A Power Integrity Wall follows the Power Wall!

Raj Nair, Anasim Corp., pi-day 3/14/2008

A decade ago, the call of the times was for solutions to address the “Power Wall”, at least at leading microprocessor institutions. Conferences discussed looming challenges with supply currents exceeding 1000's of amperes, and temperatures exceeding, what was it, nuclear nozzles, or surfaces of galactic stars, and academia as well as niche industrial efforts rushed after cryogenic or microfluidic cooling systems as well as designs approaching electrical power stations on a chip. Today, power consumption is the single dominant design constraint for integrated circuits, but less noticed, and even less respected is power integrity despite its undeniable role in determining power and energy consumption. Most of us notice that if we dim the lights in our entertainment rooms too much, sharp changes in brightness of our television screens hurt our eyes. Our attempt to reduce lighting energy consumption therefore depends directly upon the level of 'light noise' we encounter. The very same is true for IC's; minimization of energy through supply voltage reduction, the most fundamental approach, depends directly upon power grid noise, or power integrity [1]. As discussed ahead, power integrity is the next dominant challenge, the call of the present for SoC's and SiP's, as power and energy continue to be dominant design constraints.

So what is Power Integrity?

In simple terms, it is how close to the ideal a given power supply is, depending upon its nature. For the supply to our homes, it is amplitude and frequency, and how steady they remain despite loading and load-shedding occurring nearby. All of us may have noticed lights dimming when a large air-conditioner or refrigerator compressor kicks in; that is a degradation in power integrity.

Integrated circuits are not much different, though they typically use DC supplies of a few volts down to a volt. Power integrity for IC's is understood through knowing the droops and overshoots produced in this constant potential as parts of the IC switch on or off, or ramp up and ramp down their function. For good power integrity, it is important that these droops and overshoots, or transient (and static) changes in the value of the supply voltage differential be kept within a small fraction, say 5%, of the supply nominal, which maintains predictable performance in the integrated circuit.

The dominant components of variation in supply voltage differential in IC power grids are I*R Drop and L*di/dt, though other noise components may at times dominate, such as propagated and reflected noise as well as resonance. Analysis tools in the EDA industry today employ variations of the I*R Drop methodology to analyze voltage reduction in various sections of an integrated circuit. The relative significance of power integrity noise components is changing rapidly as scaling continues, primarily because most integrated circuits today function through binary computation involving circuits switching currents, and doing so in synchrony with clock signals. As we head deeper into nanoscale processes, these switching speeds continue to increase rapidly, correspondingly increasing on-chip di/dt, and it is now imperative that we inspect 'total power integrity' including L*di/dt and other electromagnetic effects rather than just I*R Drop or its derivatives.

Loop Inductance, L*di/dt, and the impact of scaling

In studies conducted about 7 years ago, published in a technology journal as 'Emerging Directions for Packaging Technologies' [4] in its Power Delivery section, I was able to demonstrate an impractical

burden placed upon support devices such as package capacitors as processors scaled inexorably. Power integrity studies showed then that the loop inductance from package capacitors into a processor device needs to scale by between the 3rd and the 5th exponent of the process scaling factor in order to maintain the same power integrity in a scaled process as in the previous generation. Loop inductance, in this instance, determined delay in response from a package capacitor to a transient charge demand by the processor, and a corresponding voltage droop in processor supply voltage.

In similar fashion, on-chip loop inductances determine the delay in response from charge stored in regions of the chip or its package to regions demanding transient charge, such as functional blocks that are switched on rapidly, or a bank of registers, flip-flops or large numbers of logic gates clocked simultaneously. A loop inductance L, in combination with a rate of rise of current, di/dt, also produces a voltage droop given by L*di/dt that adds to any instantaneous voltage drop due to wire resistance in the power distribution network. It is instructive to go through an exercise to determine how scaling impacts L*di/dt droop (noise) in successive generations. To do so, we will make a few assumptions, based on current industry trends, regarding scaling and its benefits and consequences:

- Capacitance-per-unit-area, C_a, scales by $\sqrt{2}$ which is roughly (1/0.7), where 0.7 is the typical fabrication process generation scaling factor (95nm to 65, 45, 32 etc.),
- Operating voltage scales by $\frac{1}{\sqrt{\sqrt{2}}}$ which reduces it only by about 16%,
- Frequency scales by $\sqrt{2}$ which provides a 40% improvement,
- Chip area scales by $\frac{1}{\sqrt{2}}$, reducing only by 30% rather than 50%, which indicates additional circuits integrated for improved performance in the scaled process

With the above assumptions, called '**Roots of Two Scaling**' for reference, one can deduce that active power, given by the equation $\alpha C V^2 f$ remains the same in the scaled process generation as in the prior generation, assuming α remains the same. In this scaling scenario, silicon area is reduced, thus reducing cost, and frequency and integration are increased, thereby increasing performance, while power penalty remains same (we assume better leakage control techniques maintain leakage power the same), maintaining the economic benefits to be accrued from following Moore's Law.

- Given above scaling, the average active current scales by the inverse of the scaling for voltage, since power remains the same, or by $\sqrt{\sqrt{2}}$.
- Since frequency scales by $\sqrt{2}$, $\frac{dI}{dT}$ scales by $\sqrt{\sqrt{2}}$ * $\sqrt{2}$.
- Also, since the chip area scales by $\frac{1}{\sqrt{2}}$, each side (assume a square chip) scales by $\frac{1}{\sqrt{\sqrt{2}}}$.
- Given the smaller dimension per side, and assuming that power buses are drawn occupying roughly the same percentage of metal resources in the scaled process generation, the power metal per side reduces by a factor of $\frac{1}{\sqrt{\sqrt{2}}}$, and including increased current-crowding, the effective inductance increases approximately by $\sqrt{\sqrt{2}}$.

Multiplying L and di/dt in the scaled process generation for this chip, we get:

➔ L*di/dt scales as $\sqrt{2} * \sqrt{2} * \sqrt{2}$ or by a **factor of 2**.

These calculations are obviously highly simplified and the scaling factors by no means accurate, but the

trend surely is. As we scale to finer process geometries, switching edge rates are faster, while devices are smaller and operating voltages lower in order to save on active and leakage energy consumption. Though power consumption may remain the same, or may even reduce, faster processes demand faster transfer of charge or higher currents through lesser available resources such as power metal and from lesser capacitance, leading to higher voltage droops.

L*di/dt noise, as seen in this simple derivation, **is doubling in every scaled process generation**, mirroring the burden placed upon support components by scaling as detailed in [4] despite constant-power scaling. Assume, for example, L*di/dt or inductive noise contributions to power network noise to be of the order of **9mv in the 180nm** generation. Following the derivation trend, the **45nm node** could see 16 times as much inductive noise, or about **144mv**, which can be about 15% of the supply voltage differential, or **3X the allowed maximum noise**. Academic research has shown a similar trend in quadratic increase of inductive noise with rise times [2] determined with careful and complex PEEC modeling of power distribution grids. I*R Drop tools completely miss this aspect of power integrity; any light at the end of this methodology tunnel is clearly the headlight of the L*di/dt train rushing exponentially towards us all.

Methodology comparison: IR Drop vs. On-chip Inductance inclusion

As before, we rely on our scaling example and derivation. A chip or block designer who sees the average supply current increasing in a scaled process generation is faced by hard decisions with regard to power distribution grid design. Sometimes relying on an over-hyped 'copy-exactly' methodology, the designer may simply allocate the same fraction of metal resources to power distribution as was done in the previous chip (that worked, so it must have been done right!). As we have deduced, that would result in twice the amount of transient inductive noise in the scaled chip. In nanoscale processes, this could be a torpedo that comes back to hit and sink the ship (chip) after fabrication.

The more conscientious designer, following best-practices IR Drop methodology, may reason that average chip currents have increased, thereby requiring more power grid metal. Faced by a choice between increasing the number of power buses, which means reducing bus pitch, or increasing the width of metal wires in a bus, and constrained by routing needs, a choice often made is to increase the metal wire width rather than crowding routing with decreased bus pitch, with IR Drop tools showing corresponding improvement in noise. Unfortunately, this solution has hardly any practical benefit, particularly where the dominant noise contribution is from L*di/dt, which is often minimally or even negatively impacted by increased metal wire width and axial separation between bus wires. Besides, high-frequency currents tend to be restricted to low-inductance regions of the power bus.

A common factor in both of the above arguably flawed approaches is that they rely disproportionately upon prior knowledge and experience and not upon comprehensive verification that eliminates gaps in methodology. Predictably, such undue reliance is bound to diminish the quality of work done and the eventual end result, for just as it is said in financial investments involving numerous uncertainties, past performance is no guarantee of future results. This inertia in the industry that works against moving to comprehensive analyses of power integrity is in large part due to a glaring absence of efficient and accurate modeling techniques and EDA tools to conduct rapid, comprehensive, true-electromagnetic simulations of IP Blocks, multi-core chips and entire power delivery systems.

Research conducted in the academia has, in fact, shown that power metal area is optimized far better through the inclusion of inductive noise in power grid simulations. A published paper on the impact of on-chip inductance on power distribution network design [3] claims about 30% improvement or reduction in metal area usage for power grids in the 90nm process, and up to 60% improvement in the 45nm process with comprehensive modeling of on-chip power grid inductance.

The Power Integrity Wall

Quadratic or exponential increases in L*di/dt noise, a near-complete absence of techniques and EDA tools that help designers of IP Cores and chips rapidly simulate and analyze their physical designs for total power integrity, lack of clear understanding of all aspects of chip power integrity (of which some are listed in [1]), and an unrelenting push to scale SoC designs to finer nanoscale dimensions is a potent combination leading to **device yield spreading** and ultimately, economic infeasibility of scaling.

The combination of semiconductor device variations in nanoscale processes and operation at lower and lower voltages for energy minimization further exacerbates yield problems. Low power and low energy design do not equate to not having problems with power integrity. It is in fact just the opposite; low energy design introduces additional complexities such as Power Gating that impact power integrity in ways that are not immediately apparent [1]. Practical chips operating at fractions of a volt with any reasonable level of integration and performance will only be feasible if the very narrow bands of noise allowable on their supplies are very well understood and comprehensively verified. The alternative is to translate all our established, area and power-efficient digital logic circuits into extremely noise-tolerant, possibly differential, current-mode circuits in nanoscale processes!

Lacking sufficient advancement in methodology and effective tools for total power integrity, we appear certain to face a scaling barrier as severe, if not more severe than the Power Wall.

A potential solution path

It is therefore imperative that all contributors to IC power integrity degradation be comprehended as SoC designs move to 65 and 45nm nodes and beyond. The inclusion of all electromagnetic effects in power distribution networks is now a critical requirement, addressed through tools employing high levels of abstraction and physics-based simulations such as Anasim Corporation's π-fp [5].

Simple, non-disruptive modifications to current design flows using such a tool are listed below:

- ➔ Begin floorplanning for power integrity at the architectural stage
- ➔ Employ low-inductance, symmetric, IP Core and global power network synthesis
- ➔ Optimize power network dimensions with true-electromagnetic simulations and determine initial on-chip decoupling capacitance strategy
- ➔ Optimize floorplan placements to minimize supply 'total noise' band; minimize operating voltage while meeting timing/performance requirements simultaneously
- ➔ Conduct static IR Drop simulations to detect any 'hot spots' and electro-migration stress points in the physical design
- ➔ Include system-level components such as package capacitors, supply connectivity etc. and re-run true-electromagnetic, dynamic noise analysis with improved block current profiles; tweak interconnect and decoupling capacitance resources accordingly

Verifications through true-electromagnetic simulations at the IP Core and full-chip/SiP/system levels can help ensure that exponentially rising L*di/dt and other noise components impacting power integrity and power/energy optimization are comprehended in the design process. Such front-end investment in comprehensive design and verification efforts are crucial to avoiding design iterations, design respins, and yield or product failures that can be orders of magnitude higher in bottomline impact.

References

[1] Raj Nair & Donald Bennett, "Power Integrity and Energy-Aware Floorplanning", SoCCentral featured article, January 2008

[2] A. Muramatsu, M. Hashimoto & H. Onodera, "Effects of On-chip Inductance on Power Distribution Grids", ISPD 2005

[3] N. Srivastava, X. Qi and Kaustav Banerjee, "Impact of On-chip Inductance on Power Distribution Network Design for Nanometer Scale Integrated Circuits", ISQED 2005

[4] Raj Nair, "Emerging Directions for Packaging Technologies - Power Delivery Considerations section", Intel Technology Journal, ITJ Vol. 6, Issue 2, May 2002

[5] Anasim π-fp, Power Integrity and Energy aware Floorplanner, Anasim Corp., 2008

Author Biography

Raj Nair is an expert consultant in IC power delivery and power integrity, and a co-founder of Anasim Corp. developing and marketing π-fp. Prior to Anasim, he founded ComLSI, Inc. that develops and licenses patents and silicon IP in advanced power delivery as well as high-speed signaling. Formerly, he was with Intel Corp., where he was responsible for strategic programs enabling CPU power delivery and power integrity innovations. He holds 36+ issued patents and has translated many of his inventions into commercial products.