Power Distribution System Design Methodology and Capacitor Selection for Modern CMOS Technology

Larry D. Smith, *Member, IEEE,* Raymond E. Anderson, *Member, IEEE,* Douglas W. Forehand, Thomas J. Pelc, and Tanmoy Roy, *Member, IEEE*

***Abstract*— Power systems for modern complementary metal-oxide-semiconductor (CMOS) technology are becoming harder to design. One design methodology is to identify a target impedance to be met across a broad frequency range and specify components to meet that impedance. The impedance versus frequency profiles of the power distribution system components including the voltage regulator module, bulk decoupling capacitors and high frequency ceramic capacitors are defined and reduced to simulation program with integrated circuit emphasis (SPICE) models. A sufficient number of capacitors are placed in parallel to meet the target impedance. Ceramic capacitor equivalent series resistance (ESR) and ESL are extremely important parameters in determining how many capacitors are required. SPICE models are then analyzed in the time domain to find the response to load transients.**

***Index Terms*— Ceramic capacitor equivalent series resistance, CMOS, power distribution system, SPICE.**

I. Introduction

DESIGN of the power distribution system (PDS) is becoming an increasingly difficult challenge for modern CMOS technology. As CMOS technology is scaled to give smaller and faster transistors, the power supply voltage must decrease. As clock rates rise and more function is integrated into micro processors (μP's) and application specific integrated circuits (ASIC's), the power consumed must increase. These trends are summarized for the 1990's in Table I. Given the voltage and power consumed, the current is calculated from Ohm's law. Assuming that only a small percentage of the power supply voltage (i.e., 5%) is allowed as ripple voltage (noise), a target impedance for the PDS is calculated. The target impedance is falling at an alarming rate, 5× per computer generation [1]

$$Z_{\text{target}} = \frac{(\text{Power Supply Voltage}) \times (\text{allowed ripple})}{\text{current}} = \frac{(5\ \text{V}) \times (5\%)}{1\ \text{A}} = 0.250\ \text{Ohms}.$$

The target impedance must be met not only at DC, but at all frequencies where current transients exist [2]. The target impedance must be met in the kHz range because the μP may be processing code for 0.5 ms, then require some information from a hard drive and have little activity for 0.5 ms and repeat the process over and over again. It must be met in the MHz range because the μP may require something from DRAM memory, work for a while, and then require something else. The target impedance also must be met near the clock frequency because the μP demands and ceases to demand current on that time scale. The PDS must be designed to handle all possible current transients. This essentially means that the PDS must be no more than about twice the target impedance over a broad frequency range (current transients are commonly about half of the maximum current).

Fig. 1 shows the target impedance versus frequency for a modern CMOS computer system. Several major components are used to meet this target impedance over the frequency range. The voltage regulator module (VRM) is effective up to about 1 kHz. Bulk capacitors supply current and maintain a low PDS impedance from 1 kHz to 1 MHz. High frequency ceramic capacitors maintain the PDS impedance from 1 MHz

TABLE I
TARGET IMPEDANCE IS DROPPING BY A FACTOR OF 5 EVERY COMPUTER GENERATION

year	voltage (volts)	power dissipated (watts)	current (amps)	Ztarget (mOhms)	frequency (MHz)
1990	5.0	5	1	250	16
1993	3.3	10	3	54	66
1996	2.5	30	12	10	200
1999	1.8	90	50	1.8	600
2002	1.2	180	150	0.4	1200

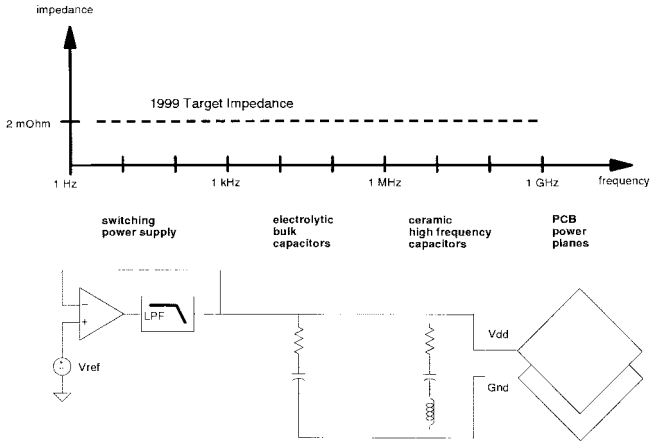


Fig. 1. Flat power supply impedance versus frequency is met by using VRM, bulk capacitor, high frequency ceramic capacitor, and power plane components.

Manuscript received January 15, 1999; revised May 10, 1999. This paper was presented at the Seventh Topical Meeting on Electrical Performance of Electronic Packaging, United States Military Academy, West Point, NY, October 26–28, 1998.

The authors are with Sun Microsystems, Inc. Palo Alto, CA 94303-4900 USA.

Publisher Item Identifier S 1521-3323(99)06576-4.

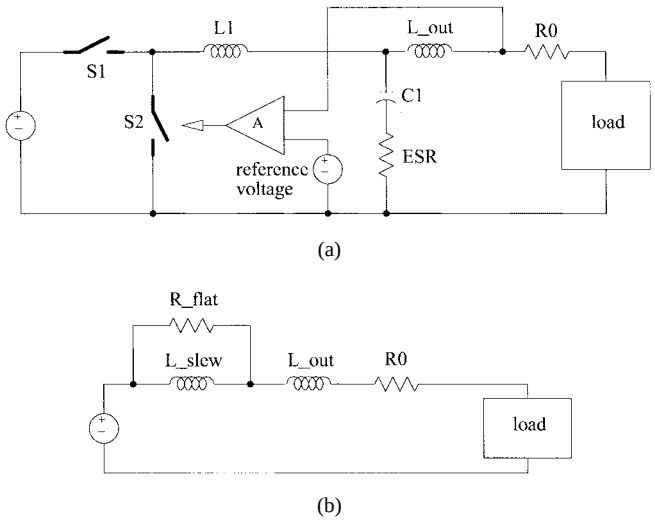


Fig. 2. (a) Simplified block diagram of buck switching regulator and (b) linear model for VRM.

to several hundred MHz. The interplane capacitance and the impedance of the printed circuit board power planes are important above 100 MHz. This paper will discuss design considerations for the VRM, bulk capacitors and high frequency ceramic capacitors. If Moore's Law and CMOS scaling is to continue, the PDS design will be an increasingly difficult design problem.

II. Voltage Regulator Module

The voltage regulator module (VRM) converts one DC voltage to another, for example 5–1.8 V. It has a reference voltage and a feedback loop. It senses the voltage near the load and adjusts the output current to regulate the voltage at the load. The bandwidth of the regulation loop is usually between 1 kHz and several hundred kHz. At frequencies above the loop bandwidth, the VRM becomes high impedance.

Fig. 2(a) is a simplified block diagram for a buck switching regulator, commonly found in VRM's. At the left of the diagram is an input voltage, assumed to be relatively constant. The function of inductor L1 is to store up energy when switch S1 is closed, and deliver current to the load. If L1 has more current than the load is demanding, S1 opens and S2 closes. Current continues to flow to the load, but in an ever diminishing amount until S2 opens and S1 closes again. There is an amplifier A with frequency compensation that senses the load voltage with respect to a reference voltage. When the load voltage is too low, it causes the switches and inductor to ramp up the current. When the load voltage is too high, it causes the switches and inductor to ramp down the current. The inductor current is integrated in C1, which smooths out the voltage. C1 has an equivalent series resistance (ESR).

The buck regulator is nonlinear because switches open and closed as a function of time. In the architecture design phase for the PDS, it is highly desirable to have a linear model for the VRM. Fig. 2(b) shows such a model consisting of an ideal voltage source and four passive elements. Simulation program with integrated circuit emphasis (SPICE) analysis of this model runs very fast in both the frequency and time domains. It is sufficiently accurate to estimate the amount of bulk capacitance necessary for the PDS. Many of the components are common to both the nonlinear and linear model. R_0 is the value of the resistor between the VRM sense point and the actual load and is usually only a few mOhms. A VRM is not capable of regulating the voltage at the actual load. One square of copper for a pair of power planes is about 1 mOhm. $L_{\rm out}$ represents the output inductance of the VRM. It may be the inductance of cables that connect the VRM to a system board or it may be the inductance of pins that connect a VRM to a μP module (about 200 and 4 nH, respectively). The maximum effective frequency for the VRM is determined by $L_{\rm out}$. $R_{\rm flat}$ represents the ESR of the capacitor associated with the VRM. Generally, the capacitor will determine the output impedance of the VRM at frequencies beyond the response time of the loop. The ideal voltage source has the value of the power supply voltage.

TABLE II
Typical Parameter Values for a VRM

parameter	value	units
R0	1	mOhm
L_out	4	nH
R_flat	30	mOhms
L_Slew	67.5	nH

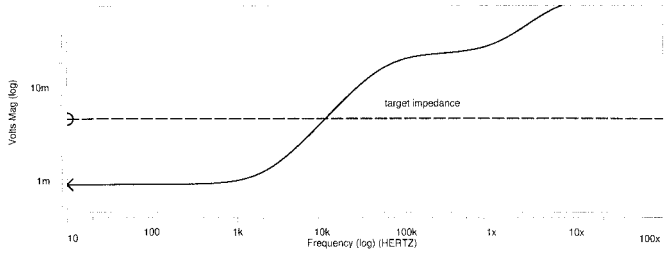
$L_{\rm slew}$ is the only element in the linear model that is not traceable back to an element in the nonlinear VRM model. The value of $L_{\rm slew}$ is chosen so that current will be ramped up in the linear model in about the same time that it is ramped up in a real VRM. It is calculated from the equation: $V = L\,di/dt$. V is the amount of voltage droop or spike that can be accepted on the PDS (i.e., 5% of 1.8 V). The maximum transient current is used for di. The total amount of time for the VRM to ramp this transient current either up or down is used for dt. A sample calculation for $L_{\rm slew}$ for a VRM that can ramp down 20 A in 15 μs is

$$L_{\rm slew} = V\frac{dt}{di} = (1.8\ \text{V} * 0.05)\frac{15\ \mu s}{20\ \text{A}} = 67.5\ \text{nH}.$$

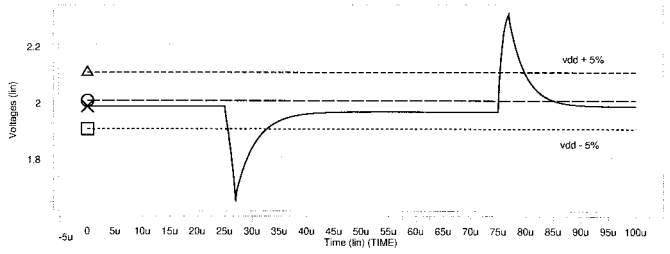
Table II gives typical parameter values for a VRM. In this case, the VRM is a small PCB attached to a processor module by a pair of large copper pins. Fig 3(a) shows the output Impedance versus frequency of the VRM using these parameters. 1 A is forced into the output terminals of the VRM. The voltage is measured and interpreted as ohms. The values of R_0 and $R_{\rm flat}$ are readily seen on the diagram. Fig 3(b) shows the time domain response of this VRM to 20 A transients occurring at 10 kHz with a 1 μs rise time. The nominal power supply voltage was 2 Volts. The voltage measured at the load is slightly below that because of the R_0 resistance. Current transients are drawn from the PDS by using a time dependant resistor. The current waveform is shown in Fig. 3(c). The PDS exceeds the $\pm 5\%$ specification limit for several μs. This may be more than 1000 clock cycles for a modern μP.

III. Bulk Capacitance

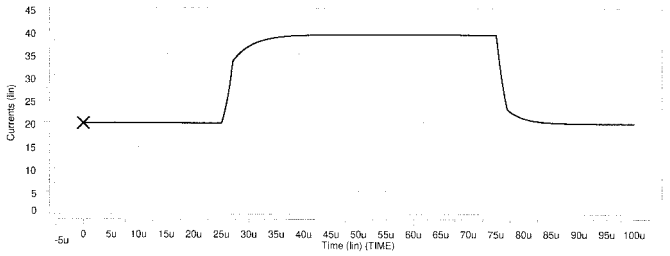
Bulk decoupling capacitors are necessary to maintain the PDS impedance at frequencies above those maintained by the



(a)



(b)



(c)

Fig. 3. (a) Output impedance versus frequency of typical VRM, (b) time response of VRM to current transient, and (c) 20 A, 50 μs current pulse with 1 μs rise and fall times.

VRM and below the frequencies where ceramic capacitors are effective. There must be sufficient capacitance in the PDS to supply the current and support voltage until the VRM can respond. That amount of capacitance is calculated from the equation $I = C\ dv/dt$. Suppose there is a 20 A current transient, the VRM responds in 15 μs, and the PDS must remain within 5% of a 1.8 V power supply. The amount of bulk capacitance required is estimated by hand calculations

$$C = I\frac{dt}{dv} = 20\ \text{A} * \frac{15\mu s}{1.8\ \text{V} * 0.05} = 3333\ \mu\text{F}.$$

This may over estimate the required bulk capacitance by a factor of 2 because the VRM is ramping up current and the average current may be half of the final value during the VRM response time. But, more capacitors may be necessary if the ESR of the bulk capacitors in parallel is higher than the target impedance. The idea is to keep the PDS below some target impedance through the transition from VRM frequencies to bulk capacitor frequencies. The performance of a VRM is somewhat determined by the loading effects of the bulk capacitance. After estimating bulk capacitance by this method, a cycle-by-cycle simulation is necessary with a detailed VRM model fine tune the combined performance of the VRM with the capacitance.

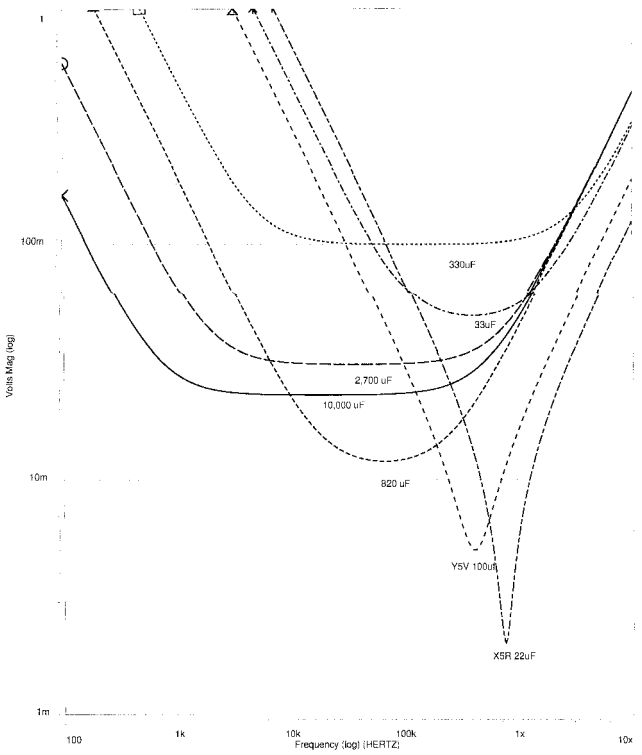
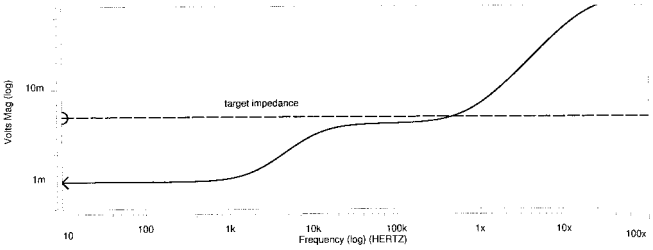
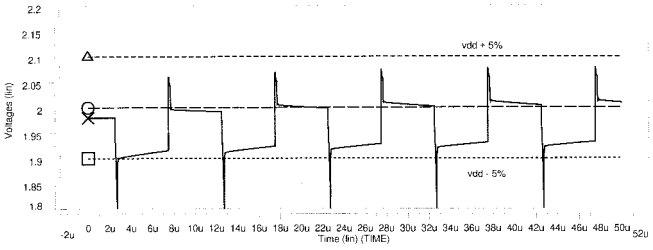


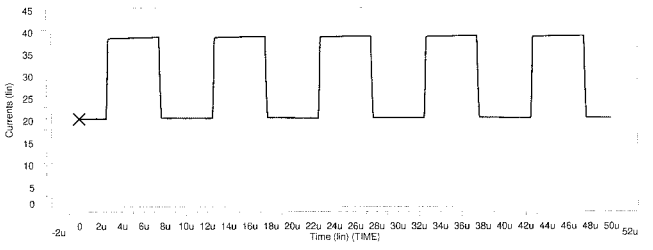
Fig. 4. Impedance of bulk capacitors versus frequency.



(a)

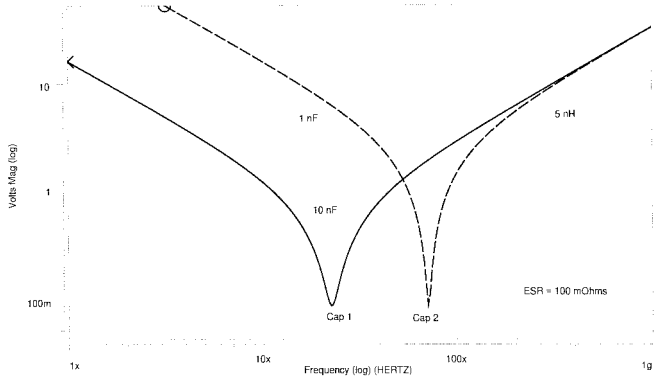


(b)

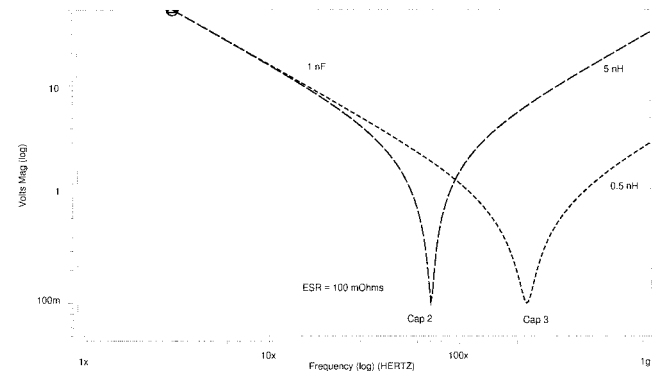


(c)

Fig. 5. (a) Output impedance versus frequency of VRM and several bulk capacitors, (b) time response of same PDS to current transient, and (c) 20 A current transients at 100 kHz with 200 ns rise and fall times.



(a)



(b)

Fig. 6. Simulation of RLC capcitor models with (a) different capacitance values and (b) differenet inductance values.

TABLE III
CAPACITOR VALUES ANALYZED IN FIG. 6

	cap 1	cap 2	cap 3
Capicitance	10 nF	1 nF	1 nF
ESR	100 mOhm	100 mOhm	100 mOhm
Inductance	5 nH	5 nH	0.5 nH

Bulk capacitors are modeled using series resistance, inductance, and capacitance (RLC) elements. Capacitance gives a slope of -20 dB per decade on a log/log impedance versus frequency (Bode) plot. Resistance gives a flat response with frequency, and inductance gives a $+20$ dB per decade slope. SPICE impedance versus frequency results for several bulk decoupling capacitors are shown in Fig. 4. The capacitive, resistive and inductive portions of these curves are readily identified. Capacitors in parallel behave the same as resistors in parallel. The impedance is cut by $\frac{1}{2}$ by the second capacitor, $\frac{1}{3}$ by the next capacitor, $\frac{1}{4}$ by the next, etc. SPICE simulation is used to determine the impedance of several parallel capacitors and determine the frequency range that is below the target impedance. The transition from VRM to bulk frequencies is observed when bulk capacitors are in the same analysis as a VRM. A properly designed PDS has an impedance that does not exceed the target impedance in the transition from VRM to bulk frequencies.

Fig. 5(a) shows the output impedance versus frequency when a VRM is placed in parallel with 5×2700 μF electrolytic bulk capacitors. The bulk capacitors provide low impedance from VRM frequencies up to almost 1 MHz. Fig. 5(b) shows a SPICE transient analysis on the same components. The PDS is loaded with 20 A of current at 100 kHz with a rise time of 200 ns as shown in Fig. 5(c).

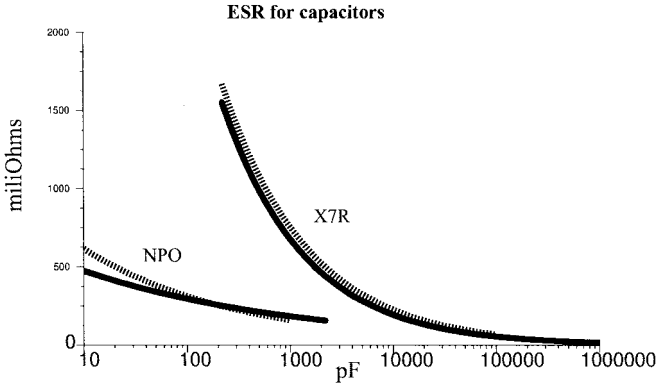


Fig. 7. ESR versus capacitor value for NPO and X7R capacitors, 80 5 and 603 size.

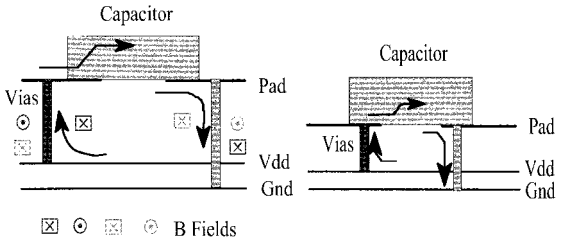


Fig. 8. Inductance of a capacitor is mostly in the pad and via structure. Inductance is minimized by minimizing the area of the current loop.

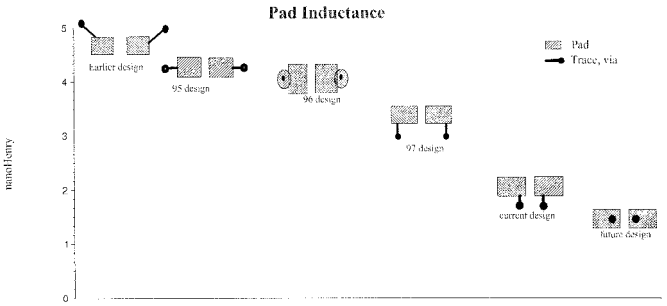
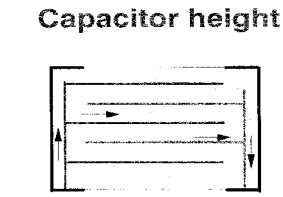


Fig. 9. Decoupling capacitor pad design progression. Pad lay out is a major contributor to capacitor inductance.

IV. HIGH FREQUENCY CERAMIC CAPACITANCE

High frequency ceramic capacitors are an increasingly important part of the PDS. Calculations for the number of capacitors necessary to maintain a target impedance are made in the frequency domain [3]. Ceramic capacitors come in several dielectric types (NPO, X7R, X5R, and Y5V) and several sizes (1206, 805, 603). NPO capacitors have the lowest ESR and best temperature and voltage properties, but are only available up to a few nF. X7R capacitors have reasonable voltage and temperature coefficients and are available from several nF to several μF. X5R is similar to X7R, but with reduced reliability and are being extended to 100 μF. Y5V



Thickness(mils)	Inductance(pH)
20	300
30	450
40	600
50	700

Fig. 10. Height adds inductance to the capacitor.

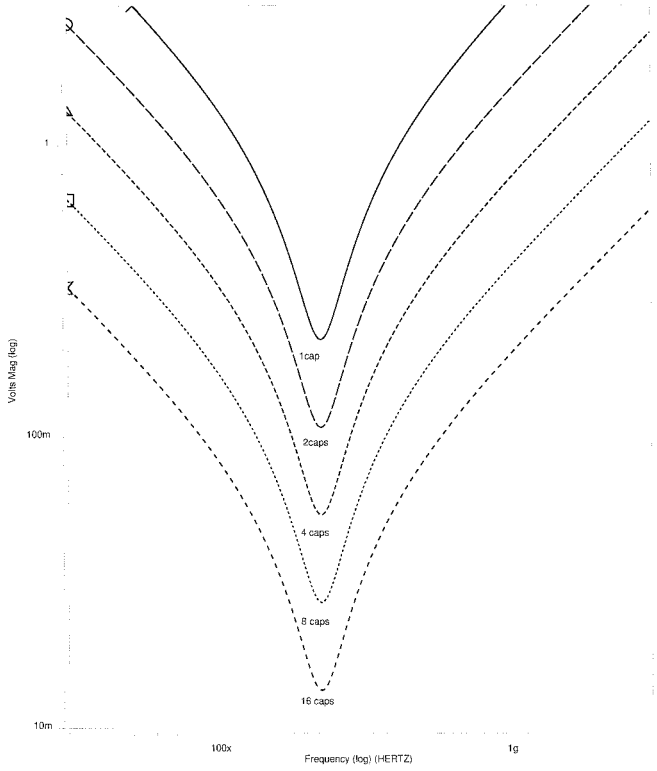


Fig. 11. Identical capacitors in parallel. The impedance is reduced by a factor of two every time the quantity is doubled.

dielectric is used to achieve high capacitance values, but has very poor voltage and temperature characteristics. Interesting thin film capacitor solutions are also becoming available [4].

A. Series RLC

A reasonable model for a ceramic capacitor is a resistance, inductance and capacitance (RLC) series circuit. Fig. 6(a) shows impedance versus frequency for capacitors 1 and 2, which are described in Table III.

The impedance profile for each capacitor comes down on two different capacitive lines. It bottoms out at the equivalent series resistance (ESR), then goes up on the line associated with the inductance. Capacitor 2 is said to have a higher Q because it is more peaked at the resonant frequency. Notice that the impedance at resonance of cap 2 is less than might have been expected if the $1/j\omega c$ capacitive line had continued.

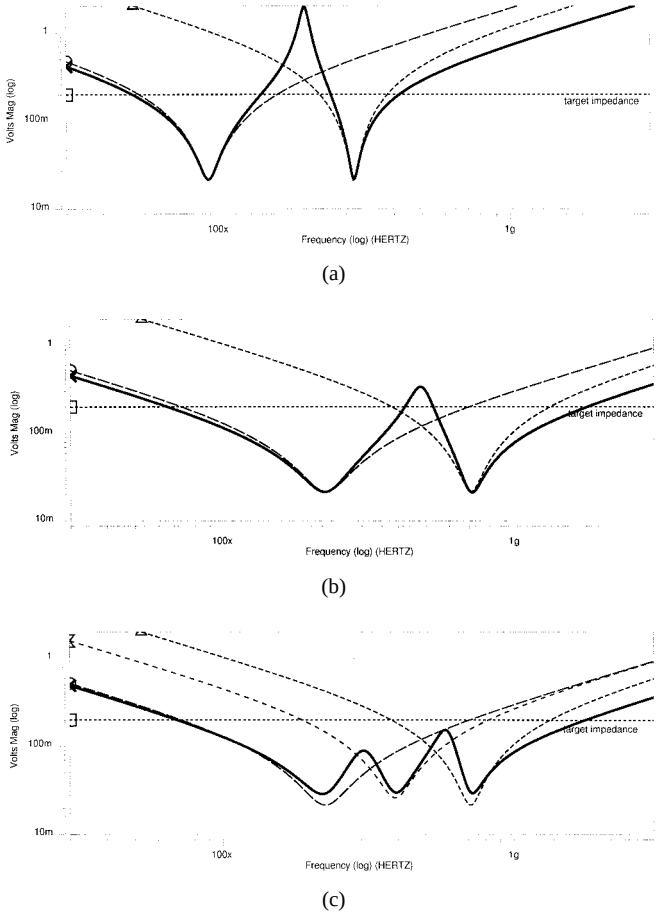


Fig. 12. Antiresonance of parallel capacitors with (a) inductive pads, (b) low inductance pads, and (c) with a third value of capacitor but the same quantity. Note that multiple values of capacitors on low inductance pads keeps the PDS below the target impedance for a broader frequency.

That is because the capacitive portion of the capacitor is in resonance with the inductive portion. Resonance occurs at the frequency

$$f_0 = \frac{1}{2\pi\sqrt{LC}}.$$

Fig. 6(b) shows impedance versus frequency for capacitors 2 and 3 which has less inductance. Once again, the ESR determines the depth of the curves. But notice that the impedance minimum for the less inductive capacitor is at much higher frequency. In fact the impedance is reduced at all higher frequencies because of the low inductance. The effectiveness of high frequency decoupling capacitors is greatly increased when the inductance is minimized.

V. ESR

An HP4291 is used to measure capacitor ESR by soldering it to an SMA connector and attaching it to a test head through an APC-7 connector. Measurements using a traditional surface mount test head may underestimate the actual ESR [5]. ESR and contact resistance are hard to separate and a portion of the ESR may be lost during the compensation process. Fig. 7 summarizes the ESR measured on capacitors from several different vendors. ESR is a strong function of dielectric type.

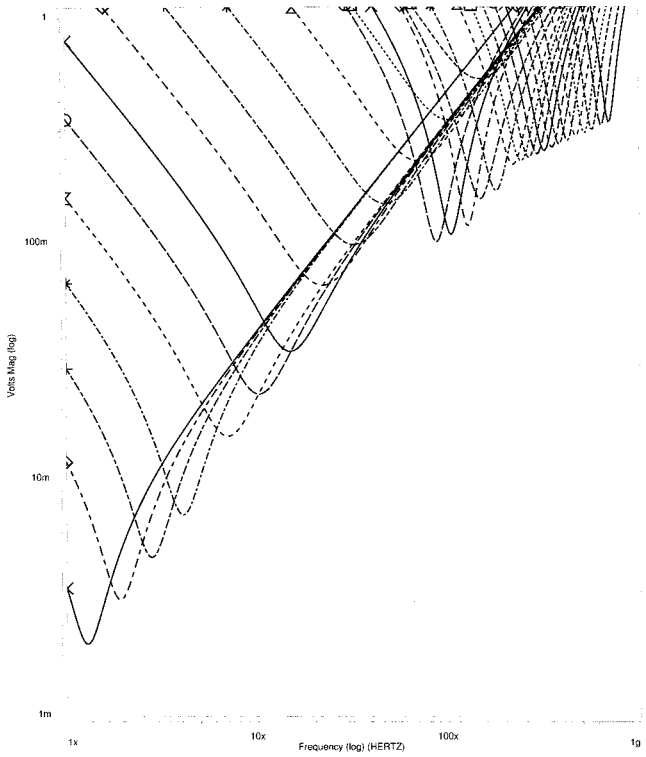


Fig. 13. Impedance versus frequency for several X7R and NPO capacitors.

Below 1 nF, it would take many X7R capacitors in parallel to achieve the impedance of a single NPO capacitor.

A. Inductance

Inductance is the amount of energy stored as magnetic flux (B field) in the environment as a result of current. A well known way to make an inductor is by a coil of wire or a loop, where flux is concentrated in the center. A good way to make the loop less inductive is to minimize the loop area. This holds true for decoupling capacitors. Fig. 8 shows how current travels in a loop when it leaves a power plane, travels through a via, pad structure, discrete decoupling capacitor, via, and returns to a ground plane. The inductance of the current loop is minimized by bringing the vias close together and reducing the height of the vias and capacitor.

Fig. 9 shows progress in decoupling capacitor pad design. In the early 1990's, decoupling capacitor inductance was dominated by pad layout. A small trace between the pads and vias adds an enormous amount of inductance. Vias should be brought as close together as possible to minimize inductance. The limiting factors are manufacturing issues at PCB fabrication and component assembly.

With optimum pad design, the dominant inductance is associated with the via and capacitor height. The vertical distance traversed by the current increases the size of the loop and therefor the inductance. The effect of capacitor height is shown in Fig. 10. By optimizing the pad design and minimizing the distance from the top of the capacitor to the power/Gnd plane pair, the inductance associated with decoupling capacitors is minimized.

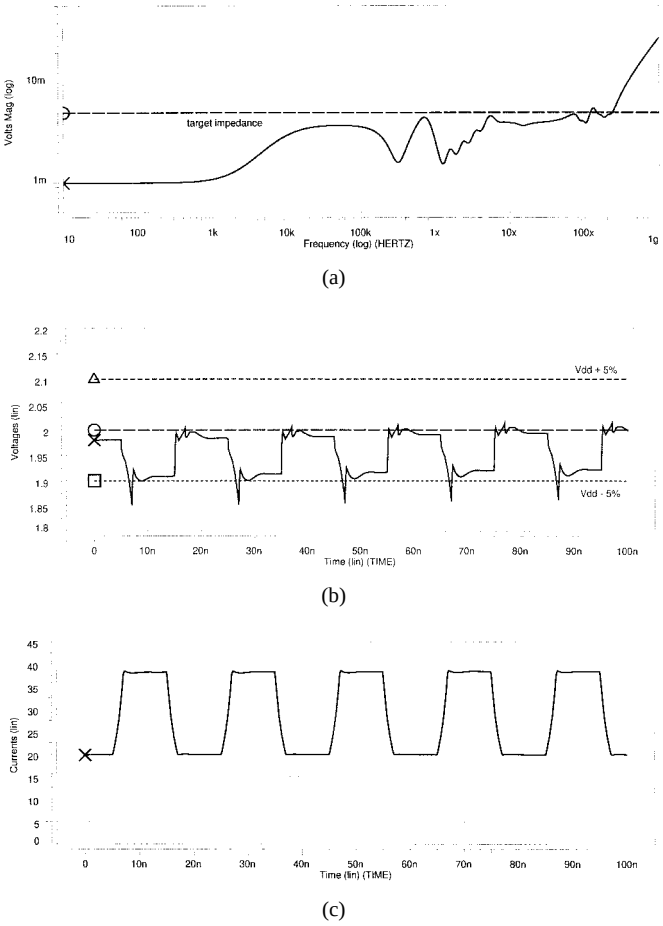


Fig. 14. (a) Impedance versus frequency of PDS with VRM, bulk and 115 ceramic capcitors, (b) time domain responce of same PDS to current transients, and (c) 50 MHz current transients with 2 ns rise times.

B. Parallel Capacitors

To achieve low PDS impedance, many capacitors are placed in parallel. Fig. 11 shows the effect of placing several identical capacitors in parallel. The impedance is reduced by a factor of two every time the number of capacitors is doubled. There is a diminishing return with each additional capacitor. It is desirable to start with low ESR capacitors in order to minimize the number of capacitors required to meet a target impedance.

Capacitors of different values in parallel produce an unwanted "antiresonance." A peaking between the two minimums of two capacitor values is shown in Fig. 12. The peak is higher than the impedance of either capacitor by itself. The peak is associated with the circuit that is formed after one capacitor has gone inductive and the other capacitor is still capacitive, the classic parallel LC tank circuit. The most effective way to reduce the height of the antiresonance is to minimize inductance. Fig. 12(b) shows the same set of capacitors mounted on low inductance pads. Notice that the peak is much lower compared to a target impedance. Large antiresonant peaks develop when low ESR capacitors are placed on inductive pads. High inductance and low resistance make a high Q circuit, $Q = L/R$. NPO capacitors have low ESR and should always be used on low inductance pads.

The antiresonance also becomes high if large gaps exist in capacitance value. Fig. 12(c) shows the same quantity of capacitors as Figs. 12(a) and (b), but they have been distributed over three values instead of two. The PDS impedance is below target over a broad frequency range. Antiresonances are effectively managed by using low inductance pad structures and many values of capacitors.

C. Ceramic Capacitor Methodology

Fig. 13 shows the impedance versus frequency for several X7R and NPO ceramic decoupling capacitors. Notice that as the capacitance value decreases and the resonant frequency goes up, the impedance minimum does not dip down as low. ESR is lower for high valued capacitors. A methodology has been developed for meeting aggressively low target impedances at high frequency. Capacitors are placed in parallel until their parallel impedance meets the target impedance. Advantage is taken of resonance where ESR is lower than the impedance that can be reached by either the capacitive or inductive curve for that capacitor. This is key in minimizing the number of capacitors necessary to keep a PDS below a target impedance. At high frequencies it becomes difficult to put enough capacitors in parallel to reach a target impedance. Capacitor effectiveness is optimized by using low inductance pads to increase the resonant frequency of a given capacitor.

Fig. 14 shows the characteristics of a PDS including the VRM, 7 bulk capacitors, and 115 ceramic capacitors. The capacitors were chosen to keep the PDS below the target impedance until nearly 200 MHz as shown on the frequency plot. The PDS has been stimulated with 20 A, 50 MHz current transients that have a rise time of 2 ns. The time domain response of the PDS shows that the voltage stays almost within 5% of the nominal V_{dd}, in spite of the very harsh transients. There is some energy above 200 MHz that is causing spikes to drop below the limit. Those spikes are dealt with most easily with power plane capacitance [6].

VI. Conclusion

Modern CMOS technology requires a low impedance power distribution system across a broad frequency range. The major components in this system include the voltage regulator module, bulk capacitors and high frequency ceramic capacitors. The impedance versus frequency of each of these components have been discussed and modeled in SPICE. A methodology for choosing the values and quantities of these components has been defined. The methodology involves placing a sufficient quantity of components in parallel to meet the target impedance in the frequency domain. Examples of frequency domain and time domain analysis has been given. If the target impedance is met in the frequency domain, noise in the time domain stays below a specified amount. Power distribution systems are efficiently designed using this methodology.

References

[1] L. D. Smith, "Packaging and power distribution design considerations for a Sun Micosystems desktop workstation," in *Proc. Elect. Perform. Elect. Packag. Conf.*, Oct. 1997.

[2] W. D. Becker *et al.*, "Modeling, simulation and measurement of mid-frequency simultaneous switch noise in computer systems," *IEEE Trans. Comp., Packag., Manufact. Technol. B*, vol. 21, pp. 286–297, May 1998.

[3] L. D. Smith, "Decoupling capacitor calculations for CMOS circuits," in *Proc. Elect. Perf. Elect. Packag. Conf.*, Nov. 1994.

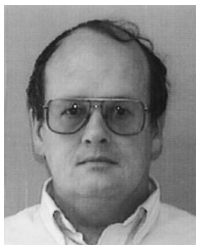
[4] K. Y. Chen, W. D. Brown, and L. W. Schaper, "Modeling and simulation of thin film decoupling capacitors," in *Proc. Elect. Perform. Electric. Packag. Conf.*, Oct. 1998.

[5] T. Roy, L. Smith, and J. Prymak, "ESR and ESL of ceramic capacitor applied to decoupling applications," in *Proc. Elect. Perform. Elect. Packag. Conf.*, Oct. 1998.

[6] L. D. Smith, R. Anderson, D. Forehand, N. O'Sullivan, and T. Roy, "Power plane analysis for CMOS power distribution systems," to be submitted.

Larry D. Smith (M'75) received the B.S.E.E. degree from the Rose Hulman Institute of Technology, Terre Haute, IN, in 1975 and the M.S. degree in material science from the University of Vermont, Burlington, in 1983.

After joining IBM in 1978, he worked in the areas of reliability, characterization, failure analysis, power supply, and analog circuit design, packaging, and signal integrity. He has continued his efforts in signal integrity at Sun Microsystems, Inc., Palo Alto, CA, since 1996. His current area of concentration is the design of power distribution systems.



Raymond E. Anderson (M'78) received the B.S. degree in computer science from National University, San Diego, CA.

After joining California Microwave in 1978, he was involved in microwave circuit design for terrestrial communication systems until 1982, and then with satellite system circuit design until 1993 with GTE Spacenet. Since joining Sun Microsystems, Inc., Palo Alto, CA, in 1993 he has been heavily involved in signal integrity issues. His areas of interest include measurement and modeling methodologies as well as the development of software analysis and synthesis tools to support signal integrity development efforts.

Douglas W. Forehand received the B.S.E.E. degree from Purdue University, West Lafayette, IN, in 1979.

He spent seven years with Computer Consoles, Inc., designing main memory and communication subsystems for their microcomputer line of products. In 1989, he joined Sun Microsystems, Inc., Palo Alto, CA, where he has worked on various board level designs with a focus on the signal integrity.

Thomas J. Pelc received the B.S.E.E. and the M.S.E.E. degrees from The Ohio State University, Columbus, in 1984 and 1986, respectively, and the M.B.A. degree from the University of La Verne, La Verne, CA, in 1991.

While at General Dynamics he worked on the Stinger Missile System performing failure analysis of targeting and guidance electronics. He is a Component Engineer at Sun Microsystems, Inc., Palo Alto, CA.

Tanmoy Roy (S'92–M'97) received the B.Tech. degree from the Indian Institute of Technology, Kharagpur, India, in 1990 and the M.S. and Ph.D. degrees in electrical engineering from Syracuse University, Syracuse, NY, 1993 and 1996, respectively.

From 1990 to 1991, he was a Software Consultant for Tata Consultancy Services, India. In 1994 and 1995, he was a Summer Intern at Schlumberger-Doll Research, Ridgefield, CT. Since 1997, he has worked as a Signal Integrity Engineer with Sun Microsystems, Inc., Palo Alto, CA. His main field of interest is numerical electromagnetics, in particular, finite-element techniques. His research deals with the design of power distribution systems.