Post-silicon power mapping techniques for integrated circuits

Sherief Reda*, Abdullah N. Nowroz, Ryan Cochran, Stefan Angelevski

School of Engineering, Brown University, 184 Hope St, Providence RI 02912, United States

1. Introduction

It is possible to sort modern chips into two main categories: power limited and hot spot limited [2]. Power-limited chips are used in battery-operated mobile devices where it is necessary to minimize the total power consumption to maximize the device's battery life. The wide range of applications (e.g., 3D games, web browsing, video decoding) that execute on mobile devices complicates the problem of power estimation. On the other hand, designers of hot-spot limited chips seek to minimize the peak junction temperatures which arise from the spatial and temporal distribution of power densities. Excessive chip temperatures impact reliability and constrain performance.

Design of power-limited and hot-spot limited chips require the computation of accurate, fine grained power estimates for the different circuit blocks. Design-time estimates are typically inaccurate in comparison to the true post-silicon power consumption for the following reasons: modeling inaccuracies during design time [3], unpredictability of runtime workloads [3,4], and manufacturing process variations [4–7]. Our objective is to provide accurate post-silicon spatial power estimates for the various circuit blocks using the runtime thermal infrared emissions emitted from the back side of semiconductor chips. Spatial power mapping from thermal emissions has emerged in the past few years through the research work of two groups [2,8]. In one approach, the spatial steady-state power consumption is estimated from temperature by minimizing total squared error [2]. A second approach uses a combinatorial optimization formulation based on genetic algorithms to find dynamic and leakage power contributions of this paper are as follows.

- We elucidate the technical challenges in thermal to power inversion. Spatial discretization and measurement errors in thermal imaging limit the ability to resolve power spatially. In addition, heat conduction has a low-pass filtering effect that attenuates the thermal impact of high-frequency spatial power variations. We investigate these phenomena theoretically and empirically and demonstrate their impact on power estimation accuracy.
- To address the outlined challenges we propose techniques from regularization theory. We propose quadratic formulations that are augmented with Tikhonov regularization methods to improve the accuracy of spatial power estimation. Our techniques limit the impact of measurement errors and improve power estimation compared to previous techniques.
- A major challenge in thermal to power inversion is to be able to validate the power estimations. Accordingly we design and implement a circuit with programmable micro-heaters that can generate power maps with various intensities and spatial constructions. We exercise the circuit with a number of different spatial power maps and then capture the thermal emissions in the mid-wave infrared range from the back-side of the chip using a cryogenic-cooled InSb-based infrared

An earlier version of this paper appeared at the International Symposium on Low Power Electronics and Design (ISLPED) 2010 [1]. This expanded article contains numerous novel material, including (1) detailed mathematical elaborations on the challenges of thermal to power inversion; (2) automatic techniques for singular value filtering in the proposed thermal to power inversion methodology; (3) a description of our thermal calibration method; (4) the implementation details of the test chip; (5) new test maps and experimental results that outclass our previous results; (6) new experiments to analyze the source of errors in measurements; and (7) results from spatial power maps with varying power intensities.

* Corresponding author.

E-mail addresses: Sherief_Reda@brown.edu (S. Reda),
abdullah_nowroz@brown.edu (A.N. Nowroz),
ryan_cochran@brown.edu (R. Cochran),
stefan_Angelevski@brown.edu (S. Angelevski).

© 2011 Elsevier B.V. All rights reserved.
camera. Our work is the first to validate its post-silicon power characterization estimates. We also trace the source of errors using statistical and frequency-domain analyses techniques.

- To address the difficulties encountered in realistic experimental setups, we provide novel techniques (1) to model the relationship between power and temperature, and (2) to calibrate thermal imaging equipment to compensate for the different material emissivities of semiconductor chips.

The organization of this paper is as follows. Section 2 discusses the motivation for this work. In Section 3 we elucidate the challenges associated with thermal to power inversion, and in Section 4 we describe our proposed methodology for thermal to power inversion that handles the outlined challenges. In Section 5 we describe our proposed test chip design and modeling techniques. In Section 6 we provide our experimental and validation results. Finally, Section 7 summarizes the main conclusions of this work.

2. Motivation

The design flow of a modern integrated circuit is highly complex. In a state-of-the-art custom chip, it is expected to spend about half of the time to market in the design phase, starting with architecture and culminating with placement, routing, timing closure, and meeting the power specifications. The initial design could take more than a year, depending on complexity and IP reuse [9]. Once completed, the initial design is shipped for fabrication and first silicon is received about three months later. At this point, the silicon still has to be debugged, nearly always requiring at least one major re-spin of the design [10]. The debug phase, including working out yield problems as volume is ramped up, can easily take as long as the initial design where in many cases the mismatch between pre-silicon and post-silicon requires major changes in the design and implementation of the chip. In a wide study from 2005, it was shown that 70% of new designs go at least one design re-spin to fix post-silicon problems and that 20% of these re-spins are due to power-related issues [10]. It is likely that these figures are much higher now as chips are more complex as they accommodate larger number of transistors. Popular examples of large mismatches between pre- and post-silicon estimates include IBM’s Cell processor, where the post-silicon power and thermal measurements led to large changes in its specifications and implementation [11].

By performing post-silicon power validation on real chips, it is possible to improve the final integrated circuit. The results of post-silicon power characterization can improve the design process during re-spins or for future designs in the following ways.

- High-level power modeling tools rely on the use of parameters that are estimated from empirical data [3,4]. The power characterization results can be used to calibrate and tune high-level power modeling tools.
- The power characterization results can drive heat sink design. Passive heat sinks remove heat indiscriminately from the die, and thus, their design is mainly driven by total power consumption. Active heat removal systems, such as thermoelectric Peltier coolers [12], can make use of the true post-silicon power characterization results to maximize their heat removal capabilities.
- To evaluate “what if?” design re-spin questions, the power characterization results can substitute the power simulator estimates and directly feed thermal simulators. For example, if the thermal characterization results are unacceptable, then the layout can be changed to reduce the spatial power densities and hot spot temperatures. The power characterization results are fed together with the new layouts to a thermal simulator to evaluate the impact of the changes.
- The post-silicon power characterization results can also force a re-evaluation of an integrated circuit specifications (e.g., operating frequencies).

The focus of this paper is to devise algorithmic and experimental techniques for post-silicon power mapping by using infrared imaging techniques, where the captured thermal emissions from the backside of the die are inverted to yield the underlying spatial power maps.

3. Identified challenges in thermal to power inversion

The physical relationship between power and temperature is described by the physics of heat transfer. Mainly, heat conduction by diffusion governs the relationship between power and temperature in the die and the heat spreader. The steady-state DC heat diffusion equation is given by

$$\nabla \cdot (k(x,y,z) \nabla t(x,y,z)) = -pt(x,y,z), \tag{1}$$

where \(t(x,y,z)\), \(p(x,y,z)\), and \(k(x,y,z)\) are the temperature, power density, and the thermal conductivity at location \((x,y,z)\) respectively [7]. The heat transferred at the boundary between the die/heat spreader and the surrounding fluid is described by Fourier’s law for heat transfer, and it is proportional to the temperature difference between the boundary of the die/heat spreader and the temperature of the surrounding fluid (air or liquid). The constant of proportionality is the heat transfer coefficient which depends on the geometry of the heat sink, the fluid used for heat removal and its convection characteristics (e.g., speed and laminarity/turbulence) [13].

In any practical implementation, the heat equation must be discretized. This discretization comes from the finite memory size of any computer and more importantly from the use of thermal imaging equipment with limited spatial resolution. In a discretized form, the continuous temperature signal is represented by a vector \(t\) that gives the true temperatures at a discrete set of backside die locations. The length of the vector \(t\) is determined by the spatial resolution of the infrared camera and the dimensions of the die. The continuous power signal is represented by a vector \(p\) that gives the power consumption of each of the circuit’s units. In such case, Eq. (1) is approximated by the following linear matrix formulation:

$$Rp = t, \tag{2}$$

where the matrix \(R\) is a linear operator that captures the impact of all modes of heat transfer. The objective of the thermal to power inversion problem is to find the vector \(p\) that leads to the temperatures \(t\). An inversion problem well-posed if it satisfies three conditions (first outlined by Hadamard [14]): existence, uniqueness, and stability. In the context of thermal to power inversion, existence means that for every measured temperature map, there exists a power map that leads to the temperature map; uniqueness means that exists one and only power map that leads to the temperature map; and stability means that small perturbations in the temperature measurements lead to small perturbations in the estimated power maps.

In the next two subsections, we identify and describe two main challenges that can lead to ill-posed thermal to power inversion problems. The first challenge arises from the physics of heat diffusion and the second challenge arises from the physics of noise and the technology used to perform thermal imaging.
3.1. First challenge: spatial filtering

The first limit in thermal to power inversion is inherent to the physics of heat transfer. It is well-established in the literature that a chip's temperature map is a low-pass filtered form of its power density map [15,16]. While the power density can spatially vary abruptly according to the chip's layout and application behavior, the temperature will always vary smoothly in space. This low-pass filtering effect is governed by Eq. (1). In this subsection we will consider a 2-D die with homogenous material, i.e., $k(x,y,z) = k$, to simplify the discussions and to focus on the key concepts. In this case, Eq. (1) simplifies to

$$k \left( \frac{\partial^2 t(x,y)}{\partial x^2} + \frac{\partial^2 t(x,y)}{\partial y^2} \right) = -p(x,y).$$  \hspace{1cm} (3)

The 2D Fourier transform of some function $t(x,y)$ is defined by

$$F_t(u,v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x,y) e^{-j2\pi uv} \; dx \; dy,$$  \hspace{1cm} (4)

where $u$ and $v$ are variables that represent the spatial frequencies in the $x$ and $y$ directions respectively. Applying the 2-D Fourier transform to both sides of Eq. (3) gives

$$k(-u^2 F_t(u,v) - v^2 F_t(u,v)) = -F_p(u,v).$$  \hspace{1cm} (5)

Thus, the power-temperature transfer function, $G(u,v)$, is equal to

$$G(u,v) = \frac{F_t(u,v)}{F_p(u,v)} = \frac{1}{k(u^2 + v^2)}$$  \hspace{1cm} (6)

It is clear from Eq. (6) that the higher frequency components of spatial power maps will be subjected to greater attenuation. This spatial filtering phenomenon is illustrated in Fig. 1, where we create checkerboard power maps with increasing spatial frequency in a test chip and measure their emissions. The maps in Fig. 1a have the same amount of total power but they differ in their spatial frequencies. Fig. 1b gives the resultant emissions demonstrating that the variations are attenuated as the spatial frequency increases. For example, the standard deviation drops from 184 mK to 111 mK and 64 mK as the spatial frequency is increased. In a simulation-based environment with double precision floating point numbers, attenuation is not an issue, but in real systems with physical and technological limits on their detectors and analog to digital converters, attenuation is a major problem as it degrades the signal to noise ratio. This degradation can attenuate the signal to a level below the detection sensitivity of the infrared imaging equipment, leading to irreversible loss of information. In the discrete form of the heat diffusion equation (Eq. (2)), the low-pass filtering effect implies that there exists at least one power vector $p_s$ that belongs to the null space of $R$. If vector $p$ satisfies Eq. (2), then $p + p_s$ also satisfies the equation. In practical terms, $p_s$ is in the null space of $R$, if $R p_s$ is below the detection sensitivity of the imaging system. Thus, the low-pass filtering effect could render the temperature to power inversion problem ill-posed as the uniqueness condition is violated.

3.2. Second challenge: noise in measurements

The second limit arises from discretization and measurement noise introduced during infrared imaging. A number of noise phenomena could lead to errors in the measurements [17]. Sources of noise include (i) dark noise caused by random generation of electron-hole pairs in quantum detectors; (ii) thermal noise caused by the agitation of charge carrier in the electronic readout circuitry; (iii) flicker noise which is inversely proportional to the emission frequency; and (iv) discretization errors introduced by the analog to digital converters of the imaging system. Mathematically, the impact of these errors can be expressed as

$$R p + e = t + e = t_m,$$  \hspace{1cm} (7)

where the vector $e$ denotes the errors in measurements introduced during imaging, and the vector $t_m$ denotes the measured temperatures. To understand the impact of measurement errors on the inversion problem, we will use the Singular Value Decomposition (SVD). SVD decomposes a matrix into a weighted sum of ordered matrices. That is

$$R = \sum_{i=1}^{\text{rank}} s_i u_i v_i^T \quad \text{and} \quad R^{-1} = \sum_{i=1}^{\text{rank}} \frac{1}{s_i} v_i u_i^T,$$  \hspace{1cm} (8)

where the $s_i$'s are the singular values, the $u_i$'s form a set of orthonormal vectors, and the $v_i$'s form a set of orthonormal vectors such that $R u_i = s_i v_i$. The operator $^T$ denotes the transpose operation. Using the non-zero singular values of the SVD of $R$ and Eq. (7), we find that

$$\text{estimated } p = R^{-1} t_m = R^{-1} (t + e) = p + \sum_{i=1}^{\text{rank}} \frac{1}{s_i} (u_i^T e) v_i.$$  \hspace{1cm} (9)

Thus, small singular values amplify the impact of noise during inversion [14]. Using Eq. (9), we can bound the error in power estimation, $\delta p$, as follows. Since $\| \delta p \|_2 \leq \| R^{-1} \|_2 \| \delta t \|_2$, where $\| \cdot \|_2$ is the $L_2$ norm. Thus,

$$\| \delta p \|_2 \leq \frac{\| \delta t \|_2}{\| \delta R \|_2} = \frac{\| \delta p \|_2}{\| R \|_2} \leq \frac{\| \delta t \|_2}{\| R \|_2} \leq \frac{\| \delta p \|_2}{\| s_{\min} \|_2},$$  \hspace{1cm} (10)

where we used the fact that $\| R \|_2 = \| s_{\max} \|_2$ and $\| R^{-1} \|_2 = 1/\| s_{\min} \|_2$ [18]. The ratio $s_{\max}/s_{\min}$ is the condition number of $R$ and it controls the propagation of errors from the measurements to the estimated power. For example, if the condition number of our chip's $R$ matrix is in the order of $10^4$ then the inversion algorithm could transform a tiny milli-Kelvin error – a typical number in cryogenic thermal quantum detectors – in temperature measurement into a significant error in power estimation. This amplification of noise leads to ill-posed inversion as the stability condition

---

**Fig. 1.** Illustrating the impact of spatial low pass filtering on the intensity and variations of thermal emissions. (a) Power maps and (b) thermal emissions.
is violated, where slight changes in the measured temperatures lead to the large changes in the estimated power.

4. Proposed methodology for thermal to power inversion

The objective of thermal to power inversion is to find the best power map \( p \) that minimizes the total squared error between the true temperatures \( t \) as computed using Eq. (2) and the measured temperatures; that is, \( \min ||Rp - t||_2^2 \). As mentioned in Section 3, this objective is challenging to achieve because (1) spatial filtering could lead to ill-posed inversion with a singular or ill-conditioned matrix \( R \) and (2) measurement noise can deviate \( t_{\text{m}} \) from the true temperature \( t \). To address these two challenges, we propose two techniques. We introduce \textit{regularization theory} techniques to reduce the impact of measurement noise, and we introduce constraints on the solution space to reduce the possibility of getting a wrong solution due to the existence of a null space for \( R \). We explain the details of these two techniques in the rest of this section.

To reduce the impact of measurement noise and the potential ill-posed nature of the problem, we propose techniques from \textit{regularization theory} [14]. Regularization techniques consider a family of approximate solutions using a positive parameter called the \textit{regularization parameter}. When the measured thermal data is noise-free, the solution converges to the true power solution as the regularization parameter goes to zero. When the measured thermal data is noisy, the solution converges to the true power solution as the regularization parameter goes to zero. Tikhonov regularization finds the power solution that gives the least total squared error while simultaneously minimizing the \( L_2 \) norm; that is,

\[
p_a = \arg\min_{p} ||Rp - t_{\text{m}}||_2^2 + \sigma ||p||_2^2,
\]

where \( \sigma > 0 \) is the regularization parameter that controls the minimization emphasis between the two terms of the objective function [19]. The first term \( ||Rp - t_{\text{m}}||_2^2 \), which gives the total squared error, controls how well the power estimates, \( p \), lead to temperatures that match the measurements (which could be noisy). If the value of this term is large then the solution is far from the true power, but a small value for this term could lead to a fitting that is driven by noise. The second term \( ||p||_2^2 \) controls the regularity of the solution. Large values for this term could lead to solutions that are dominated by high-frequency measurement noise. Using one of our captured thermal traces, the trade-off between the two terms as a function of \( \sigma \) is illustrated in Fig. 2.

Increasing the value of \( \sigma \) traces the curve from the top-left corner to the lower-right corner, and thus the trade-off curve is typically referred to as the \textit{L-curve} [19]. Small values for \( \sigma \) can lead to solutions dominated by noise, while large values for \( \sigma \) lead to more regularized solutions.

To develop further insight into the role of the regularization parameter \( \sigma \), we utilize the SVD, but it is first necessary to re-cast the objective of Eq. (11) as a least squares problem as follows:

\[
p_a = \arg\min_{p} \left\| \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} p - \begin{bmatrix} t_{\text{m}} \\ 0 \end{bmatrix} \right\|_2^2.
\]

Thus, the solution to the least square estimation problem is

\[
\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \begin{bmatrix} p_a \\ \sigma \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \begin{bmatrix} t_{\text{m}} \\ 0 \end{bmatrix}.
\]

\[
p_a = (R_1^T R_1 + \sigma I)^{-1} R_1^T t_{\text{m}}.
\]

The SVD expansion of Eq. (8) can be written as \( R = U S V^T \), where \( U \) and \( V \) are unitary matrices formed from the \( u_i \) and \( v_i \) vectors respectively, and \( S \) is a diagonal matrix with the diagonal elements are the non-zero singular values \( s_i \). Using the SVD,

\[
p_a = (U S^2 V^T)^{-1} V S^2 U^T t_{\text{m}} = V (S^2 S + \sigma I)^{-1} S^2 U^T t_{\text{m}} = V \text{diag} \left( \frac{s^2_i}{s^2_i + \sigma} \right) U^T t_{\text{m}} = \sum_{i} \frac{s^2_i}{s^2_i + \sigma} v_i u_i^T t_{\text{m}}.
\]

Eq. (13) can be further analyzed as follows:

\[
p_a = (V S^2 U^T S^2 V^T + \sigma V V^T)^{-1} V S^2 U^T t_{\text{m}} = V (S^2 S + \sigma I)^{-1} S^2 U^T t_{\text{m}}
\]

Comparing Eq. (14) against Eq. (9), we find that each singular value in the SVD decomposition of Eq. (9) is multiplied by a factor \( s^2_i/(s^2_i + \sigma) \) known as \textit{Tikhonov attenuation factor} [20,19]. Fig. 3 gives the value of the Tikhonov attenuation factor \( s^2_i/(s^2_i + \sigma) \) as a function of the singular value \( s_i \) for \( \sigma = 0.5 \). The figure shows that the attenuation factor essentially functions as a \textit{filter function} that filters out singular components that are small relative to \( \sigma \) and passes singular components that are large relative to \( \sigma \) [20]. Singular values of zero value are totally eliminated. The attenuation of the small singular components makes the inverse problem more well-conditioned and controls the error propagation as governed by Eq. (10). Good values for \( \sigma \) can be found by inspecting the L-curve. One possibility is to use the corner of the curve [19]. The corner is the point on the L-curve with the maximum curvature. In Fig. 2 the corner occurs at \( \sigma \) is equal to 1.6.

![Fig. 2. Tradeoff between \(|Rp - t|\) and \(|p|\) as a function of the regularization parameter \( \sigma \).](image)

![Fig. 3. Attenuation of the singular values as a function of \( \sigma \). The regularization parameter \( \sigma = 0.5 \).](image)
To reduce the possibility of getting a wrong solution due to the existence of a null space for \( R \), we introduce additional constraints on the solution space. One possible constraint is that the sum of the elements of the power map must be equal to the total power consumption of the chip \( p_{\text{total}} \) and that these elements are nonnegative; i.e.,
\[
|p|_1 = \sum_{i} p_{i} = p_{\text{total}} \quad \text{and} \quad p \geq 0,
\]
(15)
where \( | \cdot |_{L_1} \) is the \( L_1 \) norm and \( p_{\text{total}} \) is the total power consumption of the chip which could be measured externally using a digital multimeter. Thus, if multiple solutions exist, then the solution with least total power error is chosen. Note that because the total (and average) power is constrained by Eq. (15), the regularization term in Eq. (11) controls the variance in spatial power estimates. In practice, any digital multimeter has a tolerance, \( \text{tol} \), in its measurements (the tolerance is typically listed in the multimeter’s data sheet), and thus it is better to replace the constraint of Eq. (15) by the following constraints:
\[
|p|_1 \leq p_{\text{total}} + \text{tol},
\]
(16)
\[
|p|_1 \geq p_{\text{total}} - \text{tol},
\]
(17)
and
\[
p \geq 0.
\]
(18)

Our overall inversion methodology is summarized in the algorithm given in Fig. 4.

**Comparison to previous approaches:** Previous numerical approaches in thermal to power inversion mainly focused on minimizing the total squared error between the temperatures computed from the power estimates and the thermal measurements [21,2,8]. These approaches can produce suboptimal results as they ignore the ill-posed nature of the problem where measurement noise and spatial diffusion reduce the accuracy of power estimation. Compared to previous approaches, our proposed numerical techniques handle many of the challenges associated with inversion. We introduce regularization theory to reduce the impact of noise and improve the numerical instability in the model matrix \( R \) by eliminating zero singular values and filter small singular values. We also introduce constraints on the solution space to eliminate the possibility of getting multiple solutions. Our constraints are customizable according to the tolerance associated with electrical current measurement equipment.

### 5. Test chip design and modeling

One of our main research goals is to assess the accuracy of thermal to power inversion. Previous studies [2,8] presented spatial power estimates for the different processor blocks “as is” [2,8]. None of the previous studies provided a thorough validation for the accuracy of their spatial power estimations. This lack of validation is the result of the experimental setup that previous studies chose. Earlier experimental setups used general-purpose processors (a dual core PowerPC 970MP [2] and an AMD Athlon 64 processor [8]) in which it is impossible to fully control the underlying spatial power consumption, and there exists no alternative means to verify the spatial power maps. To address this issue, we designed our experimental test chip with special emphasis on the ability to estimate the spatial power consumption through two different means, and hence we are able to provide estimations for the accuracy of our proposed power characterization methodology.

#### 5.1. Test chip design and implementation

The basic unit of our circuit is a *programmable micro-heater*, which consists of a number of ring oscillators (ROs) that are controlled by flip-flops that determine the operational status of the micro-heater. We create two kinds of micro-heater designs:

1. **Bi-level micro-heaters:** A bi-level micro-heater consists of nine 15-stage ROs together with one flip-flop that controls their operational status. If the D Flip-flop (DFF) holds a binary value of 1 then the heater is turned on; otherwise, it is turned off. When enabled, each micro-heater consumes 25 mW. Using the programmable heater, a grid that consists of 10 \( \times \) 10 micro-heaters is created as shown in Fig. 5a. In the grid structure, the output of each DFF is connected to the input of the DFF of the consecutive heater forming a scan chain.

2. **Multi-level micro-heaters:** A multi-level micro-heater consists of ROs that can be programmed to one of the following configurations: 9 51-stage ROs with a power consumption of 25 mW; 18 25-stage ROs with a power consumption of 60 mW; 27 19-stage ROs with a power consumption of 102 mW; and 36 13-stage ROs with a power consumption of 142 mW. Thus, each micro-heater}

---

**Procedure:** Thermal to power inversion method  
**Input:** Thermal map \( t_m \), \( p_{\text{total}} \), \( R \)  
**Output:** \( p \)

1. Given \( t_m \) and \( R \), construct the L-curve
2. Identify \( \alpha \) from the corner of the L-curve
3. Compute \( R_s = \sum s_i u_i v_i \)
4. Solve the quadratic program: \( \min ||R_s p - t_m||^2 \) such that \( ||p||_1 \leq p_{\text{total}} + \text{tol}, ||p||_1 \geq p_{\text{total}} - \text{tol} \) and \( p \geq 0 \)
5. Return the solution of the quadratic program \( p \)

---

**Fig. 4.** Proposed thermal to power inversion methodology.

**Fig. 5.** Grid of microprogrammable heaters. (a) Grid of bi-level micro-heaters and (b) grid of multi-level micro-heaters.
block offers five different power levels: 0, 25, 60, 102, and 142 mW. The DFFs associated with each micro-heater determine its status. Using the programmable micro-heater blocks, a grid that consists of $6 \times 6$ blocks is created as shown in Fig. 5b. In the grid structure, the output of each DFF is connected to the input of the DFF of the consecutive heater forming a scan chain.

In both designs, the advancements of the programming bits in the chain is controlled by the clock signal. To create any desired power pattern, we inject control bits into the flip-flops of the micro-heaters to selectively turn on the micro-heaters that correspond to the required power pattern. Our experimental novelty of using a chip of programmable micro-heaters enables us to achieve the following two experimental goals which have not been attained in previous works:

1. The grid structure of the micro-heaters, where every micro-heater can be selectively controlled, enables us to create any desired spatial power map on a real chip. Previous works used processors in which independent control of various processor blocks is impossible. The programmable nature of the grid enables the generation of a large number of different maps that can be used to test the accuracy of the thermal to power inversion methodology.

2. The regular and homogenous structure of the micro-heater grid enables us to estimate the power consumption of each micro-heater by simply measuring the total power consumption of the grid and dividing it by the number of enabled micro-heaters. The locations of the enabled heaters are known by construction. Hence, we are able to construct the spatial power map through an alternative path to infrared emissions. Previous works lacked this ability to validate their spatial power estimations through different means.

For implementation, we choose a 90 nm Altera Stratix II (EP2S180) field programmable gate array (FPGA) with 180,000 logic elements with total die dimensions of 23 mm $\times$ 24 mm. The regular fabric of the FPGA ideally fits our design. For our experiments, we use a relatively homogenous section of the die that spans 7.2 mm $\times$ 7.9 mm for the bi-level 10 $\times$ 10 grid as shown in Fig. 6a, and a section that spans 8.8 mm $\times$ 8.9 mm for the multi-level 6 $\times$ 6 grid. The micro-heater blocks are mapped to the logic array blocks at the precise grid locations using Altera’s Quartus II placement assignment editor. In order to capture the chip’s thermal emissions it was necessary to remove the heat spreader. While removing the heat spreader is going to change the spatial thermal behavior, the change in spatial thermal emissions does not change the underlying dynamic power consumption ($fCV^2$) [22], which is weakly dependent on temperature. The spatial power consumption remains relatively intact, and the new interactions between temperature and power are captured in the learned $R$ matrix as described in the next paragraph.

5.2. Measuring operator matrix ($R$) of the test chip

Given the test chip, it is necessary to measure the operator matrix $R$. The matrix $R$ can be measured in a column-by-column basis as follows. Enabling only one micro-heater at a time is mathematically equivalent to setting the vector $p$ to be equal to $[0 \ldots p_k \ldots 0]^T$, where $p_k$ is the power consumption of the $k$th enabled block. For each micro-heater location, we enable the selected micro-heater as shown in Fig. 7a and record the emitted temperatures across the field as shown in Fig. 7b. If $t_k$ denote the thermal emissions captured from enabling the $k$th micro-heater, then column $k$ of matrix $R$ is equal to $t_k/p_k$. We automate the whole process in order to measure the columns of $R$ with fast turn-around time.

We utilized the programmability of FPGAs to estimate the model matrix $R$. In custom chips with generic designs, the matrix $R$ can be estimated in the same conceptual way but through a different implementation approach [2,8]. One approach is to turn off the chip, and scan a laser beam with known power density to deliver the power from the outside to the regions of interest. The scanning of the laser system can be automated by using a pair of galvo-directing mirrors [2]. Our method uses the programmable nature of our design to get the same results of the expensive laser scanning system but in a much cheaper way. Another approach is to use the actual design and layout of the chip to conduct a fluid dynamic simulation coupled with a heat diffusion simulation to estimate the matrix $R$ [2,8].

There is always a possibility that errors might occur during the estimation of $R$. If $R$ is estimated from direct measurement, then measurement noise can introduce errors, and if $R$ is estimated from simulations, then unrealistic simulation assumptions can lead to errors. Our experimental results in the next section show that the overall power estimation error arising from our inversion procedure is relatively small.

6. Validation and experimental results

To test our post-silicon power characterization methodology, we put together the following experimental setup which is shown in Fig. 8.

- To capture the thermal emissions from the back-side of our die, we use a FLIR SC5600 infrared camera with a mid-wave
levels of the camera’s A/D converters) to temperatures. One problem is that radiation intensity necessary to convert photon measurements recorded by the infrared camera can measure the temperature through the chip under test [8]. The camera’s cryogenic cooled InSb detectors have a sensitivity of about 15 mK noise equivalent temperature difference (NETD).

- We use an Agilent E3634A power supply to supply and measure the total power consumption of FPGA chip.

6.1. Thermal calibration

Measuring the temperature is slightly complicated because an infrared camera is really a photon detector that measures the infrared radiation intensity at different parts of the chip. Thus, it is necessary to convert photon measurements recorded by the camera to temperatures. One problem is that radiation intensity is not constant among different materials even if they are at the same temperature. Perfect radiation emitters are black bodies with an emissivity of 1. The emissions of real materials are a fraction of the black-body level, and each material is characterized with an emissivity value, which is defined as the ratio of that material’s thermal emission to that of a perfect black-body at the same temperature [17]. As integrated circuits are comprised of a mixture of different materials (e.g., copper, silicon and dielectrics) with varying spatial material densities, the radiation intensities of different parts of the chip could be different even if the chip is held at isothermal temperature by external means.

To handle the emissivity problem, previous approaches coated the backside of the die with a material of constant emissivity [2]. The downside of this approach is that it obstructs the silicon transparency and only measures the heat emissions resulting from the projections of the internal emissions on the backside of the die. Previous approaches were thus forced to thin the backside silicon to reduce the amount of internal projections. To get a more accurate thermal imagery, we avoid coating the backside of the silicon and instead devise a pixel-by-pixel calibration process that translates the captured emissions (as measured by the digital levels of the camera’s A/D converters) to temperatures.

The relationship between the digital level $D_i$ and temperature $t_i$ of a pixel $i$ can be modeled by an exponential function

$$D_i = a_i e^{b_i t_i},$$

where $a_i$ and $b_i$ are per-pixel coefficients that are function of many factors including emissivity, path to length, and integration time [23]. The exponential relationship arises from the physics of photon detectors in which the current of an infrared-sensitive diode depends exponentially on the incident radiation. If there were no emissivity differences between the different pixels then $a_i$ and $b_i$ would be chip-wide constants, but instead they must be computed for each pixel. For example, Fig. 9 shows two curves relating the temperatures to the measured digital levels for two different pixels on the test chip. Our pixel-by-pixel calibration procedure is simple. We first turn the chip off and then force it to an isothermal status through external means,1 and then measure the temperature of the chip through a thermocouple as shown in Fig. 8. Two thermocouples placed at opposite ends of the die could be used to verify chip-wide steady-state attainment. Once steady-steady is reached, the digital levels of all pixels are captured using the camera. This process is repeated for a few temperatures and then the calibration curves (as the ones given in Fig. 9) are constructed, and the $a_i$ and $b_i$ for every pixel $i$ are found through curve fitting. Fig. 10 contrasts the raw digital thermal levels before calibration and the thermal images after calibration. The images show that the calibration process successfully removes imaging artifacts introduced by emissivity variation.

We conduct and report the results of four experiments:

1. The first experiment compares the estimation accuracy of our proposed method against previous techniques using the bi-

---

1 For high-power chips, thermally controlled infrared transparent oil can be used to force the isothermal status. Interested readers are referred to our earlier publication for oil-based setup [24].
level micro-heater block grid. We analyze the sources of errors using statistical and frequency-domain techniques.

2. The second experiment assesses the effectiveness of our method using random spatial maps constructed using the bi-level micro-heater grid.

3. The third experiment evaluates the accuracy of our power estimation method as a function of the spatial frequency of the power maps.

4. The fourth experiment assesses our power mapping method using the multi-level micro-heater block grid. We demonstrate that our method is capable of handling underlying power maps with various intensities and spatial constructions.

**Experiment 1.** In the first experiment we assess the accuracy of our thermal to power inversion methodology by evaluating its power estimates for a number of reference spatial power maps generated using the bi-level micro-heater grid. The locations of the activated micro-heaters of these maps are illustrated in Fig. 11a, and the measured temperature maps after emissivity calibration are illustrated in Fig. 11b. Given our low-power micro-heater designs, the temperature range, i.e., the difference between the maximum temperature and the minimum temperature in each map, is about 1.5–2 °C. The power maps estimated by just minimizing total squared error as proposed by previous approaches [2] are given in Fig. 11c; and the estimated power maps from our methodology are given in Fig. 11d. The power mapping results in Fig. 11c and d are rounded to the nearest power level. We report the estimation error in percentage which is equal to the sum of the absolute differences between the power estimates and their true values divided by the total power. The average error from using previous approaches is 4.95%, while the average error from using our proposed approach is 1.5%. We also report the number of block heaters that were not estimated correctly (either turned on and estimated to be off or vice versa). Previous techniques give a total of nine incorrectly estimated blocks, whereas our technique give six incorrectly estimated blocks for reduction of 30%. By visually comparing the injected power maps and the estimated maps, we notice that our technique recovers the injected maps to a very good extent.

To understand the source of errors, we conduct the following two analyses. First, we simulate the resultant temperatures if the true power maps are used as inputs; the simulation is basically the result of multiplying the injected power maps by the matrix \( R \). We plot in Fig. 12a the residual error between the measured temperatures and the simulated measurements. We verify that the errors form a normal distribution using the Kolmogorov–Smirnov test, and we compute the standard deviation for each residual distribution. The standard deviations are given as labels (\( \sigma \)) in Fig. 12a. For instance, the first pattern, which has an error of 2.0% in its power map estimates, has a residual standard deviation of 9 mK which means that the large majority of errors (97%) are between \( \pm 18 \) mK. The residual errors fall within the sensitivity limitation (15 mK) of our camera detectors. The first pattern has the highest standard deviation in residuals which implies that it had the highest amount of noise. Our second analysis technique computes Discrete Cosine Transform (DCT) of the true power maps and report them in Fig. 12b (the top-left corner is the lowest frequency). The DCT quantifies the spatial frequencies of an input power map. For each power map, we compute the percentage of
the power signal energy that is present in higher frequencies of the DCT. The third pattern, which has 12.8% error in its power map estimates, has the largest amount of energy, 35%, in the high-frequency range. Thus, we attribute the power estimation error to such high-frequency energy content.

Experiment 2. In the second experiment we use random spatial power maps for testing our methodology. In addition to the maps of Experiment 1. The use of random maps provides a more complete assessment of our proposed method. In Fig. 13a we provide the random maps tested, and the resultant thermal images are provided in Fig. 13b. Fig. 13c gives the resultant power maps from previous techniques, while Fig. 13d gives the results from using our inversion methodology. The results show estimation errors of 0% and 11.8% from our techniques versus 2.1% and 15.7% from previous techniques. We also report the number of block heaters that were not estimated correctly. Previous techniques give a total of 9 incorrectly estimated blocks, whereas our technique give six incorrectly estimated blocks for reduction of 30%. Overall, the results of experiments 1 and 2 show an average error of about 4.4% from our techniques versus 6.6% from previous techniques, where our technique gives a total of 13 incorrectly estimated blocks and previous techniques give 19 incorrectly estimated blocks. Thus, our technique gives 31% improvement in power mapping.

To further understand the reason for the errors in some of the maps, we conduct analyses similar to the ones of Experiment 1. First, we compute the residual error between the simulated

Fig. 13. Accuracy of estimating spatial power estimates from thermal emissions of random power maps. (a) Injected power patterns, (b) resultant temperature measurements, (c) estimated power maps using previous methods, and (d) estimated power maps using proposed method.
temperatures and the measured temperatures and plot the results in Fig. 14a. We verify that the errors form a normal distribution, and we compute the standard deviations for the distributions which are given as labels ($\sigma$) in Fig. 14a. For instance, the second pattern, which has a large error of 11.8% in its power map estimates, has a residual standard deviation of 18 mK. Second, we compute the Discrete Cosine Transform (DCT) of the true power maps and give them in Fig. 14b (the top-left corner is the lowest frequency). Again, the second pattern shows the largest amount of energy, 38%, in its higher-frequency components. Thus, we attribute both the noise and high-spatial frequencies to the large error in power estimation.

**Experiment 3.** To further gain insight into the behavior of thermal to power inversion, we assess the accuracy of our methodology as a function of the spatial frequency of power maps. As discussed earlier in Section 3, the nature of heat conduction on chips leads to a low-pass filtering effect. Hence, we create checkerboard maps of increasing spatial frequencies as illustrated in Fig. 15a, which lead to the thermal emissions illustrated in Fig. 15b. The estimated spatial power maps are given in Fig. 15c. The average errors are: 0.0%, 0.0%, and 11.5%. To provide further analysis into the source of error, we plot the 2-D DCT of the power maps in Fig. 16. The figures clearly show the trend of increased frequencies in the power pattern, where the third pattern has the highest frequency and error. The results agree with our earlier discussions in Section 3 that concluded that
increasing the spatial frequencies of power maps can lead to a deterioration in the accuracy of thermal to power inversion due to the impact of low pass filtering.

**Experiment 4.** In this experiment, we provide results using the multi-level micro-heater grid design. In contrast to the first three experiments where the power of each micro-heater was only restricted to two levels (0 and 25 mW), the micro-heaters in the constructed spatial power maps of this experiment have power levels of multiple intensities (0, 25, 60, 102, and 142 mW). Our inversion procedure can naturally handle any number of levels, and the results of this experiment confirm this capability. Fig. 17a gives the constructed multi-level power maps. The resultant thermal maps are given in Fig. 17b, and the estimated power maps from our inversion procedure are given in Fig. 17c. The power mapping results are rounded to the nearest power level. The power estimation errors of the four maps are 4.76%, 0.00%, 1.81% and 3.10%, with an average of 2.41%. The magnitudes of the power estimation errors from the multi-level power mapping experiment are comparable to the magnitude of the errors in the first three experiments using bi-level micro-heaters. The results of this experiment confirm that our technique is capable of handling a large range of power intensities and spatial power maps.

**7. Conclusions and future work**

In this paper, we have presented a new methodology for spatial post-silicon power characterization using the infrared emissions from the back of the silicon die. We have elucidated the various challenges that underlie thermal to power inversion. We have demonstrated mathematically and experimentally how low-pass filtering, discretization, and measurement errors could all compromise the accuracy of power estimation. We have proposed new techniques from regularization theory to improve the accuracy of temperature to power inversion. Furthermore, we have provided experimental techniques to compensate for the varying emissivity of different chip materials and to measure the thermal resistance model matrix. We designed a highly modular, programmable test chip to create sets of known power maps. Our test chip and realistic infrastructure enabled us to validate our methodology by comparing its power estimates against the known injected spatial power maps. Compared to previous approaches, our experiments demonstrate consistent improvement in power estimation accuracy where we improve the power mapping accuracy by 30%. We have also analyzed the residual errors between the simulated temperatures and the measured temperatures to understand the error sources. Our statistical and frequency-domain analyses quantify the roles of noise and spatial filtering on the accuracy of power estimates.

For future work, we will extend our approach to handle transient analysis. The relationship between temperature and power in transient analysis can be described using state space models. In particular, the temperature vector \( T(k+1) \) at time \( k+1 \) is linked to the temperature vector \( T(k) \) at time \( k \) and the power consumption \( P(k) \) at time \( k \) by \( T(k+1) = AT(k) + BP(k) \). Transient analysis can follow a similar approach to the one described in this paper, where \( T(k+1) \) and \( T(k) \) are measured using the camera, and \( A \) and \( B \) are learned in a similar way to the approach used to learn \( R \).

**Acknowledgments**

This work is partially supported by a DoD ARO grant number W911NF-09-1-0320, NSF grants number 1115424 and 0952866, and an equipment grant from Altera Corporation. We would like to thank J. Renau from UCSC for the insights on the setup of infrared imaging systems, J. Mundy from Brown University for the helpful discussions on inversion problems, P. Temple from Brown University for constructing an acquisition system to automate data collection, and the anonymous reviewers who helped improve the contents of this paper.

**References**


