Exploratory Power Noise Models of Standard Cell 14, 10, and 7 nm FinFET ICs

Ravi Patel, Kan Xu, and Eby G. Friedman
University of Rochester, Rochester, NY, USA
rapatel, kxu8, friedman@ece.rochester.edu

Praveen Raghavan
IMEC Research, 3001 Leuven, Belgium
Praveen.Raghavan@imec.be

ABSTRACT

The physical dimensions of standard cells constrain the dimensions of power networks, affecting the on-chip power noise. An exploratory modeling methodology is presented for estimating power noise in advanced technology nodes. The models are evaluated for 14, 10, and 7 nm technologies to assess the impact on performance. Scaled technologies are shown to be more sensitive to power noise, resulting in potential loss of performance enhancements achieved by scaling. Stripes between local track rails is evaluated as a means to reduce power noise, exhibiting up to 56.5% improvement in power noise for the 7 nm technology node. A strong dependence on the width of a stripe is observed, indicating that fewer wide stripes are more favorable then many thin stripes. As a promising alternative material for power network interconnects, graphene is shown to exhibit good potential in reducing power noise. The effects of different scaling scenarios of local power rails on power noise are also discussed.

Keywords

Power noise model; advanced technology nodes; on-chip power networks; on-chip interconnect resistance

1. INTRODUCTION

Technology development for modern integrated circuits is a complex process that balances speed and power consumption with circuit density, manufacturability, and process yield [11]. Classical scaling theory predicts that the global interconnects are the dominant limitation to performance as compared to local interconnects, which scales with lithographic improvements [1, 2, 11, 13]. Local interconnect scaling, however, is unlikely to significantly lower RC delay. The resistivity of individual metal lines increases superlinearly as the minimum feature size shrinks [3].

Physical design information is typically unavailable during technology development. Lithographic constraints, device parameters, and standard cells are typically developed without a clear understanding of post-manufactured IC behavior. As a result, the resistance of a local power network is often constrained early in the design process. To maintain system performance under these constraints, aggressive global power grid design methodologies are required to reduce power noise [5]. To assess the impact of technology development, a procedure is described here for analyzing current distribution in standard cell CMOS circuits with little initial information of the process technology, enabling early projections of power noise and current distribution.

The paper is organized as follows. Background information is presented in Section 2. The modeling approach is discussed in Section 3. Power noise for different technology nodes, stripe conditions, interconnect materials, and power rail scaling scenarios are discussed in Section 4, followed by some conclusions in Section 5 and acknowledgments in Section 6.

GLSVLSI '16, May 18-20, 2016, Boston, MA, USA

DOI: http://dx.doi.org/10.1145/2902961.2903035

2. BACKGROUND

The trend of resistivity changes in on-chip interconnects as technology is scaled is discussed in Section 2.1. The structure and impedance characteristics of power networks are presented in Section 2.2. The topology of a standard cell circuit influences the design and switching activity of the power network, and, therefore, an overview of the structure is provided in Section 2.3.

2.1 On-chip interconnect resistance

As interconnects scale to more advanced technology nodes, highly resistive interconnects degrade performance improvements due to faster devices and greater integration. The resistivity of copper, used in traditional on-chip interconnects, sharply increases as the metal line pitch decreases [14]. Local power and ground rails therefore become significantly more resistive due to the "resistivity wall" phenomenon resulting in a more challenging power network design process.

Replacing copper with low resistivity materials in power grids is one way to reduce the effect of the "resistivity wall." Due to the excellent conductivity of both heat and electricity, graphene has been considered for different applications [8, 10]. Silver is another material whose bulk level resistivity is lower than copper. The thin film resistivity of two materials, silver and graphene, has been investigated, respectively, in [6, 9]. A comparison of the resistivity of different material with interconnect width scaling is illustrated in Figure 1. The thin film resistivity of silver increases significantly at 50 nm, and eventually becomes larger than copper as the metal line pitch is scaled to 10 nm. Graphene exhibits

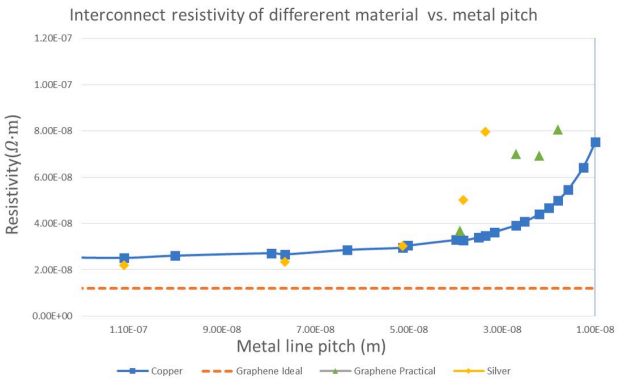

Figure 1: Interconnect resistivity of different materials versus line width.

higher resistivity than copper when the metal line pitch is small (from 40 nm to 10 nm).

2.2 Power grids

The resistance of power metal lines is also affected by the structure of the power grids. A power grid is a hierarchical structure consisting of power (V_{DD}) and ground (V_{SS}) backmetal pads, a global interdigitated mesh, and local power and ground rails, as illustrated in Figure 2. Four backmetal pads produce an effective global V_{DD} or V_{SS} mesh, as illustrated in Figure 2(a). Each pad is connected to several pairs of power and ground pairs (P/G pair) within a global power mesh. Each metal layer in the mesh consists of parallel P/G pairs separated from adjacent pairs by tens of micrometers. Metal layers are oriented orthogonal to the adjacent layers to create a mesh structure. The impedance of the global mesh, therefore, typically exhibits low resistance and significant inductance.

Standard cell tracks are patterned beneath the grid with local power and ground rails *(track rails)* placed horizontally between each P/G pair, as illustrated in Figure 2(b). The track rail impedances are dominated by the metal resistance and decoupling capacitance. On-chip power noise is due to switching on the track rails with the largest contribution arising from the clocked gates and buffers placed throughout the IC [7].

2.3 Standard cell design

An individual standard cell track is structured as a row with a substrate region patterned between the power and ground rails, as illustrated in Figure 2. Gates within a technology library are structured to fit within a track with transistors patterned within the substrate. Standard cells exhibit several characteristics that affect the power network. Most notably, the size and material of the power and ground rails within the standard cells affect the resistance of the metal line. The height of a standard cell is typically controlled by lithographic limits introduced by double and quadruple patterning processes [15]. Standard cell gates are mirrored to ensure that two tracks share a common power rail, doubling the effective current load on the line. After the gates are placed, interconnections are routed among the internal gates, constraining the available metal resources.

3. CIRCUIT MODELS

The overall grid model consists of a global mesh model, a local rail model, and a load model, as illustrated in Figure 3.

The global grid is modeled by an interdigitated mesh with the parameters described in [4]. The mesh size is determined by an effective mesh based on the space between the pads, as illustrated in Figure 2(a). The model considers the physical area, supply current, and stage delay for each process technology (14 nm, 10 nm, and 7 nm). The global mesh, track rail, and load models are discussed in the following sections.

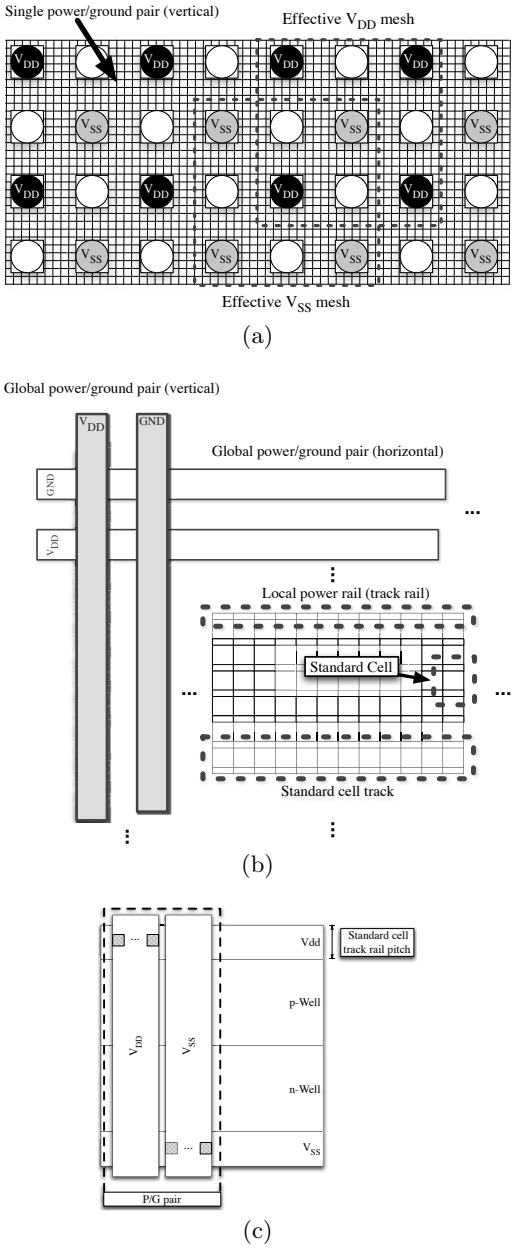

Figure 2: Topology of a standard cell power network with a) pad to global mesh, b) local rails attached to the tracks, and c) an individual standard cell connected to a global power/ground pair.

3.1 Load model

The peak power noise is dependent on the clock network The load model is based on the current demands of a register and the adjacent gates within a standard cell track. An individual load on a track rail is modeled as a current source with a triangular load characteristic [12], as illustrated in

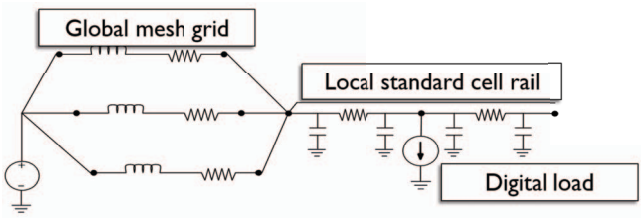

Figure 3: Model of power network

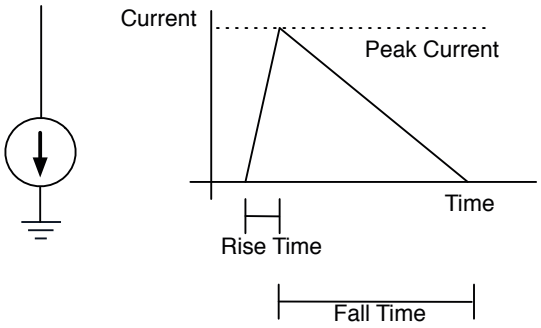

Figure 4: Model of an individual load on the power network. The rise and fall times are extracted from a loaded inverter and peak current, as described by (2).

Figure 4. The timing parameters of the model are extracted from a fan-out 4 (FO4) loaded inverter.

Those gates spatially adjacent to the register are also likely to switch at approximately the same time as the register and contribute to the local current draw. If an adjacent gate at the load switches before the track rail is recharged to the supply voltage, the magnitude of the noise increases. If the gate does not switch before the voltage is restored to V_{DD}, the gate does not contribute to the peak noise. Recharging determines the noise window (t_{window}) during which the loads that switch within the window are summed and the gates that switch outside of the window are ignored. The noise window, which determines the recharge time of a track rail, is approximated by three RC time constants,

$$t_{window} \approx 3\frac{N_{cell}^2}{4} R_{cell}(C_{cell} + C_{decap}), \quad (1)$$

where N_{cell} is the number of cells between each P/G pair, R_{cell} and C_{cell} are, respectively, the resistance and capacitance of the track rail within a standard cell, and C_{decap} is the decoupling capacitance per cell.

An adjacent logic gate only switches if the gate delay is within the noise window of the load. The delay of the adjacent gates is approximated by the delay of an inverter. The load current is

$$I_{Load} = \frac{\alpha 2 t_{window}}{t_{inv1}} I_{inv1} + 2I_{invd4}, \quad (2)$$

where t_{window} is the noise window, t_{inv1} and I_{inv1} are, respectively, the delay and peak current of a 1x inverter, I_{inv4} is the peak current of a 4x inverter, and α is the switching factor of the circuit.

3.2 Rail model

Each local rail is modeled as a distributed resistor-capacitor with multiple loads, with the length of the rail determined by the space between two P/G pairs in the global power

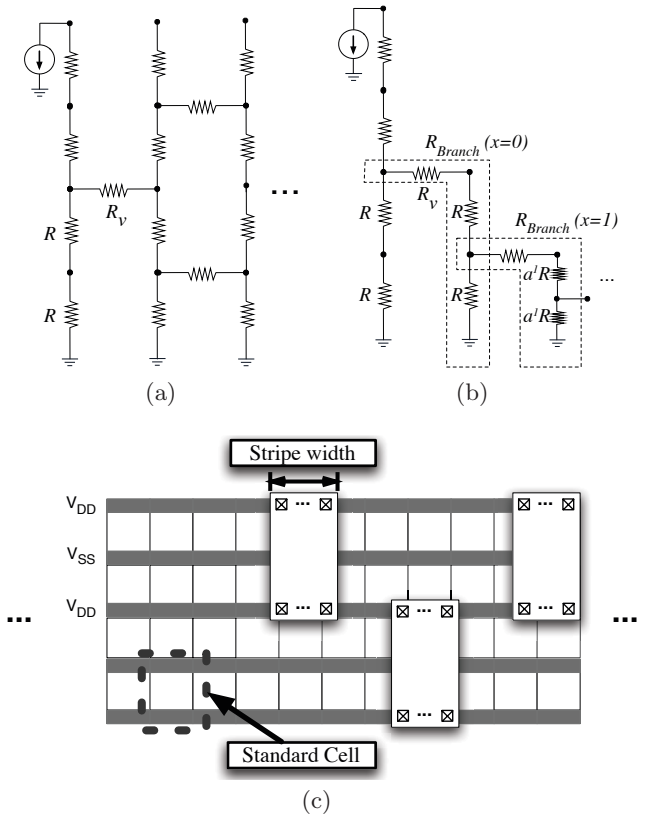

Figure 5: Impedance of a) striped power rails, b) R_{branch} approximation of impedance, and c) physical structure of a stripe.

network. At least one load is placed at the center of the rail to model a single register in the worst case position. The number of loads and the space between loads are determined by the target clock frequency of the circuit. An individual logic gate is modeled with an inverter delay (t_{inv}) where the logic depth (D) at a given frequency (f_{clock}) is

$$D = \frac{1}{f_{clock} t_{inv}(1+U)}, \quad (3)$$

where D is the delay uncertainty. The logic depth is the number of gates between adjacent loads on a rail. The width of an inverter is used to estimate the size of a standard cell. The physical distance between loads on a local rail is therefore known. Based on this assumption, the total number of active loads and the impedance between each active load can be estimated. The logic depth is also used to determine the decoupling capacitance,

$$C_{decap} = C_{gate}(1-\beta)D, \quad (4)$$

where C_{gate} is the gate capacitance of an inverter, and β is the fill factor of the standard cell layout. The fill factor is a common metric which is the fraction of silicon area occupied by the standard cells.

3.3 Striping of the power rail

Each track rail is typically distinct. Recently, however, connections between adjacent track rails have been used to reduce the local rail resistance and any associated power noise, as illustrated in Figure 5(a). These connections between power ground rails, called stripes, ensure that loads on the adjacent rails interact. For any interaction, however, the

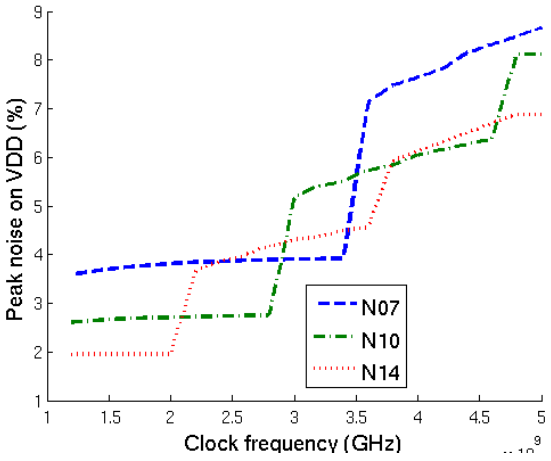

Figure 6: Local peak power noise in 14 nm, 10 nm, and 7 nm technologies with increasing clock frequency.

worst case noise is equivalent to the case of a single track rail without striping. The maximum reduction in power noise from striping occurs when the adjacent rails do not have loads that switch at the same time. These two conditions, therefore, bound the potential noise generated by a circuit. The number of interacting rails is determined by approximating a set of rails as a resistive tree, as illustrated in Figure 5(b). The resistance from the center load to the edge of the track rail is

$$R_{branch} = R_v + a^x * R + \left(\frac{1}{a^x * R} \frac{1}{R_{branch}(x+1)} \right)^{-1}, \quad (5)$$

where R_v is the resistance of a stripe, x is the number of additional branches, and a is the scaling factor of the resistance. As x increases, the error decreases. Note that (5) is used to estimate the maximum number of rails that minimizes the error. A distributed resistance across the rail is included in the model.

4. EVALUATION OF POWER NOISE

The model has been evaluated for power networks in 14 nm, 10 nm, and 7 nm CMOS FinFET technologies. The global power grid dimensions are determined from a 14 nm circuit. The global grid pitch is scaled to 10 nm and 7 nm. Model generation and simulation are based on MATLAB and Cadence Spectre. A power noise analysis with different technology nodes, stripes conditions, interconnect materials, and interconnect scaling scenarios is presented in the following sections.

4.1 Different technology nodes

The local V_{DD} rails exhibit a peak power noise that ranges from 3% to 10% of V_{DD} with a trend of increasing power noise with technology scaling. As the clock frequency supported by the track increases, the power noise increases in discrete steps, as illustrated in Figure 6. Each step is due to an increase in the number of loads that simultaneously switch on a track rail, which corresponds to a relative decrease in logic depth. Local noise levels also increase with each technology node, although the magnitude of the noise is strongly dependent on the clock frequency and number of loads per rail. At lower frequencies with only a single load switching per rail, N10 and N07 exhibit, respectively, power noise increases of 0.7% and 1.8% as compared to N14. With higher frequencies with two loads per rail, the power noise increases by, respectively, 1.8% and 4.1%. This behavior is expected as the width of a standard cell gate is proportionally larger with scaled technologies, producing a larger track rail resistance per cell.

To measure the effects of power noise on circuit performance, a five stage ring oscillator (RO) is driven with power noise injected into both the power and ground rails. The per cent reduction in ring oscillator frequency is depicted in Figure 7. As the power noise increases with frequency, the performance of the ring oscillator decreases. As expected, the RO performance increases with each technology generation and drops in discrete amounts with increasing clock frequency. Notably, the magnitude of the decrease in oscillator frequency is much higher in N07 than in N10 and N14, indicative of the increasing sensitivity to power noise with device scaling. At frequencies above 3 GHz, the performance of the N07 ring oscillator drops below the performance of the N10 ring oscillator operating at a lower clock frequency. Intuitively, the delay of an N07 circuit degrades, losing the advantages of scaling. Maintaining performance requires a proportionally smaller P/G pitch that is more aggressive than a linearly scaled grid.

4.2 Stripes across power rails

To reduce local power noise, an individual track rail can use multiple stripes to adjacent rails, each with a variable width. The noise exhibited by a 3.6 GHz circuit with striping for variable width and count is illustrated in Figure 8. For reference, the peak noise of a 3.6 GHz circuit without striping for the N14, N10, N07 technology nodes are, respectively, 4.6%, 5.7%, 7.1%. The stripe count is the number of stripes per track rail, and the stripe width is the pitch of a stripe with additional via contacts. Both the stripe count and stripe width are normalized to the minimum metal pitch of the technology node.

Introducing striping reduces power noise by almost a factor of two for each technology, with a slight increase in noise reduction with each technology generation. The maximum stripe width and count, with nine stripes at a stripe width of ten, is impractical in conventional circuits for any technology node. In these cases, ten cells are between each stripe, and each stripe is approximately the size of four inverter cells. These additional interconnects cause significant routing congestion and area overhead.

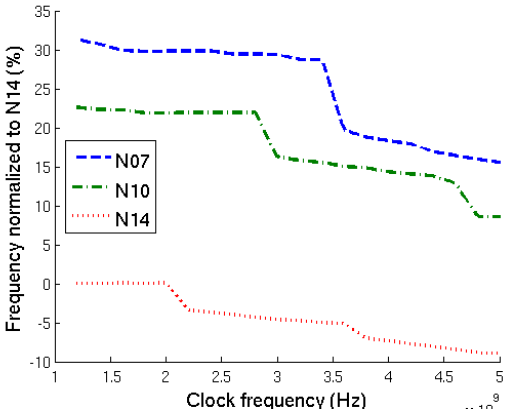

Figure 7: Per cent performance decrease of average power noise on five stage ring oscillator in 14 nm, 10 nm, and 7 nm technologies normalized to an N14 ring oscillator.

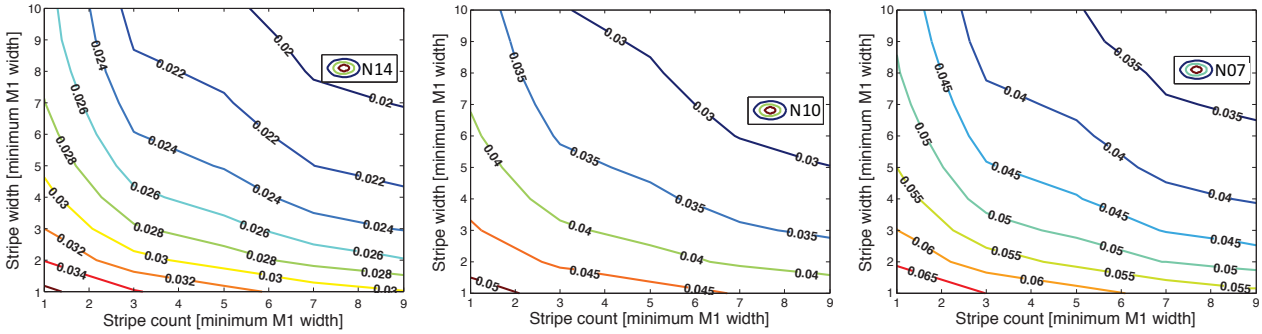

Figure 8: Effect of track stripe count and stripe width on noise for 3.6 GHz track rails in a) 14 nm, b) 10 nm, and c) 7 nm technologies.

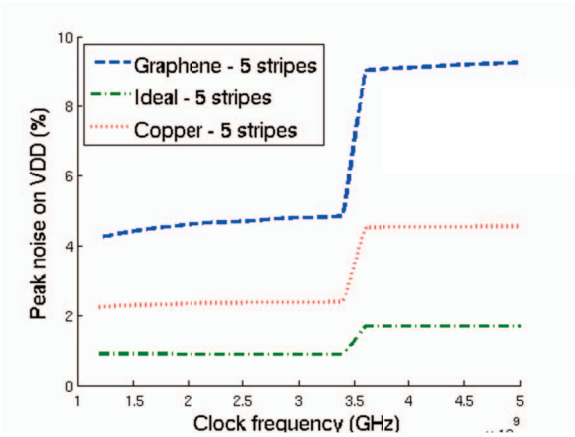

Figure 9: Peak power noise in graphene, ideal graphene, and copper power grids with increasing clock frequencies in 7 nm technology.

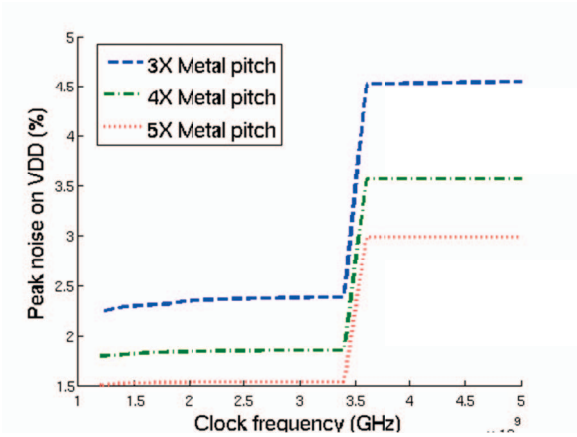

Figure 10: Peak noise in 3X, 4X, and 5X minimum metal pitch interconnect scaling scenarios with increasing clock frequencies in 7 nm technology.

Much of the benefit, however, can be achieved by utilizing wide stripes. A single stripe with a stripe width of ten can reduce the power noise by almost a third for N14, N10, and N07. This reduction in noise is due to the relatively large resistance of the via contacts for each stripe. As the stripe width increases, additional via contacts can be added, reducing the effective resistance of the stripe, thereby lowering the resistance of the path to the power supply. At stripe counts greater than five, there are diminishing returns on the reduction in power noise. In this case, a stripe width above six reduces much of the power noise without incurring excessive overhead.

4.3 New material interconnects

Another way to reduce power noise is exploiting lower resistivity material in power grids. As illustrated in Figure 1, the resistivity of graphene is comparable to copper in deeply scaled metal lines. Although integrating graphene with CMOS technology is not yet practical, graphene as an interconnect replacement would provide significant improvement in reducing power noise.

Power noise is evaluated for the 7 nm technology with five stripes across the power ground rails for three different materials. The resistivity of graphene is extracted from experimental data based on the local interconnect width used in 7 nm technology [9]. The theoretically lowest resistivity achievable by graphene is described here as ideal graphene and is used for comparative purposes. The third material is copper. As illustrated in Figure 9, a large difference in power noise between graphene and copper is exhibited since the difference in resistivity is large in the 7 nm interconnect technology. A 59.1% reduction of peak noise is achieved with ideal graphene as compared with copper. The bottleneck is the vias between two adjacent metal layers, which is highly resistive in advanced technology nodes as compared to the resistance of the metal lines.

4.4 Power rail scaling

For a mesh global power grid, the pitch of each power/ground pair is decreased with technology scaling. The width of global power/ground interconnect is however fixed in advanced technology nodes to prevent an increase in impedance within the global power grid. Enlarging the width of the global power grid significantly increases on-chip area while also introducing larger parasitic capacitances between adjacent metal layers. For interdigitated local power rails, the pitch of the adjacent power and ground rail is proportional to the gate pitch to match the standard cell height for each technology node. The width of the local power rail is proportional to the minimum metal pitch of each technology. This scaling process is a primary source of power noise due to IR drops.

In evaluating power noise, the local power rail width is set to three times the minimum metal pitch of each technology. With technology scaling, a tradeoff should be considered between the footprint of the local power rails and the

impedance characteristics of the power rails to satisfy power noise budgets in advanced technology nodes. A smaller standard cell height makes it possible to increase the on-chip area of the local power rails while maintaining performance improvements over previous technology nodes. As illustrated in Figure 10, a 32.4% reduction in peak noise is exhibited after increasing the power rail width from three times to five times the minimum metal pitch in 7 nm technology. As compared with Figure 9, the reduction in power noise with larger power rail widths is lower than exploiting new interconnect materials. Changing the metal width is however more practical since it does not require novel fabrication and integration technologies. Increasing the width of the local power rail degrades the performance due to the large area overhead of the local metal layer.

5. CONCLUSIONS

An exploratory modeling methodology is proposed for assessing power network noise in standard cell digital circuits. Models are evaluated for 14, 10, and 7 nm technologies to evaluate noise trends. It is shown that local resistive noise increases with technology scaling. The effects of the local stripes are evaluated on power grids, exhibiting a 2x reduction in local power noise. Trends evaluating the effects of noise on performance are also discussed. The sensitivity to power noise is shown to increase with clock frequency and scaling, potentially reducing performance gains achievable in deeply scaled technology nodes. Performance degradation in 7 nm is sufficiently severe that no delay advantage to linear scaling is exhibited. Below 10 nm, more aggressive power networks with a shorter distance between adjacent P/G pairs are required to compensate for the increase in noise. Exploiting new materials in on-chip interconnect exhibits good potential to lower power noise in advanced technology nodes. Tradeoffs between power noise and performance need to be carefully considered when scaling local power rails.

6. ACKNOWLEDGMENTS

This research is supported in part by the Binational Science Foundation under Grant No. 2012139, the National Science Foundation under Grant Nos. CCF-1329374, CCF-1526466, and CNS-1548078, the IARPA under Grant No. W911NF-14-C-0089, and by grants from Cisco Systems and Intel.

7. REFERENCES

[1] H. B. Bakoglu and J. D. Meindl. Optimal interconnection circuits for vlsi. *IEEE Transactions on Electron Devices*, 32(5):903–909, May 1985.

[2] M. T. Bohr et al. Interconnect scaling-the real limiter to high performance ulsi. In *Proceedings of the IEEE International Electron Devices Meeting*, pages 241–244, December 1995.

[3] J. A. Davis et al. Interconnect limits on gigascale integration (gsi) in the 21st century. *Proceedings of the IEEE*, 89(3):305–324, March 2001.

[4] R. Jakushokas and E. G. Friedman. Multi-layer interdigitated power distribution networks. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 19(5):774–786, May 2011.

[5] R. Jakushokas, M. Popovich, A. V. Mezhiba, S. Köse, and E. G. Friedman. *Power Distribution Networks with On-Chip Decoupling Capacitors, Second Edition*. Springer Science & Business Media, 2011.

[6] F. Lacy. Developing a theoretical relationship between electrical resistivity, temperature, and film thickness for conductors. *Nanoscale Research Letters*, (1), December 2011.

[7] S. Lin and N. Chang. Challenges in power-ground integrity. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pages 651–654, November 2001.

[8] J. S. Moon and D. K. Gaskill. Graphene: Its fundamentals to future applications. *IEEE Transactions on Microwave Theory and Techniques*, 59(10):2702–2708, October 2011.

[9] R. Murali, K. Brenner, Y. Yang, T. Beck, and J. D. Meindl. Resistivity of graphene nanoribbon interconnects. *IEEE Electron Device Letters*, 30(6):611–613, June 2009.

[10] P. Pasanen, M. Voutilainen, M. Helle, X. Song, and P. J. Hakonen. Graphene for future electronics. *Physica Scripta*, 2012(T146):014025, January 2012.

[11] E. Salman and E. G. Friedman. *High Performance Integrated Circuit Design*. McGraw Hill Professional, 2012.

[12] E. Salman, E. G. Friedman, R. M. Secareanu, and O. L. Hartin. Worst case power/ground noise estimation using an equivalent transition time for resonance. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 56(5):997–1004, May 2009.

[13] K. C. Saraswat and F. Mohammadi. Effect of scaling of interconnections on the time delay of vlsi circuits. *IEEE Journal of Solid-State Circuits*, 17(2):275–280, April 1982.

[14] W. Steinhögl, G. Schindler, G. Steinlesberger, M. Traving, and M. Engelhardt. Comprehensive study of the resistivity of copper wires with lateral dimensions of 100 nm and smaller. *Journal of Applied Physics*, 97(2):023706–1— 023706–7, January 2005.

[15] K. Vaidyanathan, S. H. Ng, D. Morris, N. Lafferty, L. Liebmann, M. Bender, W. Huang, K. Lai, L. Pileggi, and A. Strojwas. Design and manufacturability tradeoffs in unidirectional and bidirectional standard cell layouts in 14 nm node. *Proceedings of SPIE*, 8327:83270K–83270K–12, March 2012.