Microelectronics Journal 46 (2015) 258–264

Contents lists available at ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

Quantifying the effect of local interconnects on on-chip power distribution

Peirong Ji [a], Emre Salman [b,*]

[a] Digital Design Group SK Hynix Memory Solutions, San Jose, California 95133, USA
[b] Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, New York 11794, USA

ARTICLE INFO

Article history:
Received 24 February 2014
Accepted 13 December 2014
Available online 4 February 2015

Keywords:
On-chip interconnect
Power delivery
Power integrity
Power supply noise

ABSTRACT

Existing methods to analyze and optimize on-chip power distribution networks typically focus only on global power network modeled as a two-dimensional mesh. In practice, current is supplied to switching transistors through a local power network at the lower metal layers. The local power network is connected to a global network through a stack of vias. The effect of these vias and the resistance of the local power network are typically ignored when optimizing a power network and placing decoupling capacitors. By modeling the power distribution network as a three-dimensional mesh, the error due to ignoring via and local interconnect resistances is quantified. It is demonstrated that ignoring the local power network and vias can both underestimate (by up to 45%) or overestimate (by up to 50%) the effective resistance of a power distribution network. The error depends upon multiple parameters such as the width of local and global power lines and via resistance. A design space is also generated to indicate the valid width of local and global power lines where the target resistance is satisfied. It is shown that a wider global network can be used to obtain a narrower local network, providing additional flexibility in the physical design process since routability is an important concern at lower metal layers. At high via resistances, however, this approach causes significant increase in the width of a global power network, indicating the growing significance of local power network and vias.

1. Introduction

Robust and reliable power delivery is a critical challenge in modern, high performance integrated circuits (ICs) due to lower operating voltages, relatively higher switching currents, and faster transition times [1,2]. Parasitic resistance and inductance within a power distribution network cause both static and dynamic voltage fluctuations, typically characterized as $I(t)R$ drop and $L\,di/dt$ bounce. These fluctuations degrade *power integrity* and have three deleterious effects on circuit operation: (1) delay uncertainty since transistor delay is a strong function of the power supply voltage, (2) degradation in gate insulator reliability due to possible overshoots, and (3) in extreme cases, logical failure.

In the frequency domain, a power distribution network is analyzed to satisfy a *target impedance* within the frequency range of interest [3,4]. The target impedance is determined from the nominal supply voltage, tolerable noise, and average load current. According to ITRS 2011, by 2026, the nominal supply voltage is expected to scale down to 0.57, 0.43, and 0.54 V, respectively, for high performance, low operating power, and low standby power ICs [5]. Furthermore, the average current drawn by the load circuit increases due to greater functionality and higher static leakage current [5]. These two trends exacerbate the process of power delivery by significantly reducing the target impedance. Thus, modern power distribution networks consume a significant portion of the on-chip metal resources [1].

Efficient and sufficiently accurate analysis of power distribution networks plays an important role not only in quantifying power supply noise, but also optimizing the physical characteristics of a power network and placing decoupling capacitors [6,7]. Despite an increasing number of studies that implies the growing importance of local interconnects, there is still a significant amount of recent work that ignores these effects in analyzing and optimizing a power network [2,8–13]. These works utilize a two-dimensional model of a global power distribution network structured as a grid, as illustrated in Fig. 1. In [10], authors investigate the effect of vias only within the global power network. Local power grid and the vias between the local and global grids are not considered.

The commonly used model as shown in Fig. 1 ignores the *local* power distribution network at the lower metal layers (such as M1 and M2) and the *stack of vias* used to connect the global and local power networks. In [14], Shelar and Patyra have investi gated the impact of local *data* interconnects on timing and power

* Corresponding author.
E-mail addresses: peirong.ji@skhms.com (P. Ji), emre.salman@stonybrook.edu (E. Salman).

http://dx.doi.org/10.1016/j.mejo.2014.12.004
0026-2692/

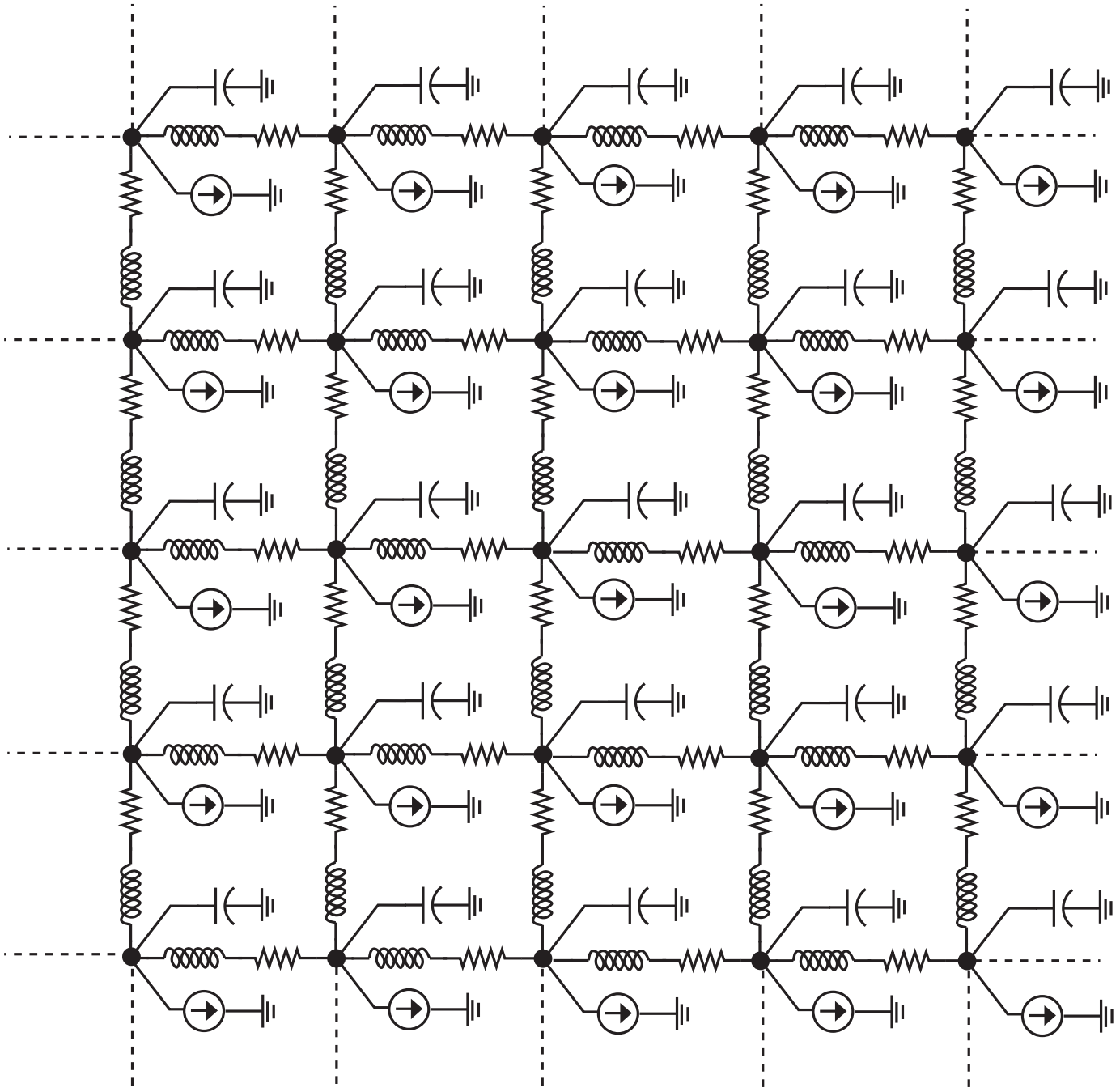

Fig. 1. Model of an interdigitated, global power distribution network that ignores the local power network and stack of vias connecting global and local power networks.

characteristics of a high performance microprocessor, demonstrating the non-negligible influence of these interconnects. The effect of local *power* interconnects and stack of vias on power integrity, and the dependence of this effect on metal widths, however, have not received much attention. In [15,16], local power network has been considered during power supply noise estimation. In [17], vias have also been included to analyze power supply noise. The primary research emphasis, however, has been on reducing the computational complexity and memory requirements of the analysis process. Error due to ignoring local power network and vias has not been investigated. This issue has become more significant since the via resistances increase with each technology node [5].

The primary contributions of this work are as follows: (1) the error in effective resistance due to ignoring local power network and vias is quantified, (2) the dependence of this error on metal width and via resistance is evaluated, (3) under constant physical area and metal coverage for the global power network, a design space is developed to determine the range of metal width where the target resistance is satisfied. The effect of via resistance on the design space is also investigated.

The rest of this paper is organized as follows. A power distribution network model considering both global and local power networks is described in Section 2. Analysis framework, methods used to achieve efficient and sufficiently accurate analysis of the model, and results quantifying the significance of local power network and vias are provided in Section 3. The paper is concluded in Section 4 and possible future work is discussed in Section 5.

2. Power delivery considering both local and global interconnects

In high performance ICs with large on-chip current demands, global power network is typically designed in an interdigitated fashion utilizing higher metal layers [2]. Local power network is located at the lower metal layers and is connected to a global power network through stack of vias. An example is illustrated in Fig. 2, where two metal layers are allocated to both global and local power networks. Note that only power lines are illustrated within the local network for clarity.

The local power network is typically determined by the physical structure of the standard cells. In standard-cell based design methodology, power trunks at the lower metal layers are placed adjacent to cell rows [18], supplying current to the adjacent logic arrays, as depicted in Fig. 3. The distance between the two power trunks is determined to permit sufficient space for routing local data interconnects. Specifically, this distance is assumed to be 100λ where λ is half of the minimum channel length.

An equivalent electrical model (three-dimensional mesh) for a power network considering both global and local networks, and stack of vias is depicted in Fig. 4. As opposed to Fig. 1, the resistances due to stack of vias and the local power distribution network are considered. Since the power network is assumed to be symmetric [7,17] (*i.e.*, resistance of each segment within the global network is equal, but different than the resistance of each segment within the local network), the equivalent model can be folded twice with respect to x- and y-axes to reduce the overall number of nodes [19]. Thus, the four power pads at the corners merge at the upper left corner whereas the sink node (originally assumed to be in the middle) is located at the lower right corner of the three-dimensional mesh. The equivalent resistance between these two nodes should be smaller than the target resistance of the power distribution network. Additional details of the analysis approach are described in the following section.

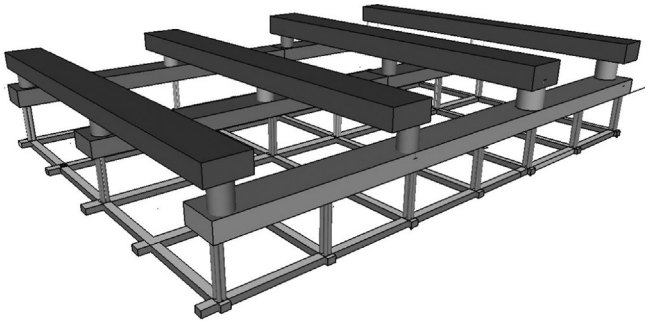

Fig. 2. Interdigitated power distribution network that includes both global and local networks, and stack of vias.

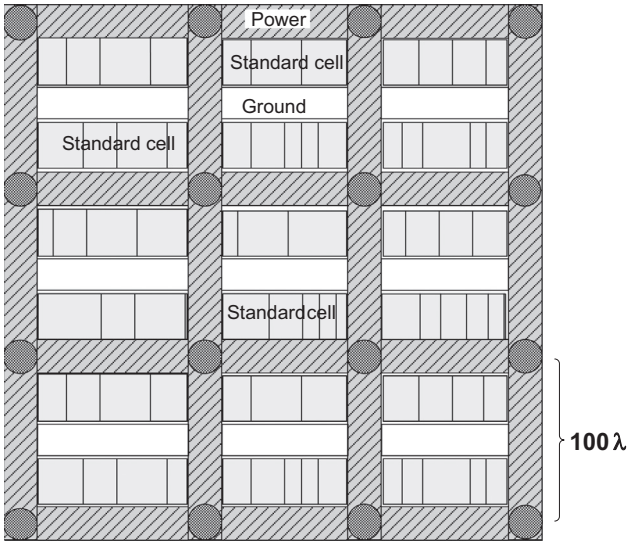

Fig. 3. Local power distribution network in standard-cell based design methodology where adjacent rows share a common power and ground line.

3. Analysis methodology and results

Analysis approach and the model parameters used to quantify the impact of local power network and stack of vias on power integrity are provided in Section 3.1. Methods adopted to efficiently and accurately analyze the overall network are summarized in Section 3.2. Results of this analyses are described in Section 3.3. Specifically, the error on equivalent resistance due to ignoring local power network and vias is determined. The design space to satisfy a target resistance and the effect of via resistance on this design space are explored in Section 3.4.

3.1. Analysis setup and model parameters

The 3-D mesh that represents global and local power networks, and the vias connecting these two networks (see Figs. 2 and 4) are analyzed assuming a constant physical area of $250 \times 250\ \mu m^2$. Assuming a flip-chip package with uniformly distributed C4 bumps, this area represents a region surrounded by four C4 bumps in each corner. Since the switching current within this region is dominantly supplied by these bumps, the overall power network can be partitioned by exploiting this locality [15].

Metal layers 7 and 8 are used for global power distribution network whereas layers 1 and 2 are used for local power distribution. Based on practical via densities provided in [20], 2000 via stacks (M1–M8) are distributed evenly throughout the power distribution network. Each via stack is modeled with an equivalent resistance [8,21]. Note that, in practice, a via stack consists of higher number of vias between two higher level metal layers as compared to two lower level metal layers due to different physical widths, producing a tapered structure. The overall via stack can be modeled with an equivalent via with equivalent resistance, as demonstrated in existing work [8,21]. For example, in [21], the equivalent via consists of 32, 16, 8, 4, and 2 vias between, respectively, M7/M6, M6/M5, M5/M4, M4/M3, and M3/M2. This structure is represented by a single resistance between M7 and M2. Since the value of this equivalent resistance (from metal 1 to metal 8) greatly varies according to the number of vias within a stack, a relatively large range (from $0.5\ \Omega$ to $7\ \Omega$) is considered in this work for two purposes: (1) to be able to cover a majority of the practical cases, and (2) understand the effect of local power network as a function of via resistance.

Technology specific parameters such as sheet resistance, minimum metal pitch, and thickness are obtained from a 32 nm technology node [22]. These parameters are listed in Table 1. Metal width and spacing are design variables. To determine these parameters, *metal coverage* is considered for the global power distribution network since clock and global data signals also need metal resources and routing space. Specifically, for each width, spacing is determined to satisfy, based on [23], 40% metal coverage. For the local power network, spacing between two power lines is determined to permit sufficient routing space for local data interconnects ($\approx 100\lambda$, as shown in Fig. 3, where λ is 16 nm for a 32 nm technology node). Metal layers 7 and 8 have the same width that varies from 1 μm to 10 μm. Similarly, the width of metal layers 1 and 2 is equal and varies from 1 μm to 5 μm. These characteristics are listed in Table 2. Note

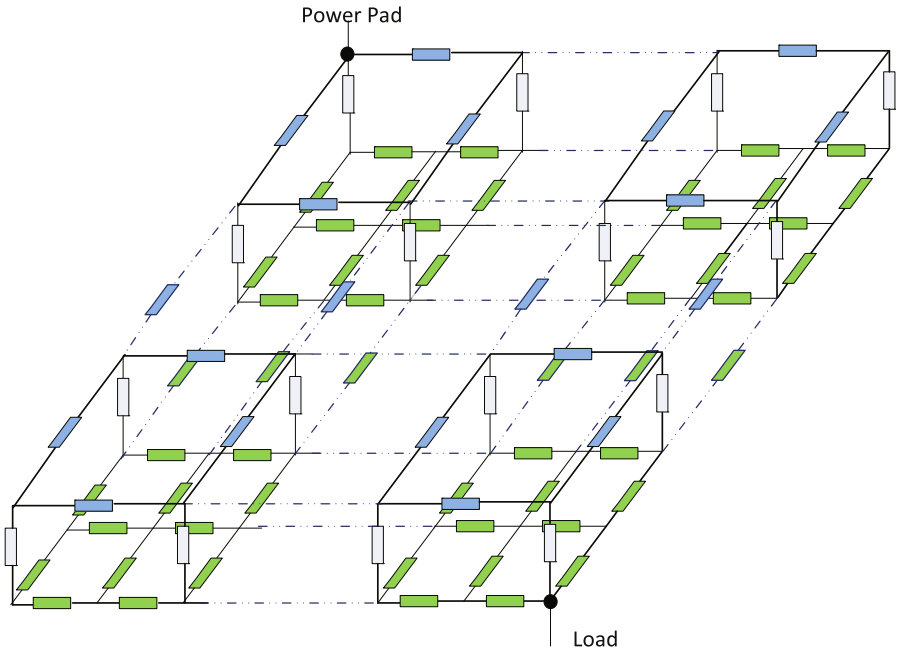

Fig. 4. Equivalent electrical model (three-dimensional mesh) for a power distribution network considering both global and local networks, and stack of vias.

Table 1
Technology specific parameters used to quantify the effect of local power network and vias [22].

Parameter	Metal 1/2	Metal 7/8
Minimum pitch (nm)	112.5	450.1
Thickness (nm)	95	388/504
Aspect ratio	1.7	1.7/1.8
Sheet resistance	0.45 Ω/□	0.11/0.085 Ω/□
Equivalent via stack resistance (M1–M8)	0.1–7 Ω	

Table 2
Width, spacing, and size for global and local power distribution networks for a physical area of $250 \times 250\ \mu m^2$ when metal coverage is constant at 40%.

M7/8			M1/2		
Width (μm)	Spacing (μm)	Size	Width (μm)	Spacing (μm)	Size
1	3.4	57×57	1	1.6	97×97
3	10.3	19×19	2	1.6	70×70
5	17.2	12×12	3	1.6	55×55
7	24.1	9×9	4	1.6	45×45
9	30.9	7×7	5	1.6	38×38

that since the physical area and metal coverage are constant, a wider metal corresponds to less number of parallel power lines. For example, when each metal has a width of 1 μm, the overall number of nodes in the power network exceeds ten thousand. Alternatively, when metals have the largest width, the number of nodes reduces to approximately one and a half thousand. Also note that the local power network is significantly denser than the global power network, as encountered in practical power networks.

3.2. Efficient and sufficiently accurate analysis

Due to a large number of nodes within the linear system, Gaussian elimination [24], as used in SPICE, is not feasible to analyze the power network. Traditional techniques such as modified nodal analysis (MNA) with Backward Euler (BE) integration and Cholesky decomposition help reducing the analysis time, but are still not sufficient for increasing number of nodes [25]. Similar to [26,27], sparsity and symmetric characteristic of the network are exploited to simultaneously apply conjugate gradient and sparse representation methods. Note that these methods have been traditionally utilized to analyze large power networks in existing literature. Using these methods, the 3-D mesh (see Fig. 4) is accurately analyzed (error within 0.15%) while decreasing the execution time by an order of magnitude as compared to SPICE. These results are illustrated in Fig. 5 where conjugate gradient method with sparse representation is compared with a direct solution that contains inverse operations.

This approach permits to efficiently investigate the effect of different physical parameters and quantify error due to ignoring local power network and vias, as described in the following section. Note that further improvements on runtime are possible through blocked compressed sparse row (BCSR) format that exploits the regular patterns within a sparse matrix to reduce the indirect memory access overhead [28].

3.3. Effect of local power network and vias

The effect of local power network and vias is investigated by analyzing the effective resistance between the power pad and sink node (see Fig. 4). Effective resistance of a power network is a first order metric that is commonly used in existing literature [3,10,17]. The methods described in the previous section are utilized to efficiently and accurately determine the effective resistance as a function of multiple parameters. For example, the variation of effective resistance as a function of local and global metal widths is depicted in Fig. 6 for four equivalent via stack resistances.

If the equivalent via stack resistance is sufficiently small [as in Fig. 6 (a)], the global and local power networks are approximately in parallel. In this case, the 2-D and 3-D mesh behave similarly when the metal 7/8 width is sufficiently high and metal 1/2 width is sufficiently low. Since the global power network exhibits significantly lower resistance at these widths, the difference between a 2-D and 3-D mesh is small. However, as the width of metal 1/2 increases, the effective resistance of the 3-D mesh becomes significantly smaller than the 2-D mesh since the resistance of the local power network becomes small enough

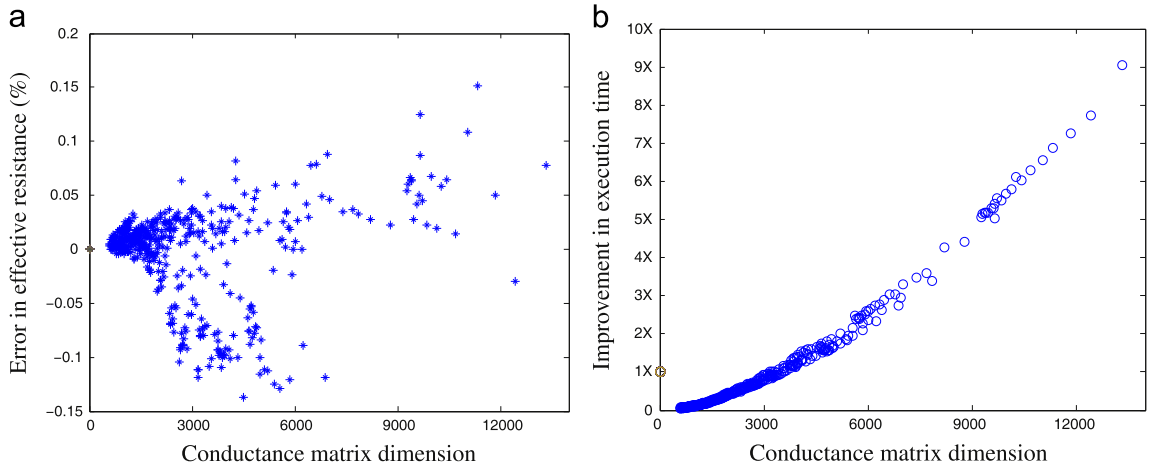

Fig. 5. Comparison of conjugate gradient method and sparse matrix representation with direct solutions: (a) accuracy and (b) computation time.

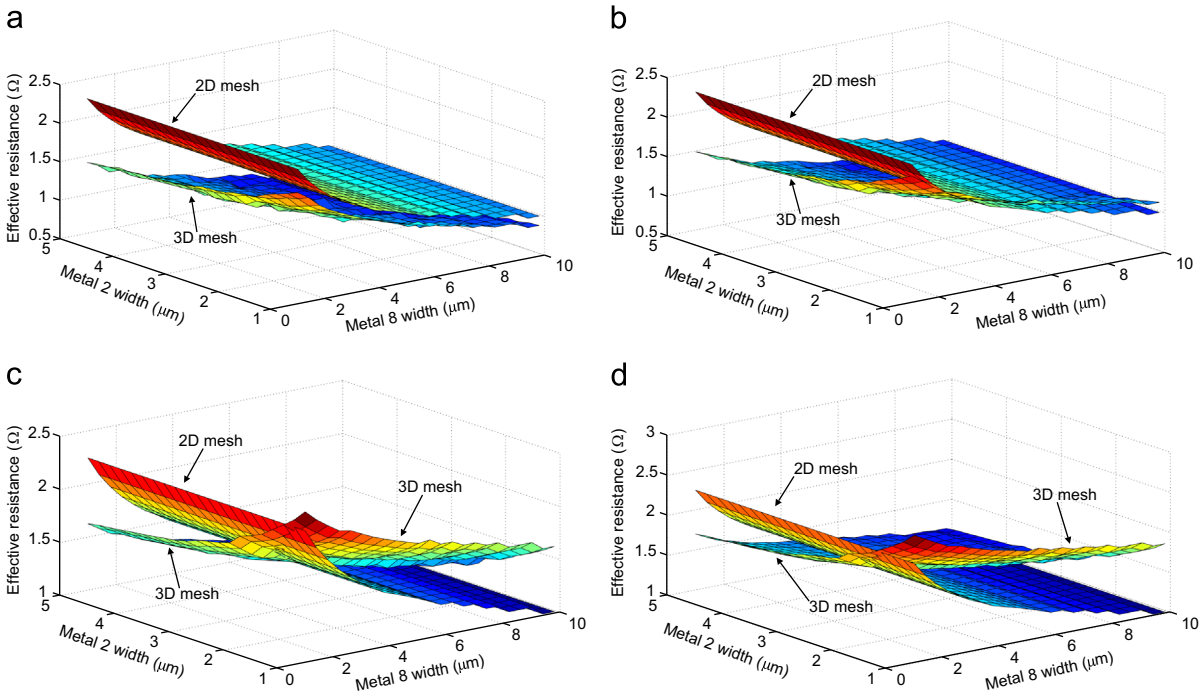

Fig. 6. Effective resistance surfaces of 2-D and 3-D power networks as a function of both global and local metal widths at equivalent via stack resistance of (a) 0.5 Ω, (b) 1 Ω, (c) 3 Ω, and (d) 5 Ω.

to change the effective resistance of the power network. *Thus, three important conclusions at low equivalent via stack resistances are*:

- The effect of the local power network and vias is negligible if the global power network is sufficiently wide and the width of the local power network is sufficiently small.
- As the width of the local power network is increased, the 2-D mesh significantly overestimates the effective resistance since the local and global power networks behave approximately in parallel and the resistance of the local power network becomes small enough to affect the overall resistance.
- A narrower local power network can be tolerated at low equivalent via stack resistances.

Alternatively, at relatively high equivalent via stack resistances ([as in Fig. 6(c) and (d)]), the effective resistance of the 3-D mesh crosses over the effective resistance of the 2-D mesh at a specific metal 1/2 width. If the width of metal 1/2 is further reduced, the 2-D mesh significantly underestimates the equivalent resistance since the local and global power networks no longer operate in parallel. *An important conclusion for high equivalent via stack resistances is*:

- If the number of vias within a single via stack and/or the overall number of via stacks is not sufficient, the local power network should be sufficiently wide to ensure that the effective resistance of the power grid does not significantly increase.

To better illustrate these conclusions, the error in effective resistance (due to ignoring local power network and vias) is depicted in Fig. 7 as a function of metal 1/2 width for two metal 7/8 widths. According to these figures, at low equivalent via stack resistances, 2-D mesh overestimates the effective resistance (note the negative error) and overestimation grows (up to 50%) as metal 1/2 width is increased. Alternatively, at high equivalent via stack resistances, 2-D mesh underestimates the effective resistance. This underestimation

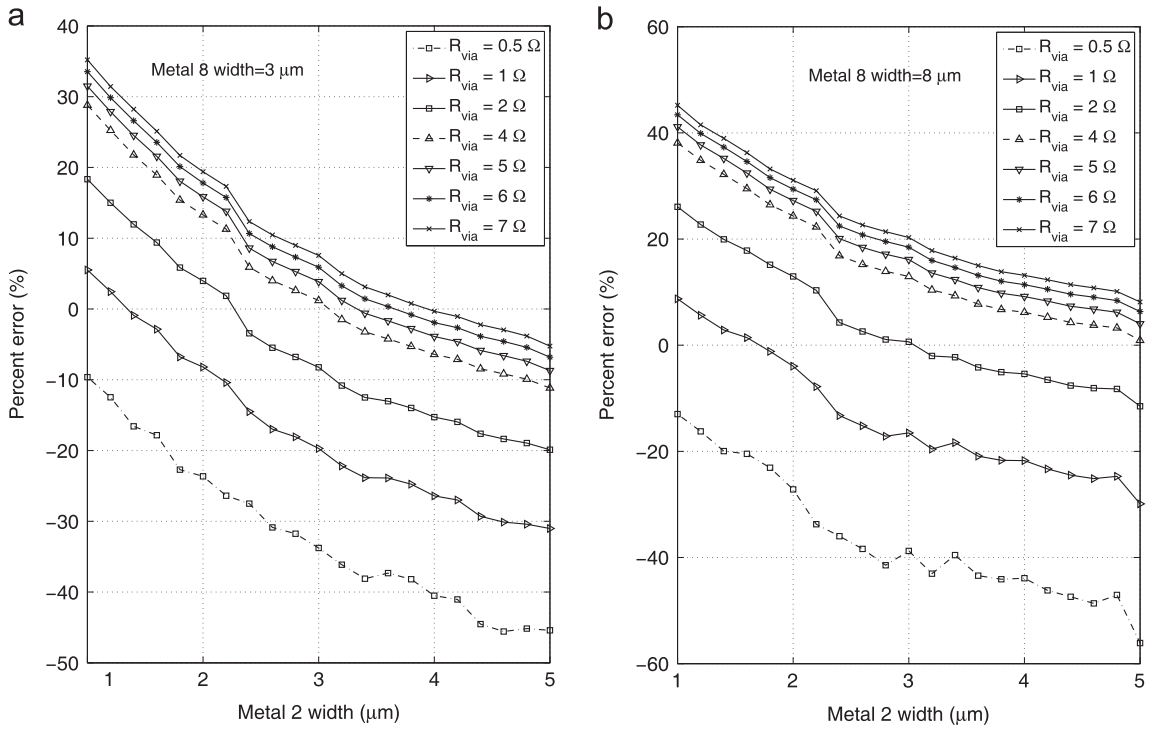

Fig. 7. Error in effective resistance (due to ignoring local power network and via resistances) as a function of metal 1/2 width at two metal 7/8 widths: (a) metal 7/8 width = 3 μm and (b) metal 7/8 width = 8 μm.

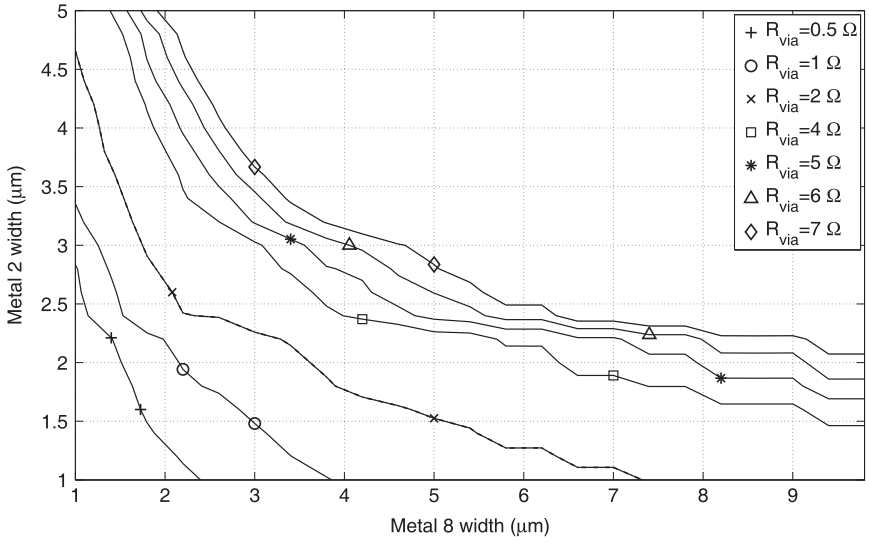

Fig. 8. Contour curves representing the design space based on a target resistance of 1.6 Ω. Region above each curve represents valid pairs of metal 1/2 and metal 7/8 widths for the corresponding via resistance.

is more significant (up to 45%) at lower metal 1/2 widths and hig her metal 7/8 widths. Note that at constant percent error, if the equivalent via stack resistance is reduced, a narrower metal 1/2 width is tolerated. Also note that, the effect of vias saturates when the equivalent via stack resistance reaches approximately 4 Ωs.

3.4. *Design space exploration in the presence of local power network and vias*

Since local power network is considered in the analysis, the design space includes not only metal 7/8 width, but also the width of metal 1/2. Thus, pairs of metal 7/8 width and metal 1/2 width that satisfy a specific *target resistance* can be identified. To determine target resistance, a power density of 50 W/cm^2 (corresponds to OpenSPARC T2 processor [29]) is considered. Scaling this power density for the physical area and assuming that the power supply voltage is 1 V, the power supply current is approximated as 31.25 mA. To limit the maximum *IR* drop to 5% of the power supply voltage (50 mV), the target resistance of the power network should be 1.6 Ω.

Design space to satisfy this target resistance can be determined by generating a contour curve from the effective resistance surfaces, as depicted in Fig. 8. This design space is generated at different equivalent via stack resistances. Any point that falls above the contour satisfies the target resistance (and therefore maximum *IR* drop) for that specific equivalent via stack resistance. Note that local

non-monotonicity observed in the contour curves is due to the location of via stacks since these locations change when the metal widths vary.

According to Fig. 8, a narrower metal 7/8 is possible at the expense of a wider metal 1/2. Alternatively, the width of metal 1/2 can be reduced by increasing the width of metal 7/8. However, if the width of the local power network (metal 1/2) is below a certain value, a significant increase is required in the width of the global power network (metal 7/8). This characteristic is exacerbated as the equivalent via stack resistance increases, demonstrating the increasing significance of local power network with higher via resistances. Note that this characteristic is also observed in Fig. 7 where the error increases with increasing via resistances. For example, in Fig. 8, if the via resistance is 4 Ω, to reduce the metal 1/2 width from 2.5 μm to 2 μm (20% decrease), the width of the metal 7/8 should be increased from approximately 3.8 μm to 6.5 μm (71% increase).

4. Conclusions

The effect of local power network and vias is investigated as a function of metal width and via resistance. A 3-D mesh based model is generated and efficiently analyzed by using conjugate gradient and sparse representation methods. It is shown that ignoring the local power network and vias can both underestimate (by up to 45%) or overestimate (by up to 50%) the effective resistance of a power distribution network depending upon the metal widths and via resistance. Specifically, if the equivalent via stack resistance is low, local and global grids behave approximately in parallel, reducing the effective resistance of the power grid. This effect becomes more dominant as the width of the local grid is increased. Alternatively, if the equivalent via stack resistance is relatively high, the two grids no longer operate in parallel and the effective resistance of the power grid increases. This effect becomes more dominant as the width of the local grid is decreased. A design space is also developed to indicate the widths of local and global power networks where the target resistance is satisfied. It is shown that a narrower local network is possible at the expense of a wider global network (and vice-versa), enabling higher flexibility for routability and electromigration constraints.

5. Future work

Future work includes considering vias and local power network while optimizing the physical characteristics of a power network such as width, pitch, and number/location of via stacks. Preliminary analyses demonstrate that number and location of via stacks (between local and global grids) should be considered during power network optimization. Placing decoupling capacitors while considering via stacks and local power network is also a focus of future study. This study is important particularly when MOS capacitors are used for decoupling since via stacks are required to connect the MOS capacitors with the local and global power networks.

References

[1] E. Salman, E.G. Friedman, High Performance Integrated Circuit Design, McGraw-Hill, New York, USA, 2012.
[2] R. Jakushokas, M. Popovich, A.V. Mezhiba, S. Kose, E.G. Friedman, Power Distribution Networks with On-Chip Decoupling Capacitors, Springer, New York, USA, 2011.
[3] L.D. Smith, R.E. Anderson, D.W. Forehand, T.J. Pelc, T. Roy, Power distribution system design methodology and capacitor selection for modern CMOS technology, IEEE Trans. Adv. Packag. 22 (3) (1999) 284–291.
[4] E. Salman, E.G. Friedman, R.M. Secareanu, O.L. Hartin, Worst case power/ground noise estimation using an equivalent transition time for resonance, IEEE Trans. Circuits Syst. I: Regul. Papers 56 (5) (2009) 997–1004.
[5] International technology roadmap for semiconductors (itrs), in: The ITRS Technology Working Groups, 2011.
[6] Y. Zhong, M.D.F. Wong, Fast algorithms for IR drop analysis in large power grid, in: Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, 2005, pp. 351–357.
[7] K. Shakeri, J.D. Meindl, Compact physical IR-drop models for chip/package co-design of gigascale integration (GSI), IEEE Trans. Electron Dev. 52 (6) (2005) 1087–1096.
[8] S.R. Nassif, Power grid analysis benchmarks, in: Proceedings of the IEEE/ACM Asia and South Pacific Design Automation Conference, 2008, pp. 376–381.
[9] Z. Zeng, P. Li, Locality-driven parallel power grid optimization, IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 28 (8) (2009) 1190–1200.
[10] S. Kose, E.G. Friedman, Effective resistance of a two layer mesh, IEEE Trans. Circuits Syst. II: Express Briefs 58 (11) (2011) 739–743.
[11] R. Jakushokas, E.G. Friedman, Power network optimization based on link breaking methodology, IEEE Trans. Very Large Scale Integr. (VLSI) Syst. 21 (5) (2013) 983–987.
[12] W. Zhang, et al., On-Chip power network optimization with decoupling capacitors and controlled-ESRs, in: Proceedings of the IEEE/ACM Asia and South Pacific Design Automation Conference, 2010, pp. 119–124.
[13] S. Zhao, K. Roy, C. Koh, Decoupling capacitance allocation and its application to power-supply noise-aware floorplanning, IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 21 (1) (2002) 81–92.
[14] R.S. Shelar, M. Patyra, Impact of local interconnects on timing and power in a high performance microprocessor, in: Proceedings of the ACM International Symposium on Physical Design, 2010, pp. 145–152.
[15] E. Chiprout, Fast flip-chip power grid analysis via locality and grid shells, in: Proceedings of the IEEE/ACM International Conference on Computer Aided Design, 2004, pp. 485–488.
[16] H. Qian, S.R. Nassif, S.S. Sapatnekar, Power grid analysis using random walks, IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 24 (8) (2005) 1204–1224.
[17] A. Dharchoudhury, et al., Design and analysis of power distribution networks in powerpc microprocessors, in: Proceedings of the IEEE/ACM Design Automation Conference, 1998, pp. 738–743.
[18] A.H. Ajami, K. Banerjee, A. Mehrotra, M. Pedram, Analysis of IR-Drop scaling with implications for deep submicron P/G network designs, in: Proceedings of the IEEE International Symposium on Quality Electronic Design, 2003, pp. 35–40.
[19] D.E. Khalil, Y. Ismail, Approximate frequency response models for RLC power grids, in: Proceedings of the IEEE International Symposium on Circuits and Systems, 2007, pp. 3784–3787.
[20] Y. Chiang, K. Banerjee, K.C. Saraswat, Analytical thermal model for multilevel VLSI interconnects incorporating via effect, IEEE Electr. Dev. Lett. 23 (1) (2002) 31–33.
[21] Y. Cai, J. Shi, S. Li, Optimization of via distribution and stacked via in multi-layered P/G networks, Integr. VLSI J. 43 (3) (2010) 318–325.
[22] S. Natarajan, et al., A 32 nm logic technology featuring 2nd-generation High-k + metal-gate transistors, enhanced channel strain and 0.171 μm^2 SRAM cell size in a 291Mb array, in: Proceedings of the IEEE International Electron Devices Meeting, 2008, pp. 1–3.
[23] A.B. Kahng, J. Lienig, I.L. Markov, J. Hu, VLSI Physical Design: From Graph Partitioning to Timing Closure, Springer, New York, USA, 2011.
[24] P.D. Mitcheson, Spice Algorithms and Internals, Technical Report, Department of Electrical and Electronic Engineering, Imperial College, 2010.
[25] T.L. Pillage, R.A. Rohrer, C. Visweswariah, Electronic Circuit and System Simulation Methods, McGraw-Hill, 1994.
[26] P. Feldmann, R.W. Freund, E. Acar, Power Grid Analysis Using a Flexible Conjugate Gradient Algorithm with Sparsification, Technical Report, Department of Mathematics, University of California at Davis, June 2006.
[27] C.H. Chou, et al., On the preconditioner of conjugate gradient method a power grid simulation perspective, in: Proceedings of the IEEE/ACM International Conference on Computer Aided Design, 2011, pp. 494–497.
[28] E.-J. Im, K. Yelick, R. Vuduc, SPARSITY: optimization framework for sparse matrix kernels, Int. J. High Perform. Comput. Appl. 18 (1) (2004) 135–158.
[29] H. Wei, et al., Cooling three-dimensional integrated circuits using power delivery networks, in: Proceedings of the IEEE International Electron Devices Meeting, 2012, pp. 14.2.1–14.2.4.