Impact of Random Dopant Placement on CMOS Delay and Power Dissipation

Xinghai Tang, Keith A. Bowman, John C. Eble*, Vivek. K. De**, and James D. Meindl

MiRC
Georgia Inst. of Tech.
Atlanta, GA 30332-0269

**Compaq Corp.*
Shrewsbury, MA 01545

***MRL.*
Intel Corporation
Hillsboro, OR 79124

Abstract†

A critical path delay distribution due to intrinsic random placement of dopant atoms in static CMOS is investigated by employing compact physical device/circuit models that enable projections for future technology generations. The trade-off between delay, power dissipation, and yield, is discussed. The intrinsic random dopant placement induces 7.5% ~ 15% critical path delay increase for the 50 nm generation depending upon the number of critical paths on a chip. To achieve a desired nominal delay, more than 5% and 9% supply voltage increases, resulting in 11% and 20% power dissipation increases, are projected corresponding to 3σ and 6σ critical path delay variations respectively for the 50 nm generation.

1. Introduction

As MOSFETs scale to the 50 nm range, more severe fluctuations in structural features induce ever larger deviations in key electrical parameters that eventually limit the number of transistors per chip that satisfy a specific performance criterion [1]. The impact of manufacturing induced *macroscopic* extrinsic parameter variations on delay and power dissipation has been studied by analytical models and numerical simulations [2-3]. However, no investigation of the impact of *microscopic* intrinsic random placement of dopant atoms on critical path delay and power dissipation has been reported.

In this paper, the impact of intrinsic random dopant placement on critical path delay is investigated through analytical models based on previously derived effective doping concentration distribution density functions [4]. The impact of random dopant placement is quantified by: 1) lowering the clock frequency or 2) increasing the supply voltage to maintain a desired yield. The trade-off between delay, power dissipation, and yield, is discussed along with projections of the delay/power penalty for the NTRS technology generations [5]. Unless otherwise stated, the input parameters used in this paper are listed in Table I.

2. Analytical Models of the Critical Path Delay Distribution

As MOSFETs continue to scale, the variations of both number and location of dopant atoms in the device active region induce drive current fluctuations. By introducing the "cube approach", viewing a MOSFET as an array of MOS capacitors separating the source from the drain, the effective doping concentration distribution due to this intrinsic fluctuation in the placement of dopant atoms is derived from fundamental device analysis [4]. Utilizing the effective doping concentration distribution, a threshold voltage distribution

† This work is supported by the Semiconductor Research Corporation (Contract: SJ-374) and the Defense Advanced Research Project Agency (Contract: F3361595C1623), and the TCAD tools from TMA Inc.

is derived [4] and compared with experimental data [6] to provide validation of the model calculations, as shown in Fig. 1. It is noteworthy that *no fitting parameter is used in model calculations*.

In high performance static CMOS, the main contribution to the power dissipation, the dynamic power P_d for charging and discharging the load capacitance, is independent of dopant placement fluctuations. Depending on the average value of threshold voltage, the static power P_s due to subthreshold leakage current is a small portion of the total power dissipation. Thus, the intrinsic dopant placement fluctuations do not appreciably affect total power dissipation. However, these fluctuations do change the drive current significantly through mobility and threshold voltage, thus causing delay variations across the chip.

The average delay of a two-input NAND gate T_G may be modeled as,

$$T_G = \left(f_{in} T_{dn}(n_a) + T_{dp}(n_d) \right) / 2 , \quad (1)$$

where f_{in} is effective fan-in factor [7], T_{dn} and T_{dp} are the n-FET and p-FET delays [8], respectively, and n_a and n_d are n-FET and p-FET effective doping concentrations respectively.

The n/p FET delay distribution density function $F_{Tdn/p}$ among a large number of nominal n/p-FETs is derived as

$$F_{T_{dn/p}} = dQ / dT_{dn/p} = f\left(n_{a/d}\right) \times dn_{a/d} / dT_{dn/p} \quad (2)$$

where Q is the probability that a n/p-FET has delay less than a specific value, $f(n_{a/d})$ is the effective doping concentration distribution density function for the n/p-FET [4], and $n_{a/d}$ is the effective doping concentration for n/p-FET [4]. The distribution density function of two-input NAND gates F_{T_G} is calculated by convolving n-FET and p-FET delays through (1) as

$$F_{T_G} = \left(f_{in} \cdot F_{T_{dn}} / 2 \right) * \left(F_{T_{dp}} / 2 \right), \quad (3)$$

since the intrinsic fluctuation in each FET is independent. The critical path delay distribution density function F_{Tcp} with n_G gates is computed by convolving the two-input NAND gate delay distribution density function F_{Td} for n_G gates. Since the delay distribution density function for each gate is identical, the critical path delay distribution density function F_{Tcp} can be calculated as a Gaussian distribution,

$$F_{Tcp} = F_{T_G}^1 * F_{T_G}^2 * \cdots * F_{T_G}^{n_G} = N\left(\mu_{Tcp}, \sigma_{Tcp}\right), \quad (4)$$

with mean $\mu_{Tcp} = n_G \mu_{T_G}$ and standard deviation $\sigma_{Tcp} = \sqrt{n_G} \sigma_{T_G}$, where μ_{T_G} and σ_{T_G} are the mean and standard deviation for the NAND gate delay distribution respectively. Fig. 2 illustrates the critical path delay distribution obtained from (4) for the 50nm and 100nm technology generations in linear and logarithmic scales respectively. The 50nm device has a more pronounced fluctuation than the 100nm generation due to the reduced device dimensions and increased nominal doping concentration [4].

3. Results and Discussions

A chip contains a set of critical paths, all of which must meet the worst case delay constraint [2-3]. Since the intrinsic fluctuation in each critical path is independent, the probability that all N critical paths satisfy the maximum delay T_{cp_max} constraint, the yield, is

$$Yield = \left(\int_0^{T_{cp_\max}} F_{T_{cp}}(t) dt \right)^N , \quad (5)$$

where t is the variable critical path delay due to random placement of dopant atoms. Since F_{Tcp} is described by a Gaussian distribution, as demonstrated in Fig. 2, T_{cp_max} is expressed in terms of σ's as,

$$T_{cp_\max} = T_{cp_\text{nominal}} + n\sigma_{Tcp} , \quad (6)$$

where $T_{cp_nominal}$ is defined as the nominal critical path delay without parameter

deviations and n is the number of standard deviations. To determine n in (6), Fig. 3 illustrates the number of standard deviations (σ) required to achieve a desired yield for a Gaussian distribution versus the number of critical paths in a chip. The table in Fig. 3 defines the number of σ's to the corresponding probability. For today's microprocessors with millions of transistors, a 3σ critical path delay variation may be an acceptable tolerance to maintain a reasonable yield (Fig. 3); however, as the number of transistors per chip continues to increase as projected by the roadmap [5], 6σ critical path delay variation may be required to maintain a reasonable yield.

Two possible options to maintain a reasonable yield within the presence of intrinsic fluctuations are: 1) to operate at a lower clock frequency corresponding to T_{cp_max}, or 2) to increase V_{dd}, resulting in larger power dissipation, that ensures $T_{cp_nominal}$ is satisfied. Thus, a power/delay trade-off can be fully analyzed.

Fig. 4 (a & b) projects the normalized maximum delay ($T_{cp_max}/T_{cp_nominal}$) for NTRS technology generations. For the 50 nm generation 7.5% (15%) of delay increase is projected for 3 (6) σ variations with W_n/L=10 and n_G=10. The intrinsic random fluctuations reduce as the MOSFET aspect ratio *(W/L)* increases. By doubling MOSFET aspect ratios (Fig. 4 a), only 1% (2%) reduction of T_{cp_max}, is achieved for 3 (6) σ variations. As the number of gates in the critical path increases from 10 to 15 (Fig. 4 b), the normalized maximum delay reduces to 1.5%(2.5%) for 3 (6) σ variations, since the ratio of standard deviation to mean for the critical path distribution is inversely proportional to the square root of n_G, as shown in (4).

The second possible option is increasing the supply voltage to compensate for the variations in delay such that $T_{cp_nominal}$ is satisfied for a desired yield. As shown in Fig. 5, for the 50 nm technology generation, more than 5% and 9% increases in V_{dd} are required for 3σ and 6σ critical path delay variations, respectively. The increased V_{dd} results in 11% and 20% more power dissipation for the same technology generation.

4. Conclusion

To maintain greater than 80% yield, the intrinsic random dopant fluctuation induces 7.5% ~ 15% critical path delay increases for the 50 nm generation depending upon the number critical paths on a chip. To achieve a nominal delay at more than 80% yield, more than 5% and 9% V_{dd} increase is required that results in 11% and 20% power dissipation increase for the 50 nm technology generation corresponding to 3σ and 6σ critical path delay variations respectively.

References:

[1] J. D. Meindl, "Low Power Microelectronics: Retrospect and Prospect," *Proc. IEEE* Vol. 83, pp. 619-635, Apr. 1995.

[2] D. J. Frank, P. Solomon, S. Reynolds, and J. Shin, "Supply and Threshold Voltage Optimization for Low Power Design," *ISLPED Dig. Tech. Papers*, pp. 317-322, 1997.

[3] M. Eisele, J. Berthold, D. Schmitt-Landsiedel and R. Mahnkopf, "The Impact of Intra-Die Device Parameter Variations on Path Delays and on the Design for Yield of Low Voltage Digital Circuits," *ISLPED Dig. Tech. Papers*, pp. 237-242, 1996.

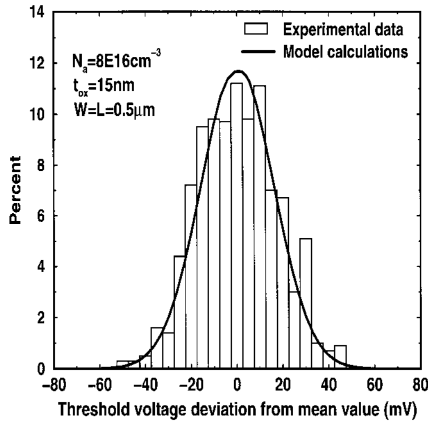
[4] X. Tang, V. K. De and J. D. Meindl, "MOSFET Fluctuation Limits on Gigascale Integrations (GSI)", *ESSDERC'98*, pp 508-511, Sep. 1998.

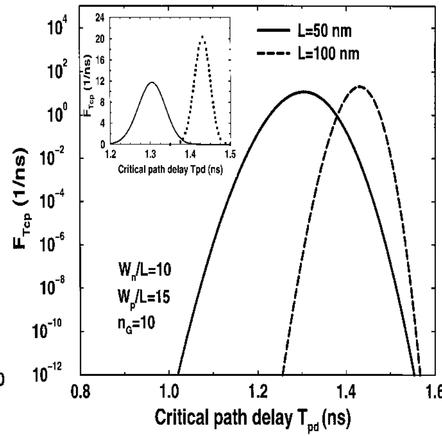
[5] SIA, Technology Roadmap for Semiconductors, 1997.

[6] D. Burnett, K. Erington, C. Subramanian, and K. Baker, "Implications of Fundamental Threshold Voltage Variations for High-Density SRAM and Logic circuits," *Symp. VLSI Tech.*, pp. 15-16, Jun. 1994.

[7] S. Sakurai and R. Newton, "Delay Analysis for Series-Connected MOSFET Circuits," *IEEE, JSSC*, vol. 26, No. 2, pp.122 – 131, Feb. 1991

[8] B. Austin, K. Bowman, X. Tang, and J. D. Meindl, "A Low Power Transregional MOSFET Model for Complete Power-Delay Analysis of CMOS Gigascale Integration (GSI),"*Proc. of the 11th Annual IEEE Intl. ASIC Conf.*, pp125-129, Sep. 1998.





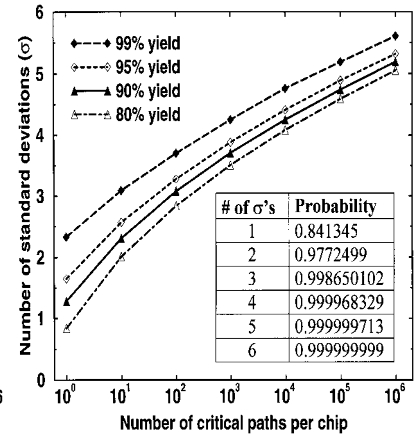
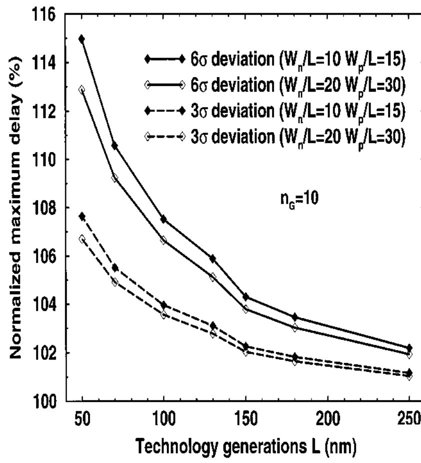
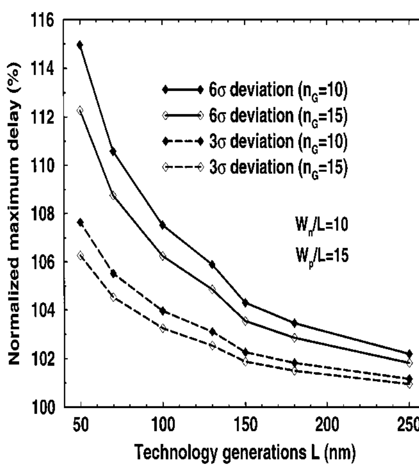


Fig. 1 Comparison of compact model calculated V_{ts} distribution with experimental data from [6].

Fig. 2 Distribution of critical path delay for L=50 and 100 nm respectively

Fig. 3. Number of σ's required to achieve 80% to 99% yield vs number of critical paths on a chip.





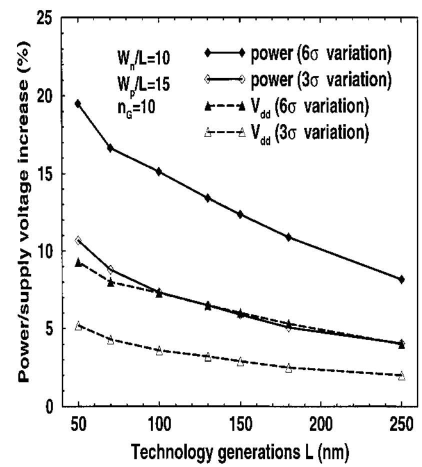


Fig. 4. Normalized maximum delay dependence on (a) MOSFET aspect ratio, and (b) number of gates in critical path projected for NTRS technology generations.

Fig. 5 Power and V_{dd} increases due to the random dopant placement. The nominal values are listed in Table I.

Table I input parameters for Fig. 2, 4, and 5. * and ** are for W_n/L=10 and 20 respectively.

Year	**1997**	**1999**	**2001**	**2003**	**2006**	**2009**	**2012**
L (nm)	250	180	150	130	100	70	50
T_{ox} (A)	45	35	28	23	17	12	8
V_{dd} (V)	2.2	1.8	1.5	1.2	1.0	0.8	0.6
V_{th} (V)	0.45	0.40	0.35	0.33	0.30	0.28	0.25
N_a ($\times 10^{18}$ cm^{-3})	0.60	0.80	0.95	1.22	1.80	2.90	4.85
C_L^* (fF)	105	95	90	80	65	55	50
C_L^{**} (fF)	157.5	142.5	135	120	97.5	82.5	75
Power/gate (μW)*	12.13	7.61	5.42	3.11	2.07	1.22	0.63