

IR Drop Analysis and Its Reduction Techniques in Deep Submicron Technology

Vanpreet Kaur

Department of Electronics Engineering, Banasthali University, Rajasthan, India

Abstract

This paper presents a detailed conceptual analysis of IR Drop effect in deep submicron technologies and its reduction techniques. The IR Drop effect in power/ground network increases rapidly with technology scaling. This affects the timing of the design and hence the desired speed. It is shown that in present day designs, using well known reduction techniques such as wire sizing and decoupling capacitor insertion, may not be sufficient to limit the voltage fluctuations and hence, two more important methods such as selective glitch reduction technique and IR Drop reduction through combinational circuit partitioning are discussed and the issues related to all the techniques are revised.

Keywords— Decoupling Capacitance, Dynamic Power, Glitch Power, IR Drop, Switching Activity

I. INTRODUCTION

IR drop is becoming an extremely important phenomenon determining the performance and reliability of VLSI designs. The IR Drop manifests itself in power/ground network of VLSI designs. The IR Drop manifests itself in power/ground distribution networks and can adversely affect the performance of signal nets including clock nets. When interconnect scaling is done aggressively, it reduces the width of interconnects, and therefore increases the resistance per unit length of the wires. This increase in resistance leads to increased IR Drop across the wires and consequently actual voltage supplied to the cells in the design is less than the required voltage. This influence the speed and functionality of the design. Since the supply voltage is also reduced with the technology scaling, the IR Drop effect becomes even more problematic with technology scaling. Since the ratio of the voltage drop to the ideal supply voltage increases, it degrades the switching speed of the CMOS gates and their DC Noise margins. An excessive voltage drop in power grid network may also result in functional failure in dynamic logic and timing violation in static logic. It has been shown that 10% voltage drop in 0.18 μ m design increases the propagation delay up to 8%.

As a result, the main challenge in the design of power distribution is to achieve a minimum acceptable voltage fluctuations across the chip while satisfying the electromigration(EM) reliability rule for power network segments and to realize such a power distribution network which results in minimum routing area of the interconnect metal layers. With technology shrinking which result in more performance to be packed in a smaller area, the problem has become worst. The power

distribution network on a chip needs to ensure circuit robustness to the average power or current requirement, as well as timing and reliability of the chip should also not be affected by the dynamic IR Drop which is caused by localized switching patterns. So, there is a need to have more advancement in IR Drop reduction techniques.

This paper discuss the static and dynamic IR Drop, effects of technology scaling, including the thermal effects, and Barrier and thin film effects on the worst case IR Drop. It is also explained how technology scaling affects the performance of the cell switching activities due to voltage drops. Finally, various IR Drop reduction techniques and design issues regarding all techniques are discussed.

The structure of paper is as follows. In section 2, Static and Dynamic IR Drop are revised which is followed by the effects of technology scaling, including the thermal effects and barrier and thin film effects, on the worst case IR Drop and effect of IR Drop on timing. Section 3 examines how the technology scaling affects the performance of cell switching activities due to the power network voltage drops. In section 4, various IR Drop reduction techniques are discussed. Finally, concluding remarks and summary are presented in section 5.

II. EFFECTS OF TECHNOLOGY SCALING ON IR DROP

2.1 OVERVIEW OF STATIC AND DYNAMIC IR DROP

Static IR Drop basically defines the average IR Drop for the design, whereas Dynamic IR Drop depends on the actual switching activity of the logic. Hence, it is vector dependent. Dynamic IR Drop depends on the switching time of the logic and is less dependent on the clock period. This nature is illustrated in figure 1. The average current depends

on the clock time period, while the Dynamic IR Drop depends on the instantaneous current which is higher when the cell is switching.

In older technologies, Static IR drop was good for sign off analysis where sufficient decoupling capacitance from the power/ground network and non switching logic were available. On the other hand, Dynamic IR drop evaluates the IR Drop caused when large amounts of circuitry switch simultaneously, causing peak current demand. The localized switching in the design may cause this high current demand within a single clock cycle (a few hundred ps) and could result in an IR Drop that causes additional setup and hold timing violations

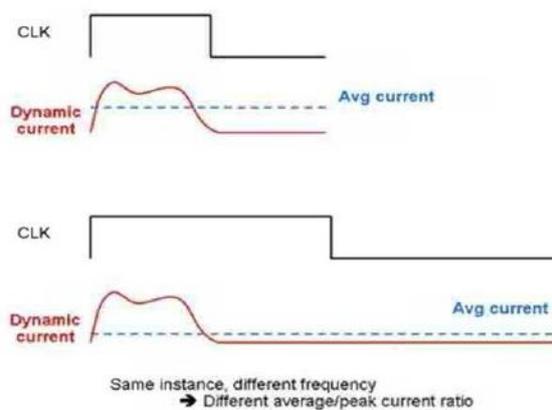


Fig. 1 Average current over a window

2.2 EFFECTS OF THIN FILM, BARRIER THICKNESS AND INTERCONNECT TEMPERATURE

As technology shrinking is in great progress, interconnects are being also scaled simultaneously. In VLSI interconnects, metal resistivity increases as scaling increases. This is because as scaling increases, the minimum dimension of the metal line becomes comparable to the mean free path of electrons due to which surface scattering starts playing a dominant role to the resistivity as compared to the contribution due to bulk scattering. Moreover, since the temperature alters the mean free path of the electrons, the temperature coefficient of resistivity α , of the thin film is also different from its bulk temperature coefficient α_0 .

As technology is scaling day by day, problems with the interconnect material is increasing. Therefore, changes in the interconnect materials are needed. So, a low resistive material such as Copper was introduced. But, there exist a problem with this material i.e. when these are diffused in the silicon devices; they degrade the performance of corresponding silicon device by introducing deep level acceptors. Hence, as a remedy for this problem, dielectrics are used but, typical dielectric materials are not effective barriers for the Copper. Copper also has poor adhesion to typical dielectric materials.

Therefore, Cu metallization requires a base layer which acts as an adhesion promoter and a diffusion barrier. The microstructure and surface condition of the barrier can strongly affect the texture and grain size of overlying Cu film, which are critical factors that determines the electromigration reliability of Cu interconnects. Hence, the presence of this barrier material for Cu interconnects increases the resistivity of the metal line. Since the resistivity of barrier material is very high as compared to the Cu, it can be assumed that Cu carries all the current. Therefore, the effective area through which the current conduction takes place reduces, or equivalently, for the same dimension, the effective resistivity increases. It is well known that interconnect resistance increases linearly with its temperature. This relationship can be expressed as:

$$R = r_0 (1 + \beta \Delta T)$$

where, r_0 is the unit length resistance at reference temperature and β is the temperature coefficient of resistance. By including the effects of scattering and thin film, this can be written as-

$$R = r_0 (\rho / \rho_0)_{\text{thin_barrier_eff}} [1 + \beta (\alpha / \alpha_0) \Delta T]$$

In order to reduce the maximum voltage drop, the global and semi global tier metals usually have widths that are many times larger than minimum width. As discussed earlier, the presence of barrier material for Cu interconnects increases the resistivity of metal line and local lines are responsible for the interconnection among silicon devices, hence these lines have contact with the silicon devices and therefore, the barrier thickness effect on resistivity can only be considered for the local lines. Hence, these lines have higher resistance than global/semi global metal lines. Therefore, global/semi global lines are the hottest lines inside a chip and effect of line temperature should be considered.

2.3 EFFECTS OF IR DROP ON TIMING

With the advent of technology scaling, in deep submicron technology, complexity in designing the high speed clock networks is increasing. As scaling increases i.e. line width decreases, the number of devices continues to grow. The difficulty in designing the chip is not only because of the increase in device count but also due to the fact that interconnects effects are dominating in determining the performance of the chip. Parasitic effects such as interconnect resistance and 3D capacitance have greatly increased the design complexity and made clock design a major challenge.

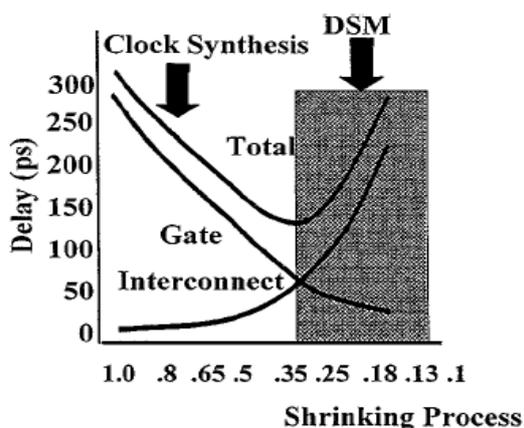


Fig. 2 Interconnect delay dominates gate delay in DSM

Figure 2 illustrates the delay effects observed in DSM. As shown in the figure, gate delay is decreasing while interconnect delay is increasing with a crossover at roughly $0.35\mu\text{m} - 0.25\mu\text{m}$. This is because, as line width decreases, resistance and capacitance increases. Hence, this introduces a significant Resistance-capacitance delay (RC delay) component. Coupling capacitance arises due to 3D effects of tall, thin metal lines that are closely spaced together. Usually, the curve applies to long nets in the design and clock nets tend to be in this category.

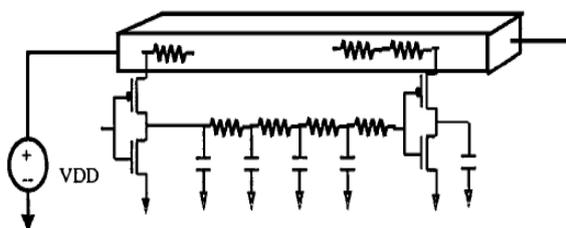


Fig. 3 cause of IR drop effects

Another key issue adding to the complexity of clock design is the IR Drop in the power supply. Figure 3 shows that IR Drop is basically due to the presence of resistance in the power distribution system. Earlier, resistance in the power grid could be safely ignored. But, in DSM, the narrow metal lines offers resistance which cannot be ignored in today's power distribution system. Usually, wider lines are used in power busses to reduce resistance, but the speed of today's clock designs requires very large buffers which draw large currents. When these currents flow from supply to the drivers, any resistance encountered in the power bus causes the voltage to drop. Therefore, the far-end inverter in figure 3 experiences a lower supply voltage than the first inverter. Since IR drop effectively reduces the supply voltage, the buffers connected to the power grid become weaker and the propagation delay

through the clock increases. Hence, this affects the speed and timing of the design

III. EFFECT OF IR DROP ON CELL PERFORMANCE AND CLOCK SKEW

In standard-cell based designs, power is distributed by putting a power trunk adjacent to a cell row. Performance of each cell connected to the local trunk segment is strongly dependent on the fluctuations of power supply (V_{dd}). To derive the sensitivity of gate delay with respect to the changes in V_{dd} , consider the delay associated with the inverter in figure 4.

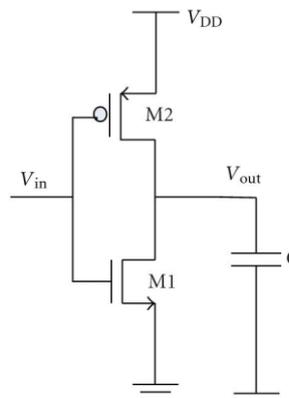


Fig. 4 CMOS inverter schematic

We can start with a simplified expression for the delay of the gate switching from low-to-high and is given by

$$D = C V_{dd} / I_{DS}$$

Where, C is the total capacitance at the output, I_{DS} is the p-channel charging current. Assuming short channel devices, it is expected that the transistor would be velocity saturated for most of the transition. A simple model for the transistor in the saturation region is given by

$$I_{DS} = W v_{sat} C_{ox} (V_{GS} - V_T - V_{DS})$$

Where, W is the channel width, v_{sat} is the carrier saturation velocity, C_{ox} is the oxide capacitance, V_{GS} is the gate-to-source voltage, V_{DS} is the drain-to-source voltage and V_T is the device threshold voltage. The gate delay sensitivity $S^D_{V_{dd}}$ with respect to the power supply V_{dd} can be expressed as follows-

$$S^D_{V_{dd}} = \frac{V_{dd} V_T - V_T^2 + E_C L V_{dd} + E_C L V_T}{(V_{dd} - V_T + E_C L)(V_{dd} - V_T)}$$

where, E_C is the critical electric field, L is the channel length ($E_C L = 1.4V$), and V_T is assumed to be $V_{dd} / 5$. Hence, with the technology scaling, the dependency of gate delay on power supply voltage fluctuations becomes more severe. Figure 5 shows the sensitivity of delay with respect to power supply variations as a function of technology node. As for example, it can be expected an 8.5% increase in gate

delay for each 10% decrease in power supply voltage in 0.18µm technology

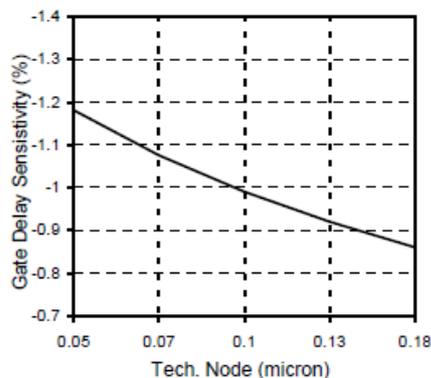


Fig. 5 Sensitivity of cell delay ($S_{V_{dd}}^D$) to the fluctuations of the supply voltage V_{dd} for different technology nodes

IV. VARIOUS REDUCTION TECHNIQUES

4.1 DECAP INSERTION

The most common method of IR drop reduction is by inserting decoupling capacitors (decaps). Decaps hold a reservoir of charge and are placed around the regions where the demand of current is high. This high demand may be due to high switching activity of the cells. Hence, when large drivers switch, decaps provide a source of current that reduces IR and LdI/dt voltage drops so as to keep the target average and peak voltages within their noise budgets. There are 2 types of decaps available to reduce IR drop. First is white-space decaps which usually consist of NMOS transistors, and are placed between blocks in the open areas of the chip. Second is standard-cell decaps which use both NMOS and PMOS devices and are placed within the logic blocks themselves.

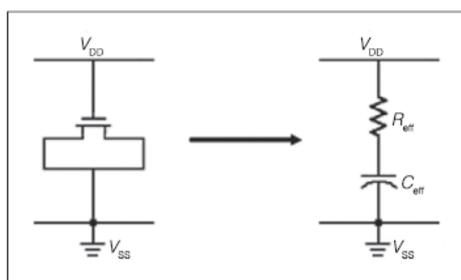


Fig. 6 Decoupling capacitor (decap) modeled as lumped RC circuit

Fig. 6 shows a decap formed by an NMOS transistor. This decap is modeled as a lumped RC circuit, where C_{eff} shows the effective capacitance which defines the charge available and R_{eff} shows the effective channel resistance which determines the transient response of charge delivery to the switching circuit.

4.1.1 PROBLEMS WITH DECAP INSERTION

At 90nm technology or below, because of decrease in oxide thickness to 2nm or less, two new major problems are arising. A large voltage across the thin oxide introduces gate tunneling leakage and potential electrostatic discharge (ESD) breakdown. Gate leakage is caused due to high voltage across the oxide layer which results in high static power consumption. White-space decaps which are located outside the standard cell blocks can be implemented using thicker oxide, hence, it reduces gate leakage and ESD, but it typically consumes three times more area. Standard cell decaps which are located within the logic blocks, require the use of thin-oxide decaps. Hence, are more prone to gate leakage problem. A possible solution for this problem can be to use only PMOS devices for standard-cell decaps, because they leak less. However, for high-performance circuits, this is not a very good solution because PMOS devices have a poorer frequency response as compared to NMOS devices.

Electrostatic discharge is a transient phenomenon of static charge transfer that can arise from human contact with an IC pin. Typically, about $0.6\mu C$ of charge is carried on a body capacitance of 100pF, generating a potential of 2kV (or higher) to discharge from the contacted IC to ground for a duration of more than 100ns. In ESD protection scheme for decaps, a series resistance R_{in} can be inserted as a protection element. This resistance limits the maximum voltage possible at its gate, where ESD damage would be more probable to occur. However, inserting this large resistor consequently reduces its frequency response and transient response. Hence, designers must take into account the trade-off between ESD protections and decap response time. To address the issue of ESD reliability, cell library developers have proposed a cross-coupled decap.

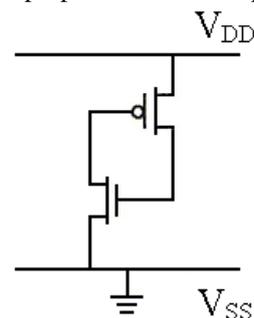


Fig. 7 cross coupled decap schematic

Figure 7 show a cross coupled schematic which reconnects the terminals of the two transistors: the drain of PMOS connects to the gate of NMOS, whereas the drain of NMOS is tied to the gate of the PMOS. This cross coupled design improves the performance of the decap by making the overall effective resistance larger without adding additional area. Since, the larger R_{eff} corresponds to a longer RC

delay; therefore, tradeoff of the design is a reduced transient response.

4.2 SELECTIVE GLITCH REDUCTION TECHNIQUE

As we know that dynamic power consumption is due to the low impedance path between the rails formed through the switching devices. When transition occurs at the output of any gate, then it can be due to two possible ways. One is the actual transmission of the input signal resulting in desired functioning of the logic gate, is also called as functional transition. Second, is due to transmission of unnecessary pulses through the logic gate resulting in undesired functioning of the gate, this is called as spurious transition. This spurious transition at the output of logic gate is an outcome of difference in arrival time of various inputs. These unnecessary signals at the output of logic gate are known as glitches. Dynamic power consumption in circuits can also be defined as product of number of transitions (N_T) and average power per transition (P_T). In this technique, the motive is to reduce N_T through glitch elimination in selected combinational cells which are contributing to peak IR drop. The peak IR drop transient time and combinational cells which are contributing to it can be determined from IR drop analysis.

Till now there have been a lot techniques developed to eliminate glitches in a logic circuit, like delay balancing, hazard filtering, gate sizing, variable input delay method etc. Among all mentioned techniques, variable input delay method is the most promising for glitch reduction on post routed layout.

4.2.1 VARIABLE INPUT DELAY METHOD

In this method, “permanently on” series transistor is inserted at the input of logic gate in order to obtain glitch free digital circuits. Consider the NAND gate shown in figure 8, the delays through different I/O paths is given as

$$d_{1 \rightarrow 3} = R_{on} \times C_{in1} + d_3$$

$$d_{2 \rightarrow 3} = R_{on} \times C_{in2} + d_3$$

Where C_{in1} and C_{in2} are the input capacitances, R_{on} is series resistance at the input of the gate. This series resistance can be manufactured by using transmission gate as a compensation circuit to compensate the effect of glitch which results due to difference in arrival times at the input of the digital logic gates.

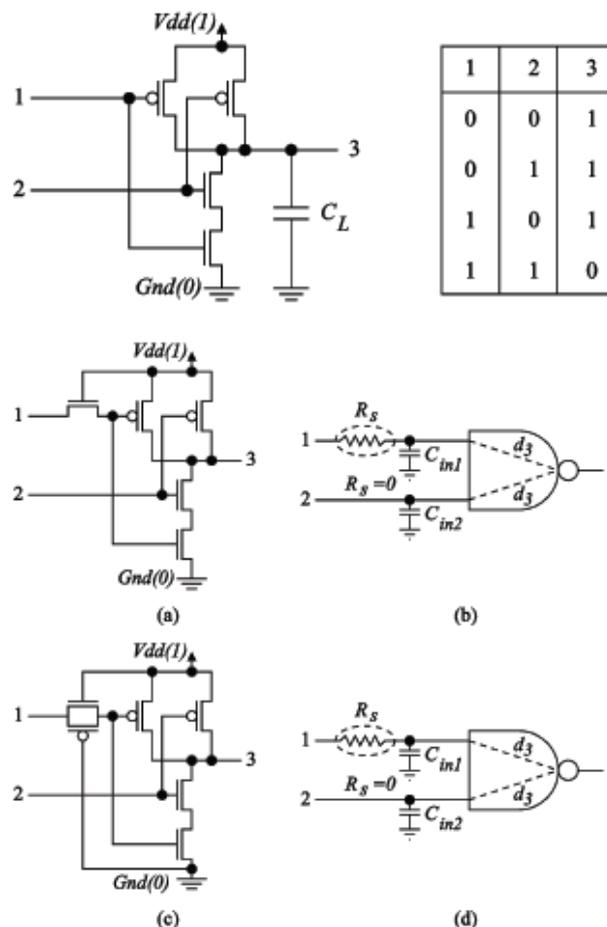


Fig. 8 schematic of variable input delay gate: a conventional 2-input NAND gate (top) and 2 ways of varying input delays by always-on NMOS transistor(center) and always-on CMOS transmission gate(bottom)

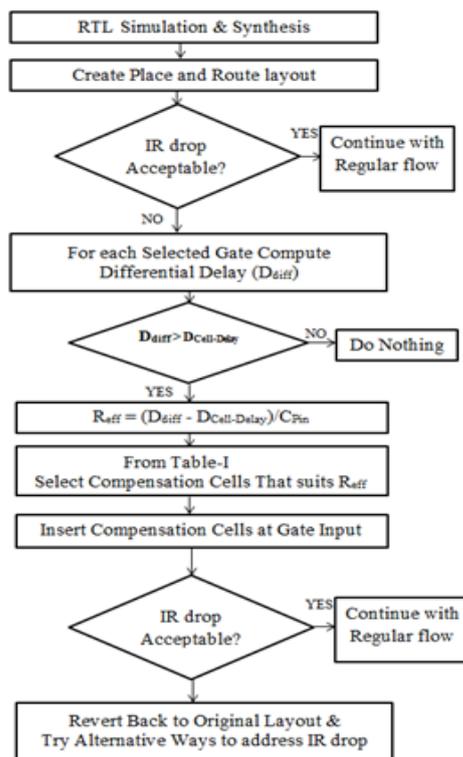


Fig. 9 flow chart of variable input delay method

This technique can be described in the flow chart as shown in figure 9. A typical implementation flow will start with a RTL simulation and moves on to place and route after successful completion of synthesis and timing. After completing place and route, gate level netlist is sent to run simulation for obtaining peak switching activity using series of test vectors. After this, we perform dynamic IR drop analysis using switching activity file for duration of peak activity window so as to capture peak IR drop numbers. If the resulted peak voltage drop values are higher than the design specification then we move on to the proposed methodology to reduce the power dissipation caused by glitch of selected combinational cells.

This selection is based on worst IR drop value resulted from the dynamic analysis. For these selected instances, the arrival time difference called differential delay (D_{diff}) is computed and compared with its cell stage delay ($D_{cell-dealy}$) of the combinational cell. If the differential delay is greater than the cell stage delay, it will have extra transitions which will form glitch and hence, will contribute to peak IR drop through more switching activity. Differential delay, cell stage delay and input pin capacitances (C_{pin}) can be computed based on Static Timing Analysis (STA) results. After computing D_{diff} , $D_{cell-dealy}$ and C_{pin} , we will compute effective resistance (R_{eff}) required which will form a time constant ($R_{eff} \cdot C_{pin}$) with the input parasitic capacitance of gate. Now using look-up table as

shown in table I, in which list of cells with corresponding equivalent series resistance is given, we can pick the correct compensation cells which compensate the delay caused due to glitches. After this, we can perform one more loop of dynamic IR drop analysis to check the effect of glitch reduction on peak IR drop value.

W/L Ratio for T-gate Structure in μm		Equivalent Series Resistance Value in Ω
NMOS	PMOS	
0.12/0.1	0.36/0.1	3.0K
0.2/0.1	0.6/0.1	1.79K
0.4/0.1	1.2/0.1	894
0.6/0.1	1.8/0.1	597
1.0/0.1	3.0/0.1	361
2.0/0.1	6.0/0.1	184
4.0/0.1	12.0/0.1	95.6
6.0/0.1	18.0/0.1	66.2
10.0/0.1	30.0/0.1	42.7
12.0/0.1	36.0/0.1	36.8

Table 1 LUT for R_{eff} used for glitch reduction

4.2.2 DESIGN ISSUE WITH THIS TECHNIQUE

The main design issue in using CMOS transmission gate as an equivalent series resistance at the input of logic gates to reduce glitches is that the effective resistance per unit length is reduced. This is because, in CMOS pass transistor, both the transistors are parallel to the current path and hence, the effective resistance is calculated as parallel combination of both nMOS and pMOS transistors, therefore, effective resistance is reduced. Hence, the transistors have to be longer to achieve the same resistance as a single nMOS resistor and hence, the area increase is higher. So, there exist a trade-off in area optimization and glitch reduction to reduce IR drop.

4.3 COMBINATIONAL CIRCUIT PARTITIONING TECHNIQUE

Earlier, in order to decrease peak switching current so as to reduce IR drop problem, focus was only on synchronous sequential logic circuits and combinational logic blocks were considered as unchangeable. Later, designers realized that combinational circuits which work alone in one clock cycle can create large current peaks and induce significant IR drops in power/ground network. For synchronous digital circuits, the technique is such that circuit is first divided into "clock regions" and then different phase clocks are assigned to these regions. This means that the designer tried to spread the original simultaneous switching activities on the time axis to reshape the switching current waveform and hence, to reduce the current peak. Those

algorithms using clock as a controlling signal to distribute the switching activity have an essential defect. These algorithms lack the ability to control the combinational circuits. Hence, there is a need to consider combinational circuits and researchers have given a novel technique for IR drop reduction in combinational circuits. This technique uses switching current redistribution (SCR) method that is based on circuit partitioning. In this method, an algorithm is defined which uses circuit slack to distribute the peak switching current.

Combinational is partitioned into sub-graphs on the basis of partitioning criterion called slack sub-graph partitioning so as to rearrange the switching time of different parts. STA tool is used to insure the actual timing constraints and critical paths, in this way, exact logic function and the highest working frequency both are preserved. A simple and proper additional delay is assigned to decomposed circuits and then, methods which modify the decomposed circuits to redistribute the switching current are compared and at the same, the logic function and the performance constraints of the circuits are also maintained.

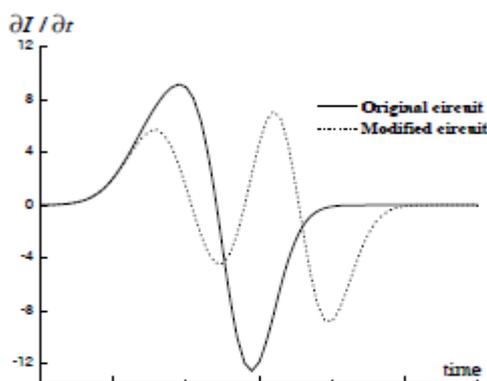
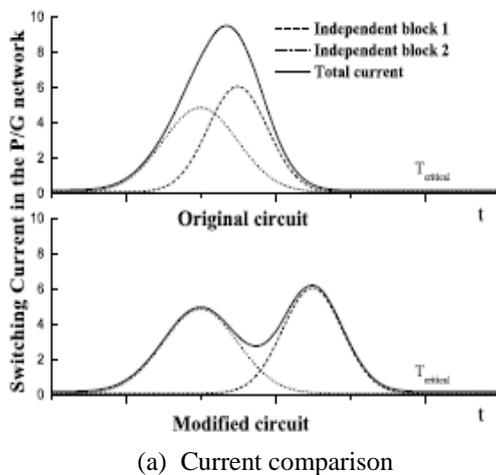


Fig. 10 switching current redistribution

In figure 10, if the combinational circuits are partitioned into independent blocks without signal dependence, their switching current can be adjusted independently. The current peak can be reduced significantly by separating the switching time of the two blocks. Since, smoothing of current waveforms reduces the di/dt , hence, Ldi/dt noise which is becoming significant when inductance of P/G network is also considered, will also get reduced. This is called as switching Current Redistribution (SCR).

The only drawback of this technique is that as the slack in the circuit is used for current distribution, the circuit is going to lose some tolerance ability to process variations which affect the path delay.

V. CONCLUSION

In this paper, I have highlighted the growing importance of the IR drop effects with technology scaling. The effect of temperature, interconnect technology scaling including resistivity increase of Cu interconnects due to electron surface scattering and finite barrier thickness has been taken into account for IR drop analysis. As timing is an important design constraint in present day designs, effect of IR drop on timing has also been considered. It has also been reviewed that gate delay is highly sensitive to the power supply voltage fluctuation in today's deep submicron technologies. Since, commonly used decap insertion methodology to reduce the IR drop is not sufficient and hence, two more techniques has been discussed. Finally, it can be concluded that more advanced techniques are required to be developed for future technologies as technology scaling is increasing day by day.

REFERENCES

- [1] Amir H. Ajami, Kaustav Banerjee, Amit Mehrotra, and Massoud Pedram, "Analysis of IR-Drop Scaling with Implications for Deep Submicron P/G network Designs", IEEE Proceedings of the forth International Symposium on Quality Electronic Design, 2003, pp. 60-67.
- [2] Resve Saleh, Syed Zakir Hussain, Steffen Rochel, David Overhauser, "Clock Skew Verification in the Presence of IR-Drop in the Power Distribution Network", IEEE Transactions on Computer-Aided Design Of Integrated Circuits and Systems, vol. 19, no. 6, June 2000, pp. 635-644.
- [3] S. Simon Wong, Changsup Ryu, Haebum LEE, Kee-Won Kwon, "Barriers for copper interconnections", IEEE/IRPS, 2001, pp. 20-27.
- [4] Kaustav Banerjee, Massoud Pedram and Amir H. Ajami, "Analysis and Optimization of Thermal Issues in high-performance

- VLSI”, IEEE Proceedings of the Forth International Symposium on Quality Electronic Design, 2004, pp. 230-237.
- [5] Xiongfei Meng, Karim Arabi and Resve Saleh, “*Novel Decoupling Capacitor Designs for sub-90nm CMOS Technology*”, IEEE proceedings of the 7th International Symposium on Quality Electronic Design, 2006, pp. 239-245.
- [6] A.Amerasekera and C. Duvvury, “*ESD in Silicon Integrated Circuits*”, 2nd ed., John Wiley & Sons, 2002.
- [7] Hai Lin, Yu Wang, Rong Luo, Huazhong Yang and Hui Wang, “*IR-drop Reduction through Combinational Circuit Partitioning*”, Springer-Verlag Berlin Heidelberg 2006 ,pp. 370-381.
- [8] Mondira Deb Pant, Pankaj Pant, Donald Scott Wills, “*On-Chip Decoupling Capacitor Optimization for Noise and Leakage Reduction*”, Proceedings of IEEE ISCAS, 2003, pp. 29-35.
- [9] Vasantha Kumar, N. S. Murthy Sharma, K. Lal Kishore, A. Rajakumari, “*Selective Glitch Reduction Technique for Minimizing Peak Dynamic IR Drop*”, Microelectronics and Solid State Electronics 2013, pp. 27-32.