

Power Noise in 14, 10, and 7 nm FinFET CMOS Technologies

Ravi Patel and Eby G. Friedman
Department of Electrical and Computer Engineering
University of Rochester
Rochester, New York 14627
Email: (rapatel, friedman)@ece.rochester.edu

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

Abstract—A methodology is described to determine the distribution of current within a power network for use in CMOS standard cell integrated circuits based on exploratory information about the power network. Models are presented to extrapolate the noise within power networks in 14, 10, and 7 nm CMOS technologies. Stripes, interconnect between local power rails, are evaluated as a means to reduce power noise, resulting in a 56.5% reduction in noise for the 7 nm CMOS technology node.

I. INTRODUCTION

DEVELOPMENT of a modern CMOS technology is a multi-objective procedure that requires tradeoffs among speed, power consumption, circuit density, manufacturability, and process yield [1], [2]. During this design process, circuit and performance information are typically unavailable. Design rules, device parameters, and synthesis libraries are often developed before the behavior of a manufactured IC is understood. This process constrains the early design of an IC power network. Overly conservative global power grid structures are therefore frequently used to manage power noise [3].

To quantify the effects of device parameters and design rules, circuit models have been developed to determine the current profile of standard cell CMOS ICs. With the limited information available in the technology development process, extrapolation of the power noise and system performance is necessary. Power networks in 14, 10, and 7 nm CMOS technology nodes are evaluated for circuit noise and clock frequency. Striping between local power rails is also evaluated as a power reduction technique.

The paper is organized as follows. Background is presented in Section II. Circuit models are reviewed in Section III. Power noise and the associated performance degradation characteristics are discussed in Section IV, followed by some conclusions in Section V.

II. OVERVIEW OF TRADITIONAL 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 1. Four backmetal pads produce an effective global V_{DD} or V_{SS} mesh, as illustrated in Figure 1a. 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 1b. The track rail impedances are dominated by the metal resistance and decoupling capacitance. On-chip power noise is due to signal switching on the track rails with the largest contribution arising from the clocked gates and buffers [4].

III. CIRCUIT MODELS

The overall grid model is comprised of a global mesh, a local rail, and a load, as illustrated in Figure 2. The global grid is modeled by an interdigitated mesh with the parameters described in [5]. The mesh size is determined by an effective mesh based on the space between pads, as illustrated in Figure 1a. 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.

A. Load model

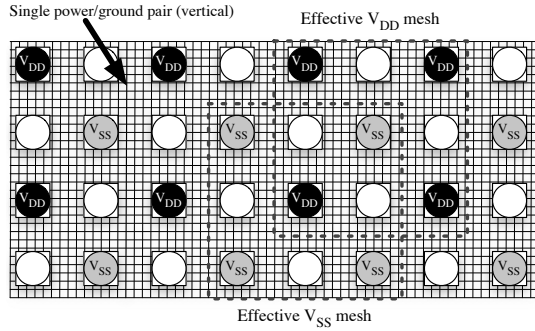
Given the dependence of the peak power noise on the clock network, the load model is based on the current demand of a register and the adjacent gates within a standard cell track. An individual load on a track rail is modeled as a current source with a triangular load characteristic [6], as illustrated in Figure 3. The timing parameters of the model are extracted from a fanout 4 (FO4) loaded inverter.

Those gates spatially adjacent to the register are likely to switch at approximately the same time as the register and contribute to the local current draw. At the load, if an adjacent gate 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. The recharging time 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

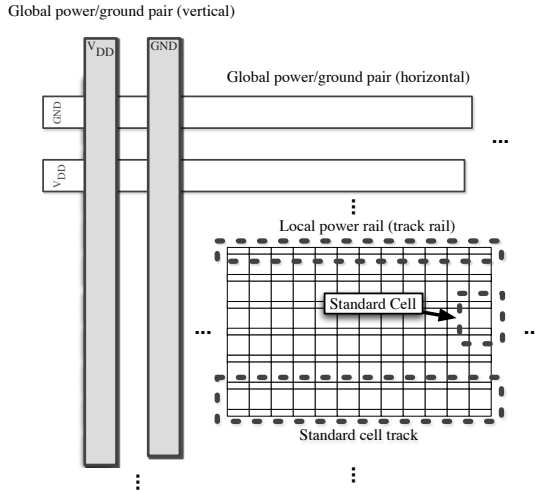
$$t_{window} \approx 3 \frac{N_{cell}^2}{4} R_{cell} (C_{cell} + C_{decap}), \quad (1)$$

where N_{cell} is the number of cells between each P/G pair, R_{cell} and C_{cell} are, respectively, the resistance and capaci-

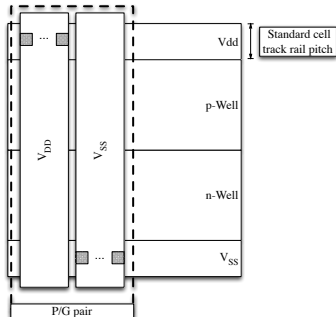
This research is supported in part by the Binational Science Foundation under Grant No. 2012139, the National Science Foundation under Grant Nos. CCF-1329374, CCF-1526466, and CNS-1548078, IARPA under Grant No. W911NF-14-C-0089, and by grants from Cisco Systems and Intel.



a



b



c

Fig. 1. 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.

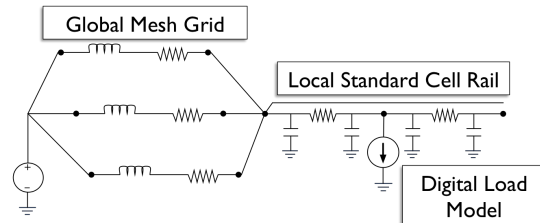


Fig. 2. Model of power network

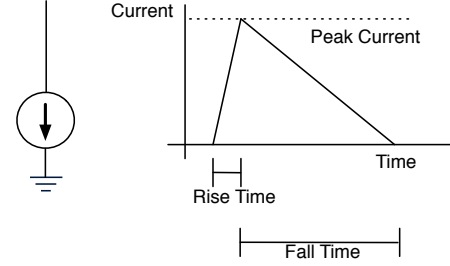


Fig. 3. Model of a current load on the power network. The rise and fall times are extracted from a loaded inverter and peak current, as described by (2).

tance 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 current load. The gate delay is approximated by the delay of an inverter. The load current is

$$I_{Load} = \frac{\alpha 2 t_{window}}{t_{inv1}} I_{inv1} + 2 I_{inv4}, \quad (2)$$

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.

B. 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 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 frequency (f_{clock}) is

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

where U 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, permitting the physical distance between loads on a local rail to be known. Based on this assumption, the total number of active loads and the impedance between each active load can be estimated. The logic depth D 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 the fraction of silicon area occupied by the standard cells.

C. 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 4a. These connections between local

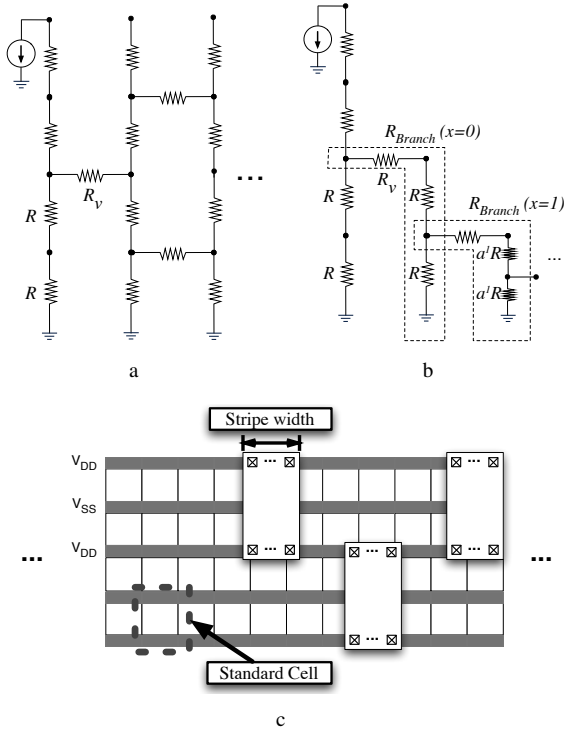


Fig. 4. Local striping in standard cells. Striping introduces a resistive interlink between the local rails which is approximated with a resistive tree. a) The impedance model of the striped power rails, b) R_{branch} approximation of the impedance, and c) physical structure of a stripe.

power ground rails, called stripes, ensure that loads on the adjacent rails interact. For any interaction, however, the 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 load on adjacent rails do not simultaneously switch. 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 4b. 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 is included in the model.

IV. EVALUATION OF POWER NOISE

The model is 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.

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

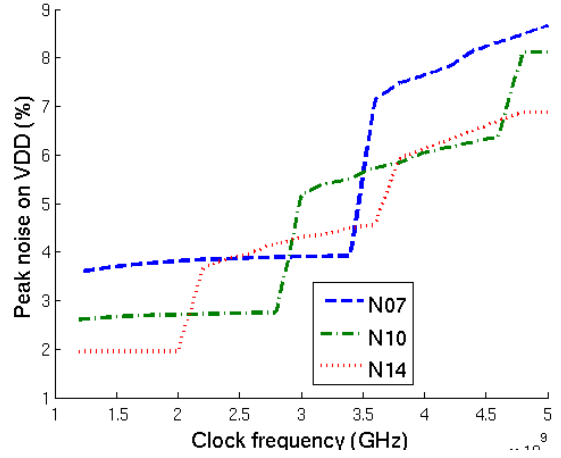


Fig. 5. Local peak power noise in 14 nm, 10 nm, and 7 nm technologies with increasing clock frequency.

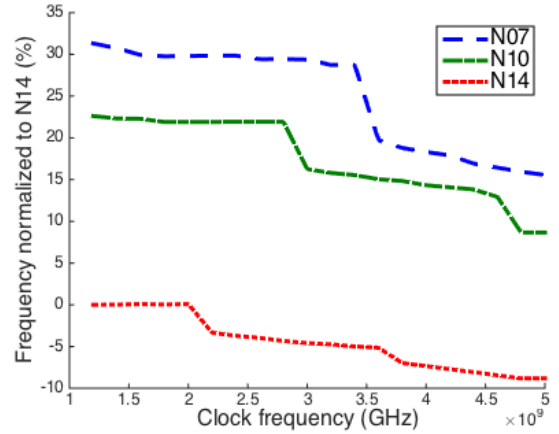


Fig. 6. Per cent decrease in performance of average power noise on a five stage ring oscillator in 14 nm, 10 nm, and 7 nm technologies normalized to an N14 ring oscillator.

discrete steps, as illustrated in Figure 5. 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. At frequencies below 2 GHz, the logic depth exceeds the total number of cells per local rail and therefore only one register switches per rail within a clock period. Reduced delay in each technology corresponds to a larger logic depth, resulting in steps in the noise level at higher frequencies for each technology node. After each noise step, the noise level increases linearly with frequency. These cases reflect those circuits where the loads are located close to the center of the track rail, and therefore exhibit a large resistive path to the power supply. Local noise levels also increase with each technology, 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. At higher frequencies with two loads per rail, the power noise increases by, respectively, 1.8% and 4.1%. This behavior occurs since the width of a standard cell gate is proportionally larger with scaled technologies, producing a higher track rail resistance per cell.

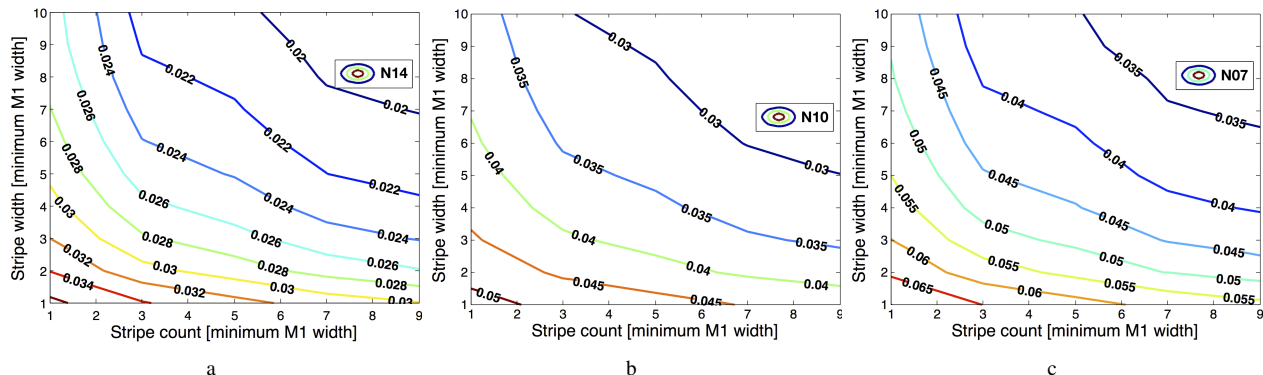


Fig. 7. Effect of track stripe count and stripe width on noise for 3.6 GHz track rails in a) 14 nm, b) 10 nm, and c) 7 nm technologies.

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 6. 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 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. 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.

TABLE I
PEAK NOISE AT 3.6 GHz WITHOUT STRIPING

	N14	N10	N07
Peak noise @ 3.6 GHz	4.6%	5.7%	7.1%

To reduce local power noise, an individual track rail can use multiple stripes connected 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 7. For reference, the peak noise of a 3.6 GHz circuit without striping is provided in Table I. 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.

Introducing striping reduces power noise by almost a factor of two for each technology node, with a slight reduction in noise 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.

Much of the benefit in lower noise, however, can be achieved by utilizing wide stripes. A single stripe with a stripe

width of ten can reduce 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, more via contacts can be added, reducing the effective resistance of the stripe, and 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. Increasing the stripe width to ten reduces noise, but also reduces the area available for standard cell placement between local rails. A stripe width that is six times the minimum pitch reduces most of the power noise without incurring excessive overhead.

V. CONCLUSIONS

Models are described to assess noise in sub-14 nm FinFET CMOS technologies, quantifying power noise trends for a range of clock frequencies. The performance impact of these noise trends are evaluated. Striping is shown to alleviate noise issues in local power networks, demonstrating a 200% reduction in power noise.

It is also shown that noise associated with IR drops increases beyond classical scaling trends. This increase in IR drops coupled with a greater sensitivity to power noise in deeply scaled technologies results in a performance drop that degrades any scaling related speed improvements. At the 7 nm technology node, no delay advantage occurs when the grid parameters are linearly scaled from prior nodes. For technologies 10 nm and below, a power network with shorter distances between P/G pairs is necessary to compensate for increased power noise.

REFERENCES

- [1] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*, McGraw Hill Professional, 2012.
- [2] M. Quirk and J. Serda, *Semiconductor Manufacturing Technology*, Prentice Hall, 2001.
- [3] 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.
- [4] S. Lin and N. Chang, "Challenges in Power-Ground Integrity," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 651–654, November 2001.
- [5] R. Jakushokas and E. G. Friedman, "Multi-Layer Interdigitated Power Distribution Networks," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 19, No. 5, pp. 774–786, May 2011.
- [6] 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*, Vol. 56, No. 5, pp. 997–1004, May 2009.