Challenges in Power-Ground Integrity

Shen Lin and Norman Chang

Apache Design Solutions, Inc., Palo Alto, CA

Abstract

With the advance of semiconductor manufacturing, EDA, and VLSI design technologies, circuits with increasingly higher speed are being integrated at an increasingly higher density. This trend causes correspondingly larger voltage fluctuations in the on-chip power distribution network due to IR-drop, L di/dt noise, or LC resonance. Therefore, Power-Ground integrity becomes a serious challenge in designing future high-performance circuits. In this paper, we will introduce Power-Ground integrity, addressing its importance, verification methodology, and problem solution.

1. Introduction

Circuits with increasingly higher speed are being integrated at an increasingly higher density. As the result, on-chip Power-Ground voltage fluctuation is significantly increasing due to IR-drop, L di/dt noise, or LC resonance. Assuming constant field scaling, the percentage of di/dt to Vdd scales as S^4 when scaling process by 1/S [1]. On the other hand, scaling imposes requirements on the reduction of the power supply voltage Vdd, which has also been decreased to aggressively save power consumption. Therefore, Power-Ground integrity becomes a serious challenge in designing future high-performance circuits. A 10% supply voltage fluctuation may translate to more than a 10% timing uncertainty. It may even cause the circuit to fail. In addition, circuit block gating is often introduced to reduce power, making sudden change of current likely and causing significant L di/dt noise. The inductance will also resonate with on-chip intrinsic and incorrectly inserted de-coupling capacitors and cause LC resonance.

In addition, large current oscillation may cause more power consumption [19] and reliability problem, such as Electromigration. In this paper, we are going to introduce these challenges of Power-Ground integrity, addressing their importance, verification methodology, and problem solution. IR-drop, L di/dt noise, LC resonance, and Electromigration will be introduced in Section 2, 3, 4, and 5, respectively. The paper concludes in Section 6.

2. IR-drop

IR-drop is mainly due to the resistance of the on-chip Power-Ground network. When large current flows through, un-acceptable voltage drop happens. The voltage drop may cause timing uncertainty and slew rate slow-down, hence affecting performance and increasing power consumption. Simulation results show that 10% supply voltage drop may cause about a 10% or more delay increase. With constant field scaling assumption, the ratio of the chip total current over Vdd scales as S^3 when scaling process by 1/S. This fast increase has made IR-drop a serious design challenge for all wire-bond and periphery-bumped chips. For example, the Power and Ground meshes completely filling top two metal layers are not enough to solve the IR-drop problem for a 12mm by 12mm chip in a 0.13u technology consuming only 6 Watts average power.

There are two kinds of IR-drop: Static IR-drop and Dynamic IR-drop. Static IR-drop is the voltage drop between Vdd and Gnd based on the average power or current while Dynamic IR-drop considers the current waveform within clock cycles and the coupling capacitance between Power and Ground, including the intrinsic decoupling capacitance and the intentionally inserted decoupling capacitors[2][3]. That means Dynamic IR-drop is an RC transient behavior. Until recently, Dynamic IR-drop was generally neglected. Designers were hoping that Dynamic IR-drop would be contained by existing decoupling capacitors and the cycle time is large enough to recharge those decoupling capacitors. Therefore, one may think as long as the Static IR-drop is within the design constraint, the chip should function without surprise.

However, as the clock frequency increases, the time to recharge those decoupling capacitors may not be sufficient. During a stream of operations that consume large current, the voltage drop may significantly increases and eventually fails the circuit. To verify the Power-Ground integrity for Dynamic IR-drop, we need to extract all the decoupling capacitance and determine the inputs that cause peak power consumption, including register states, and then perform a transient simulation on the resulting RC network.

Static IR-drop may be remedied by sizing up Power-Ground wire widths and by decreasing their pitch. However, without good planning and optimization, the over-design Power-Ground network may create congestion problems for the later signal routing because valuable routing resources are over-used. Therefore, several Power-Ground network optimization works were proposed in the literature [4][5][6][7].

Dynamic IR-drop problem need to be remedied by both wire-sizing and decoupling capacitor insertion. Decoupling capacitor can be made effectively using NMOS gate-oxide capacitance [2][3], but its insertion has to be considered early in the chip floor-planning stage and area should be reserved; otherwise, it would be too difficult to squeeze out space after placement or routing is completed.

IR-drop analysis is comparatively easier because both the R and the RC matrices are positive definite. Hence, the Conjugate-Gradient Iterative method with Incomplete Cholesky Pre-conditioner may be efficiently applied to solve the voltage values[8]. Also, the loading devices and the Power-Ground network may be solved separately and iteratively. The solution converges in a few iterations. Later in the paper, we will see that none of these are valid when inductance is considered. Multi-grid[9] and Macro-modeling[10] were also proposed to further speed-up the IR-drop analysis.

3. L di/dt Noise

A sudden change of the current flowing through a wire will induce abrupt voltage changes on that wire and its neighboring wires due to inductance. If these wires are part of the on-chip Power-Ground network, the induced voltage fluctuation is called L di/dt noise (often called *Delta-I* noise). In circuit, the current spikes are caused by gate switching; therefore, we may derive the chip's di/dt measure as

$$\frac{di}{dt} \approx \frac{I}{\tau} = \frac{I\,V_{dd}}{\tau\,V_{dd}} = \frac{P}{\tau\,V_{dd}} \approx \frac{P\,f_c}{V_{dd}} \quad , \qquad \text{(Eq. 1)}$$

where I is the total average current flowing into the chip, P is the chip's power consumption, and τ is the typical rise-time of a gate, which multiplied by the chip operating frequency f_c should normally be a constant.

A general IC design trend is that the chip power P is constantly increased because of denser circuits and aggressively higher operating frequencies f_c. And, to save power, V_{dd} is constantly decreased. Therefore, all elements on the right-hand side of Eq. 1 point out that the di/dt value should be significantly increasing. The ratio of L di/dt measure over Vdd equals to

$$\frac{L\,\frac{di}{dt}}{V_{dd}} \approx \frac{L\,P\,f_c}{V_{dd}^2} \quad . \qquad \text{(Eq. 2)}$$

Up to now, the on-chip Power-Ground inductance has been ignored. However, our measurement results show that the inductance from one side of a 8mm by 8mm mesh to the other side may be larger than 2nH. With the introduction of lower resistivity metal, such as Copper, to boost chip's performance, the impact of wires' inductive impedance may become comparable with the resistive portion if not larger. Therefore, the on-chip Power-Ground inductance should not be ignored in the L di/dt analysis especially for wire-bond or periphery-bumped chips.

The analysis of L di/dt noise is much more difficult because the Modified Nodal matrix with L [11] is no longer positive definite. Most iterative linear solvers fail if the number of inductors is large. Also, the strong negative-feedback within an LC loop makes separate solving devices and Power-Ground network never converge. We will observe numerical oscillation at the results, often seen in the Explicit Numerical Integration[11]. Furthermore, we may no longer assume the existence of a Universal Ground node on the chip and every grounded capacitors are connected to this ground node as in RC analysis. This Universal Ground is used to model the chip substrate. Since high frequency currents, important in L di/dt noise, will not travel through the resistive substrate for return, the coupling capacitors between Power and Ground should be explicitly connected in between those two networks. They are all floating capacitors. Without a grounded capacitor connected at each node, the Modified Nodal matrix may not be diagonal dominant, and hence, the convergence for iterative solver is even more difficult.

Partial mutual inductive coupling has long range effect [12] as the coupling between two parallel wire segments decays very slowly with the increase of their separation. Therefore, almost all parallel wires on chip couple to each other, making the L matrix a huge dense matrix and difficult to solve. By observing that the susceptance (the inversion of L) decays very fast with separation, the susceptance method [13] and the double inversion method [14] were proposed to prune the couplings and to speed-up the solving process.

Another difficulty in L di/dt analysis is in determining the loading current waveform while using piecewise current sources to approximately model the devices [15][16]. Fig. 1 shows the simulated current waveform through a CMOS driver. Two points are important, the peak and the knee. These two points primarily determine the L di/dt noise value. However, without knowing the loop inductance, they cannot be determined. Since the

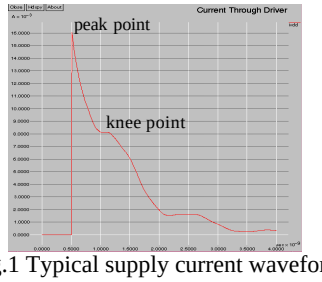

Fig.1 Typical supply current waveform

integration of the current up to the peak point is very small, knowing the average power or current of the cycle

does not help. However very often, it is the only information known to designers.

A better way to approximately model devices is by using Time-Varying Resistors (TVR) as shown in Fig. 2. Their resistance values are piecewise functions of time and the times of break points are determined according to timing analysis or design constraints. At the time instance when a group of devices turn ON, the value of the TVR to model them change from a very large OFF resistance to a ON resistance. The ON resistance value may be determined from the slew rate design constraint, equal to the multiplication of the ON resistance and the total loading C. The loading C may be determined based on the power number or by scanning the devices and interconnects. Another advantage of employing TVR is that

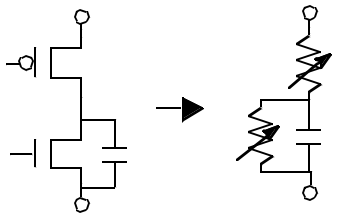

Fig. 2 Using TVR to mimic MOSFET switching

intrinsic decoupling capacitances are automatically modeled. The capacitor connected to Vdd through a ON TVR is an intrinsic decoupling capacitor.

L di/dt noise may be tamed by inserting decoupling capacitors near hot spots, by turning on current-hungry blocks gradually[17], or by using advanced IC packaging, such as center-bumped flip-chip. Since gate-oxide decoupling capacitors may reduce die yield, their amount needs to be optimized. Assuming that at the very instance when the sudden current spike happens, the supply current is completely provided by the decoupling capacitor C_D due to inductance, we have

$$C_D \times V_{dd} = (C_{Load} + C_D) \times (1 - \alpha) V_{dd} \quad , \qquad \text{(Eq. 3)}$$

or,

$$\frac{C_D}{C_{Load}} = \frac{1 - \alpha}{\alpha}, \qquad \text{(Eq. 4)}$$

where α is the allowable percentage L di/dt noise between Power and Ground, and C_D is the required decoupling capacitance, including intrinsic decoupling capacitance. Therefore, if the L di/dt noise should not be more than 10% of Vdd, decoupling capacitance of more than 9x the loading is needed.

4. LC Resonance

L di/dt noise is a high frequency phenomenon. LC resonance, on the other hand, impacts a circuit even when it is operated at low frequencies. Fig. 3 shows a *RLC* circuit. When a step input current *I* is applied, we have

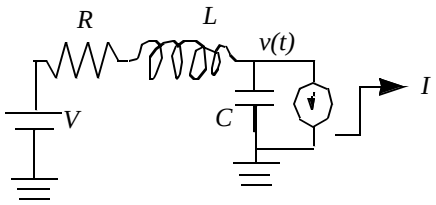

Fig. 3 RLC circuit with step current

$$v(t) \approx V - I \times R - I \times \sqrt{\frac{L}{C}} \times e^{-\frac{R}{2L}t} \sin(2\pi f_r t - \theta) \qquad \text{(Eq. 5)}$$

when $$\sqrt{\frac{L}{C}} > \frac{R}{2} \qquad \text{(Eq. 6)}$$

and the oscillation frequency $$f_r = \frac{1}{2\pi\sqrt{LC}} \qquad \text{(Eq. 7)}$$

The transient oscillating term diminishes at the speed of *exp(-Rt/(2L))*. It takes several nano seconds to decrease to 36.78% with the R and L values of typical wire-bond or periphery-bumped chips of **Al** Power-Ground wires and may take even longer for **Cu** technology. If before this transient term completely diminishes another current pulse is injected synchronously, this term is magnified and takes even longer to diminish. Once this process repeats, the circuit resonates with the frequency f_r. Although Eq. 5 assumes a step current *I*, the oscillating condition Eq. 6, also called under-damped condition, and the oscillating frequency Eq. 7 are still valid if *I* is not.

Another significant difference between L di/dt noise and LC resonance is that L di/dt noise may be *localized*. That is the L di/dt noise may be confined in an area by inserting sufficient decoupling capacitors surrounding it. However, LC resonance waves travel through the whole chip. Since its frequency is low, on-chip decoupling capacitors are not effective to stop the wave propagation.

Designers often mistakenly think that the resonance noise needs to be as large as almost V_{th} to corrupt the chip's logic function. The truth is that it may take a lot less if the chip has analog parts, such as the sense amplifier in a SRAM block. If the two sensing signals are driven by one device with Vdd swinging up and the other with Vdd swinging down, very little noise is sufficient to cause the amplifier to mis-read the result.

LC resonance has been considered a very difficult challenge faced by circuit designers because apart from the rare mistake by tying a chip's core Vdd with its dirty I/O Vdd together, an obvious fix, in general, does not exist. Multiple silicon spins may not be avoided if resonance noise is observed at the chip tester. Designers are forced to blindly modify the chip, and hence, its *L* and *C*, hoping to move the resonance frequency out of synchronization. Fortunately with the full chip *RLC* P/G simulation becoming possible recently, most this kind of try-and-sees may be verified first by circuit simulations.

Decreasing the Power-Ground inductance is more effective in remedying the problem because not only has it changed the resonance frequency but also increase the transient diminishing rate. And, if cost is not a constraint, employing advanced packaging technology, such as center bumped packaging, is the most effective cure. Those center P/G bumps adhere the on-chip Power-Ground network directly to the package Power-Ground plane or mesh, which has a lot less inductance. Other remedies include increasing the Power-Ground damping by connecting series resistors at decoupling capacitors or at the network [1]. Since these techniques are trading IR-drop, L di/dt noise, or the chip's performance, verification is required.

LC resonance verification is very difficult. Besides the challenges in inductance simulation mentioned in the previous Section, it poses two significant requirements. One is that the simulation time needs to be very long in order to see whether resonance happens because the resonance frequency is low and the initial numerical transients need to diminish. For example, to make sure the chip would not resonate at 66 MHz, almost 100 nano second simulation time is required. Another is many different input vectors need to be simulated. This is different from L di/dt verification where only very few peak current inputs are likely to cause problems and designers normally have clear idea of them. However, the resonance may be induced by many possible input patterns.

5. Electromigration

Metal electromigration results from a conductor carrying too much current and hence, the displacement of metal atoms due to electron-flux. It will cause shorts or opens in metal lines, and affects the chip's reliability. The Black's Mean-Time-to-Failure (MTTF) [18] equation below governs this phenomenon.

$$\mathrm{MTTF} = AJ^{-2}\exp\left(\frac{\Delta H}{kT}\right)f(W), \qquad \text{(Eq. 8)}$$

where J is the average current density of the wire and T is its temperature. Since MTTF is an exponential function of temperature, silicon thermal analysis is needed besides circuit simulation to determine J.

6. Conclusions

With the advance of semiconductor manufacturing, EDA, and VLSI design technologies, circuits with increasingly higher speed are being integrated at an increasingly higher density. This trend causes correspondingly larger voltage fluctuations in the on-chip power distribution network which may cause timing uncertainty, design failure, more power consumption [19], and reliability problems. We introduced four challenges in Power-Ground integrity: Static/Dynamic IR drop, L di/dt noise, LC resonance, and Electromigration. Designers need EDA supports, such as *RLC* extraction, modeling, and simulation, to verify and solve the problems.

References

[1] Patrik Larsson, "Resonance and Damping in CMOS Circuits with On-Chip Decoupling Capacitance," IEEE Trans. on CAS-I, pp. 849-858, Aug. 1998.

[2] R. Allmon et al., "System, Process, and Design Implications of a Reduced Supply Voltage Microprocessor," IEEE ISSCC, pp. 48-49, 1990.

[3] D. W. Dobberpuhl et al., "A 200-MHz 64-b Dual-issue CMOS Microprocessor," IEEE J. Solid-State Circuits, pp. 1056-1063, Nov. 1992.

[4] S. Chowdhry and M. Breuer, "Optimization Design of Reliable IC Power Networks Having General Graph Topologies," 26th DAC, pp. 787-790, 1989.

[5] T. Mitsuhashi and E. Kuh, "Power and Ground Network Topology Optimization for Cell Based VLSI," 29th DAC, pp. 524-529, 1992.

[6] X. Tan, C. Shi, et al. "Reliability-Constrained Area Optimization of VLSI Power/Ground Network via Sequence of Linear Programming", 36th DAC pp. 78-83, 1999.

[7] X. Tan and C. Shi, "Fast Power/Ground Network Optimization Based on Equivalent Circuit Modeling," 38th DAC, pp. 550-554, 2001.

[8] G. Golub and C. Van Loan, "Matrix Computations," 3rd, Johns Hopkins University Press, 1996.

[9] S. Nassif and J. Kozhaya, "Fast Power Grid Simulation," 37th DAC, pp. 156-161, 2000.

[10] M. Zhao, R. Panda, et al. "Hierarchical Analysis of Power Distribution Networks," 37th DAC, pp. 150-155, 2000.

[11] L. Chua and P. Lin, "Computer-Aided Analysis of Electronic Circuits," Prentice Hall, 1975.

[12] C. K. Cheng, J. Lillis, Shen Lin, and Norman Chang, "Interconnect Analysis and Synthesis," John Wiley & Sons, 1999.

[13] A. Devgan, H. Ji, W. Dai, "How to Efficiently Capture On-Chip Inductance Effect: Introducing a New Circuit Element K," Proc. ICCAD, 2000.

[14] M. Beattie and L. Pileggi, "Modeling Magnetic Coupling for On-Chip Interconnect," DAC, 2001.

[15] H. Chen, et al. "Power Supply Noise Analysis Methodology for Deep-Submicron VLSI Chip Design," 34th DAC, 1997.

[16] R. Panda, D. Blaauw, et al. "Model and Analysis for Combined Package and On-Chip Power Grid Simulation," Int. Symp. on Low Power Electronics and Design, pp. 179-184, 2000.

[17] Z. Tang, N. Chang, S. Lin, et al. "Inductive Noise Reduction based on Instruction Prediction," Int. Symp. on Quality Electronic Design, 2001.

[18] J. Black, "Electromigration Failure Modes in Aluminum Metallization for Semiconductor Devices," Proc. of IEEE, p. 1587, 1969.

[19] T. Burd and R. Brodersen, "Design Issue for Dynamic Voltage Scaling", Int. Symp. on Low Power Electronics and Design, pp. 9-14, 2001.