Estimation of Transient Voltage Fluctuations in the CMOS-Based Power Distribution Networks

Kevin T. Tang and Eby G. Friedman
Department of Electrical and Computer Engineering
University of Rochester
Rochester, New York 14627-0231

Abstract— **Decreased power supply levels have reduced the tolerance to voltage changes within power distribution networks in CMOS integrated circuits. High on-chip currents, required to charge and discharge large on-chip loads while operating at high frequencies, produce significant transient *IR* voltage drops within a power distribution network. These transient *IR* voltage drops can affect the propagation delay of a CMOS logic gate, creating delay uncertainty within data paths. Analytical expressions characterizing these transient *IR* voltage drops are presented in this paper. Circuit- and layout-level design constraints are also discussed to manage the peak value of the transient *IR* voltage drops.**

I. Introduction

As modern VLSI technology moves into the very deep submicrometer (VDSM) regime, millions of transistors will be integrated onto a single chip, operating at frequencies greater than a gigahertz. The die size is expected to increase from 385 mm^2 in 2001 to 620 mm^2 by 2009 while average on-chip currents will increase from 70 amperes in 2001 to 190 amperes by 2009 [1]. Power distribution networks in high complexity CMOS integrated circuits must be able to provide sufficient current to support average and peak power demand within all parts of an integrated circuit [2], [3], [4]. The large chip dimensions and average currents require special design strategies to maintain a constant voltage supply within a power distribution network [5], [6].

The voltage supply is expected to decrease from 1.5 volts in 2001 to 0.9 volts by 2009 [1], reducing the tolerance to voltage changes within a power distribution network. Because of the lossy characteristics of the metal interconnections in CMOS integrated circuits, *IR* voltage drops within a mesh power distribution structure are no longer negligible [7], [8]. For example, metal 4 in the Alpha 21164 provides the power supply for each element within the entire integrated circuit [9]. The thickness of metal 4 is 1.53 μm and the pitch is 6.0 μm [3]. The average on-chip current is about 15 amperes. The current density is approximately 1.2 $mA/\mu m^2$ and the current is about 5.5 mA for a 3.0 μm wide line. For a 9.0 mm long aluminum power line with a resistivity of 4.0 $\mu\Omega$-cm, a parasitic line resistance of 59 Ω results in an average *IR* voltage drop of 0.33 volts, which is about 10% of the voltage supply (3.3 volts for the Alpha 21164). Transient *IR* voltage drops which occur during logic transitions in a synchronous CMOS integrated circuit are even greater than these average *IR* voltage drops.

Therefore, significant transient *IR* voltage drops can occur in a synchronous CMOS integrated circuit [10]. These *IR* voltage drops can create delay uncertainty within data paths due to momentary changes in the power supply voltage, making the maximum and minimum propagation delays difficult to estimate, such as needed in clock skew scheduling in the design of high performance clock distribution networks [11]. Additional power planes are an effective design technique to reduce transient *IR* voltage drops, decreasing the parasitic resistance associated with a power distribution network.

An analysis of transient *IR* voltage drops is presented in this paper. The MOS transistors are characterized by the nth power law I-V model [12]. Analytical expressions describing these transient *IR* voltage drops are developed based on an assumption of a fast ramp input signal. The peak *IR* voltage drops are shown to occur when the input signal completes a transition. The peak value of the transient *IR* voltage drops based on the analytical expression is within 6% as compared to SPICE. Circuit- and layout-level design constraints are also addressed to manage the maximum *IR* voltage drops. Analytical expressions characterizing the output voltage and propagation delay of a CMOS logic gate are presented for a capacitive load. A propagation delay model based on these analytical expressions is within 5% as compared to SPICE while an estimate without considering transient *IR* voltage drops can reach 20% for a 20 Ω power line.

Analytical expressions characterizing transient *IR* voltage drops are developed in Section II. A comparison of the analytical result with SPICE and discussions of circuit- and layout-level constraints are presented in Section III. The effects of transient *IR* voltage drops on the output voltage and propagation delay of a CMOS logic gate are addressed for a capacitive load in Section IV followed by some concluding remarks in Section V.

II. Modeling of Transient *IR* Voltage Drops

Transient *IR* voltage drops are caused by a large number of logic gates switching close to the same time in a synchronous integrated circuit. For a switching CMOS logic gate, the current through the power lines is assumed to be m times greater than the current through a single CMOS logic gate. This assumption is equivalent to m simultaneously triggered logic gates connected to the same power line.

An analytical expression characterizing the transient *IR* voltage drops on the ground rail is developed in this section for a high-to-low output transition. $R_{V_{ss}}$ is the parasitic resistance of the ground rail. In order to derive an analytical expression characterizing the transient *IR* voltage drops on the ground rail, the short-circuit current is neglected based on an assumption of a fast ramp input signal [13] with a transition time of τ_r, permitting the current through the PMOS transistor to be neglected.

Once the input voltage exceeds the threshold voltage

This research was supported in part by National Science Foundation under Grant No. MIP-9610108, the Semiconductor Research Corporation under Contract No. 99-TJ-687, a grant from the New York State Science and Technology Foundation to the Center for Advanced Technology–Electronic Imaging Systems, and by grants from Xerox Corporation, IBM Corporation, Intel Corporation, Lucent Technologies Corporation, and Eastman Kodak Company.

0-7803-6685-9/01/$10.00©2001 IEEE

of the NMOS transistor, the NMOS transistor turns ON and is assumed to operate solely in the saturation region during the input transition. The drain-to-source current in this region is

$$I_{DS} = B_n(V_{in} - V_{TN} - mR_{V_{ss}}I_{DS})^{n_n}. \quad (1)$$

Assuming that $mR_{V_{ss}}I_{DS}$ is less than $V_{in} - V_{TN}$, the drain-to-source current can be approximated as

$$I_{DS} = \frac{B_n(V_{in} - V_{TN})^{n_n}}{1 + mR_{V_{ss}}n_nB_n(V_{in} - V_{TN})^{(n_n-1)}}. \quad (2)$$

Therefore, the transient *IR* voltage drops for $\tau_n \leq t \leq \tau_r$ are

$$V_{IR} = mR_{V_{ss}}\frac{B_n(\frac{t}{\tau_r}V_{dd} - V_{TN})^{n_n}}{1 + mR_{V_{ss}}n_nB_n(\frac{t}{\tau_r}V_{dd} - V_{TN})^{(n_n-1)}}, \quad (3)$$

where $\tau_n = \frac{V_{TN}}{V_{dd}}\tau_r$. The transient *IR* voltage drops reach the maximum value at $t = \tau_r$,

$$V_{IR,max} = mR_{V_{ss}}\frac{B_n(V_{dd} - V_{TN})^{n_n}}{1 + mR_{V_{ss}}B_n(V_{dd} - V_{TN})^{n_n-1}}. \quad (4)$$

The NMOS transistor is assumed to remain saturated when the input transition is completed. The drain-to-source current is a constant, independent of the output voltage. The transient *IR* voltage drops in this region are the same as $V_{IR,max}$.

After τ_{sat}, the NMOS transistor operates in the linear region. In order to derive a tractable expression, the drain-to-source current is characterized by γ_nV_{DS}, where γ_n is the effective output conductance of the NMOS transistor. The transient *IR* voltage drops in this region can be characterized as

$$V_{IR} = V_{IR,max}e^{-\alpha(t-\tau_{sat})}, \quad (5)$$

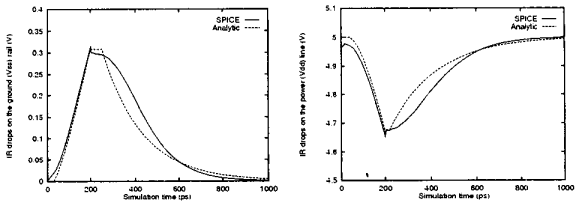
where $\alpha = \frac{\gamma_n}{C_L(1+mR_{V_{ss}}\gamma_n)}$ and C_L is the load capacitance. However, the effective output conductance of a MOS transistor also depends upon the output voltage in the linear region, changing from γ_{nsat} to $2\gamma_{nsat}$ [12]. In order to accurately characterize the transient *IR* voltage drops in the linear region, a value of γ_n is chosen between γ_{nsat} and $2\gamma_{nsat}$.

III. Characteristics of Transient *IR* Voltage Drops

The waveform and peak value of the transient *IR* voltage drops based on (4) are compared to SPICE in Section III-A. Circuit- and layout-level design constraints to manage the peak value of the transient *IR* voltage drops are discussed in Sections III-B and III-C, respectively.

A. Comparison with SPICE

A comparison of the analytical expression characterizing the waveform of the transient *IR* voltage drops with SPICE is shown in Fig. 1 for both the V_{ss} and V_{dd} rails. The transient *IR* voltage drops along the V_{ss} rail increase the potential on the V_{ss} rail, while the transient *IR* voltage drops along the V_{dd} rail decrease the potential on the V_{dd} rail, as shown in Fig. 1. Thus, the overall voltage swing is decreased, degrading system speed. Transient *IR* voltage drops on the power supply rails increase the effective gate voltage required to turn on the MOS transistors ($V_{GS} = V_{TN} + V_{IR}$ for the NMOS transistor). Note that the analytical prediction is quite close to SPICE. The difference is caused by the effective output conductance of the MOS transistors changing from γ_{sat} to $2\gamma_{sat}$ in the linear region [12].



(a) $R_{V_{ss}} = 20.0\,\Omega$, $C_L = 0.1$ pF, and $\tau_r = 0.20$ ns

(b) $R_{V_{dd}} = 20.0\,\Omega$, $C_L = 0.1$ pF, and $\tau_f = 0.20$ ns

Fig. 1. Comparison of the analytical waveform of transient *IR* voltage drops with SPICE for $w_n = 1.8\,\mu$m, $w_p = 3.6\,\mu$m, and $m = 10$.

The results of comparing the peak value of the transient *IR* voltage drops with SPICE are listed in Table I for a high-to-low output transition with $w_n = 1.8\,\mu$m, $w_p = 3.6\,\mu$m, and $C_L = 0.1$ pF. The peak value of the transient *IR* voltage drops based on (4) is within 6% as compared to SPICE.

TABLE I
COMPARISON OF PEAK *IR* VOLTAGE DROPS ON THE V_{ss} RAIL WITH SPICE

$R_{V_{ss}}$ (Ω)	τ_r (ps)	m	Analytic (V)	SPICE (V)	δ (%)
40.0	100	20	0.971	0.968	0.3
		15	0.786	0.800	1.8
		10	0.569	0.595	4.4
	150	20	0.971	0.945	2.8
		15	0.785	0.780	0.6
		10	0.568	0.578	1.7
	200	20	0.971	0.931	4.3
		15	0.785	0.767	1.0
		10	0.568	0.569	0.2
30.0	100	20	0.785	0.786	0.1
		15	0.626	0.641	2.3
		10	0.445	0.468	4.9
	150	20	0.785	0.766	2.5
		15	0.626	0.624	0.3
		10	0.445	0.455	2.2
	200	20	0.785	0.754	4.1
		15	0.626	0.614	2.0
		10	0.445	0.447	0.4
20.0	100	20	0.568	0.571	0.5
		15	0.445	0.458	2.8
		10	0.311	0.329	5.5
	150	20	0.568	0.556	2.2
		15	0.445	0.445	0.0
		10	0.311	0.319	2.5
	200	20	0.568	0.546	4.0
		15	0.445	0.439	1.4
		10	0.311	0.313	0.6
Maximum error					5.5
Average error					2.0

B. Circuit-level constraints

Assuming the maximum *IR* voltage drops should be less than a critical voltage V_c, the product of m and $R_{V_{ss}}$ must satisfy

$$mR_{V_{ss}} \leq \frac{V_c}{B_n(V_{dd}-V_{TN})^{n_n} - V_c B_n (V_{dd}-V_{TN})^{n_n-1}}. \quad (6)$$

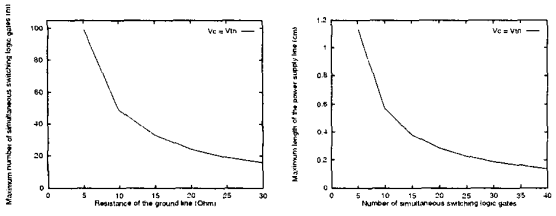
The constraint defined in (6) demonstrates that the product of m and $R_{V_{ss}}$ should be less than a constant determined by the right hand side of (6). Therefore, the maximum parasitic resistance of a power rail can be determined for a fixed m; the maximum number of simultaneously triggered logic gates can also be determined for a target power rail resistance of $R_{V_{ss}}$ as shown in Fig. 2(a).

C. Layout-level constraints

For a metal interconnection, the parasitic resistance can be expressed as

$$R = \rho \frac{l}{wt}, \quad (7)$$

where ρ is the resistivity of the material, and l, w, and t are the length, width, and thickness of the metal line, respectively. In practical CMOS integrated circuits, the current density must be less than a limit set by the electromigration constraint [8]. Therefore, for a metal interconnection with a fixed thickness, the minimum width and maximum length of the metal line can be determined by combining (6) and (7). The maximum length of the power supply rail with $w = 3.0\,\mu$m, $t = 1.53\,\mu$m, and $\rho = 4.0\,\mu\Omega$-cm is shown in Fig. 2(b).



(a) Maximum number of simultaneously switching logic gates versus the ground line resistance.

(b) Maximum length of the power supply rail versus the number of simultaneously switching logic gates with $w = 3.0\,\mu$m, $t = 1.53\,\mu$m, and $\rho = 4.0\,\mu\Omega$-cm.

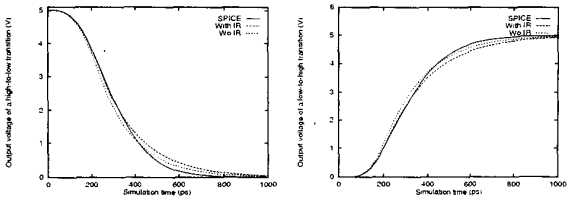
Fig. 2. Maximum number of simultaneously switching logic gates and maximum length of the power supply rail for $V_c = V_{TN}$.

Both (6) and (7) provide design guidelines for managing the transient *IR* voltage drops within a power distribution network. The use of additional power planes is an effective design technique to reduce the peak value of the transient *IR* voltage drops, significantly reducing the parasitic resistance associated with a power distribution network.

IV. Output Voltage and Propagation Delay

The effect of transient *IR* voltage drops on the output voltage and propagation delay of a CMOS logic gate is discussed in this section. Analytical expressions characterizing the output voltage waveform are developed for a capacitive load. The propagation delay of a CMOS logic gate based on these analytical expressions is also compared with SPICE.

Analytical expressions characterizing the output voltage of a CMOS logic gate driving a capacitive load are listed in Table II based on an assumption of a fast ramp input signal. τ_{sat} is the time when the NMOS transistor starts to operate in the linear region and is determined from (9). The analytical results are compared to both SPICE and the analytical prediction without considering *IR* voltage drops in Fig. 3. Note that transient *IR* voltage drops affect the propagation delay of a CMOS logic gate. Therefore, delay uncertainty caused by transient *IR* voltage drops should be included when analyzing the timing of critical data paths [11].



(a) $R_{V_{ss}} = 20.0\,\Omega$, $C_L = 0.1$ pF, and $\tau_r = 0.20$ ns

(b) $R_{V_{dd}} = 20.0\,\Omega$, $C_L = 0.1$ pF, and $\tau_r = 0.20$ ns

Fig. 3. Comparison of the analytical output voltage with SPICE. $w_n = 1.8\,\mu$m, $w_p = 3.6\,\mu$m, and $m = 10$.

The propagation delay of a CMOS logic gate driving a capacitive load can be approximated by (9) or (10). The drain-to-source saturation voltage is typically greater than $0.5V_{dd}$, therefore, the high-to-low propagation delay of a CMOS logic gate can be expressed as

$$\tau_{P_{HL}} = \frac{C_L(1+mR_{V_{ss}}\gamma_n)}{\gamma_n} ln \frac{2(V_{sat}+V_{IR,max})}{V_{dd}} + \tau_{sat} - \frac{\tau_r}{2}. \quad (11)$$

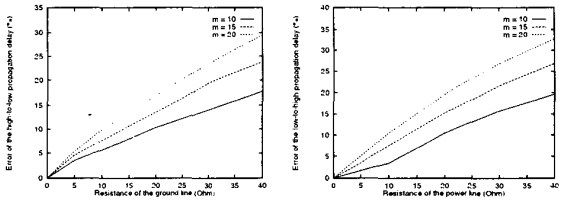
The error of the propagation delay model of a CMOS logic gate without considering transient *IR* voltage drops is illustrated in Fig. 4. A comparison of the propagation delay based on (11) with both SPICE and an estimate without considering transient *IR* voltage drops is listed in Table III for a high-to-low output transition. Note that the maximum error of the proposed delay model is within 3% as compared to 20% when transient *IR* voltage drops are not considered (for a $20\,\Omega$ power line which is equivalent to a resistance of a 0.23 cm long power line with $w = 3.0\,\mu$m, $t = 1.53\,\mu$m, and $\rho = 4.0\,\mu\Omega$-cm).

V. Conclusions

An analytical model and design constraint expressions characterizing transient *IR* voltage drops are presented in this paper. The peak *IR* voltage drops occur when the input signal completes a transition (for a fast ramp input signal). The peak value of the transient *IR* voltage drops based on the proposed analytical expression is within 6%

TABLE II
ANALYTICAL EXPRESSIONS CHARACTERIZING THE OUTPUT VOLTAGE WITH *IR* VOLTAGE DROPS FOR A CAPACITIVE LOAD

Operating region	Analytical expressions
$[\tau_n, \tau_r]$	$V_o = V_{dd} - \frac{\tau_r B_n}{C_L(n_n+1)V_{dd}}(\frac{t}{\tau_r}V_{dd} - V_{TN})^{(n_n+1)} + \frac{mR_{V_{ss}}B_n^2\tau_r}{2C_L V_{dd}}(\frac{t}{\tau_r}V_{dd} - V_{TN})^{2n_n}$ (8)
$[\tau_r, \tau_{sat}]$	$V_o = V_o(\tau_r) - \frac{B_n(V_{dd}-V_{TN})^{n_n}}{C_L(1+mR_{V_{ss}}nB_n(V_{dd}-V_{TN})^{(n_n-1)})}(t-\tau_r)$ (9)
$t \geq \tau_{sat}$	$V_o = (V_{sat} + V_{IR,max})e^{-\frac{\gamma_n}{C_L(1+mR_{V_{ss}}\gamma_n)}(t-\tau_{sat})}$ (10)



(a) Error of high-to-low propagation delay

(b) Error of low-to-high propagation delay

Fig. 4. Error of the propagation delay model of a CMOS logic gate without considering transient *IR* voltage drops with $\tau_r = 150$ ps, $w_n = 1.8\ \mu$m, and $w_p = 3.6\ \mu$m.

as compared to SPICE. Circuit- and layout-level design constraints are also addressed to manage the maximum value of the transient *IR* voltage drops, providing guidelines for the design of power distribution networks.

TABLE III
HIGH-TO-LOW PROPAGATION DELAY WITH *IR* VOLTAGE DROPS FOR A CAPACITIVE LOAD

Condition			SPICE	Analytic			
$R_{V_{ss}}$ (Ω)	τ_r (ps)	m	(ps)	Wo IR (ps)	δ (%)	Wi IR (ps)	δ (%)
40.0	100	20	228	160	29.8	231	1.3
		15	212	160	24.5	209	1.4
		10	196	160	18.4	191	2.6
	150	20	235	166	29.4	239	1.7
		15	218	166	23.9	216	0.9
		10	202	166	17.8	198	2.0
	200	20	240	172	28.3	248	3.3
		15	225	172	23.6	223	0.9
		10	208	172	17.3	203	2.4
30.0	100	20	212	160	24.5	209	1.4
		15	200	160	20.0	196	2.0
		10	187	160	14.4	182	2.7
	150	20	217	166	23.5	216	0.5
		15	206	166	19.4	202	2.0
		10	193	166	14.0	189	2.1
	200	20	223	172	22.9	223	0.0
		15	211	172	18.5	208	1.4
		10	199	172	13.6	196	1.5
20.0	100	20	194	160	17.5	190	2.1
		15	187	160	14.4	183	2.1
		10	178	160	10.1	175	1.7
	150	20	200	166	17.0	197	1.5
		15	192	166	13.5	189	1.6
		10	185	166	10.3	181	2.2
	200	20	206	172	16.5	203	1.5
		15	200	172	14.0	196	2.0
		10	190	172	10.5	187	1.6
Maximum error				29.8		3.3	
Average error				20.9		1.0	

REFERENCES

[1] Semiconductor Industry Association, "The National Technology Roadmap for Semiconductors," 1997.

[2] D. W. Dobberpuhl *et al.*, "A 200-MHz 64-bit Dual-Issue CMOS Microprocessor," *IEEE Journal of Solid-State Circuits*, Vol. SC-27, No. 11, pp. 1555–1565, November 1992.

[3] W. J. Bowhill *et al.*, "Circuit Implementation of a 300-MHz 64-bit Second-Generation CMOS Alpha CPU," *Digital Technical Journal*, Vol. 7, No. 1, pp. 100–118, 1995.

[4] D. W. Bailey and B. J. Benschneider, "Clocking Design and Analysis for a 600-MHz Alpha Microprocessor," *IEEE Journal of Solid-State Circuits*, Vol. SC-33, No. 11, pp. 1627–1633, November 1998.

[5] SRC, "SRC Physical Design Top Ten Problem," Technical Report, SRC Physical Design Task Force, Semiconductor Research Corporation, November 1998.

[6] SRC, "Design Sciences TAB: Physical Design Task Force Report," Technical Report, Semiconductor Research Corporation Physical Design Task Force, Semiconductor Research Corporation, April 1997.

[7] Y.-M. Jiang and K.-T. Cheng, "Analysis of Performance Impact Caused by Power Supply Noise in Deep Submicron Devices," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 760–765, June 1999.

[8] W. S. Song and L. A. Glasser, "Power Distribution Techniques for VLSI Circuits," *IEEE Journal of Solid-State Circuits*, Vol. SC-21, No. 1, pp. 150–156, February 1986.

[9] P. E. Gronowski *et al.*, "High-Performance Microprocessor Design," *IEEE Journal of Solid-State Circuits*, Vol. SC-33, No. 5, pp. 676–686, May 1998.

[10] K. T. Tang, *On-Chip Interconnect Noise in High Performance CMOS Integrated Circuits*. PhD Thesis, University of Rochester, July 2000.

[11] E. G. Friedman, *High Performance Clock Distribution Networks*. Kluwer Academic Publishers, 1997.

[12] T. Sakurai and A. R. Newton, "A Simple MOSFET Model for Circuit Analysis," *IEEE Transactions on Electron Devices*, Vol. ED-38, No. 4, pp. 887–894, April 1991.

[13] N. Hedenstierna and K. O. Jeppson, "CMOS Circuits Speed and Buffer Optimization," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. CAD-6, No. 2, pp. 270–280, March 1987.