a combination of a K-means clustering and an AR model to identify a compact model for fast thermal simulation. This approach is effective only when starting from an highly accurate thermal model of the HW. Moreover, missing exogenous terms in both these approaches neglects the direct link between power dissipated and temperature requiring the identification of a set of different models for capture the different functional units and program phases characteristics. Bartolini et al. [12] present a distributed model learning approach based on a set of ARX models. Each core executes its own model identification routine generating a local thermal model. The model is used internally, in each core, by the local model-predictive controller. Even if this methodology is promising, thanks to its run-time low overhead and scalability, it has been applied only on a simulator and is based on the assumption that per-core power traces and thermal sensor outputs are accurate and without noise. Indeed, standard ARX models are suitable to represent process noise, but are based on the assumption that input and output data are accurate and not affected by measurement noise [19], [20].

All the above methods have never been applied to real chips, with strongly non-ideal thermal sensors, working in a real chassis with complex cooling. To achieve that, we propose to represent the chip as a set of *MISO ARXs with noisy output* to cope with both the process noise and the output measurement noise. With this setting, the related identification problem can be mapped into an Errors-In-Variables (EIV) framework [21], and we propose to solve it by a an ad-hoc algorithm based on the Frisch scheme [22], which has a good tradeoff between estimation accuracy and computational complexity. The rationale behind the Frisch scheme consists in searching for the solution of the identification problem within a locus of solutions which are compatible with the second order statistics of the noisy data [22]. To select a single solution among all compatible ones a suitable criterion must be introduced [22]–[25]. The selection criterion considered in this paper consists in computing the equation error of the ARX with noisy output model and in comparing its sampled statistical properties with the expected theoretical ones. We show the effectiveness of the proposed solution by applying it to the identification of the thermal model of an actual SCC chip, by using experimental data.

The paper is organized as follows. In Section II the adopted SCC platform is described to highlight the available measurements, the signals adopted for identification and information on its floor-plan to be exploited for distributed modelling. In Section III the structure of the distributed models adopted to capture the chip thermal behavior is presented and the identification problem is defined. In Section IV the proposed identification algorithm is presented and in Section V the results obtained applying it to SCC experimental data are reported. Some concluding remarks are given in Section VI.

II. SCC ARCHITECTURE

The SCC processor has 24 dual-core tiles arranged in a 6x4 mesh. Each core is a P54C core. Each tile integrates two thermal sensors based on a couple of ring oscillators, one positioned in proximity of the router and the other positioned close to the top core L1 cache. These thermal sensors are originally uncalibrated. We used the calibration procedure presented in [26] to obtain a meaningful temperature for each sensor. Calibrated thermal sensors outputs show the presence of significant white noise [26]. Each P54C core has two performance counters. These counters can be programmed to track various architectural events (such as number of instructions or cache misses) at periodic intervals. Performance counters can be accessed from the specific core they are located at by reading the dedicated registers. The Board Memory Controller (BMC) includes a

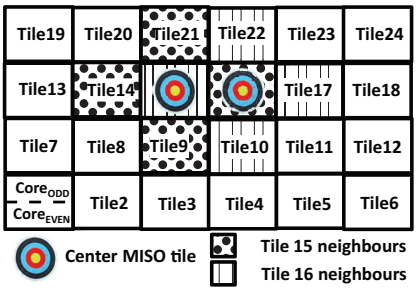

Fig. 1. SCC topology

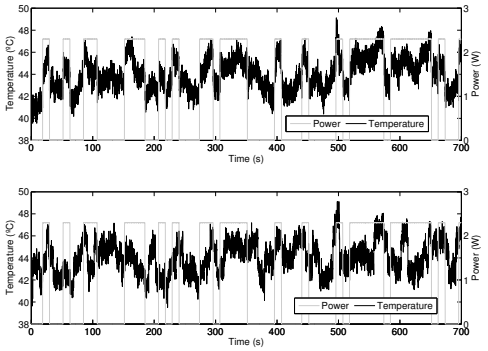

Fig. 2. Examples of power and temperature traces of tile 15 and tile 16

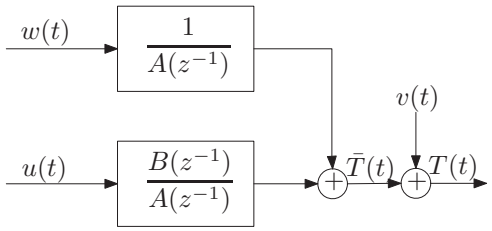

Fig. 3. ARX model with additive output noise.

power sensor capable of measuring the full SCC chip power consumption and an ambient temperature sensor. A per-core power estimation for SCC can be obtained by using a power model. This has been obtained by correlating the full-chip power measurement with each core activity and operating point (i.e. frequency) measured through the HW performance counters, as presented in [27].

Given these HW features, we used the SCC to generate a set of traces suitable for the thermal model identification. We designed a set of bash scripts that use POSIX signals to start and stop synchronously a given workload/power virus[2] on the different SCC cores while at the same time collecting the HW monitors outputs (i.e. performance counters and thermal sensors). As a consequence of that, the framework can apply a given Pseudo-Random Binary Sequence (PRBS) stress workload to the SCC cores. The performance counters outputs are then transformed through the power model in a set of per-tile power traces and used as input vector for the model identification problem. The output vector is instead composed by the calibrated thermal sensor output for each tile[3]. Fig. 1 shows the SCC tile topology and some other details exploited later on. Fig. 2 shows a sample of power and temperature traces for two specific tiles, namely 15 and 16. According to [26], a relevant white noise can be noted on the temperature readings given by the adopted sensors. In [28], it has been shown that, because of the presence of a significant additive noise, standard ARX models are not suitable to represent the SCC thermal dynamics.

III. DISTRIBUTED MODELLING AND IDENTIFICATION OBJECTIVE

According to the purpose of defining a distributed modelling, in this Section we propose to represent the SCC device by means of a collection of interacting ARX models *with noisy output* (referred as *noisy ARX* for the sake of brevity), characterized by the following features.

– For each tile[4] a MISO noisy ARX model is considered, as illustrated in Fig. 3; where $\bar{T}(t)$ is the actual tile temperature, $T(t)$ is the available noisy measurement, $w(t)$ is the process noise, $v(t)$ is the measurement noise. The input $u(t)$ is a vector whose entries are the dissipated power $P(t)$ related to tile activity and the temperatures $T_{n1}(t), T_{n2}(t), \ldots T_{nq}(t)$ probed on the tiles belonging to the tile neighborhood.

– The neighborhood of each tile is defined as the set of tiles sharing an edge with the considered one as in Fig.1. Then, the number of neighbors q can range from 2 to 4.

– It is worth noting that noisy inputs should be considered as well as noisy output, since the neighbor temperature measurements are adopted as inputs in the model of each tile. This would lead to a full Errors-In-Variables (EIV) framework [25]. Nevertheless, since the inputs effects are filtered by the system thermal dynamics, it looks reasonable to assume a very low effect of the noise corrupting the temperature readings (i.e. by assuming the noise is entering in the system as well as the true signal, the changes in the output are expected to be negligible). The validity of such hypothesis will be checked later on the actual data by suitably assessing the estimation results.

– Tests will be carried out under constant ambient temperature (T_{amb}) conditions so that all the temperatures considered in the MISO ARX models are actually temperature gap w.r.t the ambient. Therefore, the identified models will be effective in predicting the difference between the tile temperatures and the ambient one, under constant (or slowly-varying) ambient temperature scenario (the most common one). Anyway, if fast[5] variations in T_{amb} can be experienced, the presented identification procedure can be adopted as well, by adding T_{amb} in the input set of the MISO noisy ARX models and using the absolute temperatures instead of the temperature

[2] *cpuburn* power virus by Robert Redelmeier: it takes advantage of the internal architecture to maximize the CPU power consumption

[3] In this paper we consider only the thermal sensor positioned close to the router since it is more central within the tile area.

[4] without loss of generality we consider tiles instead of cores

[5] *fast* w.r.t. the chip thermal dynamics, i.e. the thermal bandwidth

gaps. In this case, a variable T_{amb} should be used in identification tests to obtain good results, according to persistency of excitation requirements [19], [20].

According to the above items, the resulting noisy ARX model of each tile is

$$A(z^{-1})\,\bar{T}(t) = B(z^{-1})\,u(t) + w(t) \tag{1}$$

$$T(t) = \bar{T}(t) + v(t). \tag{2}$$

The input $u(t)$ is the r–dimensional signal, with $r = q + 1$,

$$\begin{aligned} u(t) &= [\,P(t)\;T_{n1}(t)\;\ldots\;T_{nq}(t)\,]^T \\ &\triangleq [\,u_1(t)\;u_2(t)\;\ldots\;u_r(t)\,]^T, \end{aligned} \tag{3}$$

and

$$\begin{aligned} A(z^{-1}) &= 1 + a_1\,z^{-1} + \cdots + a_n\,z^{-n} \\ B(z^{-1}) &= \begin{bmatrix} B_1(z^{-1}) & B_2(z^{-1}) & \cdots & B_r(z^{-1}) \end{bmatrix} \\ B_i(z^{-1}) &= b_{i1}\,z^{-1} + \cdots + b_{in}\,z^{-n}, \quad i = 1, \ldots, r \end{aligned} \tag{4}$$

where $A(z^{-1})$ and $B_i(z^{-1})$ are polynomials in the backward shift operator z^{-1} (i.e $z^{-1}\,x(t) = x(t-1)$).

With the above model (1)-(4) at hand, the following standard assumptions (clearly reasonable for the case under investigation) and the identification problem can be set.

A1. The dynamic system (1) is asymptotically stable, i.e. all zeros of $A(z^{-1})$ are inside the unit circle.

A2. The input power $P(t)$ as well as the the neighbor temperatures $T_{n1}(t)\;\ldots\;T_{nq}(t)$ are persistently exciting of sufficiently high order.

A3. The driving noise $w(t)$ and the sensor noise $v(t)$ are zero–mean ergodic white processes with unknown variances σ_w^{2*}, σ_v^{*2}.

A4. $w(t)$ and $v(t)$ are mutually uncorrelated and uncorrelated with $u(t)$.

Identification Problem. Estimate the model order, n, the coefficients of $A(z^{-1})$, $B_i(z^{-1})$, $(i = 1\ldots, r)$ and the noise variances σ_w^{2*}, σ_v^{2*}, starting from N input–output samples $u(1), \ldots, u(N), T(1), \ldots, T(N)$.

Remark 1: It is worth noting that the process and output noises can be managed by using ARMAX or Box-Jenkins models, as well. Then, standard identification techniques could be applied. Nevertheless, in that setting, process and measurement noises cannot be distinguished, while, with the proposed approach, properties of the process and measurement noises can be obtained. This would be very helpful to design Kalman observers for filtering and prediction purposes.

IV. THE PROPOSED IDENTIFICATION METHOD FOR ARX WITH NOISY OUTPUT MODELS

The proposed identification procedure is based on two main parts. In the first, the most important one, the polynomial coefficients and the noise variances are estimated by assuming a given model order, n, and exploiting a Frisch scheme–like approach. In the second, a procedure is defined to use the first part with different values of the model order n aiming to determine the correct one, along with the polynomial coefficient and the noise variances, to solve the *Identification Problem.*

A. Frisch scheme–like identification with fixed model order

In this Subsection the actual temperature $\bar{T}(t)$ and the measured temperature $T(t)$ will be denoted as $\bar{y}(t)$ and $y(t)$ respectively since they represent the actual and measured outputs. Let us introduce the vectors

$$\begin{aligned} \bar{\varphi}(t) = [\,&-\bar{y}(t)\;\ldots\;-\bar{y}(t-n)\,u_1(t-1)\;\ldots\;u_1(t-n) \\ &u_2(t-1)\;\ldots\;u_2(t-n)\;\ldots\;u_r(t-1)\;\ldots\;u_r(t-n)\,]^T \end{aligned} \tag{5}$$

$$\varphi(t) = [\,-y(t)\;\ldots\;-y(t-n)\,u_1(t-1)\;\ldots\;u_r(t-n)\,]^T \tag{6}$$

$$\tilde{\varphi}(t) = [\,-v(t)\;\ldots\;-v(t-n)\;\underbrace{0\;\ldots\;0}_{rn}\,]^T \tag{7}$$

$$\varphi_w(t) = [\,w(t)\;\underbrace{0\;\ldots\;0}_{(r+1)n}\,]^T \tag{8}$$

and the parameter vector

$$\begin{aligned} \bar{\theta}^* &= \begin{bmatrix} 1\,a_1\,\cdots\,a_n\;b_{11}\,\cdots\,b_{1n}\,\cdots\;b_{r1}\,\cdots\,b_{rn} \end{bmatrix}^T \\ &\triangleq \begin{bmatrix} 1\;\theta^{*T} \end{bmatrix}^T. \end{aligned} \tag{9}$$

From (1) and (2) it follows that

$$\left(\bar{\varphi}^T(t) + \varphi_w^T(t)\right)\,\bar{\theta}^* = 0 \tag{10}$$

$$\varphi(t) = \bar{\varphi}(t) + \tilde{\varphi}(t). \tag{11}$$

Define the covariance matrices

$$\Sigma = E\,[\varphi(t)\,\varphi^T(t)] \tag{12}$$

$$\bar{\Sigma} = E\left[\left(\bar{\varphi}(t) + \varphi_w(t)\right)\left(\bar{\varphi}(t) + \varphi_w(t)\right)^T\right], \tag{13}$$

where $E[\,\cdot\,]$ denotes the mathematical expectation. It is worth to stress that, because of Assumptions A2–A3, the covariance matrix of the noisy data Σ is positive definite. From (11) and Assumptions A3–A4 it follows that

$$\Sigma = E\,[\bar{\varphi}(t)\,\bar{\varphi}^T(t)] + E\,[\tilde{\varphi}(t)\,\tilde{\varphi}^T(t)], \tag{14}$$

where

$$E\,[\tilde{\varphi}(t)\,\tilde{\varphi}^T(t)] = \begin{bmatrix} \sigma_v^{2*}\,I_{n+1} & 0 \\ 0 & 0_{rn\times rn} \end{bmatrix}. \tag{15}$$

Since $E\,[\bar{y}(t)\,w(t)] = \sigma_w^{2*}$ it is also easy to show that

$$\bar{\Sigma} = E\,[\bar{\varphi}(t)\,\bar{\varphi}^T(t)] - \mathrm{diag}\,[\sigma_w^{2*}\;\underbrace{0\;\cdots\;0}_{(r+1)n}], \tag{16}$$

and, because of (10)

$$\bar{\Sigma}\,\bar{\theta}^* = 0. \tag{17}$$

Finally, by combining (14) and (16) it is possible to write

$$\Sigma = \bar{\Sigma} + \tilde{\Sigma}^*, \tag{18}$$

where

$$\tilde{\Sigma}^* = \begin{bmatrix} \sigma_v^{2*} + \sigma_w^{2*} & & 0 \\ & \sigma_v^{2*}\,I_n & \\ 0 & & 0_{rn\times rn} \end{bmatrix}. \tag{19}$$

Though the covariance matrix of the noisy data Σ can be estimated from the available data it is not possible to obtain directly an estimate of θ^* via (17)–(18) since the noise variances σ_w^{2*} and σ_v^{2*} are unknown. As it will be shown in the following, the identification of ARX models with additive output noise can be mapped into an errors–in–variabels identification problem in the Frisch scheme context. The rationale behind the Frisch scheme consists in searching for the solution of the identification problem within a locus of solutions which are compatible with the covariance matrix of the noisy data Σ [22].

The locus of solutions of the Frisch scheme

Now, we will focus our attention on the problem of determining the family of all non–negative definite diagonal matrices $\tilde{\Sigma}$ of type

$$\tilde{\Sigma} = \text{diag } [\sigma_v^2 + \sigma_w^2 \ \underbrace{\sigma_v^2 \ \cdots \ \sigma_v^2}_{n} \ \underbrace{0 \ \cdots \ 0}_{rn}] \tag{20}$$

such that

$$\Sigma - \tilde{\Sigma} \geq 0, \qquad \min \text{ eig } (\Sigma - \tilde{\Sigma}) = 0. \tag{21}$$

This problem consists in determining the set of points $P = (\sigma_v^2, \sigma_w^2)$ belonging to the first quadrant of $\mathcal{R}^2$ satisfying (20)–(21), i.e. leading to positive semidefinite matrices $\bar{\Sigma}(P) = \Sigma - \tilde{\Sigma}(P)$ with one eigenvalue equal to zero. This set is decribed by the following results.

Lemma 1: Consider the following partition of Σ

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yu} \\ \Sigma_{yu}^T & \Sigma_{uu} \end{bmatrix}, \tag{22}$$

where Σ_{yy} is a $(n+1)\times(n+1)$ matrix. This partition follows easily from (6) and (12). The maximal admissible value of σ_v^2 compatible with (21) is given by

$$\sigma_{v,\max}^2 = \min \text{eig } \left(\Sigma_{yy} - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{yu}^T\right). \tag{23}$$

Proof. The maximal value of σ_v^2 compatible with (21) is obtained by setting $\sigma_w^2 = 0$. In this case

$$\Sigma - \tilde{\Sigma} = \begin{bmatrix} \Sigma_{yy} - \sigma_v^2 \, I_{n+1} & \Sigma_{yu} \\ \Sigma_{yu}^T & \Sigma_{uu} \end{bmatrix} \tag{24}$$

so that,

$$\begin{aligned} &\det (\Sigma - \tilde{\Sigma}) \\ &= \det (\Sigma_{uu}) \det (\Sigma_{yy} - \sigma_v^2 \, I_{n+1} - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{yu}^T). \end{aligned} \tag{25}$$

Note that, because of Assumption A2, Σ_{uu} is a positive definite matrix. Moreover, since Σ is positive definite, the matrix $\Sigma_{yy} - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{yu}^T$ is positive definite too, see Lemma A.3 in [20]. Conditions (21) are thus satisfied when σ_v^2 has the value $\sigma_{v,\max}^2$ given by (23).

$\diamond$

Theorem 1: Consider the following partition of Σ

$$\Sigma = \begin{bmatrix} \sigma_y^2 & \rho^T \\ \rho & R \end{bmatrix}, \tag{26}$$

where σ_y^2 is a scalar. The set of all matrices $\tilde{\Sigma}$ compatible with conditions (21) is defined by the couples $P = (\sigma_v^2, \sigma_w^2)$ satisfying the relations

$$\sigma_v^2 \in [0, \sigma_{v,\max}^2] \tag{27}$$

$$\sigma_w^2(\sigma_v^2) = \sigma_y^2 - \sigma_v^2 + \rho^T \theta(\sigma_v^2) \tag{28}$$

where

$$\theta(\sigma_v^2) = -\left(R - \tilde{R}(\sigma_v^2)\right)^{-1} \rho \tag{29}$$

and

$$\tilde{R}(\sigma_v^2) = \text{diag } [\underbrace{\sigma_v^2 \ \cdots \ \sigma_v^2}_{n} \ \underbrace{0 \ \cdots \ 0}_{rn}]. \tag{30}$$

Proof. Consider a generic value of $\sigma_v^2 < \sigma_{v,\max}^2$ belonging to the interval (27) and the couple $P = (\sigma_v^2, \sigma_w^2)$, where σ_w^2 has to be determined. We have

$$\bar{\Sigma}(P) = \Sigma - \tilde{\Sigma}(P) = \begin{bmatrix} \sigma_y^2 - \sigma_w^2 - \sigma_v^2 & \rho^T \\ \rho & R - \tilde{R}(\sigma_v^2) \end{bmatrix}, \tag{31}$$

with $\tilde{R}(\sigma_v^2)$ given by (30). Since σ_v^2 belongs to the set (27), $R - \tilde{R}(\sigma_v^2)$ is positive definite (see Lemma 1). The determinant of $\bar{\Sigma}(P)$ is given by

$$\begin{aligned} \det (\bar{\Sigma}(P)) &= \det \left(R - \tilde{R}(\sigma_v^2)\right) \\ &\times \left(\sigma_y^2 - \sigma_w^2 - \sigma_v^2 - \rho^T \left(R - \tilde{R}(\sigma_v^2)\right)^{-1} \rho\right), \end{aligned} \tag{32}$$

so that conditions (21) are satisfied if and only if σ_w^2 coincides with (28).

$\diamond$

Corollary 1: The values $(\sigma_v^{2*}, \sigma_w^{2*})$, associated with the true variances of $v(t)$ and $w(t)$, belong to the solution set described by Theorem 1 and the corresponding coefficient vector coincides with θ^*, i.e. $\sigma_w^2(\sigma_v^{2*}) = \sigma_w^{2*}$ and $\theta(\sigma_v^{2*}) = \theta^*$.

Proof. The result follows easily from (17) and (18).

Remark 2: The above results show that the covariance matrix of the noisy data Σ can be associated with the solution set $[0, \sigma_{v\,\max}^2]$. In the sequel, this interval will be denoted by $\mathcal{I}(\Sigma)$.

Every value of $\sigma_v^2 \in [0, \sigma_{v,\max}^2]$ leads to a possible solution $\sigma_v^2, \theta(\sigma_v^2), \sigma_w^2(\sigma_v^2)$ of the noisy ARX identification problem where $\theta(\sigma_v^2)$ and $\sigma_w^2(\sigma_v^2)$ are given by (29) and (28). Because of Corollary 1, the determination of σ_v^{2*} inside the interval $\mathcal{I}(\Sigma)$ leads, in the asymptotic case, to the solution of the identification problem with fixed model order, n. When only a finite number of data, N, is available, it is possible to replace the covariance matrix Σ with the sample estimate

$$\hat{\Sigma} = \frac{1}{N-n} \sum_{t=n+1}^{N} \varphi(t) \, \varphi^T(t) \tag{33}$$

and it is necessary to introduce a suitable criterion to select a single model among the family of possible solutions, i.e to select a single point within $\mathcal{I}(\hat{\Sigma})$.The criterion described in the following consists in computing the equation error of the noisy ARX model and in comparing its sampled statistical properties with the expected theoretical ones.

The selection criterion

By inserting (2) in (1) it is easy to obtain

$$A(z^{-1})\,y(t) - B(z^{-1})\,u(t) = A(z^{-1})\,v(t) + w(t) \triangleq e(t), \tag{34}$$

where the stochastic process $e(t)$ can be considered as the equation error of the noisy ARX model. Because of Assumptions A3–A4, $e(t)$ has a finite number of autocovariances $r_e(\tau) = E\,[e(t)\,e(t-\tau)]$, given by

$$r_e(0) = \sigma_v^2 \sum_{i=0}^{n} a_i^2 + \sigma_w^2 \tag{35}$$

$$r_e(\tau) = \sigma_v^2 \sum_{i=0}^{n-\tau} a_i\,a_{i+\tau} \quad \text{for } \tau = 1,\ldots,n \tag{36}$$

$$r_e(\tau) = 0 \quad \text{for } \tau > n, \tag{37}$$

where $a_0 = 1$. Consider now a generic point $\sigma_v^2 \in \mathcal{I}(\hat{\Sigma})$ and the corresponding values $\theta(\sigma_v^2), \sigma_w^2(\sigma_v^2)$. It is possible to compute the corresponding residuals (see (34))

$$\hat{e}(t,\sigma_v^2) = \hat{A}(z^{-1})\,y(t) - \hat{B}(z^{-1})\,u(t), \quad t = n+1,\ldots,N \tag{38}$$

where $\hat{A}(z^{-1})$ and $\hat{B}(z^{-1})$ are constructed with the entries of $\theta(\sigma_v^2)$, and, subsequently, the sample autocovariances

$$\hat{r}_e^N(\tau,\sigma_v^2) = \frac{1}{N-n-\tau} \sum_{t=n+\tau}^{N} \hat{e}(t,\sigma_v^2)\,\hat{e}(t-\tau,\sigma_v^2), \tag{39}$$

$\tau = 1,\ldots,n$. It is also possible to compute, by using σ_v^2, $\theta(\sigma_v^2)$ and (36), the theoretical quantities $\hat{r}_e(\tau,\sigma_v^2)$. By introducing the vector

$$\xi(\sigma_v^2) = \begin{bmatrix} \hat{r}_e^N(1,\sigma_v^2) - \hat{r}_e(1,\sigma_v^2) \\ \hat{r}_e^N(2,\sigma_v^2) - \hat{r}_e(2,\sigma_v^2) \\ \vdots \\ \hat{r}_e^N(n,\sigma_v^2) - \hat{r}_e(n,\sigma_v^2) \end{bmatrix}, \tag{40}$$

and the cost function

$$J(\sigma_v^2) = \|\xi(\sigma_v^2)\|_2^2 = \xi^T(\sigma_v^2)\xi(\sigma_v^2), \tag{41}$$

The identification problem, with fixed n, can be solved by minimizing $J(\sigma_v^2)$ along $\mathcal{I}(\hat{\Sigma})$. In fact, due to the ergodicity property

$$\lim_{N\to\infty} \hat{\Sigma} = \Sigma, \quad \lim_{N\to\infty} \mathcal{I}(\hat{\Sigma}) = \mathcal{I}(\Sigma) \tag{42}$$

and, because of Corollary 1

i) $J(\sigma_v^2) \geq 0, \quad \sigma_v^2 \in \mathcal{I}(\Sigma)$

ii) $J(\sigma_v^{2*}) = 0$.

The whole identification procedure can be summarized as follows.

Identification algorithm.

1) Compute, on the basis of the available observations, the estimate $\hat{\Sigma}$ (33) of Σ and consider the corresponding partitions (22), (26)

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{yy} & \hat{\Sigma}_{yu} \\ \hat{\Sigma}_{yu}^T & \hat{\Sigma}_{uu} \end{bmatrix}, \qquad \hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_y^2 & \hat{\rho}^T \\ \hat{\rho} & \hat{R} \end{bmatrix}.$$

2) Compute the value $\hat{\sigma}_{v,\max}^2$ by means of (23).
3) Start from a generic point σ_v^2 belonging to $\mathcal{I}(\hat{\Sigma})$.
4) Compute the parameter vector

$$\theta(\sigma_v^2) = -\left(\hat{R} - \tilde{R}(\sigma_v^2)\right)^{-1} \hat{\rho}. \tag{43}$$

5) Compute the residuals (38) and the sample autocovariances (39). Compute also the theoretical quantities (36) by using σ_v^2 and $\theta(\sigma_v^2)$ and construct the vector (40).
6) Compute the cost function $J(\sigma_v^2)$ (41).
7) Move to a new value $\sigma_v^2 \in \mathcal{I}(\hat{\Sigma})$ corresponding to a decrease of $J(\sigma_v^2)$.
8) Repeat steps 4–7 until the point $\hat{\sigma}_v^2$ associated with the minimum of $J(\sigma_v^2)$ is found. The corresponding parameter estimate is $\hat{\theta} = \theta(\hat{\sigma}_v^2)$.
9) Compute the estimate of the driving noise variance by using (28).

$$\hat{\sigma}_w^2 = \hat{\sigma}_y^2 - \hat{\sigma}_v^2 + \hat{\rho}^T\,\hat{\theta}. \tag{44}$$

Remark 3: The noisy ARX model (1)-(4) can be identified by using classic identification methods like the prediction error method (PEM) and the instrumental variable estimator (IV). The PEM approach, that can rely on the state space representation described in the next subsection, has a computational complexity which is remarkably higher than that of the proposed Frisch scheme method. This aspect can be relevant for the considered application. The IV estimator, on the contrary, is computationally very efficient but its estimation accuracy is in general lower than that of the Frisch scheme [21]. Moreover, the IV method does not provide an estimate of the noise variances which is required for order estimation and model validation (see the next subsection).

B. Model order and actual parameter identification

According to common identification approach, the basic idea to derive the model order, n, is to repeat the procedure of Subsection IV-A increasing it, starting from $n = 1$, until a suitable yes/no test, applied to the current estimation results, is satisfied. This will give the actual model order, along with the actual polynomial coefficients and noise variances to tackle the *Identification Problem*. The adopted yes/no test is the whiteness of the residuals generated by the identified model used as a predictor of the output. In standard methods as least squares, this is rather straightforward, since the innovation can be directly computed along the identification procedure and then its whiteness can be easily tested. In contrast, for the framework and the method proposed in this Section, such a property does no longer hold. Then, the assessment is carried out evaluating the whiteness of the innovation generated by a Kalman optimal predictor based on the following state space representation of the identified noisy ARX model

$$x(t+1) = A\,x(t) + B\,u(t) + G\,w(t+1) \tag{45}$$

$$y(t) = C\,x(t) + v(t) \tag{46}$$

where

$$A = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & \ddots & \ddots & \\ \vdots & \vdots & & \ddots & \\ \vdots & \vdots & & & 1 \\ -a_n & 0 & \cdots & \cdots & 0 \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & \cdots & b_{r1} \\ b_{12} & \cdots & b_{r2} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ b_{1n} & \cdots & b_{rn} \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \qquad G = C^T.$$

In fact, in the above representation the first component of the state vector $x(t)$ is the actual tile temperature $\bar{y}(t)$, i.e. $\bar{y}(t) = C\,x(t)$. The innovation signal $\varepsilon(t)$ is thus given by the difference

$$\varepsilon(t) = y(t+1) - \hat{\bar{y}}(t+1|t), \tag{47}$$

where $\hat{\bar{y}}(t+1|t)$ is the prediction of the actual temperature.

Remark 4: Validation of the identified models is another important step in real-life applications. It usually consists of applying the estimated models to new data sets w.r.t. the ones adopted in the identification procedure, and checking their performance by some tests. The yes/no test presented in this Subsection can be exploited for validation, as well.

V. EXPERIMENTAL RESULTS

According to the procedure shown in Section IV and the SCC testing framework reported in Section II, the thermal model identification for tiles 15 and 16 of the considered SCC platform has been carried out on a 700 s-long training set, with a sampling time $T_s = 100$ ms. The resulting model order is $n = 2$ and the identified poles are $p_{1,15} = 0.9412$, $p_{2,15} = 0.0854$ for tile 15 and $p_{1,16} = 0.8803$, $p_{2,16} = 0.1522$ for tile 16. It is worth noting that the poles are stable and positive real, in accordance with the features of thermal systems deriving from the Second Law of Thermodynamics. $\chi^2_{0.01}(8)$ whiteness tests on the innovations obtained by suitable Kalman predictors (see previous Section) have been performed [20] following the procedure of Subection IV-B. The obtained test quantities are 7.7 and 14.2 for the tiles 15 and 16 respectively while the maximal admissible value to validate the models is 20.1. Achieving the compliance with the yes/no whiteness test with a rather small model order ($n = 2$) and with stable positive real poles confirms, from a practical view point, the validity of the adopted structure, where measurement noise in the outputs is directly taken into account; while it can be neglected at the input side, as assumed in Section III. The measurement and process noise variances estimated with the proposed method are $\sigma^2_{v,15} = 0.2461$, $\sigma^2_{w,15} = 0.0294$, for tile 15, and $\sigma^2_{v,16} = 0.2202$, $\sigma^2_{w,16} = 0.0446$, for tile 16. The measurement noise variances are quite large, as assumed in Section III and they give quite small SNRs (8.9 dB and 9.2 dB for tile 15 and 16 respectively).

Fig. 4 reports the predicted temperatures and the related innovations (i.e. difference between measurements and predictions) for tile 15 in the identification test. The figure related to tile 16 is similar and has not been reported. The mean values of the innovations are negligible, as expected

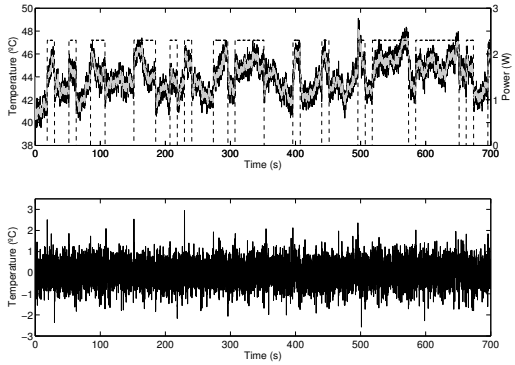

Fig. 4. Tile 15 identification traces. Upper picture: input power (dashed), measured temperature (black) and one-step-ahead predicted temperature (gray). Lower picture: innovation (difference between measured temperature, $T(t)$, and predicted one, $\hat{T}(t)$)

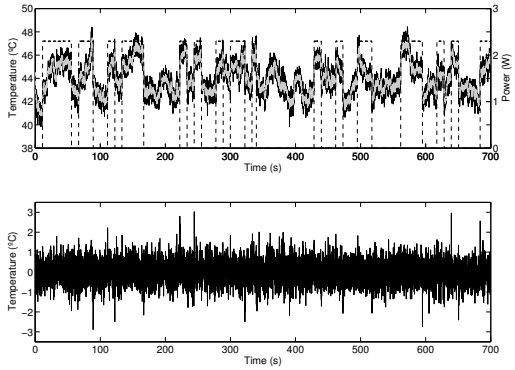

Fig. 5. Tile 15 validation traces. Upper picture: input power (dashed), measured temperature (black) and one-step-ahead predicted temperature (gray). Lower picture: innovation (difference between measured temperature, $T(t)$, and predicted one, $\hat{T}(t)$)

(0.0016 °C for tile 15 and 0.0013 °C for tile 16) while their variances are 0.3393 and 0.3300. It is important to note that the variances values are very close to the steady-state theoretical ones $\sigma^2_{\varepsilon,15} = 0.3392$ and $\sigma^2_{\varepsilon,16} = 0.3320$. This is reasonable for identification traces, but not so foregone since the adopted method is not a prediction error method [19]. By looking at the upper pictures of Fig. 4 it can be noted that the Kalman predictor, based on the identified model, can filter out a relevant part of the measurement noise. This is the reason why the innovation is quite large (see the lower pictures and recall the above mentioned variances).

In order to validate the identification robustness, the above mentioned Kalman predictors have been applied to data sets different from the training ones. Fig. 5 reports the predicted temperatures and the related innovations for tile 15 in a validation test. The figure related to tile 16 is similar and has not been reported. The mean values of the innovation are small (0.0010 °C for tile 15 and 0.0010 °C for tile 16) while their variances are 0.3419 and 0.3608. These values are very close to the ones obtained in the identification

test; this means that the prediction performances are robust w.r.t. the data sets. The whiteness of the innovations have been checked by means of a $\chi^2_{0.01}(8)$ test obtaining the quantities 8.65 and 21.1. Since the maximal admissible value is 20.1, both results can be considered as good. This, again, highlights the effectiveness of the proposed solution.

VI. CONCLUSIONS

We presented a Frisch-based identification algorithm to achieve distributed, compact and scalable thermal modelling of multi and many-core processors. This method allows to distinguish process noise from output measurement noise. Such a feature is particularly interesting for using the identification results in Kaman filters and predictors. The proposed solution has been successfully applied to an Intel's SCC experimental processor. The identified models can be effectively adopted in Model Predictive Control for performance optimization, under temperature capping contraints. The proposed solution is valid in general for any platform with features similar to SCC. Since the SCC characteristics are becoming a standard for advanced processors, the presented solution is promising for application on next generation multi and many-core computing systems.

VII. ACKNOWLEDGMENTS

This work was supported, in parts, by the EU FP7 ERC Project MULTITHERMAN (GA n. 291125). Technical support from Intel Labs. Braunschweig is also acknowledged.

REFERENCES

[1] S. Borkar, Design challenges of technology scaling, *IEEE Micro*, vol. 19, no. 4, pp. 23–29, 1999.

[2] G. E. Moore, Cramming more components onto integrated circuits, *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.

[3] K. Asanovic et al., The landscape of parallel computing research: a view from Berkeley, *Technical Report UCB/EECS-2006-183*, EECS Department, University of California, Berkeley, December 2006.

[4] M. B. Taylor, Is dark silicon useful? Harnessing the four horsemen of the coming dark silicon apocalypse, *Proc. of 49th ACM/EDAC/IEEE Design Automation Conference (DAC 2012)*, San Francisco, California, June 2012, pp. 1131–1136.

[5] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam and D. Burger, Dark silicon and the end of multicore scaling, *IEEE Micro*, vol. 32, no. 3, pp. 122–134, 2012.

[6] E. Rotem, Power management architecture of the 2nd generation intel core microarchitecture, formerly codenamed sandy bridge, *Proc. of Hotchips*, 2011.

[7] I. Chowdhury et al., On-Chip cooling by superlattice-based thin-film thermoelectrics, *Nature Nanotechnology*, vol. 4, pp. 235–238, 2009.

[8] R. Grimes, E. Walsh and P. Walsh, Active cooling of a mobile phone handset, *Appl. Thermal Eng.*, vol. 30, no. 16, pp. 2363–2369, 2010.

[9] J. Kong, S. W. Chung and K. Skadron, Recent thermal management techniques for microprocessors, *ACM Comput. Surv.*, vol. 44, no. 3, 2012.

[10] J. Charles et al., Evaluation of the Intel Core i7 Turbo Boost feature, *Proc. of IISWC 2009*, Austin, Texas, October 2009, pp 188–197.

[11] Y. Wang, K. Ma and X. Wang, Temperature-constrained power control for chip multiprocessore with online model estimation, *Proc. of ISCA 2009*, Austin, Texas, June 2009.

[12] A. Bartolini, M. Cacciari, A. Tilli and L. Benini, Thermal and energy management of high-performance multicores: Distributed and self-calibrating model-predictive controller, *IEEE Trans. Parall. Distr. Syst.*, vol. 24, no. 1, pp. 170–183, 2013.

[13] A. Tilli, E. Garone, M. Cacciari and A. Bartolini, Thermal models characterization for reliable temperature capping and performance optimization in multiprocessor systems on chip, *Proc. of 2012 American Control Conference (ACC 2012)*, Montréal, Canada, June 2012, pp. 4721–4726.

[14] J. Howard et al., A 48-Core IA-32 message-passing processor with DVFS in 45nm CMOS, *Proc. of 2010 IEEE International Solid-State Circuits Conference (ISSCC 2010)*, San Francisco, CA, February 2010, pp. 108–109.

[15] R. Cochran and S. Reda, Consistent runtime thermal prediction and control through workload phase detection, *Proc. of 47th ACM/IEEE Design Automation Conference (DAC 2010)*, June 2010, pp. 62–67.

[16] A. Coskun, T. Rosing and K. Gross, Utilizing predictors for efficient thermal management in multiprocessor SoCs, *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 10, pp. 1503–1516, 2009.

[17] D.-C. Juan, H. Zhou, D. Marculescu and L. Xin, A learning based autoregressive model for fast transient thermal analysis of chip-multiprocessors, *Proc. of 17th Asia and South Pacific Design Automation Conference (ASP-DAC 2012)*, Sidney, Australia, February 2012, pp. 597–602.

[18] S. Sharifi and T. S. Rosing, Accurate direct and indirect on-chip temperature sensing for efficient dynamic thermal management, *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 10, pp.1586–1599, 2010.

[19] L. Ljung, *System Identification – Theory for the User* Prentice Hall, Englewood Cliffs, NJ, 1999.

[20] T. Söderström and P. Stoica, *System Identification*, Prentice Hall, Cambridge, UK, 1989.

[21] T. Söderström, Errors–in–Variables methods in system identification, *Automatica*, vol. 43, no. 6, 2007, pp. 939–958.

[22] R. Guidorzi, R. Diversi, U. Soverini, The Frisch scheme in algebraic and dynamic identification problems, *Kybernetika*, vol. 44, no. 5, pp. 585–616, 2008.

[23] R. Diversi, R. Guidorzi and U. Soverini, Blind identification and equalization of two–channel FIR systems in unbalanced noise environments, *Signal Processing*, vol. 85, no. 1, pp. 215–225, 2005.

[24] R. Diversi, R. Guidorzi and U. Soverini, Identification of autoregressive models in the presence of additive noise, *International Journal of Adaptive Control and Signal Processing*, vol. 22, no. 5, pp. 465–481, 2008.

[25] R. Diversi, R. Guidorzi and U. Soverini, Identification of ARX and ARARX models in the presence of input and output noises, *European Journal of Control*, vol. 16, no. 3, pp. 242–255, 2010.

[26] A. Bartolini et al. A system level approach to multi-core thermal sensors calibration, *Proc. of PATMOS 2011*, Madrid, Spain, September 2011, pp. 22–31.

[27] M. Sadri, A. Bartolini and L. Benini, Single-chip cloud computer thermal model, *Proc. of 17th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC 2011)*, Paris, France, September 2011.

[28] R. Diversi, A. Bartolini, A. Tilli, F. Beneventi and L. Benini, SCC thermal model identification via advanced bias-compensated least-squares, *Proc. of DATE 2013 (Design, Automation & Test in Europe)*, Grenoble, France, March 2013.