A Spectral Approach to Scalable Vectorless Thermal Integrity Verification

Zhiqiang Zhao
Electrical and Computer Engineering
Michigan Technological University
Houghton, MI
qzzhao@mtu.edu

Zhuo Feng
Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, NJ
zfeng12@stevens.edu

Abstract—**Existing chip thermal analysis and verification methods require detailed distribution of power densities or modeling of underlying input workloads (vectors), which may not always be feasible at early-design stage. This paper introduces the first vectorless thermal integrity verification framework that allows computing worst-case temperature (gradient) distributions across the entire chip under a set of local and global workload (power density) constraints. To address the computational challenges introduced by the large 3D mesh-structured thermal grids, we propose a novel spectral approach for highly-scalable vectorless thermal verification of large chip designs. Our approach is based on emerging spectral graph theory and graph signal processing techniques, which consists of a thermal grid topology sparsification phase, an edge weight scaling phase, as well as a solution refinement procedure. The effectiveness and efficiency of our approach have been demonstrated through extensive experiments.**

Index Terms—**spectral graph sparsification, algebraic multigrid, vectorless verification**

Aggressive VLSI technology scaling has lead to dramatically increased power densities as well as significantly elevated temperature on chip, which imposes ever-increasing challenges in design of nowadays integrated circuit (IC) systems [11]. To achieve desired level of chip reliability and functionality, compute-intensive full-chip thermal analysis and verification become indispensable. Existing methods for chip thermal analysis and verification require the underlying workloads or power densities to be known in advance [6], [8], [13], which may not always be practical. For example, at early chip design phase it is usually not possible to obtain very accurate estimation of underlying power densities since accurate modeling for workloads may not be necessarily available at the early design phase. As a result, traditional thermal analysis methods may not always provide useful guidance for verifying and improving the design reliability and performance that can be significantly impacted by extreme (worst-case) chip thermal profiles, such as worst-case temperature or thermal gradients across the chip.

In this work, we propose the first vectorless thermal integrity verification framework that is motivated by existing vectorless power grid integrity verification problems [2], [5], [14]–[16]. The goal of vectorless thermal verification is to provide a scalable solution for estimating nearly-worst-case thermal profiles under various complex power density or workload uncertainties and constraints. Existing vectorless verification methods need to set up linear programs (LPs) for finding the worst-case vectors that will lead to the extreme thermal profiles, which requires computing thermal sensitivities with respect to each individual power source. However, the 3D meshes in thermal grid verification tasks are much more challenging to tackle than the 2D meshes in power grid verification [14], [15] due to the super-linear complexities of existing vectorless verification methods.

To address the computational challenges in vectorless ther-

mal integrity verification, we propose to aggressively simplify the 3D thermal grids during vectorless verification while assuring the approximation accuracy via spectral graph sparsification and iterative edge weight scaling. To this end, motivated by emerging graph signal processing research [12] we propose a mathematically rigorous method to match full chip temperature distributions that can be understood as the "low-frequency" graph signals on thermal grids obtained after applying a "low-pass" graph filter on the original input power sources. Our thermal grid simplification task will aim to minimize the number of edges in the sparsified thermal grid that can still precisely preserve slowly-varying, "low-frequency" temperature distribution across the entire thermal grid. Such simplified thermal grids will allow finding worst-case thermal profiles in a highly efficient way without losing accuracy. The proposed vectorless thermal integrity verification method is highly scalable and thus can be adopted in either early chip design phase or final chip verification phase. The main contribution of this work has been briefly summarized as follows:

(1) For the first time, we propose a vectorless thermal integrity verification framework that allows estimating nearly-worst-case thermal profiles under various kinds of complex workload or power density uncertainties and constraints.

(2) To make the proposed method scalable to large problems, we introduce a multilevel vectorless verification framework that is significantly accelerated by a novel thermal grid simplification method motivated by emerging spectral graph sparsification and graph signal processing research.

(3) We demonstrate extensive experimental results on two chip designs with various problem sizes and power density (workload) constraints, as well as the flexible tradeoffs between the verification cost and solution quality enabled by the proposed vectorless verification method.

I. BACKGROUND

A. Vectorless Thermal Integrity Verification

Give many workloads or power source configurations, traditional thermal analysis and verification methods usually need to run thermal analysis for every workload to determine the extreme thermal distributions across the chip, which can be extremely expensive for large designs. In this work, we introduce a vectorless thermal integrity verification framework, where constraints are introduced to capture the uncertain workloads or power source configurations. Given a 3D thermal grid and a specific workload (power density) distribution, the chip steady-state thermal analysis can be performed by solving following equation:

$$T\,x\,=\,b, \qquad (1)$$

where T is the thermal conductance matrix of the 3D thermal grid, b is the right-hand-side (RHS) vector modeling the underlying power sources, and x is the unknown temperature vector to be computed.

The proposed vectorless thermal integrity verification seeks to identify the maximum temperature or temperature gradients under various workload constraints similar to prior vectorless power grid integrity verification problems [16]. Two types of constraints are considered in this work: local constraints for setting the lower and upper bounds of the power density for each source and global constraints for setting the lower and upper bounds for blocks of sources. The proposed vectorless thermal integrity verification tasks compute the worse-case temperatures across the chip by solving the following linear program (LP) for each individual node i:

$$\text{maximize}: \quad t_i = e_i^\top T^{-1} b, \quad \text{for} \ i = 1, \ldots, n \quad (2)$$

subject to the following constraints on power densities:

$$\begin{aligned}\text{Local Constraints}: \quad b^L &\leq b \leq b^U, \\ \text{Global Constraints}: \quad B^L &\leq M b \leq B^U,\end{aligned} \quad (3)$$

where n is the number of nodes in the 3D thermal grid, e_i is an elementary unit vector with $i - th$ entry to be 1 and others being zeros. Since the thermal conductance matrix T is an M-matrix, the T^{-1} only contains non-negative sensitivity values. The b^L (B^L) and b^U (B^U) represent the lower bounds and upper bounds of individual power sources (blocks), while M is an $m \times n$ matrix that only contains 0s and 1s for defining m global (block) constraints. After getting the worst-case vector b_{wst} through the above optimization procedure, we can simply compute the worst-case temperature distribution t_{wst} using $t_{wst} = T^{-1} b_{wst}$.

B. Vectorless Thermal Verification Challenges

The adjoint temperature sensitivity with respect to each power source will be needed for setting up the LP problems in (2) for vectorless thermal verification. For example, the adjoint sensitivity vector s for computing node temperature t_i considering all power sources in b can be calculated by solving the linear system of equations $T s = e_i$. Once the matrix factorization for T is computed, adjoint thermal sensitivity vectors for individual node temperatures can be efficiently obtained by reusing the matrix factors.

However, **factorization of the thermal matrix** obtained from 3D mesh-structured grids can be much more costly than factorizing the conductance matrices for power grid vectorless verification tasks [15], [16], due to the much faster growing computational complexity of existing direct solution methods, such as LU and Cholesky decomposition methods [1]. For example, our results show that factorizing a matrix with one million rows (columns) using the state-of-the-art Cholesky solver [1] may take over 30 minutes and consume 18GB memory.

Meanwhile, since the adjoint sensitivity vector is needed for solving the following LP problem:

$$maximize: \quad t_i = \sum s_i b_i, \quad (4)$$

very high computational complexity will be expected when a large number of uncertain power sources (variables) are involved.

C. Graph Signal Processing and Spectral Sparsification

There is an analogy between traditional signal processing or classical Fourier analysis and graph signal processing [12]: **1)** The signals at different time points in classical Fourier analysis correspond to the signals at different nodes in an undirected graph; **2)** The more slowly oscillating functions in time domain correspond to the graph Laplacian eigenvectors associated with lower eigenvalues or the more slowly varying (smoother) components across the graph. The recent spectral graph sparsification process [3], [4] aims to maintain as few as possible edges for preserving the slowly-varying or "low-frequency" signals of the original graphs, which therefore can be regarded as a "low-pass" graph filter. As a result, spectrally-sparsified graphs will be able to preserve the eigenvectors associated with low eigenvalues more accurately than high eigenvalues.

To aggressively simplify the 3D thermal grids and thereby addressing the computational challenges in vectorless integrity verification without sacrificing the approximation accuracy, emerging graph signal processing and spectral graph sparsification research can be leveraged [4], [12]. Since full chip temperature distributions can be considered as the "low-frequency" graph signals on thermal grids obtained after applying a "low-pass" graph filter on the original input power sources, the spectrally-sparsified thermal grids will very well preserve the temperature distributions.

II. A SPECTRAL APPROACH TO VECTORLESS THERMAL INTEGRITY VERIFICATION

A. Spectral Sparsification and Scaling of 3D Thermal Grids

To substantially reduce the cost for the matrix factorization and LP solution phases, we exploit a perturbation-based spectral graph sparsification engine [3], [4] to dramatically sparsify the topology of the original 3D thermal grid during the vectorless thermal verification. The spectral sparsification step can effectively control the thermal grid densities while maintaining good spectral approximation quality that is critical for accurate vectorless verification tasks: the sparsified thermal grids have tree-like structures that will immediately reduce the matrix factorization time while preserving the effective thermal resistances between nodes. It is noted that preserving effective resistance is equivalent to preserving the adjoint sensitivities to be applied for setting up LPs. Therefore, the adjoint sensitivity for each LP task can be computed in a more efficient way without sacrificing the final solution quality (e.g. worst-case vector). Additionally, the sparsified thermal grids will have many low-degree nodes that can be potentially merged together to further reduce the number of variables in LPs, which can also effectively reduce the cost for solving LPs in vectorless verification tasks.

To further improve the approximation quality of the sparsified thermal grid, we introduce an iterative edge weight scaling scheme to gradually scale up the edge weight in the sparsified thermal grid, which has been described in Algorithm 1. This scheme will compensate for the thermal conductance loss due to the missing edges by matching the "low-frequency" behaviors of the original thermal grids, which is motivated by recent graph signal processing techniques [12].

We define $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ to be the n eigenvalues of the Laplacian matrix L_G for a connected graph G with the corresponding eigenvectors denoted by $u_1, u_2, \cdots, u_n$. The

Algorithm 1 Algorithm for Iterative Edge Weight Scaling
___
Input: The error tolerance ϵ, the number of partitions k, the original graph $\mathbf{G}$ and the initial spectrally-sparsified graph $\mathbf{P^{(0)}}$.
Output: The spectrally-sparsified graph $\mathbf{P}$ with scaled edge weights.

1: Generate a random vector $\mathbf{b}$ that is orthogonal to the all-one vector.
2: Partition the original graph $\mathbf{G}$ into $\mathbf{k}$ blocks $\mathbf{P_1}, \mathbf{P_2}, \cdots, \mathbf{P_k}$ using multilevel graph partitioning method [7].
3: Construct matrices $\mathbf{T_G} = \mathbf{L_G} + \mathbf{g_{min}I}$ and $\mathbf{T_{P^{(0)}}} = \mathbf{L_{P^{(0)}}} + \mathbf{g_{min}I}$ by adding a small value $\mathbf{g_{min}}$ similar to the ambient thermal conductance to each diagonal entry of $\mathbf{L_G}$ and $\mathbf{L_{P^{(0)}}}$ for graph signal filtering purpose.
4: Solve $\mathbf{T_G x} = \mathbf{b}$ and $\mathbf{T_{P^{(0)}}}\widetilde{\mathbf{x}} = \mathbf{b}$ and compute $\mathbf{err} = \frac{\|\mathbf{x} - \widetilde{\mathbf{x}}\|}{\|\mathbf{x}\|}$.
5: **while err $> \epsilon$ do**
6:   **for** partition $\mathbf{P_i}$, $\mathbf{i = 1, ..., k}$, **do**
7:     calculate $\mathbf{y_i} = \sum_{\mathbf{t} \in \mathbf{P_i}} \mathbf{x[t]}$, $\widetilde{\mathbf{y}}_\mathbf{i} = \sum_{\mathbf{t} \in \mathbf{P_i}} \widetilde{\mathbf{x}}[\mathbf{t}]$, and $\alpha_\mathbf{i} = \frac{\widetilde{\mathbf{y}}_\mathbf{i}}{\mathbf{y_i}}$ for all nodes;
8:   **end for**
9:   **for** all edges $(\mathbf{p}, \mathbf{q}) \in E_s$ **do**
10:     **if** $\mathbf{p}, \mathbf{q} \in P_i$, scale up $\mathbf{w_{p,q}}$ by a factor of $\alpha_\mathbf{i}$;
11:     **if** $\mathbf{p} \in P_i$ and $\mathbf{q} \in P_j$, scale up $\mathbf{w_{p,q}}$ by a factor of $(\alpha_\mathbf{i} + \alpha_\mathbf{j})/\mathbf{2}$;
12:   **end for**
13:   Update $\widetilde{\mathbf{P}}$, $\mathbf{L}_{\widetilde{\mathbf{P}}}$ and $\mathbf{T}_{\widetilde{\mathbf{P}}}$ with the latest edge weights;
14:   Solve $\mathbf{T}_{\widetilde{\mathbf{P}}} \widetilde{\mathbf{x}} = \mathbf{b}$ and update the mismatch $\mathbf{err} = \frac{\|\mathbf{x} - \widetilde{\mathbf{x}}\|}{\|\mathbf{x}\|}$;
15: **end while**
16: Return the latest spectrally-sparsified graph $\mathbf{P}$.
___

spectral decomposition of the Laplacian matrix of graph G can be expressed as follows:

$$L_G = \sum_{i=1}^{n} \lambda_i \, u_i \, u_i^\top. \tag{5}$$

Adding a small grounded thermal conductance g_{min} to each node in graph G or equivalently a small element g_{min} to each diagonal element in L_G leads to:

$$T_G = L_G + g_{min}I = \sum_{i=1}^{n} (g_{min} + \lambda_i) \, u_i \, u_i^\top, \tag{6}$$

where the identify matrix $I = \sum_{i=1}^{n} u_i \, u_i^\top$. When expressing a random vector b using Laplacian eigenvectors, we have:

$$b = \sum_{i=1}^{n} \beta_i u_i^\top. \tag{7}$$

Solving $T_G x = b$ is equivalent to computing $x = T_G^{-1} b$, which can be further expressed as:

$$x = (L_G + g_{min}I)^{-1} b = \left(\sum_{i=1}^{n} (g_{min} + \lambda_i) \, u_i \, u_i^\top \right)^{-1} b$$

$$= \sum_{i=1}^{n} \frac{u_i \, u_i^\top b}{g_{min} + \lambda_i} = \frac{1}{g_{min}} \sum_{i=1}^{n} \frac{\beta_i u_i}{1 + r \, \lambda_i} \tag{8}$$

where $r = 1/g_{min}$. (8) indicates that when using a small g_{min}, the eigenvectors associated with small eigenvalues or only "low-frequency" components in b will remain in x; on the other hand, a relatively large g_{min} ($r \approx 0$) will allow more higher frequencies to be included in x and thus lead to $x \approx b$.

Based on the above analysis, we can consider T_G^{-1} as a "low-pass" filter matrix for filtering graph signals b: by properly choosing g_{min} values it is possible to filter out graph signal's "high-frequency" (highly-oscillating) components and only keep "low-frequency" components in x. Since chip temperature distributions mainly contain "slowly-varying" ("low-frequency") components due to relative small ambient thermal conductance values, it becomes possible to exploit emerging spectral sparsification techniques [3], [4] to only maintain a small number of edges in the sparsified thermal grids while still preserving accurate thermal profiles, since spectrally-sparsified graphs can very well preserve "low-frequency" graph signals. Based on the above intuition, Algorithm 1 is proposed for scaling up edge weights in the sparsified thermal grid by matching the "low-frequency" responses filtered by the original thermal grids.

B. Spectral Solution Refinement.

We define $0 = \tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \cdots \leq \tilde{\lambda}_n$ to be the eigenvalues of L_P for sparsified graph P with the corresponding eigenvectors $\tilde{u}_1, \tilde{u}_2, \cdots, \tilde{u}_n$. Assume that k smallest eigenvalues and corresponding eigenvectors of L_G can be well preserved in L_P, while the remaining higher eigenvalues and eigenvectors are not. Then the spectral decomposition of L_P can be approximately written as:

$$L_P \approx \sum_{i=1}^{k} \lambda_i \, u_i \, u_i^\top + \sum_{i=k+1}^{n} \tilde{\lambda}_i \, \tilde{u}_i \, \tilde{u}_i^\top. \tag{9}$$

Based on (9) and (8), the solution error due to spectral sparsification and scaling becomes:

$$\Delta x = x - \tilde{x} \approx \sum_{i=k+1}^{n} \left(\frac{u_i \, u_i^\top b}{g_{min} + \lambda_i} - \frac{\tilde{u}_i \, \tilde{u}_i^\top b}{g_{min} + \tilde{\lambda}_i} \right), \tag{10}$$

indicating that the solution error using spectrally-sparsified graphs can be expressed as a linear combination of high eigenvectors corresponding to large Laplacian eigenvalues. In other words, the error is a linear combination of high frequency signals on graphs, which can be efficiently filtered out by using "low-pass" graph signal filters [12]. To further improve the solution obtained on sparsified thermal grids, weighted Jacobi iterative method is adopted in this work, which has been described in Algorithm 2. The inputs of our algorithm include the original thermal conductance matrix T_o that can be decomposed into a diagonal matrix D_o and the remainder matrix R_o, the solution vectors $\tilde{x}_1, ..., \tilde{x}_k$ obtained on the sparsified thermal conductance matrix T_s, the RHS vectors $b_1, ..., b_k$ as well as the weight factor γ and iteration number N_{max} for signal filtering.

___
Algorithm 2 Solution Refinement Algorithm
___
Input: $\mathbf{T}_o = \mathbf{D}_o + \mathbf{R}_o$, $\tilde{x}_1, ..., \tilde{x}_k$, $b_1, ..., b_k$, γ, N_{max};
1: For each of the approximate solution vectors $\tilde{x}_1, ..., \tilde{x}_k$, do
2: **for** $i = 1$ to N_{max} **do**
3:   $\tilde{x}_k^{(i+1)} = (1 - \gamma)\tilde{x}_k^{(i)} + \gamma \mathbf{D}_o^{-1}(b_k - \mathbf{R}_o \tilde{x}_k^{(i)})$
4: **end for**
5: Return the smoothed solution vectors $\tilde{x}_1, ..., \tilde{x}_k$.
___

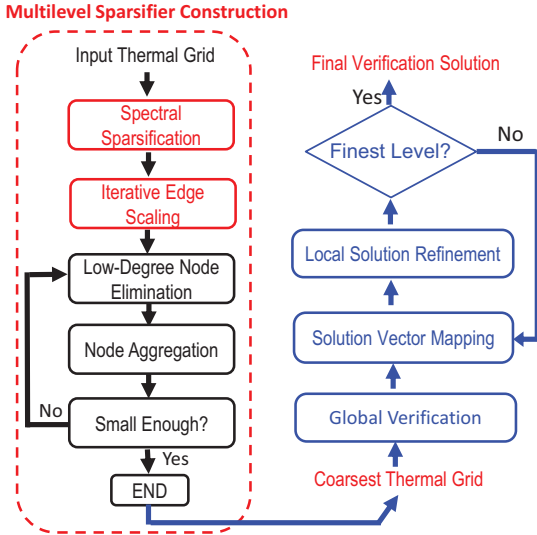**Multilevel Sparsifier Construction**

Fig. 1. Multilevel vectorless thermal integrity verification.

C. Multilevel Verification Framework

In this work, we propose a multilevel vectorless thermal verification framework shown in Figure 1. Our approach is based on the latest graph-theoretic algebraic multigrid (AMG) research works [9], [17] for generating coarse-level (sparsified) thermal grids according to the original thermal grid problem, as well as the recent multilevel vectorless power grid integrity verification framework [2], [16]. In the following, we will describe a two-level vectorless verification approach, while multilevel schemes can be defined in a similar way.

Two-level local and global constraints mapping. Local power constraints can be directly mapped from fine level h to coarse level H using AMG's restriction operator R_h^H obtained as follows:

$$Upper\ bound: \quad b_H^U = R_h^H\, b_h^U,$$
$$Lower\ bound: \quad b_H^L = R_h^H\, b_h^L,$$

where b_H^U, b_H^L, b_h^U and b_h^L denote the upper bound and lower bound of power sources for coarse and fine grids, respectively. The global constraints mapping can be defined in a similar manner by choosing the global constraints on the coarse grid to be the sum of each block's lower and upper bounds on the fine grid.

Two-level Solution mapping and refinement. To reduce the verification cost on the coarse level, the global critical region C_{glb} will be identified based on the adjoint sensitivity threshold [2], such that C_{glb} will only include the most important power sources. Given a sensitivity threshold ϵ_{th}, we will only include the power sources that have adjoint sensitivity values greater than ϵ_{th} into C_{glb} for setting up LPs:

$$maximize: \quad t_{wst} = \sum_{\forall\, b_i \in C_{glb}} s_i\, b_i \quad (11)$$

subject to local and global constraints:

$$b^L \leq b \leq b^U, \quad B^L \leq M\, b \leq B^U. \quad (12)$$

The solution b_{wst}^H will be further mapped back to the fine level using the AMG prolongation operator R_H^h by $\widetilde{b}_{wst}^h = R_H^h\, b_{wst}^H$. To control the error introduced during the mapping

process, a local solution refinement procedure at the fine level will be applied to incrementally improve the solution quality on the fine grid by setting up a new LP for a much smaller local critical region.

Algorithm flow and complexity. The detailed multilevel vectorless thermal integrity verification algorithm has been described in Algorithm 3, while the key steps for each level grid have been described as follows:

(1) Scale up the sensitivity threshold $\epsilon_{loc} = \beta\,\epsilon_{glb}$ with the scaling factor $\beta > 1$ to obtain a much smaller local critical region C_{loc}.

(2) Set up a new LP problem for the local critical region C_{loc} to obtain the solution vector $\bar{b}_{wst}^h$.

(3) Update solution for C_{loc} with $\bar{b}_{wst}^h$; Reuse the interpolated solution $\widetilde{b}_{wst}^h$ for the sources that belong to C_{glb} but not C_{loc}.

The complexity for setting up multilevel problems is $O(m)$ where m denotes the number of thermal resistors in the chip thermal model. The complexity for thermal grid spectral sparsification and edge scaling is $O(m \log n)$ with n denoting the number of nodes in the 3D thermal grid, which is nearly linear. The cost for solving LPs will depend on the algorithm to be adopted as well as the sizes of critical regions for setting up the LPs, which can be well controlled by taking advantage of the proposed multilevel verification framework. It should be noted that by leveraging the proposed solution refinement procedure, only ultra-sparse (tree-like) spectral sparsifiers of the original 3D thermal grids are needed for vectorless verification, which can significantly improve the overall algorithm scalability, as shown in our experiment results in Section III.

Algorithm 3 Multilevel Vectorless Thermal Integrity Verification

---

Input: original thermal grid, user-defined local and global power constraints $\mathbf{b^U}$, $\mathbf{b^L}$ and $\mathbf{M}$, initial normalized sensitivity threshold $\epsilon_{\mathbf{th}}$, and sensitivity scaling factor $\beta > \mathbf{1}$
Output: worst-case thermal profile of the original thermal grid.

1: Extract spectrally sparsified grid for the original thermal grid.
2: Update sparsified grid using iterative edge weight scaling method (Algorithm 1).
3: Multilevel coarse grid construction:
   (a) Construct all hierarchy levels from finest to coarsest level;
   (b) Get local and global power constraints $\mathbf{b^U}$, $\mathbf{b^L}$ and $\mathbf{M}$ for each level using AMG mapping operators.
4: Factorize each coarse-level grid for adjoint sensitivity calculation.
5: Perform global verification at the coarsest level $\mathbf{K}$:
   (a) Find global critical region $\mathbf{C_{glb}^K}$ for a given sensitivity threshold $\epsilon_{\mathbf{K}}$, and set up LP to get worst case vector $\mathbf{b_{wst}^K}$
6: Perform solution mapping and refinement on finer to finest levels:
7: $\mathbf{k} \leftarrow \mathbf{K}$
8: **while** $\mathbf{k > 1}$ **do**
9:    Interpolate solution vector to finer level by: $\bar{\mathbf{b}}_{\mathbf{wst}}^{\mathbf{k-1}} = \mathbf{R_k^{k-1} b_{wst}^k}$
10:    Set sensitivity threshold $\epsilon_{\mathbf{k-1}} = \beta\epsilon_{\mathbf{k}}$ and identify $\mathbf{C_{loc}^{k-1}}$.
11:    Setup a new LP for $\mathbf{C_{loc}^{k-1}}$ to obtain solution vector $\bar{\mathbf{b}}_{\mathbf{wst}}^{\mathbf{k-1}}$.
12:    Combine the latest LP and interpolated solutions to form $\mathbf{b_{wst}^{k-1}}$.
13:    $\mathbf{k} \leftarrow \mathbf{k - 1}$
14: **end while**
15: Calculate the worst-case thermal distribution using the worst-case power source vector.

III. EXPERIMENTAL RESULTS

In this section, we present the experiment results of the proposed vectorless thermal verification method for two microprocessor designs [8]. The design details of the two micro-

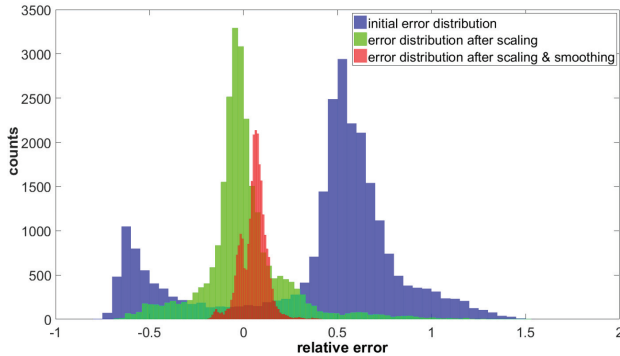| Design	Processor A	Processor B
Power Consumption (W)	28	50
Die Area (mm^2)	195	302
Num. of Metal Layers	4	6
Num. of Material Layers	11	15
Equivalent Heat Transfer	3.3 (heat sink)	3.3 (heat sink)
Coefficients ($10^3 W/m^2 K$)	2.0 (package)	2.0 (package)

Fig. 2.   Relative Error Distributions.

processors are shown in Table I. The heat conductance paths are modeled using equivalent heat transfer coefficients. The proposed multilevel vectorless thermal integrity verification method has been implemented in MATLAB and C++. The LP problems are solved by the state-of-the-art LP solver [10] and all experiment results have been obtained using a single CPU core of a computing platform running 64-bit RHEL 6.0 with 2.67GHz 12-core CPU and 48GB DRAM memory.

To demonstrate the effectiveness of the proposed edge scaling and solution refinement schemes, four solution (temperature) vectors are calculated for a 3D thermal grid and its spectral sparsifiers: **(a)** the true solution vector obtained using the original thermal grid, **(b)** the approximate solution vector computed using the sparsifier without edge scaling, **(c)** the approximate solution vector obtained using the sparsifier with edge scaling, **(d)** as well as the refined (smoothed) solution vector using the sparsifier with edge scaling. Meanwhile, we plot histogram distributions of relative errors of the solution vectors (b)-(d) by comparing them against the true solution vector (a), as shown in Figure 2. We can see that the solution errors between the sparsifiers and the original graph can be significantly reduced by leveraging the proposed iterative edge scaling scheme, and further mitigated by the proposed solution refinement procedure.

As we mentioned in the previous sections, the spectral graph sparsification method can well preserve the low frequency components of the original thermal grid solutions, which will allow achieving high-quality solutions for vectorless verification tasks. Figures 3 and 4 show the worst-case thermal distributions of processors A and B using **(a)** the direct method, **(b)** the multilevel vectorless verification method w/o sparsification, and **(c)** the multilevel vectorless verification method w/ spectral sparsification, respectively. As observed, the three worst-case thermal distributions are very close to each other, indicating that rather accurate vectorless verification results can be obtained using spectrally-sparsified thermal grids.

Vectorless thermal integrity verification results using dif-

ferent methods are shown in Table II, where $N.\#$ and $P.\#$ denote the numbers of thermal grid nodes and power sources, respectively. "Single Level", "Multilevel w/o Sparsification", and "Multilevel w/ Sparsification" denote the vectorless thermal verification methods using single-level (direct), multilevel w/o and w/ spectral sparsification schemes, respectively; T^*_{chol}, T^*_{sol}, and T^*_{lp} denote the Cholesky factorization time [1], the adjoint sensitivity calculation time, and the LP solving time [10] for each of the three verification methods, respectively. Except for T^*_{chol}, all other time are computed by summing up the runtimes for verifying 100 randomly chosen nodes. Err denotes the average relative solution error compared to the single-level method and σ^2 denotes the relative condition number (spectral similarity) used for similarity-aware spectral sparsification [4] of 3D thermal grids.

It is observed that both the matrix factorization and adjoint sensitivity calculation procedures in the "Multilevel w/ Sparsification" method are consistently much faster than the other two methods, especially for larger test cases. While the "Multilevel w/o Sparsification" is the slowest method due to the fast growing matrix densities at coarse levels. The overall LP solution time T_{lp} for the "Multilevel w/ Sparsification" method is also the smallest, indicating that the proposed method can effectively reduce the number of decision variables in LP and thus result in much lower computational cost in vectorless thermal verification tasks.

Meanwhile, the proposed method scales very comfortably with even very large 3D thermal grids, since the total verification time increases almost linearly with the 3D thermal grid sizes, as shown in Figure 5.

IV. CONCLUSION

We present a highly-scalable multilevel vectorless thermal integrity verification framework for computing chip worst-case thermal profiles without knowing exact distribution of underlying power sources or workloads. Recent theoretical results in spectral graph sparsification and graph signal processing techniques enable us to develop much faster and more scalable vectorless thermal integrity verification algorithms, while achieving flexible tradeoffs between computing efficiency and solution quality. Extensive experiment results for two modern chip designs have been demonstrated, indicating that the proposed scalable vectorless verification method can always efficiently obtain highly-accurate worst-case thermal profiles for large chip designs.

V. ACKNOWLEDGMENTS

REFERENCES

[1] T. Davis. *CHOLMOD: sparse supernodal Cholesky factorization and update/downdate*. [Online]. Available: http://www.cise.ufl.edu/research/sparse/cholmod/, 2008.

[2] Z. Feng. Scalable multilevel vectorless power grid voltage integrity verification. *IEEE Trans. on VLSI Systems*, 21(8):1526–1539, August 2013.

[3] Z. Feng. Spectral Graph Sparsification in Nearly-Linear Time Leveraging Efficient Spectral Perturbation Analysis. In *Proc. IEEE/ACM DAC*, 2016.

[4] Z. Feng. Similarity-aware spectral sparsification by edge filtering. In *Design Automation Conference (DAC), 2018 55nd ACM/EDAC/IEEE*. IEEE, 2018.

[5] N. Ghani and F. Najm. Fast Vectorless Power Grid Verification Using an Approximate Inverse Technique. In *Proc. IEEE/ACM DAC*, pages 184–189, 2009.

[6] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. R. Stan. Hotspot: A compact thermal modeling methodology for early-stage vlsi design. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 14(5):501–513, 2006.
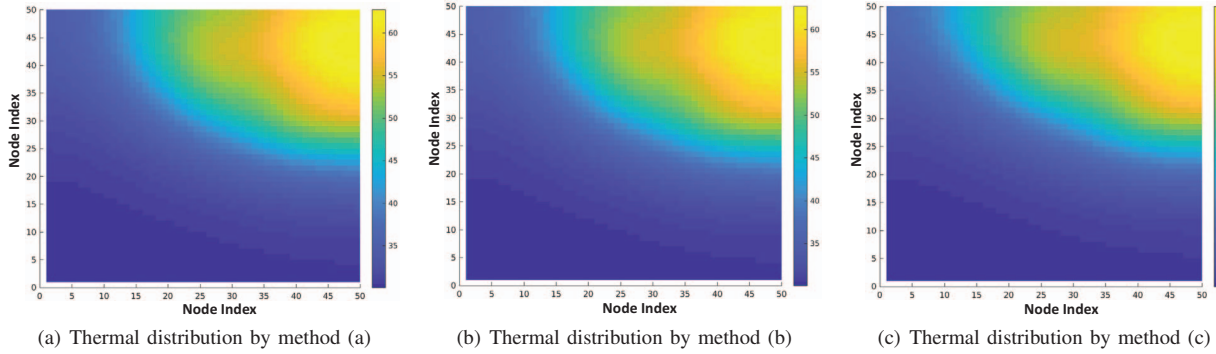
(a) Thermal distribution by method (a)   (b) Thermal distribution by method (b)   (c) Thermal distribution by method (c)

Fig. 3. Worst-case temperature distributions of processor A

(a) Thermal distribution by method (a)   (b) Thermal distribution by method (b)   (c) Thermal distribution by method (c)
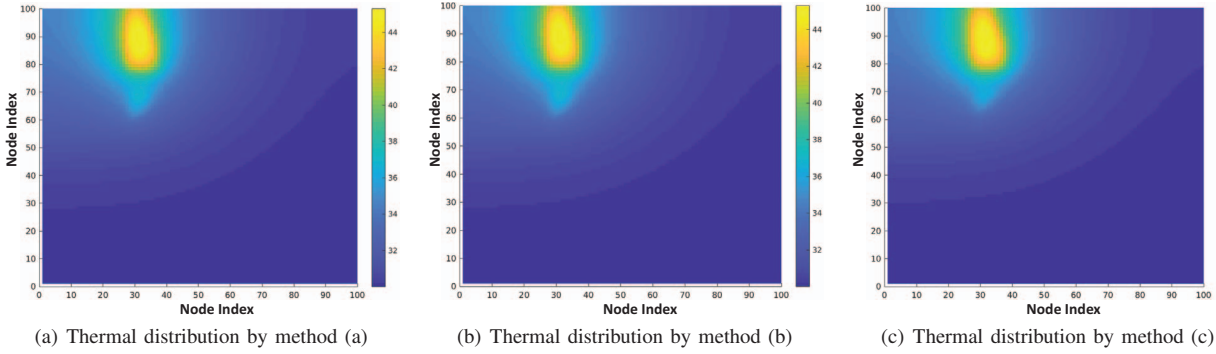
Fig. 4. Worst-case temperature distributions of processor B

TABLE II
RESULTS OF THE PROPOSED MULTILEVEL VECTORLESS THERMAL INTEGRITY VERIFICATION METHOD (TWO-LEVEL SCHEME IS USED).

|	Grids Specs.		(a) Single Level			(b) Multilevel w/o Sparsification				(c) Multilevel w/ Sparsification				
CKT	$N.\#$	$P.\#$	$T^o_{chol}\left(\frac{T^o_{chol}}{T^s_{chol}}\right)$	$T^o_{sol}\left(\frac{T^o_{sol}}{T^s_{sol}}\right)$	$T^o_{lp}\left(\frac{T^o_{lp}}{T^s_{lp}}\right)$	$T^m_{chol}\left(\frac{T^m_{chol}}{T^s_{chol}}\right)$	$T^m_{sol}\left(\frac{T^m_{sol}}{T^s_{sol}}\right)$	$T^m_{lp}\left(\frac{T^m_{lp}}{T^s_{lp}}\right)$	$Err(\%)$	T^s_{chol}	T^s_{sol}	T^s_{lp}	$Err(\%)$	σ^2
T1	25K	2.5K	0.94s (31X)	2.30s (18X)	2.71s (1.4X)	1.24s (41X)	3.21s (25X)	3.12s (1.6X)	1.0%	0.03s	0.13s	2.02s	5.0%	2,073
T2	0.1M	10K	5.89s (20X)	14.79s (20X)	10.36s (1.6X)	8.12s (28X)	20.48s (28X)	5.80s (0.85X)	2.1%	0.29s	0.72s	6.81s	3.8%	2,400
T3	0.2M	10K	24.26s (22X)	55.20s (19X)	20.08s (4.6X)	33.91s (30X)	86.07s (30X)	25.90s (6.0X)	4.0%	1.13s	2.90s	4.33s	4.0%	1,435
T4	0.4M	40K	38.03s (9X)	99.91s (10X)	60.56s (4X)	50.91s (11X)	131.97s (13X)	24.15s (1.7X)	2.0%	4.61s	10.46s	14.48s	5.0%	2,193
T5	0.9M	90K	110.17s (6X)	262.53s (6X)	159.83s (18X)	148.11s (8X)	335.97s (8X)	21.43s (2.3X)	1.0%	20.09s	43.50s	9.36s	2.0%	2,469
T6	1.6M	0.16M	1.18Ks (23X)	33.60Ks (201X)	0.87Ks (6.5X)	1.25Ks (24X)	33.99Ks (204X)	0.79Ks (5.9X)	1.0%	51.70s	167.00s	133.83s	1.0%	2,141
T7	2.0M	0.20M	1.32Ks (20X)	32.27Ks (172X)	1.76Ks (9.7X)	1.42Ks (22X)	28.91Ks (154X)	1.70Ks (9.1X)	1.0%	65.23s	187.36s	181.76s	2.0%	3,073
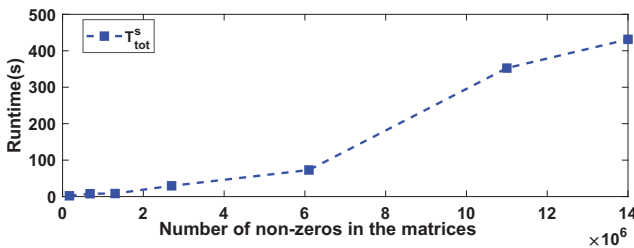
Fig. 5. Total verification time with the number of non-zeros in matrices.

[7] D. LaSalle, M. M. A. Patwary, N. Satish, N. Sundaram, P. Dubey, and G. Karypis. Improving graph partitioning for modern graphs and architectures. In *Proceedings of the 5th Workshop on Irregular Applications: Architectures and Algorithms*, page 14. ACM, 2015.

[8] P. Li, L. T. Pileggi, M. Asheghi, and R. Chandra. Ic thermal simulation and modeling via efficient multigrid-based approaches. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 25(9):1763–1776, 2006.

[9] O. Livne and A. Brandt. Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver. *SIAM Journal on Scientific Computing*, 34(4):B499–B522, 2012.

[10] G. Optimization. Gurobi optimizer [online]. available: www. gurobi. com, 2016.

[11] M. Pedram and S. Nazarian. Thermal modeling, analysis, and management in vlsi circuits: Principles and methods. *Proceedings of the IEEE*, 94(8):1487–1501, 2006.

[12] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

[13] K. Skadron, M. R. Stan, K. Sankaranarayanan, W. Huang, S. Velusamy, and D. Tarjan. Temperature-aware microarchitecture: Modeling and implementation. *ACM Transactions on Architecture and Code Optimization (TACO)*, 1(1):94–125, 2004.

[14] X. Xiong and J. Wang. An Efficient Dual Algorithm for Vectorless Power Grid Verification under Linear Current Constraints. In *Proc. IEEE/ACM DAC*, pages 837–842, 2010.

[15] J. Yang, Y. Cai, Q. Zhou, and W. Zhao. A selected inversion approach for locality driven vectorless power grid verification. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 23(11):2617–2628, 2015.

[16] Z. Zhao and Z. Feng. A spectral graph sparsification approach to scalable vectorless power grid integrity verification. In *Proceedings of the 54th Annual Design Automation Conference 2017*, page 68. ACM, 2017.

[17] Z. Zhao, Y. Wang, and Z. Feng. Samg: Sparsified graph-theoretic algebraic multigrid for solving large symmetric diagonally dominant (sdd) matrices. In *Proceedings of the 36th International Conference on Computer-Aided Design (ICCAD)*. ACM, 2017.