Large-Scale Flip-Chip Power Grid Reduction with Geometric Templates

Zhuo Feng,
Department of Electrical and Computer Engineering
Michigan Technological University
Email: zhuofeng@mtu.edu

***Abstract*—Realizable power grid reduction becomes key to efficient design and verification of nowadays large-scale power delivery networks (PDNs). Existing state-of-the-art realizable reduction techniques for interconnect circuits, such as TICER algorithm, can not be well suited for effective power grid reductions, since reducing the mesh-structured power grids by TICER's nodal elimination scheme may introduce excessive number of new edges in the reduced grids that can be even harder to solve than the original grid due to the drastically increased sparse matrix density. In this work, we present a novel geometric template based reduction technique for reducing large-scale flip-chip power grids. Our method first creates geometric template according to the original power grid topology and then performs novel iterative grid corrections to improve the accuracy by matching the electrical behaviors of the reduced template grid with the original grid. Our experimental results show that the proposed reduction method can reduce industrial power grid designs by up to 95% with very satisfactory solution quality.**

I. Introduction

As the relentless technology scaling reaches into the sub-16nm regime, integrated circuit (IC) designers are facing phenomenal growth of design complexity in power delivery network (PDN) designs. Analysis, verifications and optimizations of such complex systems become increasingly important and critical for designing reliable and efficient VLSI systems. Realizable power grid reduction is essential for efficient large-scale PDN optimization and verifications, which is expected to reduce the original large scale systems into much smaller ones that can be further combined with other subsystems (nonlinear digital and analog circuit blocks) for faster full-chip simulations.

Existing realizable parasitics reduction methods fall into the following categories: high order $Y-\Delta$ transformations [1], nodal eliminations [2], [3], and algebraic multigrid (AMG) reduction [4]. Unfortunately, most of prior realizable reduction techniques are focused on reducing interconnect circuits with tree-like structures seen in typical digital circuit analysis problems [1], [3], which can not be readily applied to mesh-structured on-chip power grid reductions.

In this work, we propose a novel geometric template based parasitics reduction approach for reducing mesh-structured flip-chip power grid circuits (nodes and elements). Our method starts from creating geometric templates based on the original power grid topology, and subsequently stamps the circuit components onto the templates to form the reduced power grid circuits. This geometric reduction process can be efficiently performed based on the chip layout information, such as the interconnect pitches, geometries and locations. We show that such a novel geometric template grid reduction approach can well approximate the electrical properties of the original grid, and is similar to geometric multigrid (GMD) reduction method [5], [6].

To further mitigate the errors introduced during the geometric template grid construction process, an efficient electrical-behavior-preserved template grid correction scheme is proposed to significantly improve the reduced grid accuracy: through the proposed iterative block wire sizing procedures for the template grid, a significantly improved reduced template grid can be obtained by matching the block power dissipations and voltage drops in the template grid with the original grid. As a result, this novel template grid correction method ensures that the proposed geometric template based power grid reduction method can achieve very desired high reduction ratio as well as accuracy.

978-3-9815370-0-0/DATE13/©2013 EDAA

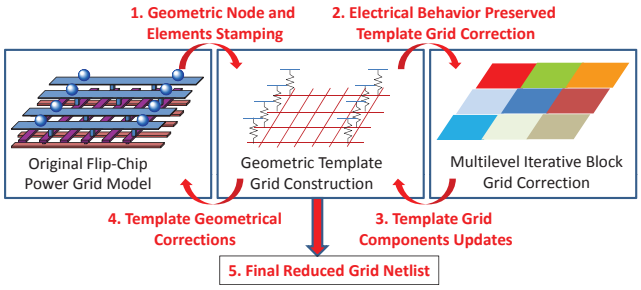

Fig. 1. Power grid reduction using geometric templates.

II. Overview of the Proposed Grid Reduction Approach

This work is largely influenced by recent computational photography research [7], [8] where many problems in low-level computer vision have been formulated using a variational approach that usually involves defining a continuous two-dimensional optimization problem, adding appropriate regularization terms, and then discretizing the problem on a regular grid.

As illustrated in Fig. 1, the basic idea of our approach is to first create a power grid template and subsequently identify the optimal circuit components (resistors, current sources, etc) on the template grid by matching key electrical behaviors such as the power dissipations and voltage drops with the original power grid circuits. Unlike previous TICER [3] or AMG based reduction methods [4] that are not aware of the final grid topologies before reductions, our new approach constructs and improves the template grid from the initial power grid template that can be easily created from the original power grid layout.

Throughout this work, we consider the power distribution network benchmarks published in [9] that are similar to the ones used in [4]. These industrial on-chip power grid designs are modeled as resistive meshes including via resistors connecting different metal layers, and VDD (GND) pads modeling the C4 bumps which are widely used in flip-chip power grid designs. Current loadings are modeled as distributed time-varying current sources that are connecting power and ground grids. Although the proposed power grid reduction framework can be potentially extended to deal with energy-storage components, we will only focus on the reduction of resistive power grids in this work.

The proposed geometric template based power grid reduction approach is very efficient and effective for reducing mesh-structured

flip-chip power grid circuits (nodes and elements). Our method includes the following key steps:

1) Construct the initial geometric template grid based on the layout information of the original power grid;
2) Stamp individual circuit components, such as circuit nodes, resistors and current excitation sources, onto the template grid;
3) Aggregate the nodes, wires and current sources of the original grid onto the template grid;
4) Perform iterative block grid corrections to match electrical behaviors of the reduced grid, such as the voltage drops and dissipated powers, with the original grid;
5) If the errors are not acceptable, correct the geometric template grid by introducing more nodes into the reduced grid template (increase the template grid dimensions), and go through the above 1-5 steps again;
6) Finalize the reduction and output the reduced power grid netlist.

The proposed geometric template based reduction approach has been more clearly illustrated in Fig. 1. The first three steps of the above geometric grid reduction process can be efficiently performed based on the chip layout information such as the power grid geometries (full grid or partial grid), pitches between wires, interconnect length, and current source amplitudes as well as their locations. It can be shown that the initial geometric template grid can well approximate the electrical properties of the original grid, which is similar to geometric multigrid (GMD) based reduction method. To further mitigate the errors introduced during the above geometric template grid construction process, a novel grid correction scheme is proposed in Step 4 to effectively reduce the template grid errors. We show that through the proposed iterative block wire sizing procedures for the geometric template grid, significant solution improvements can be achieved by matching the block voltage drops and block power dissipations in the template grid with the original grid. As a result, this novel grid correction scheme ensures that the geometric template based power grid reduction method can always achieve very desired high reduction ratio as well as accuracy.

III. Power Grid Reduction Using Geometric Templates

A. Geometric Template Construction

The geometric template for the reduced grid can be created based on the layout information as well as the desired reduction ratio and approximation accuracy. Including more nodes into the template grid allows for better approximation of the original power grid but results in lower grid reduction ratio and efficiency. As shown in Fig. 1, given any initial geometric template grid, if the electrical behaviors of the reduced grid, such as the voltage profiles as well as the power dissipations, are not matching well with the original grid, the geometric template should be improved through correcting the template grid by either adding more nodes through adjusting the template grid pitches, or by properly sizing the wires on the template grid to change the electric conductances. For the industrial power grid designs provided in [9], we find that *2D regular geometric template* grids work properly, as shown in Fig. 2, while for other grid designs the shapes of geometric template grids can be determined according to the full grid layout geometries, topologies and current loading distributions.

The idea of our approach is to stamp the original power grid elements onto a regularized template grid, which can well preserve the electrical property of the original grid on the template grid. It seems that developing a general way for creating geometric templates for power grid designs can be very challenging. Fortunately, industrial power grid designs [9] typically exhibit globally uniform grid structures, while maintaining irregular patterns for local grids. Consequently, we propose an efficient yet effective technique to generate the 2D geometric template grid using typical wire pitch

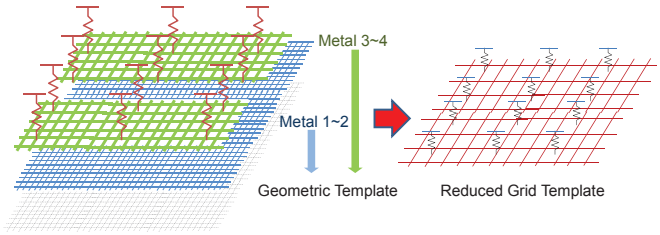

Fig. 2. Grid reduction using geometric template grids.

values of the bottom metal layer grid and subsequently stamp the original multi-layer irregular 3D power grid elements (resistors, current and voltage sources) onto the template grid, which is very similar to the grid regularization process used in [6], [10], [11]. In this manner, the electrical properties of the original power supply network can be well preserved on the geometric template grid.

It should be emphasized that the purpose of the above element stamping procedure is to get a relatively good template grid, though the template grid does not need to fully maintain the original power grid structure. For instance, assuming that the original 3D power grid design is "reasonably good" without wrongly-connected layers, since the power grid via resistances are typically much smaller than the grid wire resistances, we can safely collapse the original 3D grid to the 2D grid template and remove the via resistors connecting different layers. As shown in [11], typical industrial power grid nodes that have been connected through via resistors have negligible voltage differences.

Consequently, we can use the following 2D regular geometric template grid construction procedure for flip-chip power grid reductions:

1) *Step 1*. By neglecting via resistances, all the metal layers in the original power delivery network are overlapped on the lateral plane, forming a collapsed 2D irregular grid.
2) *Step 2*. By examining the original grid pitches in lateral directions, proper pitch values are selected based on user-defined grid reduction ratio, such that locations of the 2D template grid nodes can be easily determined.
3) *Step 3*. Circuit elements such as resistors, VDD/GND pads, and current sources are stamped onto the regularized template grid. Elements that span multiple template grid nodes are decomposed into several smaller pieces during the stamping, while current and voltage sources are stamped according to their geometrical locations.

It is observed in extensive experiments on industrial power grid designs [9] that using the above simple 2D regular geometric template grid construction method, over 90% reduction ratio can be achieved for flip-chip power grid reductions that involve hundreds of C4 bumps (ports) without introducing significant errors. For instance, the average voltage errors of the reduced grids (after performing the following proposed iterative grid correction steps) can be less than $2mV$ when using 90% to 97% reduction rates.

B. Template Grid Correction Scheme

It is usually difficult to predict the errors introduced in the geometric template grid construction process shown in Fig. 2, since the large number of nodes and circuit components of the original power grid are aggregated (stamped) onto much fewer nodes and circuit elements of the reduced geometric template grid. In order to more accurately preserve the electrical behaviors of the original grids on the reduced grid, we propose a novel iterative template grid correction scheme, as shown in Fig. 3.

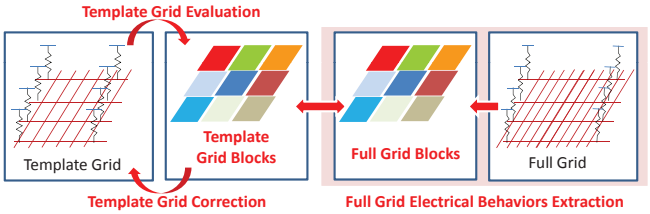

Fig. 3. Iterative geometric template grid correction scheme.

1) Preservation of Electrical Behaviors on Reduced Geometric Template Grid: Consider that the original grid and reduced grid template can be split into multiple blocks based on locality properties [12] of flip-chip power grid designs [1]. As a result, each individual block of the original grid will see a corresponding block in the geometric template grid. It is expected that if each block of the reduced template grid exhibits the same electrical behaviors, such as the power dissipations and voltage drops, of the corresponding original grid block, the reduced template grid will be a good approximation of the original grid, since the powers dissipated in wires are directly related to current amplitudes while the node voltage drops are related to the node voltage. On the other hand, if the electrical behaviors of two grids do not match well, the reduced grid template may not be a good approximation of the original system, and consequently there is a need to correct/improve the reduced template grid by sizing its wires or including more template grid nodes, as depicted in Fig. 1 and Fig. 3.

In the following sections, we assume the template grid can be further improved to reach a desired accuracy level without introducing new nodes, though extra nodes can be added if the accuracy level does not meet the requirements, as illustrated in Fig. 1.

Consider the reduced geometric template grid and the original power grid split into the same number of blocks. We propose a novel optimization framework to size the template grid wires such that for each of its blocks, the power dissipation and voltage drop are similar as the ones of the corresponding fine grid block. More specifically, assuming the template grid is a 2D grid with horizontal ($X-$ direction) and vertical wires ($Y-$ direction), we can match the block power dissipations of the horizontal and vertical wires, respectively. Simultaneously, we can match the average node voltage drop of every block with the fine grid block. Once these important electrical behaviors are well preserved on the reduced grid, the overall power grid reduction accuracy can be significantly improved.

2) Optimization Approach to Geometric Template Grid Correction: To gain better insight for the proposed geometric template grid correction algorithm, we first neglect the coupling effects between different grid blocks in the following analysis. Subsequently, the power dissipation of a power grid block can be easily computed by

$$P_{Joule} = \sum G_{ij}(x_i - x_j)^2 = x^T G x, \tag{1}$$

where $x = G^{-1}b$, which is followed by

$$P_{Joule} = b^T G^{-T} G G^{-1} b = b^T G^{-1} b. \tag{2}$$

So once we size up all the wires in the block by a *block scaling factor* τ, the block voltage drop after the wire sizing become $x' = G^{-1}b/\tau$, while the updated power dissipation becomes

$$P'_{Joule} = b^T G^{-1} b/\tau = P_{Joule}/\tau. \tag{3}$$

[1]"Split" here does not mean the separation of the large grid into isolated smaller ones. Instead, it indicates that if a node/wire belongs to one specific region, it falls into that block.

The above derivations indicate that we can use the block power dissipation (voltage drop) ratios of the reduced geometric grid template and the original power grid to determine the block scaling factors τ. Based on the above idea, once we split the block power dissipations into horizontal and vertical wire power dissipations, we can use horizontal and vertical block scaling factors τ^H and τ^V to size up or down the block template grid wires.

Considering the coupling effects between blocks due to the inter-block connections, we formulate the geometric template grid correction scheme into the following nonlinear optimization problem that will try to best preserve the block power dissipations as well as block average voltage drops (average voltage of the nodes in that block) on the geometric template grid.

$$\begin{aligned} &minimize: \phi(\vec{\tau}) = \sum_{i=1}^{m} \left({x_{Ri}}^T G_{Ri}^V x_{Ri} \tau_i^V - {x_i}^T G_i^V x_i \right)^2 + \\ &\alpha \sum_{i=1}^{m} \left({x_{Ri}}^T G_{Ri}^H x_{Ri} \tau_i^H - {x_i}^T G_i^H x_i \right)^2 + \beta \sum_{i=1}^{m} \left(\frac{|x_{Ri}|_1}{n_{Ri}} - \frac{|x_i|_1}{n_i} \right)^2 \end{aligned} \tag{4}$$

subject to the constraints:

$$\left[G_R^H(\vec{\tau}) + G_R^V(\vec{\tau}) \right] \begin{bmatrix} x_{R1} \\ \vdots \\ x_{Rm} \end{bmatrix} = b_R, \left[G^H + G^V \right] \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = b, \tag{5}$$

where G_{Ri}^H (G_i^H) is the conductance matrix of the reduced template (full) grid block i for horizontal wires, G_{Ri}^V (G_i^V) is the conductance matrix for vertical wires, m is the number of blocks, $\vec{\tau} = \left[\tau_1^V, \tau_1^H, ..., \tau_m^V, \tau_m^H\right]^T$ is the vector including block scaling factors to be computed through the nonlinear optimization problem (4) , α and β are weighting coefficients, and n_{Ri} (n_i) is the number of nodes in the template (fine) grid block. It should be noted that the reduced grid conductance matrix $G_R(\vec{\tau}) = G_R^H(\vec{\tau}) + G_R^V(\vec{\tau})$ is parameterized, and the block grid solution x_{Ri} will be a linear function of the vector $\vec{\tau}$ that includes all scaling factors.

3) A Practical Template Grid Correction Flow: The above optimization formula enables to optimize template grid wires such that it can preserve key electrical properties of the original power grid. However, there is no efficient method for solving the above nonlinear optimization problems. Fortunately, for flip-chip power grids, locality effects [12] enable us to relax the above optimization problem, resulting in the following iterative template grid correction scheme that can achieve the similar goal of (4):

1) Partition the template and fine grids into several blocks such that each block includes a few C4 bumps (VDD/GND pads).
2) Compute the DC solution of the original power grid once, and compute each block's power dissipation and average voltage drop.
3) Compute the DC solution of the geometric template grid, and compute each block's power dissipation and average voltage drop.
4) Size up/down the block horizontal and vertical wires of the template grid using the block scaling factors computed by

$$\tau_i^H = \left(\frac{{x_{Ri}}^T G_{Ri}^H x_{Ri}}{{x_i}^T G_i^H x_i} \right) \omega + \left(\frac{|x_{Ri}|_1 N_i}{|x_i|_1 N_{Ri}} \right) (1-\omega), \tag{6}$$

$$\tau_i^V = \left(\frac{{x_{Ri}}^T G_{Ri}^V x_{Ri}}{{x_i}^T G_i^V x_i} \right) \omega + \left(\frac{|x_{Ri}|_1 N_i}{|x_i|_1 N_{Ri}} \right) (1-\omega). \tag{7}$$

5) Repeat steps 3 and 4 until convergence. Output the new geometric template grid (reduced power grid) circuit.

In the above iterative scheme, ω can be set to 0.5 to balance the scaling factors of block power dissipations and average node voltage. We try to iteratively match the block power dissipations as well as voltage drops since in many cases, the block power dissipation or

TABLE I

RESULTS OF GEOMETRIC-TEMPLATE POWER GRID REDUCTION METHOD. "# OF NODES" DENOTES THE NUMBER OF NODES (RESISTORS) AFTER POWER GRID REDUCTION (WITH REDUCTION RATIO), "# OF BLOCKS" AND "# OF ITER." DENOTE THE NUMBERS OF BLOCK PARTITIONS AND ITERATIONS FOR THE GEOMETRIC TEMPLATE GRID CORRECTION STEP, "ERR." DENOTES THE AVERAGE ERRORS AFTER GRID REDUCTION (ERROR IMPROVEMENTS COMPARED TO THE REDUCTION METHOD W/O CORRECTION), AND "RUNTIME" DENOTES THE TOTAL POWER GRID REDUCTION TIME.

CKT	Power Grid Reduction Set A					Power Grid Reduction Set B				
	# of Nodes (Redu.)	# of Blocks	# of Iter.	Err. (Imp.)	Time	# of Nodes (Redu.)	# of Blocks	# of Iter.	Err. (Imp.)	Time
ibmpg3	$78,400$ (82%)	121	10	$2.0\ mV$ $(5X)$	$10.5\ s$	$38,400$ (91%)	121	15	$2.5\ mV$ $(4.5X)$	$9.7\ s$
ibmpg4	$38,400$ (92%)	121	12	$0.1\ mV$ $(2.5X)$	$8.5\ s$	$18,496$ (94%)	121	14	$0.1\ mV$ $(3.0X)$	$8.5\ s$
ibmpg5	$46,656$ (92%)	121	10	$0.8\ mV$ $(4.0X)$	$6.4\ s$	$23,104$ (96%)	121	15	$1.0\ mV$ $(5.0X)$	$6.5\ s$
ibmpg6	$78,400$ (91%)	169	10	$1.5\ mV$ $(5.6X)$	$12.0\ s$	$48,384$ (94%)	169	14	$2.2\ mV$ $(4.2X)$	$13.5\ s$
ibmpg7	$69,696$ (90%)	169	10	$1.2\ mV$ $(4.0X)$	$14.3\ s$	$23,104$ (97%)	169	13	$1.6\ mV$ $(3.8X)$	$13.4\ s$
ibmpg8	$57,600$ (92%)	169	12	$1.4\ mV$ $(3.4X)$	$13.9\ s$	$28,244$ (96%)	169	14	$1.9\ mV$ $(3.4X)$	$13.7\ s$

TABLE II

DETAILS OF SIX LARGEST IBM BENCHMARKS IBMPG3 TO IBMPG8 [9]. NUMBER OF NODES, RESISTORS, METAL LAYERS AS WELL AS C4 BUMPS (PORTS) HAVE BEEN SHOWN.

CKT	# of Nodes	# of Res.	# of Lay.	# of C4s
ibmpg3	440,616	724,184	5	461
ibmpg4	478,095	779,946	6	650
ibmpg5	581,473	871,182	3	177
ibmpg6	862,418	1,283,371	3	249
ibmpg7	742,330	1,194,370	6	461
ibmpg8	742,330	1,194,506	6	461

voltage drop can be too small to be used for finding the scaling factors. We want to emphasize that a user supplied damping factor should be used when sizing the template wires. When $|\tau_i^H - 1| > bound$ ($|\tau_i^V - 1| > bound$), we simply use the block scaling factor $1 + bound$ if $\tau_i^H > 1$ ($\tau_i^V > 1$) or $1 - bound$ if $\tau_i^H < 1$ ($\tau_i^V < 1$) to size the block wires. This factor damping scheme can greatly improve the convergence of the proposed template grid correction iterations (we use $bound = 0.1$ in this work). The convergence is reached when the relative differences in (4) of the electrical behaviors of the reduced grid and full grid are small enough (e.g. less than 0.5%). We have implemented and tested the above iterative geometric template grid correction scheme for all industrial designs in [9] and observed that solution errors can be quickly reduced by a factor of $3X$ after only ten iterations.

IV. EXPERIMENTAL RESULTS

In this section, we validate the proposed geometric template reduction methods considering several large IBM power grid benchmarks [9]. More details about the benchmarks are shown in Table II. The proposed template-based power grid reduction method has been implemented in C++ and CUDA [13]. We use two-level iterative template grid correction method for all test cases. When comparing the solution of the reduced template grid with the original grid solution, we geometrically interpolate the reduced template grid results to the original power grid node solution, and subsequently measure the solution differences. The power grid analysis for the full grids and the reduced grids are performed on CPU-GPU platforms based on the iterative algorithms proposed in [6], [10]. The hardware platform is a Linux PC with Intel Core 2 Quad CPU running at 2.66 GHz clock frequency with an NVIDIA GTX 285 GPU. All runtime results are measured in seconds.

We show in Table I comprehensive results of the proposed power grid reduction method for all the six largest IBM power grid benchmarks. Two grid reduction ratios are applied when reducing the power grid problems: reduction set A includes results using 82% to 92% reduction ratios, and reduction set B includes results using 91% to 97% reduction ratios. As observed, a larger power grid reduction ratio will require more template grid correction iterations due to the increased number of optimization variables, and result in slightly greater solution errors. It is also observed that using the proposed template grid construction and correction scheme, we can reduce a half-million power grid design in less than ten seconds on our CPU-GPU heterogeneous computing platforms. It is not surprising to notice that after reducing 90% of the grid nodes and wires, the solution can still be very accurate on the reduced power grid. From the DC solution errors, we found that the naive geometric grid reduction without using iterative template grid corrections can cause up to $5X$ greater solution errors.

V. CONCLUSIONS

In this work, we propose a novel geometric template based power grid reduction approach for reducing mesh-structured flip-chip power grid circuits (nodes and elements). It can be shown that such a geometric template grid reduction approach can well approximate the electrical properties of the original grid, and is similar to geometric multigrid (GMD) reduction method. To further eliminate the errors introduced during the naive geometric template grid reduction process, a novel iterative electrical-behavior-preserved grid correction scheme is proposed to significantly mitigate the reduced grid errors. Our experimental results show that the proposed reduction method can reduce industrial power grid designs by up to 95% without loss of accuracy.

REFERENCES

[1] Z. Qin and C. Cheng. Realizable parasitic reduction using generalized Y-Δ transformation. In *Proc. of IEEE/ACM DAC*, pages 220–225, 2003.

[2] C. S. Amin, M. Chowdhury, and Y. Ismail. Realizable RLCK circuit crunching. In *Proc. of IEEE/ACM DAC*, pages 226–231, 2003.

[3] B. N. Sheehan. Realizable Reduction of RC Networks. *IEEE Trans. on Computer-Aided Design*, 26(8):1393–1407, 2007.

[4] H. Su, E. Acar, and S. R. Nassif. Power grid reduction based on algebraic multigrid principles. In *Proc. IEEE/ACM DAC*, pages 109–112, 2003.

[5] J. N. Kozhaya, S. R. Nassif, and F. N. Najm. A multigrid-like technique for power grid analysis. *IEEE Trans. on Computer-Aided Design*, 21(10):1148–1160, 2002.

[6] Z. Feng and P. Li. Multigrid on GPU: tackling power grid analysis on parallel SIMT platforms. In *Proc. IEEE/ACM ICCAD*, pages 647–654, 2008.

[7] R. Szeliski. Locally adapted hierarchical basis preconditioning. In *Proc. of ACM SIGGRAPH*, pages 1135–1143, 2006.

[8] D. Krishnan and R. Szeliski. Multigrid and multilevel preconditioners for computational photography. In *Proc. of SIGGRAPH Asia*, to appear in 2011.

[9] S. R. Nassif. *IBM power grid benchmarks*. [Online]. Available: http://dropzone.tamu.edu/ pli/PGBench/, 2008.

[10] Z. Feng and Z. Zeng. Parallel multigrid preconditioning on graphics processing units (GPUs) for robust power grid analysis. In *Proc. IEEE/ACM DAC*, pages 661–666, 2010.

[11] Z. Feng, Z. Zeng, and P. Li. Parallel on-chip power distribution network analysis on multi-core-multi-GPU platforms. *IEEE Trans. on Very Large Scale Integration (VLSI) Systems, to appear*, pages 1823 – 1836, 2011.

[12] E. Chiprout. Fast flip-chip power grid analysis via locality and grid shells. In *Proc. IEEE/ACM ICCAD*, pages 485–488, 2004.

[13] NVIDIA Corporation. *NVIDIA CUDA Programming Guide*. [Online]. Available: http://www.nvidia.com/object/cuda.html, 2007.