Macro-Modeling Concepts For The Chip Electrical Interface

Brian W. Amick
Sun Microsystems Inc.
Austin, TX
Brian.Amick@Sun.com

Claude R. Gauthier
Sun Microsystems Inc.
Sunnyvale, CA
Claude.Gauthier@Sun.com

Dean Liu
Sun Microsystems Inc.
Sunnyvale, CA
Dean.Liu@Sun.com

ABSTRACT
The power delivery network is made up of passive elements in the distribution network, as well as the active transistor loads. A chip typically has three types of power supplies that require attention: core, I/O, and analog. Core circuits consist of digital circuits and have the largest current demand. In addition to all of the system issues/models for the core, modeling the I/O subsystem has the additional requirement of modeling return paths and discontinuities. The analog circuits present yet a different challenge to the macro-modeling of the supply network because they place a tight demand on supply variations. This paper presents a design methodology on how to generate macro-models of the entire chip electrical interface. This methodology can be used by the chip, package, and system designers and is being used to design high-reliability servers.

Categories and Subject Descriptors
C.5.3 **[Computer System Implementation]**: VLSI Systems.

General Terms
Performance, Design, Reliability.

Keywords
VLSI Power Distribution, Inductance, High Speed Microprocessor Design, Analog and I/O Power Delivery.

1. INTRODUCTION
Delivering power to VLSI chips is an increasingly difficult challenge as voltage scaling decelerates, the supply current and power dissipation accelerate, and the device count increases. As voltage scales, the tolerances on the power supply get tighter, while the noise sources actually increase in magnitude. Processor power transients can be large in magnitude and occur at arbitrary frequencies though the largest power transients occur at relatively low frequencies, necessitating a model of the entire system. I/Os do not escape this challenge as I/O counts and frequencies increase. The high-frequency signal return current associated with each of these I/O circuits must travel through the entire network from chip to package to board.

A potential road-block to meeting this objective is the fact that most of the data required for design verification is available only when the design is completed, and the system is ready to ship. Also, when the data becomes available, the simulation problem can be intractable, owing to the amount of elements and information present.

Power-delivery models for core, IO, and analog supplies requires modeling of the entire system (typically circuit-board/package/die) to capture the basic first order effects. This paper will address the modeling constructs required to understand these effects early in the design cycle in such a way that the design can be influenced. Examples will be presented that illustrate the impact of the various components of the core, I/O, and analog supply models.

DAC 2002, June 10-14, 2002, New Orleans, Louisiana, USA.

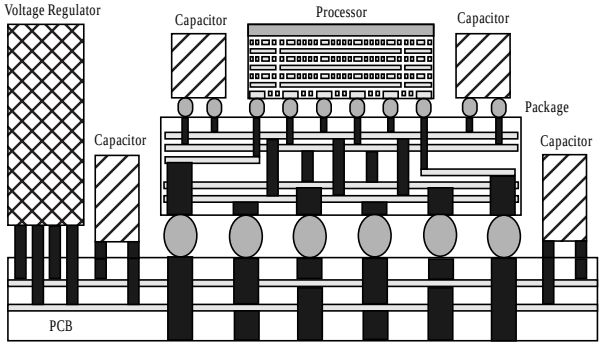

Figure 1. The Power Delivery System

2. CORE POWER MODELING
A microprocessor power delivery system is illustrated in Fig. 1. The major components include the voltage regulator, PCB stack-up, capacitor (type, quantity, and placement), the package type/stack-up, package capacitors, and on-die capacitance. These major issues are resolved by creating a system-level model of the core power distribution. For low/mid-frequencies, one can model this complex network as shown in Fig. 2. In Fig. 2 the parasitic inductance comes from: bumps, package and board vias, planes and perforated planes, decoupling capacitor parasitics, and the voltage regulator. The capacitance comes from on-chip capacitance (non-switching gates and decoupling capacitance) and discrete off-chip capacitors. The model is accurate in the 1 to 200MHz range and will serve to develop the system, package, and on-chip capacitor requirements. This type of model is not meant to see the effects of localized, very high-frequency on-chip noise, however it can be expanded into 2D/3D models to extend the effective frequency range upwards.

The off-chip inductance (L) and on-chip capacitance (C) form a LC tank circuit whose impedance peaks at a the chip/package resonant frequency (between 10MHz-100MHz in most systems). Depending on system configuration the peak can be quite large. The resonance can be excited by instruction patterns that cause large transients in power. However, the excitation is not limited to a sine or square wave, as any large power transient will contain energy at the frequencies of interest. If this resonance is not removed, then large power transients will degrade the supply significantly over multiple cycles, thereby reducing performance and reliability. The most substantial voltage droop in a microprocessor power delivery system is associated with this frequency band. The following sections addresses how to estimate the various parasitics in Fig. 2, and formulate a methodology for designing core power distribution networks using a macro-model of the chip electrical interface.

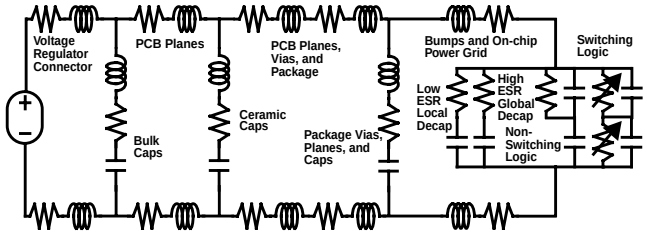

Figure 2. Core supply equivalent circuit (low/mid frequencies)

2.1 Decoupling Capacitors

Voltage regulators contain a feedback mechanism that allows them to respond to changes at the load. However the bandwidth of these regulators are typically limited to the 10-kHz to 1-MHz range [1], above which they behave like inductors. The inductor value can be obtained from the voltage regulator vendor and is highly dependent on the design of the connector. Therefore, a hierarchy of decoupling is used to provide power at higher frequencies.

Discrete capacitors are far from ideal as each capacitor contains an equivalent series resistance (ESR) and an equivalent series inductance (ESL). Each capacitor is a series resonant circuit that is most effective at its resonant frequency (f_O) where it appears resistive. Below f_O the response is capacitive and beyond f_O the inductance dominates. Capacitor effectiveness is limited by the capacitor's ESL and the inductance of the path to the capacitor [2]. The ESL can be reduced by using many capacitors is parallel. In order to explore the reduction of the path inductance, a model of the package and PCB power planes must be derived.

2.2 Package and Circuit Board Power Planes

A typical system contains both non-perforated and highly perforated power planes. The non-perforated sections of power planes are the sections between capacitors and the package. Perforated power planes are those that lie under the package where hundreds or thousands of vias cut through the planes. The loss of metal increases the inductance and resistance, which can be represented using a derating factor based on the amount of metal lost. Power planes have R, L, and C components, however for low/mid-frequencies the C can typically be ignored as this distributed capacitance will be relatively small compared to explicate capacitance. Different approaches to finding the spreading inductance can be found in [3] and [4].

Multiple planes in parallel can be treated as resistors in parallel, given an alternating power-ground stack-up. For example, two pairs of power planes have one third the inductance of a single pair because of the addition of a dielectric layer between the pairs of planes. Therefore, the key to managing inductance lies in using multiple pairs of planes, with thin spacing. A very important package consideration is the use of capacitors which will require a low impedance path through the horizontal package power planes. These planes are modeled just like perforated and unperforated planes on the PCB. With the use of package capacitors, it is possible to eliminate a high-Q chip/package resonance.

2.3 Vias, Pins, & Bumps

Hundreds or thousands of vias are used to bring the core power from the PCB power planes to the pads connecting to the package and between different layers in the package. The inductance is proportional to the via height and inversely proportional to the total number of vias. Typically core power is brought in through the middle section of the package pin array (to alleviate I/O routing issues). The layer location of PCB power planes controls the via lengths.

In general packages are a series of vertical via layers and pins that connect the PCB to the chip along with some horizontal planes of metal connecting the vias inside of the package. For the model shown in Fig. 2, the package planes can be ignored in the vertical path from chip to PCB. Along the way the via dimensions change to reflect the connection dimensions at the package ends. The total minimum thickness of the package is dominated by the number of routing layers needed to route the IO signals or by mechanical reasons. It often takes several layers for the IO signals to escape towards the edge of the package.

The vertical bump/via/pin inductances can be derived from approximations which relate the self and mutual inductance of a pair of parallel filaments [4]. The mutual inductance (M) of two equal parallel straight filaments is given by:

$$M = \frac{u_o}{2\pi} l \left(\ln\left(\frac{l}{sep} + \sqrt{\left(1 + \frac{l^2}{sep^2}\right)} \right) - \sqrt{\left(1 + \frac{sep^2}{l^2}\right)} + \frac{sep}{l} \right) \quad (1)$$

where l is the via length and *sep* is the separation distance between the center of the vias. To obtain the self inductance, we replace *sep* by the radius of the wire (or half the diameter, d) and add the internal inductance of the wire ($u_r/4$) we obtain the self inductance (L):

$$L = \frac{u_o}{2\pi} l \left(\ln\left(2\frac{l}{d} + \sqrt{\left(1 + \frac{4l^2}{d^2}\right)} \right) - \sqrt{\left(1 + \frac{0.25d^2}{l^2}\right)} + 0.5\frac{d}{l} + \frac{u_r}{4} \right) \quad (2)$$

It is of note that at higher frequencies, where the skin effect is predominant, the internal inductance term drops out.

Using EQ 2, the self inductance of each via/pin in the package can be calculated along with the mutual inductance, using EQ 1, between all pairs of vias/pins in the package. Once the inductive terms of the package vias/pins are computed the network is collapsed into a two-port network as illustrated in Fig. 3(a) by placing all power inductors in parallel and then in series with the parallel combination of ground inductors. The network can be excited with a 1-A_{AC} current source, and by measuring voltage at a given frequency, ω, we can solve for the equivalence inductance by applying Ohms law. Using this result, an equivalent single inductor and resistor can be found for arbitrary patterns such as the ones shown in Fig. 3(b) [5]. This methodology can be applied to the several alternatives for socketing the microprocessor including LGA, BGA, or µPGA. From a modeling perspective, the major difference is brought about by the pin/ball lengths and the pin/ball pattern used.

2.4 Chip Model

The processor determines the power transient behavior and contains a substantial amount of on-die capacitance. The on-die capacitance comes from non-switching logic, switching logic, bypass capacitance that is local to the logic, and global (in the channels) capacitance (as illustrated in Fig. 2). The local capacitance can be represented as a capacitor with a low ESR. Inductance is usually ignored on-chip in this analysis as off-chip inductances and on-chip resistance dominate at low/mid frequencies. The ESR for local capacitance is usually dominated by the channel resistance and increases as the channel length increases. Global capacitance resides in the channels, often far away from the logic. At this distance, the ESR is dominated by the power grid, and the capacitor is modeled with a much higher ESR than its local counterpart.

Digital logic is represented by a series pair of parallel capacitor and resistor. For non-switching logic, one of the two resistors is set to a very high value and the other is set to give a RC time constant equal to a fanout-of-four (FO4) inverter. Switching logic (including leakage and crow-bar current) can be represented by either switch-

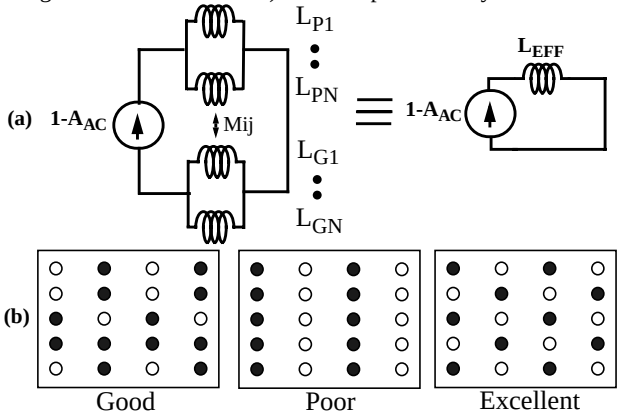

Figure 3. (a) The pattern can be simulated in the AC domain by mapping the L's, M's to a passive circuit. (b) Effective inductance is strong function of placement.

ing the resistors in another copy or copies of the circuit used for non-switching logic. Leakage current is modeled by making the higher "off" value resistor just large enough to still draw the correct amount of leakage. Crow-bar current is modeled by having both resistors in a low resistance state for a short amount of time. Finally by using multiple copies of the circuit, an expected current profile, from an architectural power model, can be modeled.

2.5 Capacitor Selection & Sizing Methodology

We have assembled all the pieces required for the sizing of capacitors at the various hierarchies. Several different types of capacitors are available, each with a different cost, area, and frequency response (due to ESR and ESL). The goal is to find a combination of the board, package and chip capacitors that reduces the overall impedance of the distribution network to within a tolerable limit, while fitting within the allotted area and for a reasonable total cost.

A model of the system similar to Fig. 2 is first constructed, with the parasitic inductance and resistance calculated based on physical geometry using the methodology described in the previous sections. The chip load is excited using an AC waveform with its frequency as the sweeping parameter. Using this model, the optimal capacitor size for each hierarchy can be found with a greedy approach. For board capacitors, the frequencies of interest range from 1-MHz to 10-MHz, while the package capacitors are responsible for frequencies from 10-MHz to 200-MHz.

Using the greedy approach, a unit quantity of the cheapest capacitor is added into the model. The impedance of the network across the frequency spectrum of interest is obtained by sweeping the frequency of the AC load using a circuit simulator. If the network impedance is above the target, another capacitor is added. Each additional capacitor lowers the impedance peak while possibly shifting its frequency. If the impedance is still above the target when the allotted decap area is full with the cheapest capacitor, then replace one capacitor with the next-most expensive type, assuming that the higher cost capacitor is more effective in lowering the overall impedance at higher frequencies. The search stops when the desired impedance is met. Algorithms for proper placement of the on-chip decoupling capacitance can be found in [6].

Fig. 4 shows the transient response to a step excitation. The absence of ringing in the design that meets the target impedance prevents even larger excitations from a repetitive source. The use of the methodology leads to a substantial increase in Vdd which would allow for higher frequency operation at a relatively low cost to the overall system.

3. I/O POWER MACRO-MODEL

I/O power modeling is an extension of the core power problem, but is slightly more complex because of the signals and return currents. Fortunately the PCB, vias, decoupling capacitors, and voltage regulator modeling constructs are the same as shown in the core power section. The current loops for I/O circuits are more complex

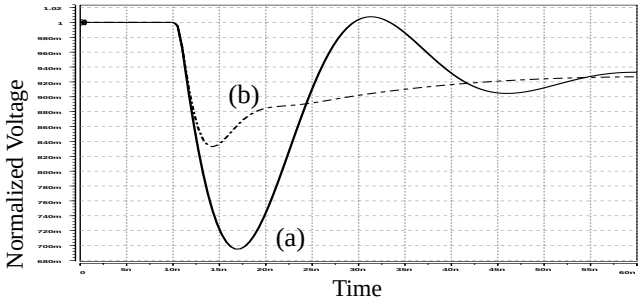

Figure 4. The effects of a 40A transient as seen from the chip (a) without a methodology for sizing capacitors and (b) using the proposed methodology to reduce the supply impedance

than that of the core because they include off-chip parasitics resulting in Vdd/Gnd bounces independent of each other across different chips. Therefore, careful attention must be paid to make sure return currents are modeled correctly. It is the flow of these return currents that causes large Vdd/Gnd bounces from either an overly inductive supply network or discontinuities in the signal return path.

In the presence of Vdd and Gnd bounces, the resulting timing and voltage margins will be reduced because the waveform will have increased jitter and reduced signal swing. In some cases the bounces, themselves, can be transmitted across a channel as noise. In addition, if the links are single-ended, the reference voltage may be inaccurate as well since it is often referenced to a different Gnd than the receiver.

To model these return currents correctly, signal traces must be referenced to the correct Vdd and/or Gnd nodes. This is done by using a 2D field solver, such as Ansoft, to extract the traces and reference planes. In addition the use of ideal ground (SPICE node 0) should be avoided in all locations except for one spot as this node is an infinite source/sink for current. This special node can allow current to avoid discontinuities in the system and jump from one point to another, masking important effects.

An example model that illustrates the major power supply effects that must be captured is shown in Fig. 5. The model represents two voltage mode drivers whose output signals are referenced to both Vdd and Gnd planes in the package, but in the PCB they are between Gnd planes only. PCB traces are typically only referenced to Gnd planes since Vdd is not routed through connectors. Note the signal being between Vdd and Gnd planes in the package is reflected in the model by the capacitors to Vdd and Gnd. Also, the extraction contains important capacitors between Vdd and Gnd, representing where signal traces are not present between reference planes.

While Fig. 5 is all that is required to accurately model the IO supply, it raises some possible issues that must be resolved. The first is the Vdd and Gnd bounce from the large supply inductance. The bounces from the supply network can be reduced by increasing the on-chip capacitance or by lowering the inductance to other decoupling capacitors (e.g., through thinner spaced planes and more pins).

The second issue is the discontinuity in the signal return when the signal travels from the package to the PCB. In the package, approximately half the return current flows in Vdd, and half through Gnd. However, in the PCB, all the return current flows through Gnd. Therefore half the return current must cross the discontinuity through the capacitance between Vdd and Gnd in the package,

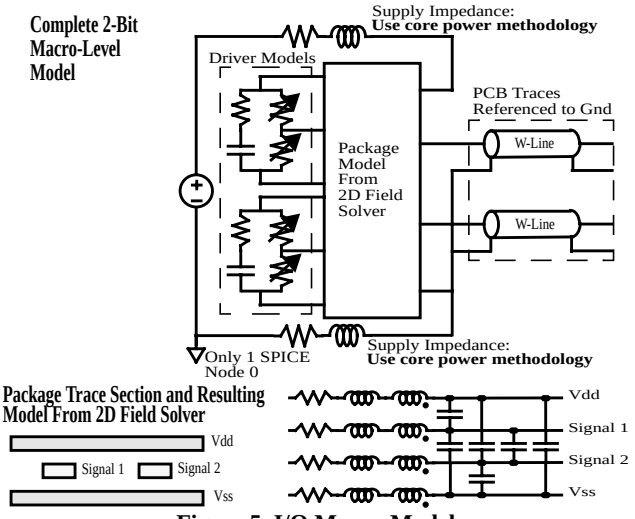

Figure 5. I/O Macro-Model

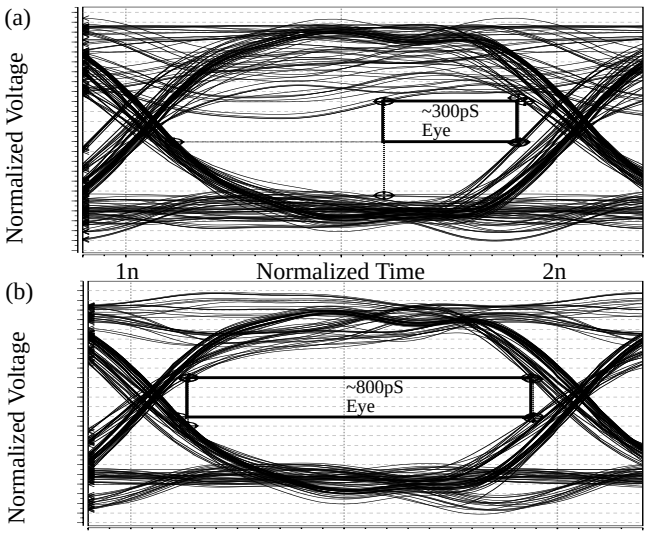

Figure 6. Effect of discontinuities to the received signal eye (a) and effect of using capacitor to satisfy discontinuities (b).

which will collapse the supply voltage in the package and on-chip. The effect, illustrated in Fig. 6a, can be avoided by placing low ESR and ESL capacitance between Vdd and Gnd near the return current discontinuity either on the PCB or on the package (illustrated in Fig. 6b). In contrast to Fig. 6a, which shows the data eye without the package capacitors, the eye opening in Fig. 6b is noticeably larger. This example model can be extended to reflect different numbers of signals by extracting more traces, scaling the non-package elements, and adding additional on-chip elements. Different configurations of capacitances, vias, and power planes, can be evaluated. The I/O macro-model allows us to quickly investigate all of the trade-offs associated with the design.

4. ANALOG POWER MACRO-MODEL

Typically microprocessors have some analog circuits such as a PLL or DLL whose effectiveness is related to the stability of the power supply. PLL and DLL jitter is dominated by power supply noise [7]. The issues addressed by the analog supply macro-model include supply isolation, signal return current strategy, and decoupling capacitor methodology. The digital circuits on a microprocessor often cause significant supply fluctuation due to large switching current. Often the fluctuation is too large for analog circuits and it is advantageous to isolate the supply of the analog circuits. The system level approach to providing supply isolation is to route a separate quiet power supply from the PCB as a trace in the package to the processor. The macro-model for this very common approach is illustrated in Fig. 7.

The model includes a PCB filter, the pin and via impedance, a transmission line to represent the package trace, the bump and via impedance, and the analog transistor load. The PCB filter is used to isolate the board and chip. The impedance for the pin, bump, and vias are calculated using the parasitic estimation methods described in the earlier section on core power macro-modeling. In the package, the analog supply is routed as a signal sandwiched between ground planes and widely spaced from other traces in order to avoid coupling from other noisy signals. If the analog supply is to be routed near other signals, then the coupling between them must be included as well.

One important issue is that in almost all microprocessor applications, circuits on analog supplies have to communicate with circuits on another supply (e.g., PLL output driving the clock tree on the core supply). If the on-chip Gnd grids are not shorted, the current loop will include the package and substrate impedances. This could

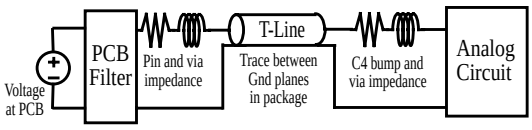

Figure 7. Analog Supply Macro-Model

result in a large portion of the signal being lost across these impedances as power supply bounce, a slowing of the signal rise/fall times, duty-cycle distortion, and jitter in the signal. Therefore the Gnd domains should be shorted in metal as metal results in a very small and controlled impedance.

The other critical issue with system level supply isolation is determining the amount of on-chip decoupling capacitance needed for the analog supply. This can be done by exciting the input to the macro-model in Fig. 7 with a noise source and finding the minimum amount of on-chip decap required for the circuits to meet required specifications. Decoupling capacitance and jitter performance can typically be traded for each other. As a final check a model should be simulated to make sure the decoupling capacitance is sufficient to limit supply noise from the return currents flowing through the supply decoupling capacitance when crossing power domains.

5. SUMMARY

Scaling trends in VLSI have created significant challenges for designers including chip power delivery and distribution as supply voltages have decreased, while transistor counts, frequency, and current have increased. A good model includes information about the entire distribution network, including the passive elements in the distribution network (decoupling capacitor hierarchies, board and package power planes, vias, and bumps), as well as the active transistor loads (core, I/O, and analog). This paper presented a methodology on how to construct the macro-models of the chip electrical interface, discussed the trade-offs involved in the design of various components, and provided some design solutions.

For I/O power, the same macro-models were used, however the model was extended to address signal return currents and potential discontinuities. Solutions to typical problems were presented within the context of the I/O macro-model. Finally, a macro-model for analog power was presented to address the major issues when analog and digital circuits reside on the same chip. Issues brought up by the model included the need for a supply isolation approach, decoupling capacitor methodology, and the interface issues associated with inter-power-domain communication. These models are currently being used to design high-reliability servers in industry.

6. ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of R. Wheeler, M. Blatt, S. Bobba, C. Durham, V. Ganesan, D. Greenhill, R. Heald, S. Pannala, P. Patel, L. Noujeim, S. Schmidt, B. Sen, G. Singh, L. Smith, J. Tam, P. Trivedi, R. Voelker, B. Wong, and G. Yee

7. REFERENCES

[1] L.D.Smith *et al.*, "Power Distribution System Design Methodology and Capacitor Selection For Modern CMOS Technology," *IEEE Transactions on Advanced Packaging*, August 1999, vol 22, no 3.

[2] Korony *et al*, "Controlling Capacitor Parasitics For High Frequency Decoupling Capacitors," *Proc. IMAPS Conference*, 2001.

[3] W. Dally and J. Poulton, "Digital Systems Engineering," Cambridge University Press, 1997.

[4] F. W. Grover, "Inductance Calculations Working Formulas and Tables," D. Van Nostrand Company, 1946.

[5] R. Wheeler, "Modeling Simultaneous Switching Noise (SSO) in the Z-axis Direction of VLSI Packages and PCB", www.wheeler.com,1999.

[6] H. H. Chen and D. D. Ling, "Power Supply Noise Analysis Methodology for Deep-Submicron VLSI Chip Design," *Proc. Design Automation Conference*, June 1997.

[7] V. Kaenel, *et al.*, "A 320 MHz, 1.5mW @ 1.35-V CMOS PLL for Microprocessor Clock Generation," *IEEE J. of Solid-State Circuits*, Vol. 31, No. 11, pp. 1715-1722, November 1996.