HAL
open science

# Globally Constrained Locally Optimized 3-D Power Delivery Networks

Aida Todri-Sanial, Sandip Kundu, Patrick Girard, Alberto Bosio, Luigi Dilillo, Arnaud Virazel

## To cite this version:

Aida Todri-Sanial, Sandip Kundu, Patrick Girard, Alberto Bosio, Luigi Dilillo, et al.. Globally Constrained Locally Optimized 3-D Power Delivery Networks. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2014, 22 (10), pp.2131-2144. 10.1109/TVLSI.2013.2283800 . lirmm-01255754

**HAL Id: lirmm-01255754**

**https://hal-lirmm.ccsd.cnrs.fr/lirmm-01255754v1**

Submitted on 26 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Globally Constrained Locally Optimized 3-D Power Delivery Networks

Aida Todri-Sanial, *Member, IEEE*, Sandip Kundu, *Fellow, IEEE*, Patrick Girard, *Senior Member, IEEE*, Alberto Bosio, *Member, IEEE*, Luigi Dilillo, *Member, IEEE*, and Arnaud Virazel, *Member, IEEE*

***Abstract*— Design of power delivery network (PDN) is a constrained optimization problem. An ideal PDN must limit voltage drop that results from switching circuits' transients, satisfy current density constraints that arise from electromigration limits, yet use only minimal metal resources so that design density targets can be met. It should also provide an efficient thermal conduit to address heat flux. Furthermore, an ideal PDN should be a regular structure to facilitate design productivity and manufacturability, yet be resilient to address varying power demands across its distribution area. In 3-D ICs, these problems are further constrained by the need to minimize through-silicon via (TSV) area and bridge power lines of different dimensions across tiers, while addressing varying power demands in lateral and vertical directions. In this paper, we propose an unconventional power grid optimization solution that allows us to resize each tier individually by applying tier-specific constraints and yet be optimal in a multitier network, where each tier is locally resized while globally constrained. Tier-specific constraints are derived from electrical and thermal targets of 3-D PDNs. Two resizing algorithms are presented that optimize 3-D PDNs standalone or 3-D PDNs together with TSVs. We demonstrate these solutions on a three-tier setup where significant area savings can be achieved.**

***Index Terms*— Algorithms, circuit analysis, optimization, power delivery networks (PDNs).**

## I. Introduction

3-D INTEGRATION presents a path toward higher performance, denser circuits, and heterogeneous implementation while delivering a smaller footprint [1]. Ironically, these advantages are also at the root of power and thermal integrity problems in 3-D ICs. Even more so for power delivery networks (PDNs).

PDNs deliver power and ground voltage from package to the devices. Variations on the supplied voltage can lead to the degradation of circuits' performance and reliability. These effects become more critical in 3-D ICs due to longer current paths and exacerbated thermal dissipation from stacking multiple tiers.

Manuscript received August 31, 2012; revised March 26, 2013; accepted September 15, 2013. Date of publication October 18, 2013; date of current version September 23, 2014. This work was supported by the French National Research Agency.

A. Todri-Sanial, P. Girard, A. Bosio, L. Dilillo, and A. Virazel are with the Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier, Montpellier 34095, France (e-mail: aida.todri@lirmm.fr; girard@lirmm.fr; bosio@lirmm.fr; dilillo@lirmm.fr; virazel@lirmm.fr).

S. Kundu is with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003-9284 USA (e-mail: kundu@ecs.umass.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2013.2283800

From early on in the design cycle, designers are concerned to know the worst case voltage drop on each tier and possibly find ways to alleviate it. While much work has been done to understand the resiliency of 2-D PDNs, designers are faced with new challenges for ensuring robustness of 3-D PDNs. Some of the unique questions related to 3-D PDNs addressed in this paper are:

1) What is the impact of physical design constraints (i.e., uniform or nonuniform dies with various power densities) on 3-D PDNs?
2) What is the impact of manufacturing constraints (i.e., thinned die or wafer) on the resiliency of 3-D PDNs?
3) What is the impact of package on voltage drop and thermal dissipation in 3-D PDNs?
4) What is the impact of through-silicon vias (TSVs)? How does TSV sizing affect power and thermal integrity of 3-D PDNs?
5) How does nonuniform temperature distribution inter and intratier affect voltage drop on PDNs?

We will take up these intricate issues in this paper, since not only are they pivotal to the design of 3-D PDNs, but they provide an important framework for evaluating the benefits and costs of various manufacturing and design constraints unique to 3-D integration. To do so, we develop a platform to perform 3-D PDN electrothermal analysis and optimization.

### A. Power and Thermal Integrity Issues on 3-D PDNs

To address these issues, we take a deep look at the source of power and thermal integrity problems in 3-D ICs. 3-D PDNs are structured as multitier networks that are bridged together using TSVs. PDN of each tier is also structured as a mutlilayer mesh grid with power lines spanning the entire tier, where top and low-level metal layers devise the global and local PDN, respectively. TSVs facilitate the continuity between the PDNs of each tier and depending on the stacking technology, they connect: 1) global to global; 2) local to local; and 3) global to local PDNs of two adjacent tiers. Fig. 1 shows a three-tier system with face-to-back TSVs connecting global to local PDNs.

Wire sizing is a widely effective applied method on 2-D PDNs to meet voltage drop and thermal constraints, however, for 3-D PDNs this problem appears more intricate due to the current and heat flow attributes in multitier networks.

In 3-D ICs, current flows hierarchically from package to the PDN of the first tier, then through TSVs to the next tier's

1063-8210 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

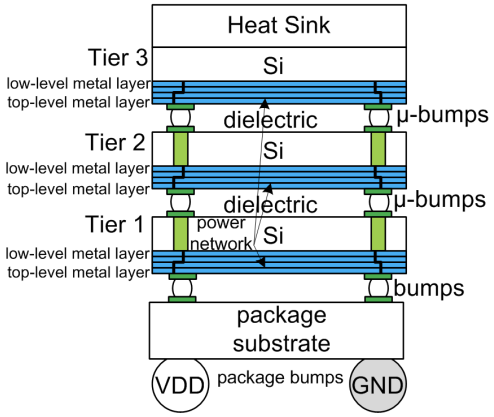



Fig. 1. Three-tier network using face-to-back TSVs.

PDN and so on, till the very last tier. Such conduct results in large amount of current flow in the PDN of the bottom tier (next to package) that reduces gradually as it reaches PDNs of top tiers. Hence, each tier depending on their stack location would require individual PDN sizing to accommodate the hierarchical current flow. Furthermore, large amounts of current flowing through parasitic branches of PDNs along with fast switching frequencies induce excessive and nonuniform voltage drops on power and ground distributions of each tier. Inconveniently, TSVs also introduce additional parasitics to power networks and contribute to voltage drop. Moreover, TSVs act as a medium for power supply noise propagation from one tier to the next. In the worst case, power supply noise of one tier can reach the adjacent tier when it is already experiencing excessive self-induced voltage drop.

Thermal effects are another important phenomena that strain 3-D PDNs. Frequent switching circuits generate heat and lead to elevated and nonuniform temperatures on each tier. Heat sink generally located on the top of the stack provides immediate cooling to the top tier, which causes additional temperature nonuniformity among tiers. Therefore, this leads to exacerbated voltage drop, as electrical resistance is temperature dependent. Besides, TSVs also propagate heat from one tier to the next, which in conjunction with voltage drop can cause detrimental effects, i.e., timing failures and reliability issues.

### B. Prior Work

In open literature, there are several works that look into electrical modeling and IR drop analysis of 3-D PDNs. In [2], authors provide detailed analytical modeling of 3-D PDNs. Authors in [3] analyze 3-D PDNs while considering TSVs inserted in array shape. In [4] and [5], authors investigate different TSVs technologies and explore novel TSV topologies. In [6], authors investigate the impact of TSVs on IR drop of 3-D PDNs. In [7], floorplanning along with power network cosynthesis is investigated. In [8], authors investigate TSV tapering for power and thermal integrity of 3-D PDNs. In [9], decoupling capacitors insertion is explored as an alternative for congestion-aware power grid optimization.

In comparison with the literature on IR drop assessment of 3-D PDNs; there is significantly fewer works on thermal analysis of 3-D PDNs. Early works in [10] and [11] identified the criticality of high temperatures in 3-D ICs and presented simple thermal models for predicting temperature. More recently, [12]–[15] developed heat transfer models for multitier designs. While these analytical models predict the temperature rise in each tier, they are not directly applicable to 3-D PDNs. Although, it is widely acknowledged the close coupling effects between electrical and thermal behavior, only a few papers have looked into electrothermal 3-D PDN analysis [16], [17].

In contrast to these previous works, we present an electrothermal optimization methodology for 3-D PDNs. To the best of our knowledge, this is the first work that investigates power and thermal integrity driven 3-D PDN optimization.

### C. Our Contributions

The contributions of this paper are threefold.

First, we present an efficient thermal model to compute temperature on 3-D PDNs by exploiting the electrical–thermal duality. Instead of splitting the metal tracks in small segments, temperature is computed on grid branches based on the temperature of vias. The resulting 3-D PDN thermal model has the same number of nodes and topology as the electrical network.

Second, we present a methodology for electrothermal optimization of 3-D PDNs. As each tier can have different topology and power density distributions, we develop a tier-based optimization method that satisfies multitier network constraints meaning that each tier is locally optimized while globally constrained. We define tier-specific constraints that are derived from multitier network's electrical and thermal targets, TSVs sizing, and power density distribution.

Third, we study the power and thermal integrity of 3-D PDNs while introducing manufacturing (stacking) and design constraints. We investigate the impact of fully or partially identical dies, wafer thinning, and package choice on electrothermal optimization of 3-D PDNs.

The rest of this paper is organized as follows. Preliminaries on the 3-D technology, TSV dimensions and power tracks are provided in Section II. Section III contains two motivational experiments to highlight that reuse of already optimized 2-D PDNs is not viable and that nonuniform temperature distribution requires circuit area increase. We describe our 3-D PDNs electrothermal analysis method in Section IV. In Section V, we describe our two proposed algorithms for performing electrothermal optimization of 3-D PDNs. Experiments are presented in Section VI. Section VII concludes this paper.

## II. Preliminaries

In high-performance designs, PDNs are commonly structured as multilayered grids from top to bottom metal layers resulting in global and local power meshes. Furthermore, PDN topology can be designed as uniform or nonuniform meshes, i.e., grids can be globally irregular and locally regular, or power-grid branches are intentionally removed to ease signal routing [18], [19]. Regardless of their structure, PDNs are designed to handle worst case switching currents while keeping voltage drop within margins.

TABLE I
PARAMETERS USED IN THIS PAPER

| PDN Parameters | | | |
|---|---|---|---|
| *Silicon Substrate* | ρ=10Ω-cm | | |
| *Power Mesh* | 300μm by 300μm | | |
| *Mesh Granularity* | 120x120 mesh | | |
| *TSVs Size* | 2μm diameter, 20μm thickness, 60μm pitch | | |
| *TSVs Density* | 10000 TSV/mm$^2$ | | |
| *C4 Bumps* | 100μm diameter, 200μm pitch | | |
| *Power Density* | 1.5μW/μm$^2$ | | |
| *Decap Density* | 5fF/μm$^2$ | | |
| **PDN Parasitics** | | | |
| | R(mΩ) | L(pH) | C(fF) |
| *C4 Package* | 10 | 6.2 | 9.5 |
| *Cu TSVs* | 44.5 | 34.2 | 35.8 |
| *Cu Metal Tracks* | 45.38 | 3.5 | 0 |
| *Decaps* | 0 | 0 | 450,000 |
| *Current (μA)* | $I_{peak}$=100μA | | $T_{rise}$=50ps |
| | $I_{leakage}$=20μA | | $T_{fall}$=100ps |
| *Supply Voltage* | $V_{DD}$=1V | Voltage Drop Margin 10% $V_{DD}$ | |
| *Frequency* | 3.2GHz | | |
| *Temperature* | $T_{ambient}$=27°C | | |

Fig. 1 shows a 3-D power network for a three-tier structure with Si substrate of $\rho = 10$ Ω-cm and connected using face-to-back TSVs. Current flows from C4 package bumps with attributes as: 200 W/cm$^2$ and 100-$\mu$m diameter on 200-$\mu$m pitch [20]. In this paper, we consider high-density face-to-back Cu TSVs of 2-$\mu$m diameter with thickness of 20 $\mu$m for densities up to 10 000 TSV/mm$^2$. As reported in [21], high-density TSVs enable to reduce the parasitic effects associated with capacitance and inductance, while increasing static resistance compared with medium-density TSVs.

PDN consists of Cu metal layers and we consider an area of 300 $\mu$m × 300 $\mu$m. Power grid model includes parasitic resistance of the metal tracks [2]. TSVs include parasitic resistance, capacitance, and inductance [21]. Package is represented by its inductive and resistive parasitics [22].

From a power grid analysis perspective, circuits draw current from the grid and are usually modeled as current sources. In this paper, switching circuits are represented as time-varying current sources in triangular-like current waveform model to present their rise time, $t_r$, fall time, $t_f$, peak time, $t_p$, current peak, $I_{\text{peak}}$, and leakage current, $I_{\text{leakage}}$. To further set up our analysis, we use parameters of 45-nm Intel Xeon [23] X5482 quad-core processor with die size of 214 mm$^2$, which consumes 150 W (1.5-$\mu$ A/$\mu$m$^2$ power density) with 1 V supply voltage.

Decoupling capacitance (decaps) is represented by intentionally inserted capacitance and nonswitching circuits decoupling capacitance. Current design practices apply on-chip decaps with range 1–30 fF/$\mu$m$^2$. We apply 5 fF/$\mu$m$^2$ per tier resulting to 450 pF per each tier. Please note that in this paper, we only consider on-die power and ground network while ignoring PCB parasitics. Table I summarizes the applied parameters.

## III. MOTIVATIONAL EXPERIMENTS

Due to 3-D integration attributes, the top tier is located next to the heat sink while the bottom tier is next to the package. Such arrangement imposes that the current flows through longer parasitic paths from package to reach the top tier in comparison with the current paths to reach the bottom tier. Hence, a voltage drop increase in the upward direction can be expected. Similarly, the top tier located next to the heat sink provides immediate cooling and relief for thermal dissipation. While the bottom tier can experience higher temperature levels, hence a progressive temperature increase in downward direction can be expected.

Electrical and thermal codependencies also contribute to these directional tendencies.

Because of the aforementioned issues, we show that already optimized 2-D PDNs cannot be directly applied to 3-D PDNs due to excessive voltage drop and elevated temperatures. Second, we show that to cope with the nonuniform voltage drop and temperature distribution would require circuit area increase to meet performance constraints, which is also unsuitable solution.

TABLE II
VOLTAGE DROP AND TEMPERATURE MEASUREMENTS

| Active Tier(s) | Max Voltage Drop | Max Temperature |
|---|---|---|
| T1 only | 96.8mV measured @ T1 | 36.80°C measured @ T1 |
| T2 only | 97.5mV measured @ T2 | 36.38°C measured @ T2 |
| T3 only | 98.9mV measured @ T3 | 28.63°C measured @ T3 |
| T1 & T2 | 113.2mV measured @ T2 | 38.13°C measured @ T1 |
| T1 & T3 | 113.5mV measured @ T3 | 36.94°C measured @ T1 |
| T2 & T3 | 117.5mV measured @ T3 | 36.65°C measured @ T2 |
| T1 T2 T3 | 135.6mV measured @ T3 | 38.33°C measured @ T1 |

### A. Reutilization of Optimized 2-D PDNs on 3-D ICs

To illustrate these effects, we perform HSPICE transient voltage drop and thermal analysis on a three-tier network. For the sake of simplicity, the underlying circuit is represented as an identical core at 45-nm technology with parameters as in Table I. The mesh grid that is initially designed and optimized for delivering supply voltage to the core while meeting electrical and thermal constraints. The same PDN is used for each tier and connected using TSVs. TSVs are inserted in array shape with pitch of 60 $\mu$m and dimensions as in Table I.

We measure voltage drop and temperature levels on each tier for various setups and results are shown in Table II.

As shown in Table II, we note that for each scenario, the worst case voltage drop is measured on the tier that is the farthest from package (tier 3, T3), while the maximum temperature is measured on the tier that is the farthest from heat sink (tier 1, T1), corresponding to two distinct tiers. This shows that to satisfy both voltage drop and thermal constraints, it requires investigating all tiers simultaneously. Furthermore, we note that for individual active tiers voltage drop is within 10% of $V_{\text{DD}}$ (96.8 mV at T1, 97.5 mV at T2,

TABLE III
BUFFER AREA INCREASE TO ACCOUNT FOR NONUNIFORM VOLTAGE DROP AND THERMAL DISTRIBUTION

| All Tiers Active | | Max Voltage Drop | Max Temperature | Delay ratio | Area Ratio |
|---|---|---|---|---|---|
| All uniformly sized buffers | T1 | 110mV | 44°C | 1.29 | 1 |
| | T2 | 135mV | 39.8°C | 1.32 | 1 |
| | T3 | 143mV | 29.4°C | 1.34 | 1 |
| After resizing buffers | T1 | 114mV | 46.27°C | 1 | 1.02 |
| | T2 | 136mV | 41.5°C | 1 | 1.04 |
| | T3 | 145mV | 29.7°C | 1 | 1.05 |

and 98.9 mV at T3) and progressively increases as more tiers become active (up to 135.6 mV when T1, T2, and T3 active). A similar behavior is observed with temperature distribution. This suggests that PDNs that were designed and optimized standalone for 2-D implementation might not necessarily be optimal for a multitier network.

One way of mitigating excessive voltage drop and temperature is by increasing metal track widths. We apply uniform upsizing to all tiers by increasing their metal track widths. Pitches of track widths are based on ITRS predictions of global wire width [24]. Voltage drop constraint is set to 10% of $V_{DD}$ drop and temperatures up to 32 °C such that timing constraints are still met. Please note that these constraints are not based on an actual design, but simply used to illustrate the increase of PDN area for satisfying these constraints. By applying uniform upsizing, we report that a PDN area increase of 52% is required to meet both voltage drop and thermal constraints. This highlights the severity of the problem and special attention should be given to 3-D PDN design since reuse of 2-D PDNs is not as straightforward.

### B. Circuit Area Increase Due to Nonuniform Voltage Drop and Temperature Rise

Another way to look at this problem is to resize the underlying circuits to account for nonuniform voltage drop and temperature distributions among tiers such that timing constraints are met.

For illustration, we insert buffer chains on each tier and measure their performance when they are uniformly sized while they experience different voltage drop and temperatures on each tier. Then, we resize the buffers such that timing constraints are met. Please note that all tiers are active and PDNs are identical and remain unchanged. Results are shown in Table III.

Initially, buffers have the same area and a delay up to 34% increase from nominal delay is measured at T3 due large voltage drop (143 mV) and temperature (29.4 °C). Buffer resizing is performed for each tier and a maximum of 5% increase in buffer area for T3 is obtained. Despite being a small area increase, for a real chip a 5% of overall circuit area increase can impose additional resources (i.e., routing, TSVs, extension of power/clock networks, and power overhead), which may be simply impractical for current designs with already tight area and power budgets. It can also impose higher manufacturing costs due to additional processing area.

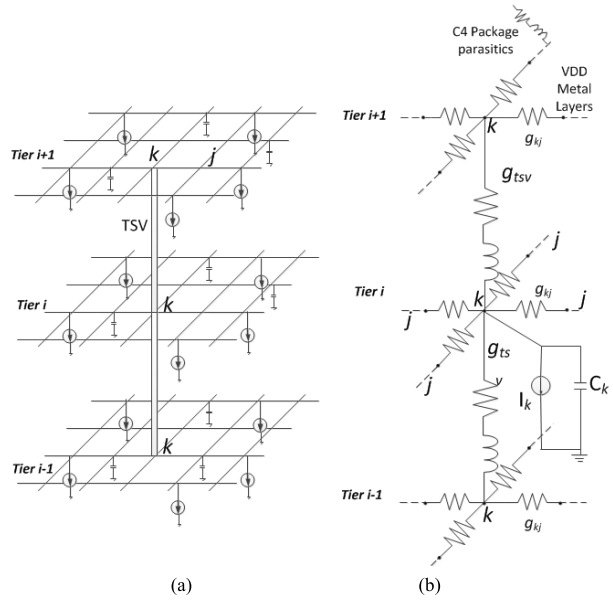



Fig. 2. (a) Illustration of 3-D PDN electrical model and (b) node $k$ on 3-D PDN.

These experiments clearly show that either reuse of already optimized 2-D PDNs or resizing of circuits to cope with nonuniform voltage drop and temperatures are unfeasible design choices for 3-D ICs. These observations further motivate us to investigate the power grid-sizing problem for multitier networks and develop an electrothermal optimization strategy for 3-D PDNs.

## IV. 3-D PDN ANALYSIS

Here, we first discuss electrical analysis followed by thermal analysis method based on electrical–thermal duality principle. We then discuss our proposed optimization method.

### A. Electrical Modeling and Analysis

*1) Electrical Modeling:* As mentioned in Section II, 3-D PDN electrical model includes package, power meshes, TSVs, decaps, and switching circuits. Topology and granularity of PDNs can vary on size and power density applied on them. In this paper, we adopt electrical models that have been widely applied in [2]–[9] and [25] and capture well PDNs behavior. In our case, we have extended these models to include TSVs and several PDNs of stacked tiers. Fig. 2(a) shows an illustration of a three-tier PDNs with TSVs, current sources, and decoupling capacitances.

*2) Electrical Analysis:* 3-D PDNs are modeled as RLC networks to represent parasitics of metal tracks for each tier, package, and TSVs where underlying circuits are modeled as time-varying current waveforms. The behavior of such network can be described by first-order differential equations formulated using modified nodal analysis method [26].

Each voltage node $k$ of tier $i$ on the multitier network can be expressed as

$$V_k^i = \frac{\sum_{\mathrm{j}\in\{\text{neighbors of k}\}} V_j^i g_{\mathrm{kj}}^i + \sum_{\mathrm{r}\in\{\mathrm{i+1,i-1}\}} V_k^r g_{\mathrm{tsv}}^{i,r} + I(s)_k^i}{\sum_{\mathrm{j}\in\{\text{neighbors of k}\}} g_{\mathrm{kj}}^i + \sum_{\mathrm{r}\in\{\mathrm{i+1,i-1}\}} g_{\mathrm{tsv}}^{i,r} + sC_k^i} \quad (1)$$

where $i$ is the current tier and $i+1, i-1$ represent top and bottom tiers. $j$ represents the neighboring nodes to node $k$ and $g_{\mathrm{kj}}$ represents the impedance admittance between the nodes $k$ and $j \cdot g_{\mathrm{tsv}}$ represents the TSV impedance admittance connecting two tiers at node $k \cdot I(s)$ represents the current waveform in $s$-domain of the underlying switching circuit at node $k$ for the current tier $i$ and $C$ is the decoupling capacitance at node $k$. Fig. 2(b) shows node $k$ on 3-D PDN.

In matrix form, node equations can be devised as

$$\left(\begin{bmatrix} G_t^1 & G_{\mathrm{tsv}}^{12} & 0 \\ G_{\mathrm{tsv}}^{21} & G_t^2 & G_{\mathrm{tsv}}^{23} \\ 0 & G_{\mathrm{tsv}}^{32} & G_t^3 \end{bmatrix} + s \begin{bmatrix} C_t^1 & 0 & 0 \\ 0 & C_t^2 & 0 \\ 0 & 0 & C_t^3 \end{bmatrix}\right) \cdot \begin{bmatrix} V_t^1 \\ V_t^2 \\ V_t^3 \end{bmatrix} = \begin{bmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{bmatrix} \quad (2)$$

where $G_{t_{n1\times n1}}^1$. $G_{t_{n2\times n2}}^2$, $G_{t_{n3\times n3}}^3$ are sparse and symmetric conductance matrices for each tier. Their sizes $n1 \times n1$ for tier 1, $n2 \times n2$ for tier 2, and $n3 \times n3$ for tier 3 can be different based on power network delivery topology and granularity of each tier. $G_{\mathrm{tsv}_{n1\times n2}}^{12}$ and $G_{\mathrm{tsv}_{n2\times n3}}^{23}$ are diagonal matrices that represent the conductance of the TSVs connecting tier 1 to 2 and tier 2 to 3, respectively.$G_{\mathrm{tsv}_{n2\times n1}}^{21}$ and $G_{\mathrm{tsv}_{n3\times n2}}^{32}$ are their transpose. We remark that several diagonal entries of $G_{\mathrm{tsv}}^{i,r}$ have zero values as there is not a TSV connected to each node of the grid given that TSV granularity is sparse on the power grid. $C_{t_{n_i\times n_i}}^i$ are diagonal matrices representing decoupling capacitances for each tier $i \cdot V_{t_{n_i\times 1}}^i$ are vectors representing node voltages and $B_{t_{n_i\times 1}}^i$ are the vectors representing current and voltage sources for each tier $i$. Using node voltage equations for each tier $i$, the current density $J_{\mathrm{branch_{kj}}}^i$ of branch $kj$ is computed as

$$J_{\mathrm{branch_{kj}}}^i = \frac{I_{\mathrm{branch_{kj}}}^i}{h_{\mathrm{kj}} \cdot w_{\mathrm{kj}}} = \frac{\left(V_k^i - V_j^i\right) g_{\mathrm{kj}}^i}{h_{\mathrm{kj}} \cdot w_{\mathrm{kj}}} \quad (3)$$

where $I_{\mathrm{branch_{kj}}}^i$ is the branch current derived from node voltages, $V_k^i V_j^i$ and their respective branch impedance, $g_{\mathrm{kj}}^i \cdot h_{\mathrm{kj}}$ and $w_{\mathrm{kj}}$ are height and width of the track where branch $kj$ resides. In this paper, we analyze 3-D power grids and current flow by considering only RL parasitics of the power metal tracks, while ignoring the current return paths.

### *B. Thermal Modeling and Analysis*

*1) Thermal Modeling:* Heat generated from transistors is dissipated through the silicon tiers, TSVs, heat sink, and package to the air. As heat sink and package are located on opposite ends, there are two heat path flows. In this paper, we consider both heat path flows and use thermal models for PDN, TSVs, package, and circuits.

In this paper, we use electrical–thermal duality principle [26] to represent a thermal network for performing fast thermal analysis. Due to this duality, temperature is analogous to voltage; heat can be modeled by current flowing through thermal resistances driven by current sources, which represent power consumption. Traditionally, thermal models are based on partitioning each power metal track into small wire segments [27]; hence it results into an extremely large thermal network. Instead, we apply an improved lumped thermal model as initially introduced for 2-D PDNs by [28], to reduce the number of thermal nodes and make thermal computation feasible.

The idea is to construct a thermal network that has the same structure and number of nodes as the electrical network. The motivation is to have a one-to-one correspondence between the electrical and thermal networks, which facilities our analytical formulation and analysis.

Thermal resistance is derived similarly as electrical resistance while using thermal conductivity coefficients. As the electrical branch resistance on the power grid is between two vias, thermal resistance is also represented similarly and derived by applying temperature at the vias as boundary condition for solving heat equation on a wire. Such representation is valid as long as the wire length between two vias is within the range of characteristic thermal length, $L_H$ [28], [29], [31]. Within this length the Joule heating generated at a wire will flow through the vias to the next metal layer rather than through the dielectric layer. Thus, wire segment temperature can be derived based on the temperature at vias. Thus, we can build a thermal model at the vias similarly as the electrical model. The final thermal network has the same structure as the electrical network, where thermal resistance is equivalent to electrical resistance [28] as

$$R_{\mathrm{thermal}} = \frac{R_{\mathrm{electrical}}}{\rho_0 k_m} \quad (4)$$

where $\rho_0$ is the resistivity at reference temperature and $k_m$ is the thermal conductivity of the metal layer. A thermal resistance represents the difference in temperature necessary to transfer a certain amount of heat and unit is °C/W or K/W. Derivation of thermal resistances represents the thermal PDN of each tier.

Thermal capacitance for each tier is derived based on the thickness and area of the die. Thermal capacitance represents the material's heat capacity that can be stored or removed to increase or decrease temperature and unit is $J$/°C or J/K. We use a distributed thermal capacitance model that is introduced on each node of thermal network. Single-die thermal capacitance [31], [32] can be derived as

$$C_{\mathrm{thermal}} = c_p \cdot t \cdot A \quad (5)$$

where $c_p$ is thermal capacitance of the material while $t$ and $A$ are die thickness and area. From [30], silicon has $c_p = 10^6$ J/m$^3$K and $\rho_0 = 10^{-2}$ mK/W.

Heat transfer from die to heat sink is modeled by a thermal resistance between every grid node to heat sink node. As a case study, we assume a heat sink with thermal resistance of 2 K/W. As circuit transients are in the order of picoseconds compared with temperature variations in seconds at the heat sink [31], we model heat sink at a constant temperature represented by a constant voltage source. For example, a heat sink with

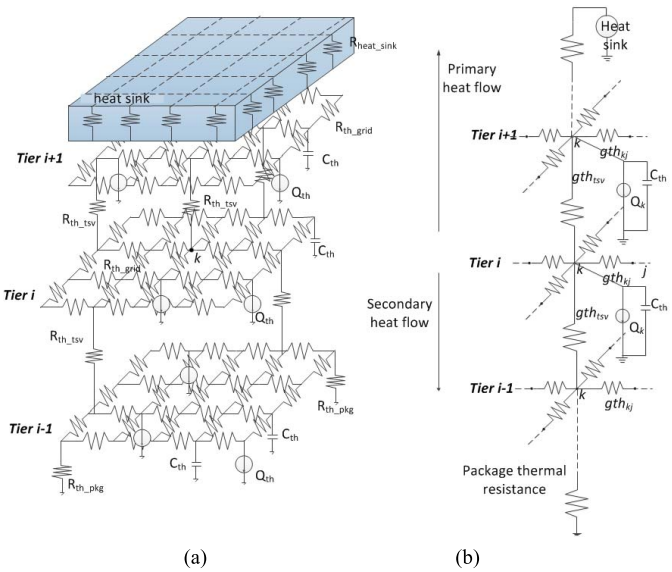



Fig. 3. (a) Illustration of 3-D PDN thermal model and (b) node $k$ on 3-D PDN.

thermal capacitance of 30 J/K and thermal resistance of 2 K/W, we can determine the time constant to cooling/heating as RC = 30 J/K ∗ 2 K/W = 60 s.

TSVs are represented as RC model based on their thermal characteristics as described in [33] and [34].

Package with C4 bumps provide a secondary heat path for dissipating the generated heat [35], [36]. As a case study, a package thermal resistance of 20 K/W is assumed.

Heat generated by switching circuits is presented as current sources proportional to their power consumption obtained from electrical analysis.

Overall, we aim to represent a complete thermal 3-D PDN model (i.e., including package, heat sink, circuits, TSVs, and power meshes) while also compact (i.e., same size and structure as electrical network) to compute temperature distribution on 3-D PDNs and feed it back to electrical analysis. Fig. 3(a) illustrates thermal 3-D PDN model.

*2) Thermal Analysis:* Similar to electrical network, a thermal network can be devised where temperature at node $k$ can be expressed as

$$T_k^i = \frac{\sum\limits_{\mathrm{j}\in\{\text{neighbors of k}\}} T_j^i g_{\mathrm{th_{kj}}}^i + \sum_{\mathrm{r}\in\{\mathrm{i+1,i-1}\}} T_k^r g_{\mathrm{th_{tsv}}}^{i,r} + Q_k^i}{\sum\limits_{\mathrm{j}\in\{\text{neighbors of k}\}} g_{\mathrm{th_{kj}}}^i + \sum_{\mathrm{r}\in\{\mathrm{i+1,i-1}\}} g_{\mathrm{th_{tsv}}}^{i,r}} \quad (6)$$

where $T_k$ represents the temperature at node $k$ and $T_j$ represent the temperatures at neighboring nodes $j \cdot g\mathrm{th_{kj}}$ represents the thermal impedance between the nodes $k$ and $j$, and $g\mathrm{th_{tsv}}$ represents the thermal impedance for a TSV. $Q_k$ represents the heat dissipation (or power consumption for electrical network) at node $k$. Thermal networks can be devised in matrix form as

$$\begin{bmatrix} P_t^1 & P_{\mathrm{tsv}}^{12} & 0 \\ P_{\mathrm{tsv}}^{21} & P_t^2 & P_{\mathrm{tsv}}^{23} \\ 0 & P_{\mathrm{tsv}}^{32} & P_t^3 \end{bmatrix} \cdot \begin{bmatrix} T_t^1 \\ T_t^2 \\ T_t^3 \end{bmatrix} = \begin{bmatrix} Q_t^1 \\ Q_t^2 \\ Q_t^3 \end{bmatrix} \quad (7)$$

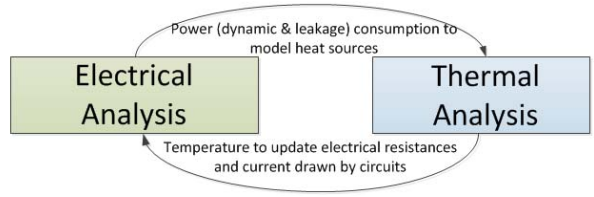



Fig. 4. Electrothermal analysis iterative process.

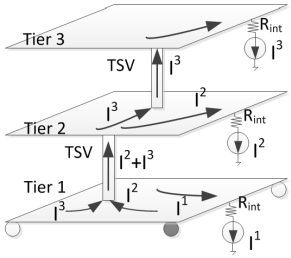
where $P_{t\,n_i x n_i}^i$ are the sparse and symmetric matrices representing thermal conductance of each tier $i$. $P_{\mathrm{tsv}_{n1\times n2}}^{12}$ and $P_{\mathrm{tsv}_{n2\times n3}}^{23}$ are the diagonal matrices of thermal conductances for TSVs connecting tier 1 to 2 and tier 2 to 3, respectively. $P_{\mathrm{tsv}_{n2\times n1}}^{21}$ and $P_{\mathrm{tsv}_{n3\times n2}}^{32}$ are their transposes. $T_{t_{n_i\times 1}}^i$ are the vectors of node temperatures for each tier $i$. $Q_{t_{n_i\times 1}}^i$ are the vectors representing the heat dissipation for each tier $i$. We also consider Joule-heating of power metal tracks due to their current flow. Joule heating power computed as $Q_{\mathrm{self}} = I^2 \cdot R_{\mathrm{electrical}}$, which is a function of thermal conductivity of wire and temperature. Self-heating power $Q_{\mathrm{self}}$ of power grid wires are considered at $Q_k$ of each node. For example, a power grid wire between two nodes (i.e., nodes $k$ and $k+1$), $Q_{\mathrm{self}}$ is proportionally distributed between these two nodes. The behavior of electrical and thermal networks can be coupled together since electrical resistivity; $\rho$ is dependent to temperature and expressed as

$$\rho = \rho_0[1 + \beta\,(T - T_0)] \quad (8)$$

where $\rho_0$ is electrical resistivity at reference $T_0$ (ambient temperature at 27 °C) and $\beta = 0.0039/°\mathrm{C}$ is the temperature coefficient for resistance. Thus, the resistance of the power metal tracks can be computed as

$$R_i = \rho \frac{L_i}{H_i \cdot W_i}[1 + \beta\,(T_i - T_0)] \quad (9)$$

where $L_i, w_i$ and $H_i$ are the length, width and height of the metal track, respectively. As temperature changes, $\mathrm{R_{electrical}}$ changes, which consequently affects voltage drop and current flow. Naturally, node voltage equation in (1), the conductance parameters (i.e., $g_{\mathrm{kj}}, g_{\mathrm{tsv}}$) are temperature dependent, which lead to voltage drop and current flow change on power wires. The temperature changes are fed to electrical analysis engine for updating impedances and circuit current model (i.e., $I_{\mathrm{leakage}}$ and $I_{\mathrm{peak}}$) to recompute voltage droop and power consumption (i.e., dynamic and leakage power). Power dissipation serves as an input to thermal analysis for deriving heat sources. Due to this closely coupled relation, electrical and thermal analyses run in several loops until final voltage drop and temperature distributions are obtained also, as shown in Fig. 4. Equations (2) and (7) can be solved fast and accurately using different linear algebra solvers (i.e., direct or iterative methods). In this paper, multigrid method is applied to reduce the size of matrices while exploiting sparsity of matrices and power grid topology [37]. Multigrid method initially relaxes the power grid by reducing its granularity from finer to coarser grid. Once the coarse grid is solved, its solution is mapped back to the original size power grid. Multigrid method has been widely and effectively employed for power grid analysis [37], [38].




(a)

$$V^3_{droop} = \frac{\max_V^{for_all_tiers}_{droop}}{3} - R^{23}_{tsv} \cdot I^3 - R^3_{int} \cdot I^3$$

$$V^2_{droop} = \frac{\max_V^{for_all_tiers}_{droop}}{3} - R^{12}_{tsv} \cdot (I^2 + I^3) - R^2_{int} \cdot I^2$$

$$V^1_{droop} = \frac{\max_V^{for_all_tiers}_{droop}}{3} - R^1_{int} \cdot I^1$$

(b)

Fig. 5. (a) Current flow on a three-tier 3-D PDN to compute voltage drop constraints and (b) voltage drop constraints for each tier.

## V. 3-D PDN Optimization Method

We propose a tier-based optimization approach for resizing the 3-D PDN to satisfy both voltage drop and thermal constraints. Given that each tier can have different PDN topology, circuit functionality, and power density, we treat them individually while imposing 3-D PDN aware power and thermal constraints. Thus, the essence of our 3-D PDN optimization method is to locally optimize each tier while applying global constraints.

Each tier's constraints are derived based on the knowledge of the 3-D PDN electrical and thermal targets, current demand and power density of each tier, and TSV sizing. We define the tier-specific voltage drop constraints for a tier $i$ as

$$V^i_{\text{droop}} = \frac{\max_V^{\text{for all tier}}_{\text{droop}}}{\text{number of tiers}} - R^{i,i+1}_{\text{tsv}} \cdot \sum_{j=i}^{\text{toptiers}} I^j - R^i_{\text{int}} I^i \quad (10)$$

where $\max_V^{\text{for all tier}}_{\text{droop}}$ represents the total amount of voltage drop allowed such as 10% of $V_{\text{DD}}$. The accumulated resistance of all TSVs connecting two adjacent tiers is represented as $R^{i,i+1}_{\text{tsv}}$ and $I^j$ as the upward current flow from package to each tier through TSVs. As face-to-back TSVs are considered, global power grid of bottom tier is connected to local power grid of top tier. Thus, resistance $R^i_{\text{int}}$ ($\Omega$) aims to represent the cumulative parasitic impedance from local to global power grid for each tier including local, intermediate, and global grids. For a three-tier system, voltage drop constraints can be derived, as in Fig. 5. In a similar way, each tier's thermal constraints can be derived with respect to number of tiers, heat dissipation and TSV sizing. We define tier-specific thermal constraint for a tier $i$ as

$$T^i_{\max} = \frac{\max_T^{\text{tiers}}_{\text{difference}}}{\text{number of tiers}} - P^{i,i+1}_{\text{tsv}} \cdot \sum_{j}^{\text{bottom tiers}} Q^j - P^i_{\text{int}} Q^i \quad (11)$$

where $\max_T^{\text{tiers}}_{\text{difference}}$ is the maximum allowed temperature difference between top and bottom tiers, $P^{i,i+1}_{\text{tsv}}$ is the accumulated thermal resistance of all TSVs connecting two adjacent tiers, and $Q^j$ represents the amount of heat generated at each tier. $P^i_{\text{int}}$ represents the cumulative parasitic thermal impedance from local, intermediate to global grid. Thus, $V^i_{\text{droop}}$ and $T^i_{\max}$ serve as tier-specific constraints, which are derived from the 3-D PDNs to account for the voltage drop and heat generated in other tiers and TSVs. Such constraints facilitate to investigate each tier's power grid sizing problem individually without imposing any power/thermal issues to the rest of the tiers. Additionally, the computational complexity of the problem is significantly reduced as the number of circuit nodes to be analyzed is same as in 2-D PDN optimization problem, whereas a full 3-D problem would have at least $m$–times more circuit nodes ($m$, number of tiers).

In the following section, we describe two proposed electrothermal algorithms.

### A. 3-D PDN Standalone Optimization

Here, we describe the optimization technique applied only to the metal tracks of the 3-D PDN.

*Problem Formulation:* Derive the metal track widths for the PDN of each tier while applying 3-D PDN aware electrothermal constraints such that the minimum PDN area is obtained.

Optimization method can be applied to any tier (regardless of stacking order) together with its specific voltage drop and thermal constraints. Please note that the optimization technique can be applied on either coarse or original size grids. The problem of PDN sizing with voltage drop and thermal constraints can be formulated as a linear optimization problem. We aim to minimize the PDN area by minimizing the summation of metal track widths as

Objective: $\min\{\sum w\}$ such that the following constraints are met:

1) Voltage drop

$$V^i_k \geq V_{\text{DD}} - V^i_{\text{droop}}$$

based on node voltage equations on (1) and 3-D PDN aware voltage drop constraint, as in (10).

2) Temperature

$$T^i_k \leq T_{\text{amb}} + T^i_{\max}$$

based on node temperature equations on (6) and 3-D PDN aware temperature constraint as in (11).

3) Track widths

$$w_{\min} \leq w \leq w_{\max}$$

track widths are within an upper and lower range.

4) Electromigration

$$J^i_{\text{branch}_{kj}} \leq J_{\max}$$

based on branch current flow, as in (3) to ensure that current density for PDN branch, $kj$ does not surpass maximum allowed current density, $J_{\max}$.

5) Optimization parameter

$$0 \leq \alpha \leq 1$$

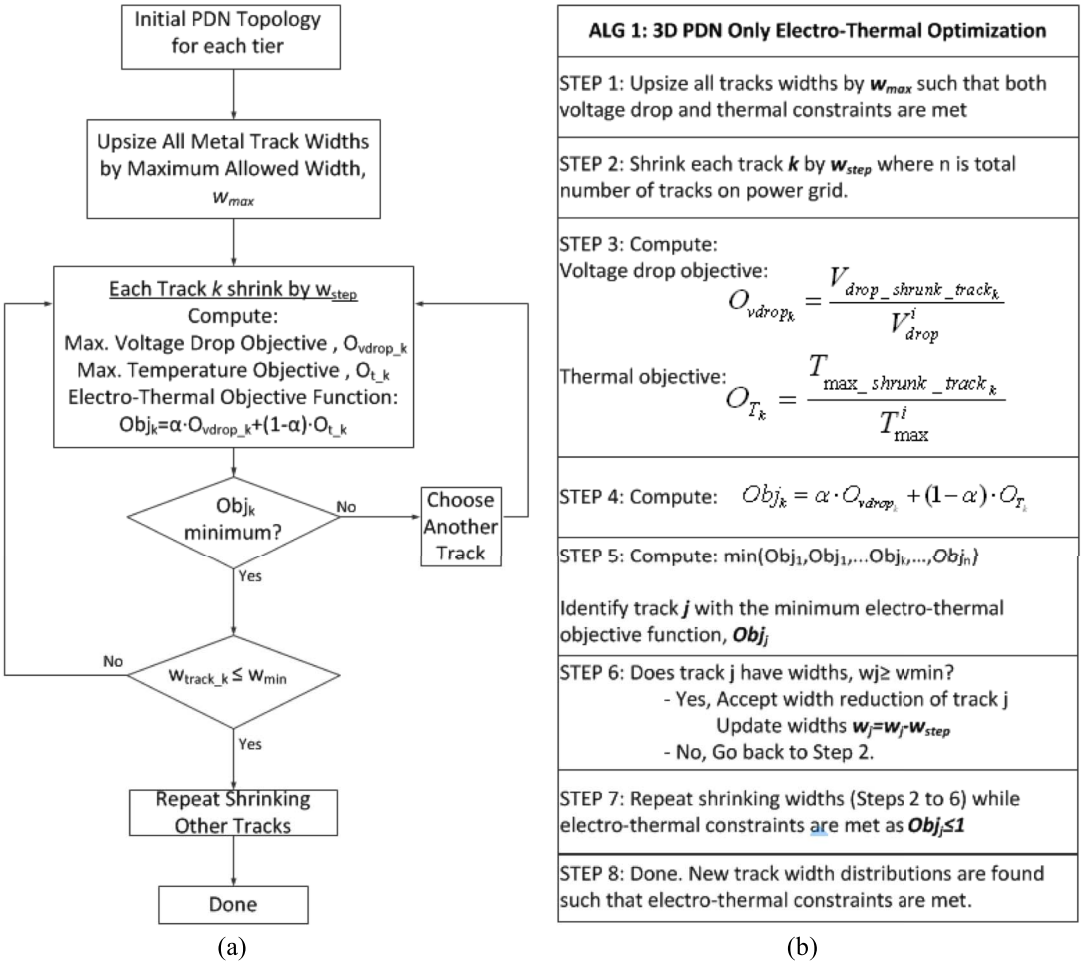



Fig. 6. (a) Flowchart of the 3-D PDN electrothermal optimization method and (b) step-by-step algorithm description.

where $\alpha = 1$ enables voltage drop optimization and $\alpha = 0$ enables thermal optimization. Constraints are weighted as $\alpha \cdot V_k^i + (1-\alpha)\, T_k^i$.

Please note that TSV sizes remain unchanged during this optimization flow. One way to satisfy all the constraints is to upsize the widths of all metal tracks on the grid. However, this would result in a significantly large power grid area that is unnecessarily over-designed. Instead, we propose to upsize all tracks by maximum allowed width, $w_{\max}$ and iteratively reduce by $w_{\text{step}}$ the widths of those tracks that have minimal impact on power and thermal integrity while still satisfying voltage drop and thermal constraints. Some tracks due to electromigration constraints may not be reduced much as they can pose current density issues. Thus, we aim to reduce the PDN area by maximizing the track widths to be downsized. Fig. 6 shows the flowchart and the steps of the optimization algorithm.

Variable $w_{\text{step}}$ is a user-selected parameter, which can impact the quality of the solution and the number of iterations. Similarly, $\alpha$ is a user-selected parameter to vary the weight between the voltage drop and the thermal optimization. Note that in Step 3, we compute voltage drop ($O_{\text{vdrop}}$) and thermal ($O_T$) objective functions as ratios to the tier-specific constraints. This allows them to be used as unitless numbers for computing the objective function in Step 4 regardless of representing two different entities.

As the final step, a voltage drop and thermal analysis is performed where all optimized PDNs are considered simultaneously. We find that tier-based electrothermal optimization guarantees power and thermal integrity of 3-D PDNs.

Initially by upsizing all power metal tracks ensure that both voltage droop and temperature constraints are met. This step also serves as the basis for convergence for our optimization method. Meaning that, the starting point of algorithm is a good (not optimal) solution, which guarantees the optimization engine to converge by further finding an optimal solution. The complexity of the algorithm is $O(n \cdot n_{\text{step}})$, where $n$ is the number of power metal tracks and n_step is the number of shrinking steps. Algorithm evaluates voltage drop and temperature each time a track is shrunk and for each shrinking step ($w_{\text{step}}$).

### B. 3-D PDN and TSV Sizing Co-Optimization

Here, we describe the electrothermal optimization method is applied to both metal track widths and TSV sizes.

*Problem Formulation:* Derive the metal track widths and TSV sizes (diameter) while applying 3-D PDN aware electrothermal constraints such that the minimum PDN area and TSVs sizes are obtained.

Only the diameter is considered as variable for sizing the TSVs, whereas their thickness and pitches are assumed constant (due to stacking approach). As mentioned in Section II,

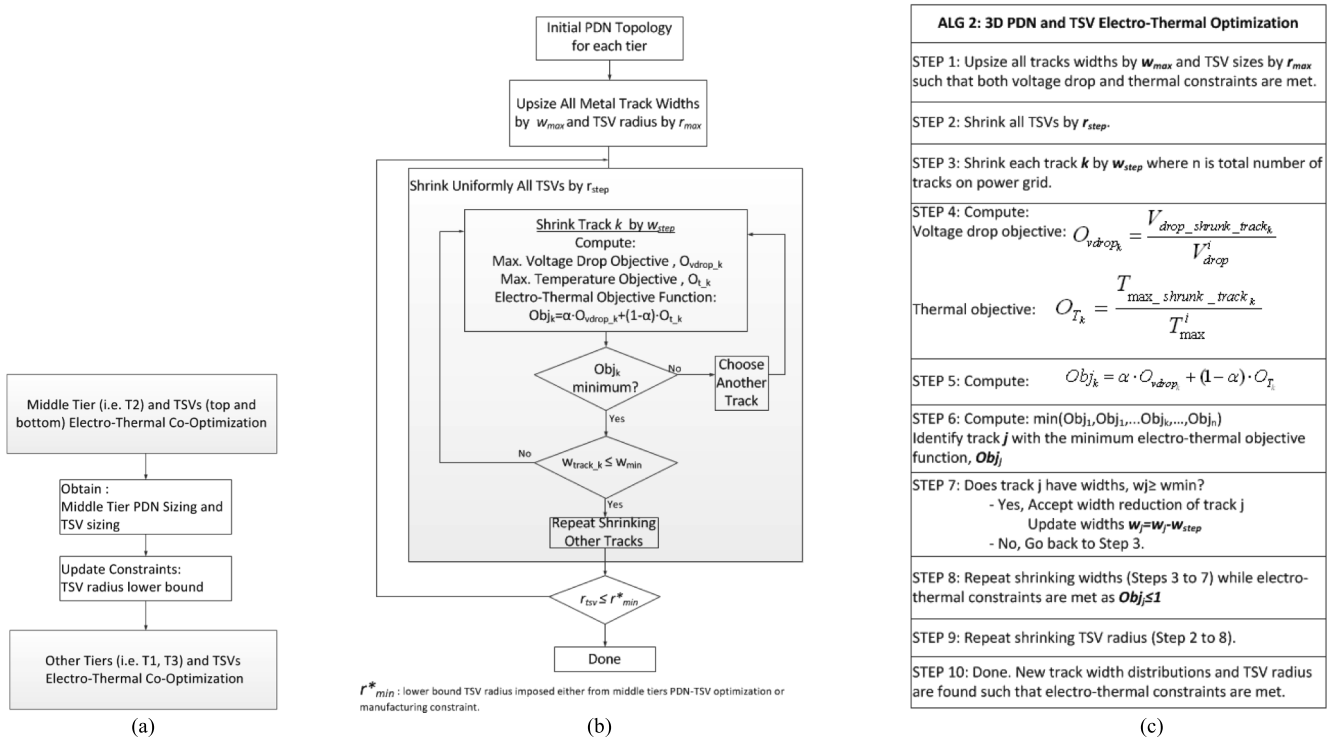



Fig. 7. (a) 3-D PDN-TSV co-optimization flow. Initially middle tiers and TSVs are co-optimized where the obtained TSV sizes serve as lower bound constraints for other PDN-TSV optimization. (b) Flowchart of the 3-D PDN-TSV electrothermal optimization method. (c) Step-by-step algorithm description.

we consider high-density TSVs inserted in array shape where the diameters can vary from 1 to 5 $\mu$m [21]. Please note we consider uniformly sized TSVs where all TSVs have the same radii. In addition, TSVs continue to remain uniformly sized even after our PDN-TSV co-optimization method.

An important aspect of the tier-based proposed algorithm is the treatment of TSVs such as how are TSVs optimized and how are they included to the PDN optimization. These are critical aspects and can impact 3-D PDNs resiliency because TSVs can be sized optimally for one tier but may create voltage drop or thermal issues for the next tier.

Most sensitive tiers to voltage drop and thermal are middle tiers as they are remotely connected to either package pins or heat sink. Thus, our electrothermal PDN-TSV optimization method is initially applied to middle tiers to account for TSV impacts. The obtained TSV sizes then serve as lower bound constraints for the rest of the tiers during their PDN-TSV co-optimization. Fig. 7 shows our PDN-TSV electrothermal co-optimization method.

The problem of PDN-TSV sizing with voltage drop and thermal constraints can be formulated as a linear optimization problem. We aim to minimize the PDN and TSV area by minimizing the summation of their widths as:

Objective: min $\left\{\sum \beta \cdot w_{\text{track}} + \sum (1-\beta) r_{\text{tsv}}\right\}$ such that the following constraints are met:

1) Voltage drop

$$V_k^i \geq V_{DD} - V_{\text{droop}}^i \tag{12}$$

based on node voltage equations on(1) and 3-D PDN aware voltage drop constraint, as in (10).

2) Temperature

$$T_k^i \leq T_{\text{amb}} + T_{\text{max}}^i \tag{13}$$

based on node temperature equations on (6) and 3-D PDN aware temperature constraint, as in (11).

3) Metal track widths

$$w_{\text{min}} \leq w \leq w_{\text{max}}$$

track widths are within an upper and lower range.

4) TSV radius

$$r_{\text{min}}^* \leq r \leq r_{\text{max}}$$

where $r_{\text{min}}^*$ can be either the minimum radius range or the imposed lower bound radius constraint from PDN-TSV optimization of the middle tiers.

5) Electromigration

$$J_{\text{branch}_{\text{kj}}}^i \leq J_{\text{max}}$$

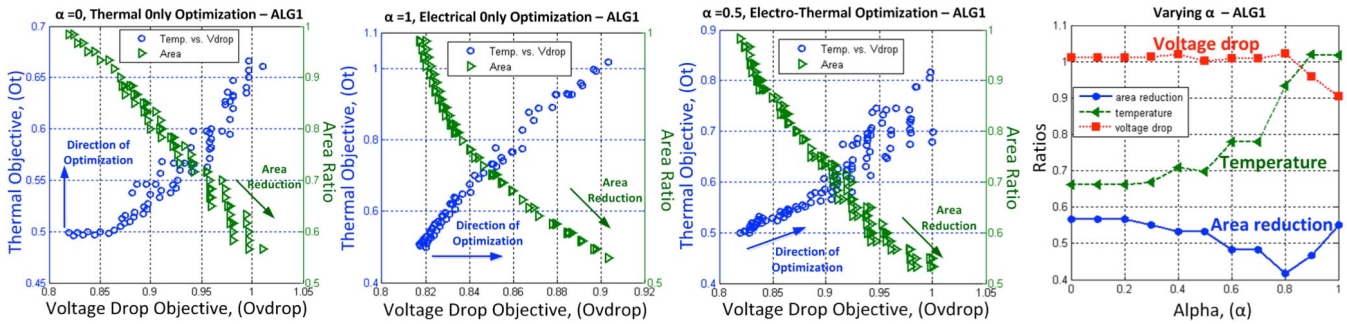
based on branch current flow, as in (3) to ensure that current density for PDN branch, $kj$ does not surpass maximum allowed current density, $J_{\text{max}}$.

6) Optimization parameter

$$0 \leq \alpha \leq 1$$

where $\alpha = 1$ enables voltage drop optimization and $\alpha = 0$ enables thermal optimization. Constraints are weighted as $\alpha \cdot V_k^i + (1-\alpha) T_k^i$.

Parameter $\beta$ is introduced to provide a weight coefficient between power track sizes and TSV radius. It can vary as $0 \leq \beta \leq 1$.




(a) (b) (c) (d)

Fig. 8. Optimization objectives for ALG 1 when (a) $\alpha = 0$, (b) $\alpha = 1$ (c) $\alpha = 0.5$, and (d) optimized power grid area with varying $\alpha$.

In summary, the proposed globally constrained locally optimized methodology is flexible to be applied to any PDN regardless of topology or technology. As heterogeneous technologies and functionalities can be implementable in 3-D, our electrothermal optimization problem helps to study each tier standalone while applying constraints that consider multitier network and TSVs.

## VI. Experiments

We have implemented our tier-based power grid-resizing algorithm in MATLAB and tested it on various 3-D PDN setups. We applied circuit parameters, as shown in Table I along with wire pitches $w_{\min}$, $w_{\max}$, and current density, $J_{\max}$ as [24]. Voltage drop and temperature constraints were set to 10% of $V_{\mathrm{DD}}$ and $\mathrm{T}_{\max} = 1\ 0$ °C difference between topmost and bottommost tiers. Ambient temperature is set to $\mathrm{T}_{\mathrm{amb}} =$ 27 °C. Please note that both $\mathrm{T}_{\max}$ and $\mathrm{T}_{\mathrm{amb}}$ are not based on any specific design or technology parameter but simply chosen as a case study. Both parameters are user dependent. Track width sizing parameter $w_{\mathrm{step}}$ and $r_{\mathrm{step}}$ are set to 5% of the original track widths and TSV radius.

In our experiments, we applied both ALG 1 and ALG 2, as shown in Figs. 6 and 7. There are no prior works that developed tools for 3-D PDNs electrothermal optimization, which is why we use uniform resizing algorithm for comparison. We conducted our experiments on a three-tier network with PDNs of mesh topology. We use three types of power density blocks distributed on a tier and through tiers: high ($h$), medium ($m$), and low ($l$) power density blocks. We studied the impact of optimization parameter $\alpha$, power density distribution among tiers and design constraints on power and thermal integrity of 3-D PDNs.

### A. Impact of Optimization Parameter, α

We applied our globally constrained locally optimized optimization method on the three-tier power network while varying the electrothermal optimization parameter, $\alpha$. $\alpha$ values vary from 0 to 1, where 0 refers to thermal only, 1 refers to electrical only and 0.5 refers to electrothermal optimization.

To capture $\alpha$ effect on the power grid area, we conducted experiments on PDNs with identical power density distributions of 0.6 $\mu$W/$\mu$m$^2$. We apply algorithms ALG1 and ALG2 and the results are shown in Figs. 8 and 9, respectively. Results are represented with respect to voltage drop objective, $O_{\mathrm{vdrop}}$, thermal objective, $O_t$, and 3-D PDN area ratio. As derived in Section V, voltage drop objective ($O_{\mathrm{vdrop}}$) is unitless number to indicate the correlation to the voltage drop constraint. For example, $O_{\mathrm{vdrop}} = 1$ indicates that voltage drop is equal to maximum allowed voltage drop of 100 mV. Similarly, thermal objective ($O_t$) is a unitless number to indicate the correlation to the maximum temperature, $\mathrm{T}_{\max}$ constraint, where $O_t = 1$ is equal to 37 °C (where $T = T_{\mathrm{amb}} + T_{\max} =$ 27 °C + 10 °C = 37 °C derived from temperature constraints as in Section V) and $O_t = 0$ corresponds to 27 °C. Area ratio displayed on the second $y$-axis (right hand side on figures) represent the area savings from the optimization algorithms in comparison with uniform resizing algorithm, where $w_{\max}$ (maximum metal track width) and $\mathrm{r}_{\max}$ (maximum TSV radius) were applied such that both electrical and thermal constraints were met. For example, area ratio = 1 represents the 3-D PDN area when maximum upsizing for ALG1 (metal tracks only) and ALG2 (metal tracks and TSVs) are applied.

In Fig. 8(a), we observe that for $\alpha = 0$ (thermal optimization only), there is up to 42% of total area reduction from uniformly upsizing all tiers by $w_{\max}$. Track widths are shrunk where thermal objective, $O_T$ varies from 0.5 to 0.68 and voltage drop objective, $O_{\mathrm{vdrop}}$ varies from 0.8 to 1, thus both electrical and thermal constraints remain satisfied. Additionally, we note that there are oscillations during the optimization. This is due to various track width distributions that provide different amount of voltage drop without much change in temperature. In the case of $\alpha = 1$ (voltage drop optimization only), there is a total of 45% of area reduction from uniform upsizing all tiers, as shown in Fig. 8(b). We note that there is a minimal impact on $O_{\mathrm{vdrop}}$ that varies from 0.82 to 0.9 while $O_T$ varies from 0.5 to 1. Fig. 8(c) shows the optimization objectives for $\alpha = 0.5$ or equally weighted electrothermal optimization. A 48% area reduction is achieved while $O_{\mathrm{vdrop}}$ varies from 0.82 to 1 and $O_T$ varies from 0.5 to 0.82.

In Fig. 8(d), we show the reduced multitier power network area for various $\alpha$ values with respect to their maximum voltage drop and temperature objectives. In general, we note that voltage is less sensitive than temperature with varying $\alpha$. This shows that the changes in power tracks widths provide significant improvement on temperature distribution

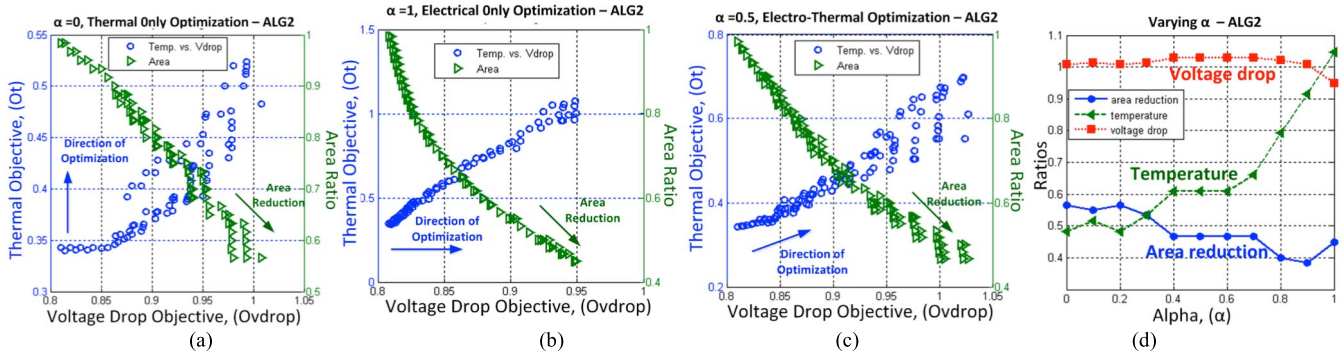



Fig. 9. Optimization objectives for ALG 2 when (a) $\alpha = 0$, (b) $\alpha = 1$, (c) $\alpha = 0.5$, and (d) optimized power grid area with varying $\alpha$.

while less difference on worst case voltage drop. This is due to wider power tracks enabling lateral heat dissipation on a tier. Whereas voltage drop modeled as IR drop, *Ldi/dt*, and *RCdv/dt* effects experiences less improvement with power tracks widths. This suggests that *Ldi/dt* and *RCdvdt* effects continue to be significant even as power track widths are varied.

In Fig. 9, we show the optimization results for ALG2, which includes the optimization of TSV sizing together with metal track widths. Overall, we notice that there is more 3-D PDN area savings obtained when both TSVs and metal tracks are considered. For example, up to 43% for $\alpha = 0$ [Fig. 9(a)], 55% for $\alpha = 1$ [Fig. 9(b)], and 53% for $\alpha = 0.5$ [Fig. 9(c)]. Fig. 9(d) shows the area for various values of $\alpha$. We note up to 62% of area savings for $\alpha = 0.9$.

These results indicate the importance of considering TSV sizing along with the design of PDN. We note that additional 3-D PDN area savings can be obtained when both metal track widths and TSV radii are considered. TSVs are crucial for delivering current and removing heat from one tier to the next. Thus, resizing TSVs provides headroom for 3-D PDNs to meet power and thermal constraints.

### *B. Impact of Identical and Nonidentical Dies*

Given that each tier can have different dies inserted on them, we study the impact of fully and partially identical dies. Each die can have different network topology, functionality, and power density distribution. We investigate three kinds of dies represented as single-cores of different power density distributions. We study their different stacking options on a three-tier setup.

Three power density distributions are: 1) $h$, high power density of 1.5 $\mu$W/$\mu$m$^2$; 2) $m$, midpower density of 1$\mu$W/$\mu$ m$^2$; and 3) $l$, low power density of 0.8$\mu$W/$\mu$ m$^2$. We note that initially all PDNs were uniformly upsized such that voltage drop and thermal constraints were met. We use ALG1 algorithm to perform electrothermal optimization. We consider six different cases (cases 1–6 in Table IV) with nonidentical dies inserted on each tier. In addition, we consider three cases with identical tiers (cases 7–9), as listed in Table IV.

For each case, we report voltage drop objective, $O_{\text{Vdrop}}$, and thermal objective, $O_T$, before and after ALG1 was applied while $\alpha = 0.5$. Some cases have $O_T$ objective of order 0.9

TABLE IV
AREA SAVINGS FOR DIFFERENT POWER DENSITY DISTRIBUTIONS

| Case # | T3-T2-T1 | $O_{\text{Vdroop}}$ | | $O_T$ | | Area | Run time (s) |
|---|---|---|---|---|---|---|---|
| | | Before | After | Before | After | Saving | |
| *1* | *h-m-l* | 0.87 | 1 | 0.46 | 0.56 | 25% | 1440 |
| *2* | *h-l-m* | 0.91 | 1 | 0.5 | 0.56 | 20% | 2160 |
| *3* | *m-h-l* | 0.91 | 1 | 0.56 | 0.61 | 18.3% | 1550 |
| *4* | *m-l-h* | 0.84 | 1 | 0.6 | 0.72 | 26.7% | 1550 |
| *5* | *l-m-h* | 0.79 | 1 | 0.63 | 0.82 | 33.3% | 1780 |
| *6* | *l-h-m* | 0.87 | 1 | 0.58 | 0.7 | 26.7% | 1200 |
| *7* | *h-h-h* | 0.98 | 1 | 0.66 | 0.7 | 8% | 900 |
| *8* | *m-m-m* | 0.88 | 1 | 0.45 | 0.9 | 36% | 780 |
| *9* | *l-l-l* | 0.86 | 1 | 0.34 | 0.82 | 41% | 1140 |

as simultaneously $O_{\text{vdrop}}$ objective reached max 1. Further optimization of thermal, $O_T$ would lead to increase of voltage drop, $O_{\text{vdrop}}$, hence optimization stops and the values of $O_T$ and $O_{\text{vdrop}}$ are reported. When nonidentical dies were used, we note that case 3 has the least area savings and this is due to high voltage drop and thermal objectives before optimization. case 2 is also similar to case 3. In case 2, high-density tier ($h$) is at T3 farthest away from package (high voltage drop) while it is located next to the heat sink (immediate cooling), thus more area saving is obtained in case 2 than 3. For cases of nonidentical dies, the most area savings are obtained in case 5 due to small $O_{\text{vdrop}}$ before optimization. case 5 appears similar to case 6. However, the location of high-density tier in case 5 cause it to have less voltage drop (as is next to package) while in case 6, it is located in the middle tier, which results in more voltage drop but at lower temperatures (closer to heat sink), thus more area saving is obtained in case 5.

When identical dies were used, we note that case 9 with low power density distributions have the least voltage drop and thermal objectives, thus the most area savings up to 41% are obtained. In contrary, identical dies with high power density distributions, case 7 lead to the least area savings, up to 8% from uniform upsizing. Such results indicate that sensitivity of 3-D PDNs to identical dies with high power density distri-

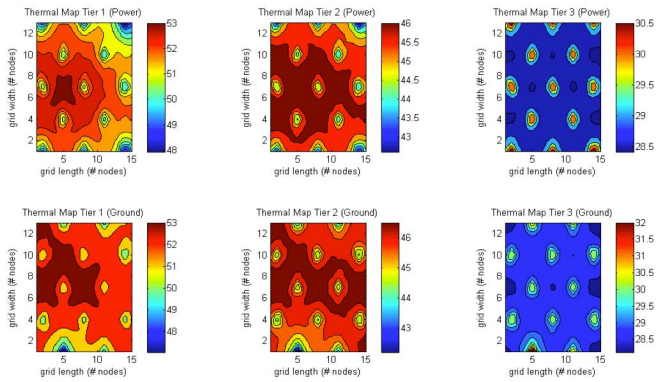



Fig. 10. Thermal maps on power and ground networks for each tier when package model is not included.

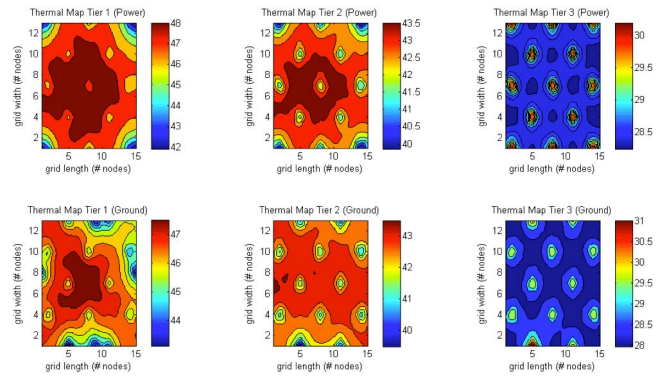



Fig. 11. Thermal maps on power and ground networks for each tier when package model is included.

butions, which might limit the benefits of 3-D integration. Furthermore, it highlights the importance of workload distribution among tiers while satisfying power and thermal constraints. We also report the runtime for each case. We observe that the initial quality of the design impacts the runtime. In addition, the number of iterations for some cases can vary due to the sensitivity of voltage drop and temperature on track width distribution.

### C. Impact of Design Constraints

*1) Package as Secondary Heat Path:* Package can have an important impact on temperature distribution across the tiers in 3-D ICs. Generated heat from devices is conducted from one silicon die to another, thermal interface material, heat spreader to the heat sink then dissipated into air. This is the primary heat flow. There is also a secondary path as heat can flow from devices, interconnects, TSVs, I/O pads, and package bumps reaching to the board. In this paper, we perform an experiment to demonstrate the impact of package as the second heat path on a three-tier network with identical power density distributions. As mentioned in Section III, package thermal model is represented by thermal resistances (lateral and vertical) based on its thermal conductivity.

Fig. 10 shows the temperature distribution on each tier for both power and ground delivery networks with no package model. We note that temperatures can vary considerably inter and intratier. Tier 3 (T3) is next to heat sink and experiences temperatures from 28.5 °C to 30 °C on power network or 32 °C on ground network. Whereas, tier 1 (T1) is next to package and undergoes temperatures from 48 °C to 53 °C on both power and ground networks. Tier 2 being in the middle of the stack experiences temperatures from 43 °C to 46 °C. We deduce that TSVs between T1 and T2 are faced with high temperature difference (up to 13 °C) between the tiers. TSVs from T2 and T3 experience temperature difference up to 2 °C between the tiers.

Fig. 11 shows the temperature distribution on each tier with the package model. Overall, the temperature levels are reduced in comparison with Fig. 10. Temperatures reach up to 31 °C for T1, 43.5 °C for T2, and 48 °C for T3. A temperature difference up to 5 °C for T3 is obtained or 10% of temperature reduction.

TABLE V
AREA SAVINGS FOR DIFFERENT POWER DENSITY DISTRIBUTIONS

| Wafer thinning | Electro-Thermal Analysis | | 3D PDN AREA | | |
|---|---|---|---|---|---|
| | $O_{Vdroop}$ | $O_T$ | Uniform Resizing | ALG1 | ALG2 |
| Top two wafers thinned at ***50μm*** | 0.82 | 0.62 | 1 | .31 | .28 |
| All wafer at ***775μm*** thickness | 0.81 | 0.54 | 1 | 0.25 | 0.22 |

Similarly, TSVs undergo slightly lower temperatures such as 10 °C between T1 and T2 and 1 °C between T2 and T3. This experiment highlights the importance of secondary heat path through package as it helps to remove some of the generated heat. In addition, neglecting the package effects can lead to inaccurate temperature estimates.

*2) Wafer Thinning:* Different manufacturing approaches for 3-D ICs are categorized by the TSV size (diameter, pitch, and aspect ratio) and wafer thickness. Wafer thickness can vary depending on the wafer-to-wafer or die-to-wafer stacking schemes. TSVs with diameter between 2 and 10 $\mu$m requires one die to be thinned between 25 and 70 $\mu$m, which makes wafer and die handling very challenging. Moreover, wafer thinning can pose thermal dissipation problems and introducing additional complexity to 3-D thermal management problem. Wafer thinning is modeled by the electrical and thermal resistance and capacitance of each tier derived based on the thickness and area of the die, as in [31] and [32].

To examine the impact of wafer thinning on the power and thermal integrity of 3-D PDNs, we study two types of wafer thickness, 50 and 775 $\mu$m. We use the three-tier network where identical power density distributions were applied to each tier. Initially, we study 3-D PDNs with equal thickness wafer of 775 $\mu$m. Then, we study wafer thinning up to 50 $\mu$m when applied to top two tiers. We investigate temperature distribution on each tier and compare the impact of wafer thinning on 3-D PDN area.

Table V summarizes our findings. Columns 2 and 3 show the voltage drop and thermal objective for the two types of wafer

thickness, respectively. We observe that wafer thickness has a considerable impact on thermal objective with $O_T = 0.54$ for thick wafers of 775 $\mu$m to $O_T = 0.62$ for thinned wafers of 50 $\mu$m. This shows that thinned wafers experience higher temperature levels. Whereas, there is little impact on voltage drop objective due to wafer thinning.

We investigate the impact of wafer thinning on 3-D PDN area by applying greedy resizing (uniformly upsize all metal tracks by $w_{\max}$), and our proposed electrothermal optimization algorithms, ALG1 and ALG2 where results are shown in columns 4–6 of Table V. The area obtained from ALG1 and ALG2 are shown in proportional (ratio) to the area obtained from greedy resizing. For wafer of 775-$\mu$m thickness, we obtain 0.25 for ALG1 or 75% of area savings from GR and 0.22 for ALG2 or 78% of area savings from GR. As ALG2 considers both metal track and TSV sizing provides more flexibility on saving 3-D PDN area. In the case of thinned wafers, we obtain 0.31 for ALG1 (or 69% of area savings from GR) and 0.28 for ALG2 (or 72% of area savings from GR). These results demonstrate that less 3-D PDN area savings are obtained for thinned wafer or larger power grids are required to satisfy power and thermal integrity.

Overall, we observe that wafer thinning increases temperature levels and poses additional thermal challenge to 3-D ICs. This is because the thermal capacity of the die to absorb heat is greatly diminished with thinning, which as consequence reduces the capability of the die to spread some of the heat. Therefore, it impacts the 3-D PDNs and their area for satisfying power and thermal constraints.

## VII. Conclusion

In this paper, we presented a voltage drop and thermal constraints driven 3-D power network delivery optimization. We have shown that already optimized 2-D power networks cannot be directly applicable as 3-D PDNs and moreover, circuits might require upsizing to meet timing constraints due to nonuniform voltage drop and temperature in a multitier system. We proposed a novel tier-based resizing technique that individually optimizes each tier while enforcing multi-tier constraints. Compact and efficient electrical and thermal models were applied for performing electrothermal analysis. Furthermore, we presented two optimization algorithms for resizing 3-D PDNs standalone or together with TSVs for satisfying power and thermal constraints.

## References


[1] J. U. Knickerbocker, P. Andry, B. Dang, R. Horton, M. Interrante, C. Patel, *et al.*, "Three-dimensional silicon integration," *IBM J. Res. Develop.*, vol. 52, no. 6, pp. 553–569, Nov. 2008.

[2] G. Huang, M. Bakir, A. Naeemi, H. Chen, and J. D. Meindl, "Power delivery for 3D chips stacks: Physical modeling and design implications," in *Proc. IEEE Electr. Perform. Electron. Packag.*, Oct. 2007, pp. 205–208.

[3] J. S. Pak, J. Kim, J. Cho, K. Kim, T. Song, S. Ahn, *et al.*, "PDN impedance modeling and analysis of 3D TSV IC by using proposed P/G TSV array model based on separated P/G TSV and chip-PDN models," *IEEE Trans. Compon., Packag. Manuf. Technol.*, vol. 1, no. 2, pp. 208–219, Feb. 2011.

[4] M. B. Healy and S. K. Lim, "A novel TSV topology for many-tier 3D power-delivery networks," in *Proc. IEEE DATE Conf. Exhibit.*, Mar. 2011, pp. 1–4.

[5] N. H. Khan, S. M. Alam, and S. Hassoun, "Power delivery design for 3-D ICs using different through-silicon via (TSV) technologies," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 4, pp. 647–658, Apr. 2011.

[6] V. F. Pavlidis and G. De Micheli, "Power distribution paths in 3-D ICs," in *Proc. ACM GLSVLSI*, May 2009, pp. 263–268.

[7] P. Falkenstern, Y. Xie, Y.-W. Chang, and Y. Wang, "Three-dimensional integrated circuits (3D IC) floorplan and power/ground network co-synthesis," in *Proc. ASP-DAC*, Jan. 2010, pp. 169–174.

[8] A. Todri, S. Kundu, P. Girard, A. Bosio, L. Dilillo, and A. Virazel, "A study of tapered 3-D TSVs for power and thermal integrity," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 21, no. 2, pp. 306–319, 2013.

[9] P. Zhou, K. Sridharan, and S. Sapatnekar, "Congestion-aware power grid optimization for 3D circuits using MIM and CMOS decoupling capacitors," in *Proc. ASP-DAC*, Jan. 2009, pp. 179–184.

[10] M. B. Kleiner, S. A. Kuhn, P. Ramm, and W. Weber, "Thermal analysis of vertically integrated circuits," in *Proc. IEDM*, Dec. 1995, pp. 487–490.

[11] T.-Y. Chiang, S. J. Souri, C. O. Chui, and K. C. Saraswat, "Thermal analysis of heterogeneous 3D ICs with various integration scenarios," in *Proc. IEDM*, Dec. 2001, pp. 31.2.1–31.2.4.

[12] A. Jain, R. E. Jones, R. Chatterjee, and S. Pozder, "Analytical and numerical modeling of the thermal performance of three-dimensional integrated circuits," *IEEE Trans. Compon. Packag. Technol.*, vol. 33, no. 1, pp. 56–63, Mar. 2010.

[13] H. Oprins, V. Cherman, M. Stucchi, B. Vandevelde, G. Van der Plas, P. Marchal, *et al.*, "Steady state and transient thermal analysis of hot spots in 3D stacked ICs using dedicated test chips," in *Proc. IEEE 27th Annu. SEMI-THERM*, Mar. 2011, pp. 131–137.

[14] H. Yu, J. Ho, and H. Lei, "Simultaneous power and thermal integrity driven via stapling in 3D ICs," in *Proc. IEEE/ACM ICCAD*, Nov. 2006, pp. 802–808.

[15] Y.-J. Lee and S. K. Lim, "Co-optimization and analysis of signal, power and thermal interconnects in 3-D ICs," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 30, no. 11, pp. 1635–1648, Nov. 2011.

[16] J. Xie, D. Chung, M. Swaminathan, M. Mcallister, A. Deutsch, J. Lijun, *et al.*, "Electrical-thermal co-analysis for power delivery networks in 3D system integration," in *Proc. IEEE Int. Conf. 3DIC*, Sep. 2009, pp. 1–4.

[17] Z. Luo, J. Fan, and S. Tan, "Localized statistical 3D thermal analysis considering electro-thermal coupling," in *Proc. IEEE ISCAS*, May 2009, pp. 1289–1292.

[18] T. Mitsuhashi and E. S. Kuh, "Power and ground network topology optimization for cell based VLSIs," in *Proc. ACM/IEEE DAC*, Jun. 1992, pp. 524–529.

[19] J. Singh and S. S. Sapatnekar, "Congestion-aware topology optimization of structured power/ground networks," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 5, pp. 683–695, May 2005.

[20] S. L. Wright, R. Polastre, H. Gan, L. P. Buchwalter, R. Horton, P. S. Andry, *et al.*, "Characterization of micro-bump C4 interconnects for Si-carrier SOP application," in *Proc. 56th Electron. Compon. Technol. Conf.*, 2006, pp. 1–8.

[21] C. Bermond, L. Cadix, A. Farcy, T. Lacrevaz, P. Leduc, and B. Flechet, "High frequency characterization and modeling of high density TSV in 3D ICs," in *Proc. IEEE Workshop SPI*, May 2009, pp. 1–4.

[22] G. Katti, M. Stucchi, K. De Meyer, and W. Dehaene, "Electrical modeling and characterization of TSVs for three-dimensional ICs," *IEEE Trans. Electron Devices*, vol. 57, no. 1, pp. 256–262, Jan. 2010.

[23] Intel Corp. (2008). *45 nm Intel Xeon Processor*, Santa Clara, CA, USA [Online]. Available: http://www.intel.com

[24] ITRS. (2011). *ITRS Report*, Geneva, Switzerland [Online]. Available: http://www.itrs.net

[25] R. L. Boylestad, *Introduction to Circuit Analysis*, Englewood Cliffs, NJ, USA: Prentice-Hall, 2000.

[26] H. Chen and D. Ling, "Power supply noise analysis methodology for deep-submicron VLSI chip design," in *Proc. ACM/IEEE DAC*, Jun. 1997, pp. 638–643.

[27] T.-Y. Wang and C. C.-P. Chen, "SPICE-compatible thermal simulation with lumped circuit modeling for thermal reliability analysis based on modeling order reduction," in *Proc. ISQED*, 2004, pp. 357–362.

[28] Y. Zhong and M. D. F. Wong, "Thermal-area IR drop analysis in large power grid," in *Proc. ISQED*, 2008, pp. 194–199.

[29] T.-Y. Chiang, K. Banerjee, and K. C. Saraswat, "Effect of via separation and low-k dielectric materials on the thermal characteristics of Cu interconnect," in *Proc. IEDM*, Dec. 2000, pp. 261–264.

[30] R. K. Endo, Y. Fujihara, and M. Susa, "Calculation of the density and heat capacity of silicon by molecular dynamics simulation," *Trans. High Temperature-High Pressures*, vol. 35, no. 5, pp. 505–511, 2003.
[31] M. Pedram and S. Nazarian, "Thermal modeling, analysis and management in VLSI circuits: Principles and methods," *Proc. IEEE*, vol. 94, no. 8, pp. 1487–1501, Aug. 2006.
[32] K. Skadron, R. Abdelzaher, and M. R. Stan, "Control-theoritic techniques and thermal-RC modeling for accurate and localized dynamic thermal management," in *Proc. Int. Symp. High-Perform. Comput. Archit.*, Feb. 2002, pp. 17–28.
[33] G. Katti, M. Stucchi, D. Velenis, B. Soree, K. De Meyer, and W. Dehaene, "Temperature-dependent modeling and characterization of through-silicon via capacitance," *IEEE Trans. Electron Device Lett.*, vol. 32, no. 4, pp. 563–565, Apr. 2011.
[34] W.-S. Zhao, X.-P. Wang, and W.-Y. Yin, "Electrothermal effects in high density TSV arrays," *Progr. Electromagn. Res.*, vol. 115, pp. 223–242, Mar. 2011.
[35] G. B. Kromann, "Thermal modeling and experimental characterization of the C4/surface-mount-array interconnect technologies," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 18, no. 1, pp. 87–93, Mar. 1995.
[36] A. Jain, R. E. Jones, R. Chatterjee, S. Pozder, and H. Zhihong, "Thermal modeling and design of 3D integrated circuits," in *Proc. 11th Intersoc. Conf. Thermal Thermomech. Phenomena Electron. Syst.*, May 2008, pp. 1139–1145.
[37] J. N. Kozhaya, S. R. Nassif, and F. N. Najm, "A multigrid like technique for power grid analysis," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 21, no. 10, pp. 1148–1160, Oct. 2002.
[38] K. Wang and M. Marek-Sadowska, "On-chip power supply network optimization using multigrid-based technique," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 3, pp. 407–417, Mar. 2005.

**Aida Todri-Sanial** (M'03) received the B.S. degree in electrical engineering from Bradley University, Peoria, IL, USA, in 2001, the M.S. degree in electrical engineering from Long Beach State University, Long Beach, CA, USA, in 2003, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, CA, USA, in 2009.

She is currently a Researcher with the French National Center of Scientific Research, Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier, Montpellier, France. She was a Research and Development Engineer with the Fermi National Accelerator Laboratory, Batavia, IL, USA. She held visiting positions with Mentor Graphics, Wilsonville, OR, USA, Cadence Design Systems, San Jose, CA, USA, STMicroelectronics, Geneva, Switzerland, and IBM's T. J. Watson Research Center, Yorktown Heights, NY, USA.

Dr. Todri was a recipient of the John Bardeen Fellowship in Engineering in 2009.

**Sandip Kundu** (M'86–SM'94–F'07) received the Ph.D. degree in electrical and computer engineering from the University of Iowa, Des Moines, IA, USA, in 1988.

He was a Professor of electrical and computer engineering with the University of Massachusetts Amherst, Amherst, MA, USA, in 2005. He was a Research Staff Member with IBM Corporation, Armonk, NY, USA, and Principal Engineer with Intel Corporation, Santa Barbara, CA, USA, from 1988 to 2005. He has published more than 200 papers on VLSI design, CAD, and architecture, holds 12 patents, and has given more than a dozen tutorials at conferences.

Dr. Kundu served as an Associate Editor of the IEEE TRANSACTIONS ON COMPUTERS AND THE IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS. He is a Distinguished Visitor of the IEEE Computer Society.

**Patrick Girard** (M'92–SM'09) received the M.S. degree in electrical engineering and the Ph.D. degree in microelectronics from the University of Montpellier, Montpellier, France, in 1988 and 1992, respectively.

He is currently the Research Director with the French National Center for Scientific Research and a Chair of the Microelectronics Department, Laboratory of Informatics, Robotics and Microelectronics, Montpellier.

Dr. Girard has served on numerous conference committees and is the Founder and Editor-in-Chief of the *ASP Journal of Low Power Electronics*. He is an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS and the *Journal of Electronic Testing – Theory and Applications* (JETTA - Springer).

**Alberto Bosio** (M'03) received the Ph.D. degree in computer engineering from Politecnico di Torino, Torino, Italy, in 2006.

He is currently an Associate Professor with the Laboratory of Informatics, Robotics and Microelectronics of Montpellier, University of Montpellier, Montpellier, France. His current research interests include computer-aided design, logic diagnosis, functional verification, and dependability.

**Luigi Dilillo** (M'03) received the Ph.D. degree in microelectronics from the University of Montpellier, Montpellier, France, in 2005.

He is currently a CNRS Researcher with the Laboratory of Informatics, Robotics and Microelectronics of Montpellier, University of Montpellier II, Montpellier. He has published articles in publications spanning diverse disciplines, including memory testing, power aware testing, and radiation effects on electronic devices.

**Arnaud Virazel** (M'03) received the M.S. degree in electrical engineering and the Ph.D. degree in microelectronics from the University of Montpellier, Montpellier, France, in 1997 and 2001, respectively.

He is currently an Assistant Professor with the University of Montpellier II, Montpellier, and with the Microelectronics Department, Laboratory of Informatics, Robotics, and Microelectronics of Montpellier, Montpellier.