

SCC Thermal Model Identification via Advanced Bias-Compensated Least-Squares

Roberto Diversi, Andrea Bartolini, Andrea Tilli, Francesco Beneventi, Luca Benini

Email: {roberto.diversi, a.bartolini, andrea.tilli, francesco.beneventi, luca.benini}@unibo.it

DEI, University of Bologna

Viale del Risorgimento 2, 40136 Bologna, Italy

Abstract—Compact thermal models and modeling strategies are today a cornerstone for advanced power management to counteract the emerging thermal crisis for many-core systems-on-chip. System identification techniques allow to extract models directly from the target device thermal response. Unfortunately, standard Least Squares techniques cannot effectively cope with both model approximation and measurement noise typical of real systems. In this work, we present a novel distributed identification strategy capable of coping with real-life temperature sensor noise and effectively extracting a set of low-order predictive thermal models for the tiles of Intel’s Single-chip-Cloud-Computer (SCC) many-core prototype.

I. INTRODUCTION AND RELATED WORK

Performance of today processors are constrained by the die temperature. High silicon temperature, hotspots and thermal gradients can lead to early chip damage, critical path delay degradation and reduced performance. To avoid imposing ultra-conservative design margins, dynamic thermal management (DTM) solutions have been widely adopted [1] to ensure, at run-time, a safe working temperature by limiting the processor performance (DVFS, core shutdown) only under critical workload phases and environmental conditions. Classical feedback control strategies, such as PID and threshold controller have been explored, but their reactive nature cannot ensure at the same time a safe working temperature and a minimal performance loss. Recently, model-predictive controllers (MPC) have been shown to be capable of achieving better results [2], [3], assuming the predictive model, used internally to project in the future the effect of a given controller action, is accurate and simple. Indeed, MPC run-time overhead rapidly increases with the complexity of the predictive model and its control performance 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 heat sink, 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 model-predictive control to be applicable in a real-life setting. In addition, particular attention has to be paid to the complexity of the learned model. In fact, low-complexity models are preferred to reduce the computational burden of model-predictive control policies.

The Single-Chip Cloud Computer (SCC) experimental processor [5] 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 consequence of that, SCC is a challenging testbench for thermal model learning strategies. Moreover, its elevate number of cores set requirements on the complexity and scalability of the proposed algorithm.

Recently, strategies for extracting compact thermal models directly from the core thermal response to a given power/workload stress input have been proposed [2], [3], [6]–[8]. The simplest ones are centralized and do not account for the multimodal nature of the thermal transient caused by the different building materials and their relative time-constants (i.e. die, heat-spreader and heat-sink) [3]. In [2], [6], a first order dynamic thermal model is learnt 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 et al. [9] shows 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 poorly scale with the increase of the number of cores, leading to complex control solutions [3].

To counteract these limitations, distributed model learning approach based on autoregressive strategies have been recently proposed. Coskun et al. [7] 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 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. [8] uses a combination of a K-means clustering and an AR model to learn a compact model for fast thermal simulation. This approach is effective only when starting from an highly accurate thermal model of the HW. Bartolini et al. [3] present a distributed model learning approach based on a set of ARX models. Each core executes its own model learning 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 [10].

Contribution

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 a distributed algorithm for ARX-like models that is capable of filter-out the output noise through an advanced bias-compensated least-squares algorithm and we show its effectiveness in identifying the thermal model of the real SCC chip, when compared to state-of-the-art standard autoregressive procedures. The main contributions are the followings.

- We present a methodology to split a complex Multiple-Input Multiple-Output (MIMO) system identification problem in a set of distributed Multiple-Input Single-Output (MISO) identification problem, one for each SCC tile.
- We show the inefficacy of standard ARX solutions in extracting valid thermal models of a real many-core device affected by noise in the thermal sensors.
- We propose an algorithm to deal with ARX models with additive output measurement noise. This is done by automatically estimating the output noise variance and by compensating it in the parameter learning Least Squares (LS) problem. We show how this algorithm can be successfully adopted in the domain multicore (with noisy temperature measurements) thermal model learning.
- The set of models learnt can be used in prediction to filter out the noise through a Kalman filter, leading to an optimal estimation of the real silicon temperature.

II. SCC ARCHITECTURE

The SCC 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 [11] to obtain a meaningful temperature for each sensor. Calibrated thermal sensors outputs show the presence of significant white noise [11]. 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 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 [12].

Given these HW features, we used 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

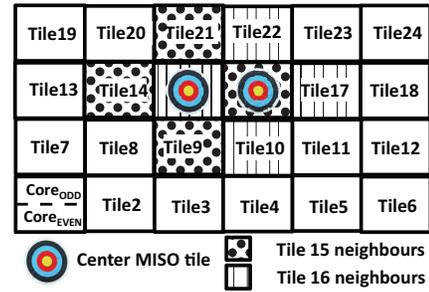


Fig. 1. SCC topology

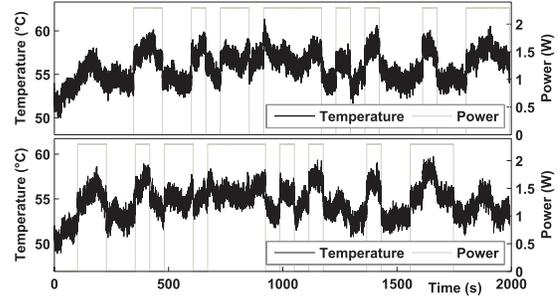


Fig. 2. Examples of power and temperature traces of tile 15 and tile 16

and stop synchronously a given workload/power virus¹ on the different SCC cores while at the same time collecting the HW monitors outputs (i.e. performance counters and thermal sensors). As consequence of that, the framework can apply a given Pseudo-Random Binary Sequence (PRBS) stress workload to the SCC cores. The typical PRBS is shown in Fig. 2 and it represents a worst case workload pattern as it triggers all the dynamics of the system. 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 tile². 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 [11], a relevant white noise can be noted on the temperature readings given by the adopted sensors.

As explained before, the identification solution presented in next Sections is distributed and thus learns the full SCC thermal model as composition of per-tile MISO thermal models.

III. IDENTIFICATION OF ARX MODELS

According to the adopted distributed approach, in this Section we propose to represent the SCC device by means of a collection of ARX models [10] with the following features.

- For each tile a MISO ARX model is considered to represent the behavior of the sensed tile temperature $\bar{T}(t)$, which is assumed to represent perfectly the *actual tile temperature*. This is a standard crucial hypothesis underlying the ARX

¹*cpuburn* power virus by Robert Redelmeier: it takes advantage of the internal architecture to maximize the CPU power consumption

²In this paper we consider only the thermal sensor positioned close to the router since it is more central within the tile area.

models in general. A second order model³ is considered according to [3], [4].

– The inputs of each ARX are the dissipated power $P(t)$ related to tile activity and the temperatures $\bar{T}_{n1}(t), \bar{T}_{n2}(t), \dots, \bar{T}_{nq}(t)$ probed on the tiles belonging to its neighborhood.

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

– Tests will be carried out under constant ambient temperature (T_{amb}) conditions. In this line, 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 fast⁴ variations in T_{amb} can be experienced, the presented identification procedures (also the new one of Section IV) can be adopted as well, by adding T_{amb} in the input set of the MISO ARX models and using the absolute temperatures instead of the temperature gaps. In this case, a variable T_{amb} should be used in identification tests to obtain good results, according to persistency of excitation requirements [10].

The ARX model of each core is then

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

where the input $u(t)$ is the r -dimensional signal

$$u(t) = [P(t) T_{n1}(t) \dots T_{nq}(t)]^T, \quad (2)$$

$\bar{T}(t)$ is the tile temperature, $w(t)$ is a white noise process and

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \dots + a_n z^{-n} \\ B(z^{-1}) &= [B_1(z^{-1}) \quad B_2(z^{-1}) \quad \dots \quad B_r(z^{-1})] \\ B_i(z^{-1}) &= b_{i1} z^{-1} + \dots + b_{in} z^{-n}, \quad i = 1, \dots, 1 + q \end{aligned} \quad (3)$$

where $A(z^{-1})$ and $B_i(z^{-1})$ are polynomials in the backward shift operator z^{-1} , i.e $z^{-1}x(t) = x(t-1)$. According to previous assumptions, the order is set to $n = 2$ and the number of neighbor temperatures q ranges from 2 to 4.

In order to estimate the parameters of tile models (1), a standard LS approach can be adopted, combined with a $\chi^2(8)$ whiteness test on residuals for validation [10].

This procedure has been applied to a couple of central tiles, namely 15 and 16, whose neighbors are highlighted in Fig.1. The following results have been derived.

– The $\chi^2(8)$ test is largely not satisfied, since the values 822 and 639 are obtained for tiles 15 and 16, respectively, while the maximum admissible one to state the whiteness with a 99% confidence level is 20.1.

– The identified parameters show a relevant negative pole, see the left part of Fig. 3. This is in contrast with the physics of thermal systems where, according to the Second Law of Thermodynamics, only real positive poles can be present.

Taking the cue from the above points, the hypothesis on the model order equal to 2 has been relaxed, to capture more complex thermal models. Then, the above identification

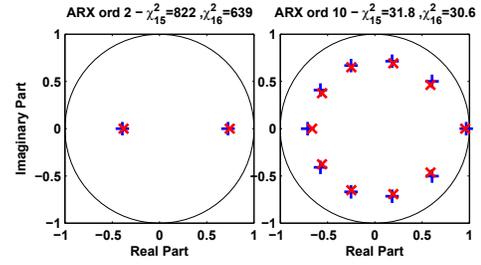


Fig. 3. Tiles 15,16 pole maps and $\chi^2(8)$

procedure has been applied iteratively, augmenting the model order up to a level satisfying the whiteness test. Nevertheless, the following negative issue emerges. An high order (10th) is required to satisfy practically the $\chi^2(8)$ test and most of the resulting model poles are complex, see right part of Fig. 3. This, again, is in contrast with the physics of thermal systems.

On the basis of the above results, we can conclude that the proposed ARX models are not suitable for representing the link between the considered inputs and the measured temperature.

IV. IDENTIFICATION OF ARX-WITH-NOISY-OUTPUT MODELS

Bearing in mind the results of previous Section and looking at the temperature measurement traces, reported in Fig.2, it looks reasonable to guess the presence of a significant additive white noise on temperature readings. Then, the ARX tile model (1) considered in Section III is augmented with an output noise, leading to the structure reported in Fig. 4, where the measured temperature, $T(t)$, is defined as

$$T(t) = \bar{T}(t) + v(t). \quad (4)$$

Note that in such formulation the interesting temperature is $\bar{T}(t)$ while the available measure is $T(t)$ and they are not assumed to be the same, in contrast with standard ARX.⁵ Thus, $w(t)$ is the driver of the significant process noise $\eta(t)$, while $v(t)$ is just a measure noise corrupting the actual readings.

In line with the previous considerations, noisy inputs should be considered as well, since the neighbor temperature measurements are adopted as inputs in the ARX model of each core (1), leading to a full Errors-In-Variables (EIV) framework

⁵It must be noted that $\bar{T}(t)$ cannot be simply reconstructed from $T(t)$ by low-pass filtering, since the white noise spectral components cannot be decoupled from the SCC thermal response dynamics.

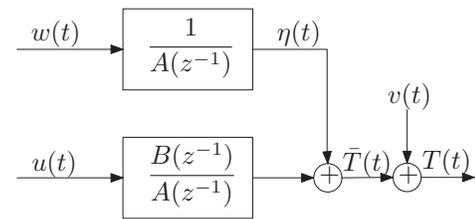


Fig. 4. ARX model with additive output noise.

³The model order specifies the number of poles/dynamics of the model.

⁴fast w.r.t. the chip thermal dynamics, i.e. the thermal bandwidth

[13]. 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 inputs. The validity of such hypothesis will be checked later on the actual data by suitably assessing the estimation results.

In order to set properly the proposed identification procedure, the following standard assumptions are considered.

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

A2. The input power $P(t)$ is persistently exciting [10].

A3. The process 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 $u(t)$.

In the following, assuming an a priori fixed model order $n = 2$, we present a technique for estimating the coefficients of $A(z^{-1}), B_i(z^{-1}), (i = 1 \dots, r)$ and the noise variances σ_w^2, σ_v^2 , starting from N input-output samples $u(1), \dots, u(N), T(1), \dots, T(N)$.

A. Bias-compensated least-squares identification

Let us introduce the vectors

$$\bar{\varphi}(t) = [-\bar{T}(t-1) \dots -\bar{T}(t-n) u_1(t-1) \dots u_1(t-n) \\ u_2(t-1) \dots u_2(t-n) \dots u_r(t-1) \dots u_r(t-n)]^T$$

$$\varphi(t) = [-T(t-1) \dots -T(t-n) u_1(t-1) \dots u_r(t-n)]^T$$

$$\tilde{\varphi}(t) = [-v(t-1) \dots -v(t-n) \underbrace{0 \dots 0}_{rn}]^T,$$

and the parameter vector

$$\theta = [a_1 \dots a_n \ b_{11} \dots b_{1n} \dots b_{r1} \dots b_{rn}]^T. \quad (5)$$

From (1) and (4) it follows that

$$\bar{T}(t) = \bar{\varphi}^T(t) \theta + w(t) \quad (6)$$

$$\varphi(t) = \bar{\varphi}(t) + \tilde{\varphi}(t). \quad (7)$$

By inserting (6) and (7) in (4) it is easy to obtain the regression form

$$T(t) = \varphi^T(t) \theta + e(t), \quad (8)$$

where

$$e(t) = v(t) - \tilde{\varphi}^T(t) \theta + w(t). \quad (9)$$

Define now the covariance matrix and cross-covariance vector

$$R_{\varphi\varphi} = E[\varphi(t) \varphi^T(t)], \quad r_{\varphi y} = E[\varphi(t) T(t)], \quad (10)$$

where $E[\cdot]$ denotes the mathematical expectation. Multiplying both sides of (8) by $\varphi(t)$ and computing their expected values we obtain

$$r_{\varphi y} = R_{\varphi\varphi} \theta + E[\varphi(t) e(t)]. \quad (11)$$

By using (7) and (9) and taking into account Assumptions A3–A4 it is easy to derive

$$r_{\varphi y} = R_{\varphi\varphi} \theta + E[\tilde{\varphi}(t) e(t)] = R_{\varphi\varphi} \theta - \sigma_v^2 J \theta, \quad (12)$$

where $J = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$ and I_n denotes the $n \times n$ identity matrix. Equation (12) can be rewritten as

$$\theta = R_{\varphi\varphi}^{-1} r_{\varphi y} + \sigma_v^2 R_{\varphi\varphi}^{-1} J \theta = \theta_{LS} + \sigma_v^2 R_{\varphi\varphi}^{-1} J \theta, \quad (13)$$

where $\theta_{LS} = R_{\varphi\varphi}^{-1} r_{\varphi y}$ is the asymptotic LS estimation of θ . It is clear that the LS estimate is biased due to the presence of σ_v^2 . However, if an estimate of σ_v^2 is available, the effect of the bias induced by the output noise $v(t)$ can be removed and an asymptotically unbiased estimate can be obtained. This leads to the iterative bias-compensated least squares estimator

$$\hat{\theta}^{k+1} = \hat{\theta}_{LS} + \hat{\sigma}_v^{2k} \hat{R}_{\varphi\varphi}^{-1} J \hat{\theta}^k, \quad (14)$$

where $\hat{\theta}_{LS}$ and $\hat{R}_{\varphi\varphi}$ are estimates obtained from the available data. Of course, it is necessary to compute, at each step, an estimate $\hat{\sigma}_v^{2k}$ of the output noise variance so that at least one more equation must be considered in addition to the $(r+1)n$ equations (13).

To this end, taking the cue from [14], consider the stochastic process $e(t)$, that can be considered as the equation error of the ARX + noise model, see (8), (9). By taking into account assumptions A3–A4 it is easy to show that the autocovariances $r_e(\tau) = E[e(t) e(t-\tau)]$ of $e(t)$ are given by

$$r_e(0) = \sigma_v^2 \sum_{i=0}^n a_i^2 + \sigma_w^2 \quad (15)$$

$$r_e(\tau) = \sigma_v^2 \sum_{i=0}^{n-\tau} a_i a_{i+\tau} \quad \text{for } \tau = 1, \dots, n \quad (16)$$

$$r_e(\tau) = 0 \quad \text{for } \tau > n, \quad (17)$$

where $a_0 = 1$. By defining the vectors

$$r_e = [r_e(1) \ r_e(2) \ \dots \ r_e(n)]^T \quad (18)$$

$$\psi = [\sum_{i=0}^{n-1} a_i a_{i+1} \ \sum_{i=0}^{n-2} a_i a_{i+2} \ \dots \ a_0 a_n]^T, \quad (19)$$

the set of relations (16) can be rewritten as

$$r_e = \sigma_v^2 \psi, \quad (20)$$

which leads to

$$\sigma_v^2 = \frac{\psi^T r_e}{\psi^T \psi}. \quad (21)$$

Note that, if an estimate of the parameter vector $\hat{\theta}$ is available, it is possible to compute an estimate of the equation error sequence from (8):

$$\hat{e}(t) = T(t) - \varphi^T(t) \hat{\theta}, \quad t = n+1, \dots, N, \quad (22)$$

and then, to compute an estimate of r_e . It is thus possible to develop an iterative least-squares based algorithm where the current estimate of the output noise variance is used to improve the estimate of the system parameters and *vice versa*. The whole identification procedure can be summarized as follows.

Algorithm

1. Compute, on the basis of the available observations $u(1), \dots, u(N), T(1), \dots, T(N)$, the sample estimates, $\hat{R}_{\varphi\varphi} = \frac{1}{N-n} \sum_{t=n+1}^N \varphi(t) \varphi^T(t)$ and $\hat{r}_{\varphi y} = \frac{1}{N-n} \sum_{t=n+1}^N \varphi(t) T(t)$.

2. Compute the least squares estimate $\hat{\theta}_{LS} = \hat{R}_{\varphi\varphi}^{-1} \hat{r}_{\varphi y}$.

3. Set $k = 0, \hat{\theta}^0 = \hat{\theta}_{LS}$.

4. Compute the sequence of equation errors $\hat{e}(t)^k = T(t) - \varphi^T(t) \hat{\theta}^k, \quad t = n+1, \dots, N$ and, subsequently, the estimates $\hat{r}_e^k(1), \dots, \hat{r}_e^k(n)$ of the autocovariances (16).

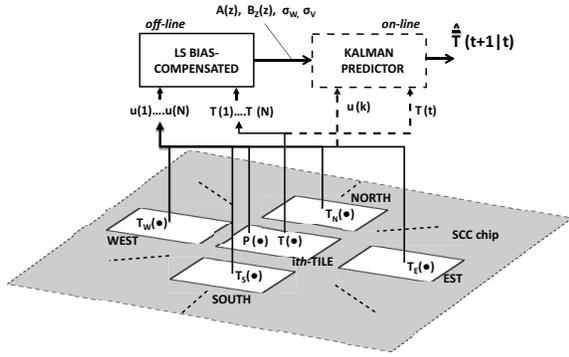


Fig. 5. How the proposed distributed identification approach can feed Kalman predictor on each tile

5. Form the vectors \hat{r}_e^k and $\hat{\psi}^k$ as in (18), (19) and compute an estimate of the output noise variance by means of $\hat{\sigma}_v^{2k} = (\hat{\psi}^T \hat{r}_e) / (\hat{\psi}^T \hat{\psi})$.
6. Update the parameter estimate as follows $\hat{\theta}^{k+1} = \hat{\theta}_{LS} + \hat{\sigma}_v^{2k} R_{\varphi\varphi}^{-1} J \hat{\theta}^k$.
7. Set $\hat{\theta}^k = \hat{\theta}^{k+1}$ and go to step (4).
8. Repeat steps 4–7 until $\frac{\|\hat{\theta}^{k+1} - \hat{\theta}^k\|}{\|\hat{\theta}^{k+1}\|} < \varepsilon$, where ε is an assigned convergence threshold.
9. Compute an estimate of the driving noise variance by using (15), i.e. $\hat{\sigma}_w^2 = \hat{r}_e(0) - \hat{\sigma}_v^2 \sum_{i=0}^n \hat{a}_i^2$.

Once the parameters have been obtained, it is crucial to assess the actual identification performance. In standard methods as LS, 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 a state space representation of the noisy ARX model [10], [15]. Referring as $\hat{T}(t+1|t)$ the Kalman prediction of the actual temperature at $t+1$ on the basis of data available up to time t , the innovation signal is the following difference $T(t+1) - \hat{T}(t+1|t)$, (recall that $T(\cdot)$ is the available measurement not the actual tile temperature, referred as $\bar{T}(\cdot)$)

It is worth noting that the Kalman solution is indispensable to estimate the current temperature $\bar{T}(t)$ (and then to predict it one-step-ahead), since the current temperature reading, $T(t)$, is affected by heavy measurement noise, hence it cannot be fully trusted. Kalman filter and predictor, when correctly parametrized, give the stochastic optimal solutions since they suitably balance their “trust in the model” and “trust in the temperature reading” according to the model dynamics and the variance of the process and measurement noises. With the presented identification method we can provide all of these parameters, as summarized in Fig. 5.

V. 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

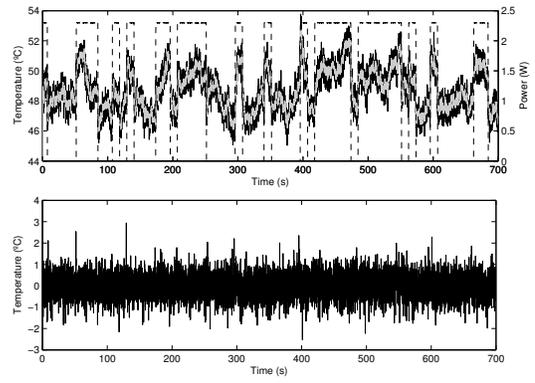


Fig. 6. Tile 15 identification traces. Upper picture: input power (dashed), measured temperature (black) and one-step-ahead predicted temperature (gray). Lower picture: innovation (difference between measured temperature, $T(t)$, and predicted one, $\hat{T}(t)$)

platform has been carried out on a 700s-long training set, with a sampling time $T_s = 100ms$. The algorithm converges after 60 iterations.

The identified poles are $p_{1,15} = 0.9599$, $p_{2,15} = 0.1090$ for tile 15 and $p_{1,16} = 0.8993$, $p_{2,16} = 0.0824$ for tile 16.

$\chi^2(8)$ tests on the innovations, obtained by suitable Kalman predictors (see previous Section), are 5.6 and 14.2 for the tiles 15 and 16, respectively. This shows the whiteness of the identification residuals (or innovations), confirming the validity of the adopted second order structures, when measurement noise in the outputs is directly taken into account; while it can be neglected at the input side, as supposed in Section IV. The measurement and process noise variances estimated with the proposed method are $\sigma_{v,15}^2 = 0.2438$, $\sigma_{w,15}^2 = 0.0342$, for tile 15, and $\sigma_{v,16}^2 = 0.2101$, $\sigma_{w,16}^2 = 0.0636$, for tile 16. These measurement noise variances lead to a poor SNRs (8.5dB for tile 15 and 9.2dB for tile 16).

Fig. 6 reports the predicted temperatures and the related innovations (i.e. difference between measurements and predictions) for tile 15 in the identification test; a similar figure for tile 16 is omitted for brevity. The mean values of the innovations are negligible, as expected ($0.0062^\circ C$ for tile 15 and $-0.0080^\circ C$ for tile 16); while the RMS are $0.5787^\circ C$ and $0.5712^\circ C$, respectively, which correspond to variances of 0.3349 and 0.3262. In the first picture of Fig. 6, it can be noted as 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 second picture and recall the above mentioned RMSs and 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. 7 reports the predicted temperatures and the related innovations for tile 15 in a validation test; again figure for tile 16 is omitted for brevity. The mean values of the innovation are small ($0.0437^\circ C$ for tile 15 and $0.0644^\circ C$ for tile 16); while the RMS are $0.5867^\circ C$ and $0.6003^\circ C$, respectively, which correspond to variances of 0.3442 and 0.3604. 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.

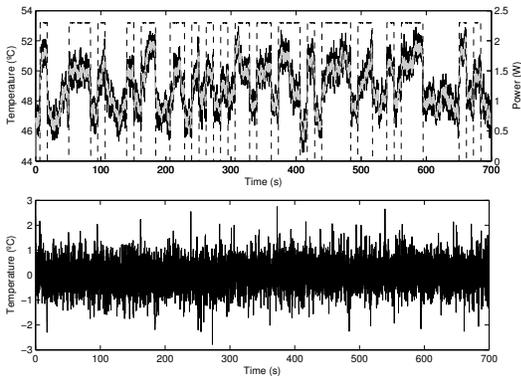


Fig. 7. Tile 15 validation traces. Upper picture: input power (dashed), measured temperature (black) and one-step-ahead predicted temperature (gray). Lower picture: innovation (difference between measured temperature, $T(t)$, and predicted one, $\hat{T}(t)$)

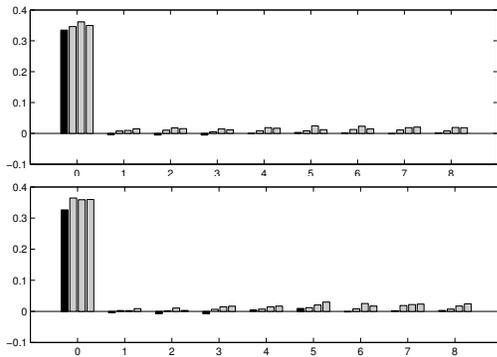


Fig. 8. Tile 15 (up) and Tile 16 (down): residual (or innovation) correlations at time lags 0 . . . 8; training set (black), validation sets (gray)

An additional test has been carried out on the validation data sets, the whiteness of the innovations have been checked obtaining good results. As a sample, in Figs. 8 the first nine autocorrelations (or autocovariances) of the residuals obtained in the original training set and other three validation sets for the tiles 15 and 16 are reported. Autocorrelations at lags larger than zero are lower than 7% of the residual variance. Hence the residuals are always almost uncorrelated, i.e. “practically white”. This, again, highlights the effectiveness of the proposed solution.

In Fig. 9, for tile 15, a simulation comparison of the 1W power step responses of the 2nd order model, obtained with the proposed solution, and the 10th order model identified by standard LS (see Section III) is reported to highlight the better physical consistency of the former. It is worth noting that the simulation test has been carried out assuming neighbor and ambient temperatures equal to zero, therefore the temperature values are small and the time constants highlighted in the 2nd order model response are rather fast when compared to the expected full SCC dynamics (with a settling time of a few hundreds of seconds).

VI. CONCLUSION

In this paper we have proposed a novel identification strategy for ARX-like models, based on an advanced bias-

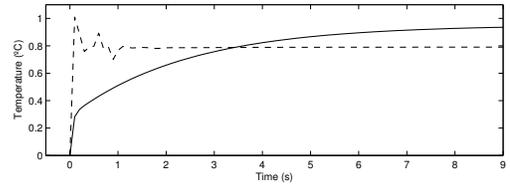


Fig. 9. Tile 15: response to a power step for the 10th-order ARX model (dashed) and the 2nd-order ARX with noisy output model (solid)

compensated least-squares algorithm, to overcome limitations due to noisy measurements in learning thermal model of real multicore devices. In the noisy scenario characterizing SCC, we have shown that the standard ARX-LS algorithm does not produce valid low-order models. The proposed solution is instead capable of filter-out the noise, learning for each tile a valid second-order thermal model. Validation tests confirm the quality of the obtained models, which can be easily integrated in a Kalman predictor to get optimal estimates of the real silicon temperatures, filtering out the thermal sensor measurement noise.

VII. ACKNOWLEDGMENTS

This work was supported, in parts, by the EU FP7 ERC Project MULTITHERMAN (GA n. 291125) and the EU FP7 Project Phidias (GA n. 318013). Technical support from Intel Labs. Braunschweig is also acknowledged.

REFERENCES

- [1] J. Kong, S. W. Chung, K. Skadron, “Recent Thermal Management Techniques for Microprocessors”. *ACM Comput. Surv.* 2010.
- [2] Y. Wang, K. Ma and X. Wang, “Temperature-Constrained Power Control for Chip Multiprocessors with Online Model Estimation”, *Proc. of ISCA*, 2009.
- [3] A. Bartolini et al. “Thermal and energy management of high-performance multicores: Distributed and self-calibrating model-predictive controller”, *IEEE Trans. Parallel Distrib. Syst.*, vol. 24(1), pp.170–183, 2013.
- [4] F. Beneventi et al. “An effective gray-box identification procedure for multicore thermal modelling”, *IEEE Trans. Comput.*, PP(99):1, 2012.
- [5] J. Howard et al. “A 48-Core IA-32 message-passing processor with DVFS in 45nm CMOS”, *Proc. of ISSCC*, pages 108–109, Feb. 2010.
- [6] R. Cochran and S. Reda, “Consistent runtime thermal prediction and control through workload phase detection”, *DAC*, 2010.
- [7] A. Coskun, T. Rosing, and K. Gross, “Utilizing predictors for efficient thermal management in multiprocessor SoCs”, *IEEE T. Comput. Aid. D.*, 28(10):15031516, oct. 2009.
- [8] D.-C. Juan et al. “A learning based autoregressive model for fast transient thermal analysis of chip-multiprocessors”, *Proc. of ASP-DAC* 2012.
- [9] S. Sharifi and T.S. Rosing, Accurate direct and indirect on-chip temperature sensing for efficient dynamic thermal management, *IEEE T. Comput. Aid. D.*, vol. 29, pp.1586–1599, 2010.
- [10] L. Ljung, *System Identification – Theory for the User* Prentice Hall, Englewood Cliffs, NJ, 1999.
- [11] A. Bartolini et al. A system level approach to multi-core thermal sensors calibration. In *PATMOS’11*, pages 22–31, 2011.
- [12] M. Sadri, A. Bartolini, and L. Benini, Single-chip cloud computer thermal model. In *Thermal Investigations of ICs and Systems Proc. of THERMINIC*, 2011.
- [13] 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, pp. 242-255, 2010.
- [14] R. Diversi, “A bias-compensated identification approach for noisy FIR models”, *IEEE Signal Processing Letters*, vol. 15, pp. 325-328, 2008.
- [15] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.