Effective Radii of On-Chip Decoupling Capacitors

Mikhail Popovich, *Member, IEEE*, Michael Sotman, Avinoam Kolodny, *Member, IEEE*, and Eby G. Friedman, *Fellow, IEEE*

***Abstract*—Decoupling capacitors are widely used to reduce power supply noise. On-chip decoupling capacitors have traditionally been allocated into the white space available on a die or placed inside the rows in standard cell circuit blocks. The efficacy of on-chip decoupling capacitors depends upon the impedance of the power/ground lines connecting the capacitors to the current loads and power supplies. A design methodology for placing on-chip decoupling capacitors is presented in this paper. A maximum effective radius is shown to exist for each on-chip decoupling capacitor. Beyond this effective distance, a decoupling capacitor is ineffective. Depending upon the parasitic impedance of the power distribution system, the maximum voltage drop seen at the current load is caused either by the first droop (determined by the rise time) or by the second droop (determined by the transition time). Two criteria to estimate the minimum required on-chip decoupling capacitance are developed based on the critical parasitic impedance. In order to provide the required charge drawn by the load, the decoupling capacitor has to be charged before the next switching cycle. For an on-chip decoupling capacitor to be effective, both effective radii criteria should be simultaneously satisfied.**

***Index Terms*—Decoupling capacitors, power distribution systems, power supply noise, signal integrity.**

I. Introduction

THE feature size of integrated circuits (ICs) has been aggressively reduced in the pursuit of improved speed, power, and cost. The scaling of CMOS is expected to continue for at least another decade. Future nanometer circuits will soon contain more than a billion transistors and operate at clock speeds well over 10 GHz [1]. Distributing robust and reliable power and ground voltages in such a high speed, high complexity environment is, therefore, a highly challenging task [2].

Decoupling capacitors are widely used to manage power supply noise. A decoupling capacitor acts as a reservoir of charge, which is released when the power supply voltage at a particular current load drops below some tolerable level. Alternatively, decoupling capacitors are an effective way to reduce the impedance of power delivery systems operating at high frequencies [3]. The location of the decoupling capacitors significantly affects the design of the power/ground (P/G) network in high performance ICs such as microprocessors. With increasing frequencies, a distributed hierarchical system of decoupling capacitors placed on-chip is needed to effectively manage the power supply noise [4].

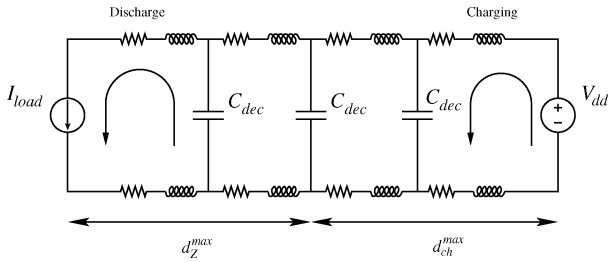


Fig. 1. Placement of an on-chip decoupling capacitor based on the maximum effective distance. To be effective, a decoupling capacitor should be placed close to the current load during discharge. During the charging phase, however, the decoupling capacitor should be placed close to the power supply to efficiently restore the charge on the capacitor. The specific location of a decoupling capacitor should therefore be determined to simultaneously satisfy the maximum effective distances $d_Z^{\max}$ during discharge and $d_{\mathrm{ch}}^{\max}$ during charging.

The efficacy of decoupling capacitors depends upon the impedance of the conductors connecting the capacitors to the current loads and power sources. During discharge, the current flowing from the decoupling capacitor to the current load results in resistive noise (IR drops) and inductive noise ($L(dI/dt)$ drops) due to the parasitic resistance and inductance of the power delivery network. The resulting voltage drop across the current load is therefore always greater than the voltage drop at the decoupling capacitor. A maximum parasitic impedance between the decoupling capacitor and the current load therefore exists for which the decoupling capacitor is effective. To be effective, a decoupling capacitor should therefore be placed close to a current load during discharge (within the maximum effective distance $d_Z^{\max}$), as shown in Fig. 1.

Once the switching event is completed, charge on the decoupling capacitor has to be restored before the next clock cycle begins. During the charging phase, the voltage across the decoupling capacitor rises exponentially. The maximum frequency at which the decoupling capacitor is effective is determined by the parasitic resistance and inductance of the metal lines and the size of the decoupling capacitor. A maximum effective distance based on the charge time, therefore, exists for each on-chip decoupling capacitor. Beyond this effective distance, the decoupling capacitor is ineffective. Alternatively, to be effective, an on-chip decoupling capacitor should be placed close to a power

Manuscript received January 23, 2007; revised March 23, 2007 and July 25, 2007. This research was supported in part by the Semiconductor Research Corporation under Contract 2004-TJ-1207, by the National Science Foundation under Contract CCR-0304574 and Contract CCF-0541206, by grants from the New York State Office of Science, Technology and Academic Research to the Center for Advanced Technology in Electronic Imaging Systems, and by grants from Intel Corporation, Eastman Kodak Company, Intrinsix Corporation, and Freescale Semiconductor Corporation.

M. Popovich is with CDMA Technologies, Qualcomm Corporation, San Diego, CA 92121 USA (e-mail: mikhailp@qualcomm.com).

E. G. Friedman is with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627 USA (e-mail: friedman@ece.rochester.edu).

M. Sotman and A. Kolodny are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: michael.sotman@intel.com; kolodny@ee.technion.ac.il).

Digital Object Identifier 10.1109/TVLSI.2008.2000454

supply during the charging phase (within the maximum effective distance $d_{\rm ch}^{\rm max}$, see Fig. 1). The relative location of the on-chip decoupling capacitors is therefore of fundamental importance. A design methodology is therefore required to determine the location of an on-chip decoupling capacitor, simultaneously satisfying the maximum effective distances, $d_Z^{\rm max}$ and $d_{\rm ch}^{\rm max}$. This location is characterized by the effective radii of the on-chip decoupling capacitors which is the primary subject of this paper. A design methodology to estimate the minimum required on-chip decoupling capacitance is also presented.

This paper is organized as follows. Existing work on placing on-chip decoupling capacitors is reviewed in Section II. The effective radius of an on-chip decoupling capacitor as determined by the target impedance in the time domain is presented in Section III. Design techniques to estimate the minimum magnitude of the required on-chip decoupling capacitance are discussed in Section IV. The effective radius of an on-chip decoupling capacitor based on the charge time in the time domain is determined in Section V. The activity radius of an on-chip decoupling capacitor in the frequency domain is analyzed in Section VI. Simulation results for typical values of on-chip parasitic resistances and inductances are presented in Section VII. Some circuit design implications are discussed in Section VIII. Finally, some specific conclusions are summarized in Section IX.

II. Background

Decoupling capacitors have traditionally been allocated on a circuit board to control the impedance of a power distribution system and suppress electro-magnetic interference (EMI). Decoupling capacitors are also employed to provide the required charge to the switching circuits, enhancing signal integrity. Since the parasitic impedance of a circuit board-based power distribution system is negligible at low frequencies, board decoupling capacitors are typically modeled as ideal capacitors without parasitic impedances. In an important early work by Smith [5], the effect of a decoupling capacitor on the signal integrity in circuit board-based power distribution systems is presented. The efficacy of the decoupling capacitors is analyzed in both the time and frequency domains. Design criteria have been developed, however, which significantly overestimate the required decoupling capacitance. A hierarchical placement of decoupling capacitors has been presented by Smith *et al.* in [6]. Smith *et al.* [6] show that each decoupling capacitor is effective only within a narrow frequency range. Larger decoupling capacitors have a greater form factor (physical dimension), resulting in a higher parasitic impedance [7]. The concept of an effective series resistance (ESR) and an effective series inductance (ESL) of each decoupling capacitor is also introduced in [6]. The authors show that by hierarchically placing the decoupling capacitors from the voltage regulator module level to the package level, the impedance of the overall power distribution system can be maintained below a target impedance.

As the signal frequency increases to several megahertz, the parasitic impedance of the circuit board decoupling capacitors becomes greater than the target impedance. The circuit board decoupling capacitors therefore become less effective at frequencies above 10 to 20 MHz. Package decoupling capacitors should therefore be utilized in the frequency range from several megahertz to several hundred megahertz [6]. In modern high performance ICs operating at several gigahertz, only those decoupling capacitors placed on-chip are effective at these frequencies.

Two types of on-chip decoupling capacitances can be described. An intrinsic decoupling capacitance (or symbiotic capacitance) is comprised of transistors, interconnect, and well-to-substrate capacitances [8]. Since the activity factor in digital circuits is typically low (10% to 30%), the intrinsic on-chip decoupling capacitance during a particular cycle is provided by the non-switching circuits. In contrast to the intrinsic capacitance, an intentional on-chip decoupling capacitance is often added. The intentional on-chip decoupling capacitance is typically an order of magnitude greater than the existing intrinsic capacitance. The intentional on-chip decoupling capacitance is therefore assumed in this paper to model all of the on-chip decoupling capacitance.

The placement of on-chip decoupling capacitors has been discussed in [9]. The power noise is analyzed assuming a resistance–inductance–capacitance (*RLC*) network model, representing a multi-layer power bus structure. The current load is modeled by time-varying resistors. The on-chip decoupling capacitors are allocated to only those areas where the power noise is greater than the maximum tolerable level. Ideal on-chip decoupling capacitors are assumed in the algorithm proposed in [9]. The resulting on-chip decoupling capacitance is therefore significantly overestimated. Another technique for placing on-chip decoupling capacitors has been described in [10]. The decoupling capacitors are placed based on activity signatures determined from microarchitectural simulations. The proposed technique produces a 30% decrease in the maximum noise level as compared to uniformly placing the on-chip decoupling capacitors. This methodology results in overestimating the capacitance budget due to the use of a simplified criterion for sizing the on-chip decoupling capacitors. Also, since the package level power distribution system is modeled as a single lumped resistance and inductance, the overall power supply noise is greatly underestimated.

An algorithm for automatically placing and sizing on-chip decoupling capacitors in application-specific integrated circuits is proposed in [11]. The problem is formulated as a nonlinear optimization and solved using a sensitivity-based quadratic programming solver. The proposed algorithm is limited to on-chip decoupling capacitors placed in rows of standard cells (in one dimension). The power distribution network is modeled as a resistive mesh, significantly underestimating the power distribution noise. In [12], the problem of on-chip decoupling capacitor allocation is investigated. The proposed technique is integrated into a power supply noise-aware floorplanning methodology. Only the closest power supply pins are considered to provide the switching current drawn by the load. Additionally, only the shortest and second shortest paths are considered between a decoupling capacitor and the current load. It is assumed that the current load is located at the center of a specific circuit block. The technique does not consider the degradation in effectiveness of an on-chip decoupling capacitor located at some distance from the current load. Moreover, only the discharge phase is considered. To be effective, a decoupling capacitor should be fully charged before the following switching cycle. Otherwise,

the charge on the decoupling capacitor will be gradually depleted, making the capacitor ineffective. The methodology described in [12] therefore results in underestimating the power supply noise and overestimating the required on-chip decoupling capacitance.

The problem of on-chip decoupling capacitor allocation has historically been considered as two independent tasks. The location of an on-chip decoupling capacitor is initially determined. The decoupling capacitor is next appropriately sized to provide the required charge to the current load. As discussed in [13], the size of the on-chip decoupling capacitors is determined by the impedance (essentially, the physical separation) between a decoupling capacitor and the current load (or power supply).

Proper sizing and placement of the on-chip decoupling capacitors however should be determined simultaneously. As shown in this paper, on-chip decoupling capacitors are only effective in close vicinity to the switching circuit. The maximum effective distance for both the discharge and charging phases is determined. It is also shown that the on-chip decoupling capacitors should be placed both close to the current load to provide the required charge and to the power supply to be fully recharged before the next switching event. A design methodology for placing and sizing on-chip decoupling capacitors based on a maximum effective distance as determined by the target impedance and charge time is the primary result presented in this paper.

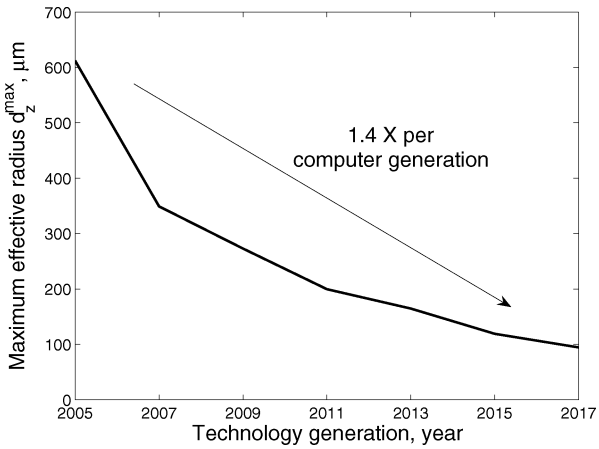


Fig. 2. Projection of the maximum effective radius as determined by the target impedance $d_Z^{\max}$ for future technology generations: $I_{\max} = 10$ mA, $V_{\rm dd} = 1$ V, and $\mathrm{Ripple} = 0.1$. Global on-chip interconnects are assumed, modeling the highly optimistic scenario. The maximum effective radius as determined by the target impedance is expected to decrease at an alarming rate (a factor of 1.4 on average per computer generation).

III. Effective Radius of On-Chip Decoupling Capacitor Based on a Target Impedance

Neglecting the parasitic capacitance [14], the impedance of a unit length wire is $Z'(\omega) = r + j\omega l$, where r and l are the resistance and inductance per length, respectively, and ω is an equivalent frequency, as determined by the rise time of the current load. The inductance l is the effective inductance per unit length of the power distribution grid, incorporating both the partial self-inductance and mutual coupling among the lines [15]. The target impedance of the metal line of a particular length is therefore

$$Z(\omega) = Z'(\omega) \times d \tag{1}$$

where $Z'(\omega)$ is the impedance of a unit length metal line and d is the distance between the decoupling capacitor and the current load. Substituting the expression for the target impedance $Z_{\rm target}$ presented in [3] into (1), the maximum effective radius $d_Z^{\max}$ between the decoupling capacitor and the current load is

$$d_Z^{\max} = \frac{Z_{\rm target}}{Z'(\omega)} = \frac{V_{\rm dd} \times \mathrm{Ripple}}{I \times \sqrt{r^2 + \omega^2 l^2}} \tag{2}$$

where $\sqrt{r^2 + \omega^2 l^2}$ denotes the magnitude of the impedance of a unit length wire, $Z_{\rm target}$ is the maximum impedance of a power distribution system, resulting in a power noise lower than the maximum tolerable level, and Ripple is the maximum tolerable power noise (the ratio of the magnitude of the maximum tolerable voltage drop to the power supply level). Note that the maximum effective radius as determined by the target impedance is inversely proportional to the magnitude of the current load and the impedance of a unit length line. Also note that the per length resistance r and inductance l account for the ESR and ESL of an on-chip decoupling capacitor. The maximum effective radius as determined by the target impedance decreases rapidly with each technology generation (a factor of 1.4, on average, per computer generation), as shown in Fig. 2 [1]. Also note that in a meshed structure, multiple paths between any two points are added in parallel. The maximum effective distance corresponding to $Z_{\rm target}$ is, therefore, larger than the maximum effective distance of a single line, as discussed in Section VII. The maximum effective radius is defined in this paper as follows.

Definition 1: The effective radius of an on-chip decoupling capacitor is the maximum distance between the current load (power supply) and the decoupling capacitor for which the capacitor is capable of providing sufficient charge to the current load, while maintaining the overall power distribution noise below a tolerable level.

IV. Estimation of Required On-Chip Decoupling Capacitance

Once the specific location of an on-chip decoupling capacitor during discharge is determined as described in Section III, the minimum required magnitude of the on-chip decoupling capacitance should be determined, satisfying the expected current demands. Design expressions for determining the required magnitude of the on-chip decoupling capacitors based on the dominant voltage droop are presented in this section. A conventional approach with the minimum supply voltage achieved at the end of a transition is described in Section IV-A. Techniques for determining the magnitude of the on-chip decoupling capacitors in the case of the first dominant droop are developed in Section IV-B. The critical length of the P/G paths connecting the decoupling capacitor to the current load is presented in Section IV-C.

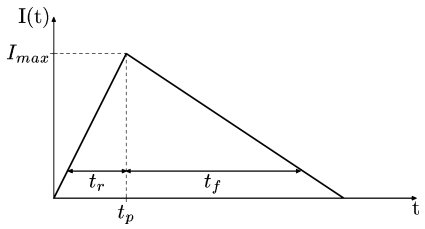


Fig. 3. Linear approximation of the current demand of a power distribution network by a current source. The magnitude of the current source reaches the maximum current $I_{\max}$ at peak time t_p. t_r and t_f denote the rise and fall time of the current load, respectively.

A. Second Dominant Droop

To estimate the on-chip decoupling capacitance required to support a specific local current demand, the current load is modeled as a triangular current source. The magnitude of the current source increases linearly, reaching the maximum current $I_{\max}$ at peak time t_p. The magnitude of the current source decays linearly, becoming zero at t_f, as shown in Fig. 3. The on-chip power distribution network is modeled as a series *RL* circuit. To qualitatively illustrate the proposed methodology for placing on-chip decoupling capacitors based on the maximum effective radii, a single decoupling capacitor with a single current load is assumed to mitigate the voltage fluctuations across the P/G terminals.

The total charge $Q_{\rm dis}$ required to satisfy the current demand during a switching event is modeled as the sum of the area of two triangles (see Fig. 3). Since the required charge is provided by an on-chip decoupling capacitor, the voltage across the capacitor during discharge drops below the initial power supply voltage. The required charge during the entire switching event is thus[1]

$$Q^f_{\rm dis} = \frac{I_{\max} \times (t_r + t_f)}{2} = C_{\rm dec} \times (V_{\rm dd} - V^f_C) \qquad (3)$$

where $I_{\max}$ is the maximum magnitude of the current load of a specific circuit block for which the decoupling capacitor is allocated, t_r and t_f are the rise and fall time, respectively, $C_{\rm dec}$ is the decoupling capacitance, $V_{\rm dd}$ is the power supply voltage, and V^f_C is the voltage across the decoupling capacitor after the switching event. Note that since there is no current after switching, the voltage at the current load equals the voltage across the decoupling capacitor.

The voltage fluctuations across the P/G terminals of a power delivery system should not exceed the maximum level (usually 10% of the power supply voltage) to guarantee fault-free operation. Thus

$$V^f_C \equiv V^f_{\rm load} \geq 0.9\, V_{\rm dd}. \qquad (4)$$

Substituting (4) into (3) and solving for $C_{\rm dec}$, the minimum on-chip decoupling capacitance required to support the current demand during a switching event is

$$C^f_{\rm dec} \geq \frac{I_{\max} \times (t_r + t_f)}{0.2\, V_{\rm dd}}. \qquad (5)$$

[1]In the general case with an *a priori* determined current profile, the required charge can be estimated as the integral of $I_{\rm load}(t)$ from 0 to t_f.

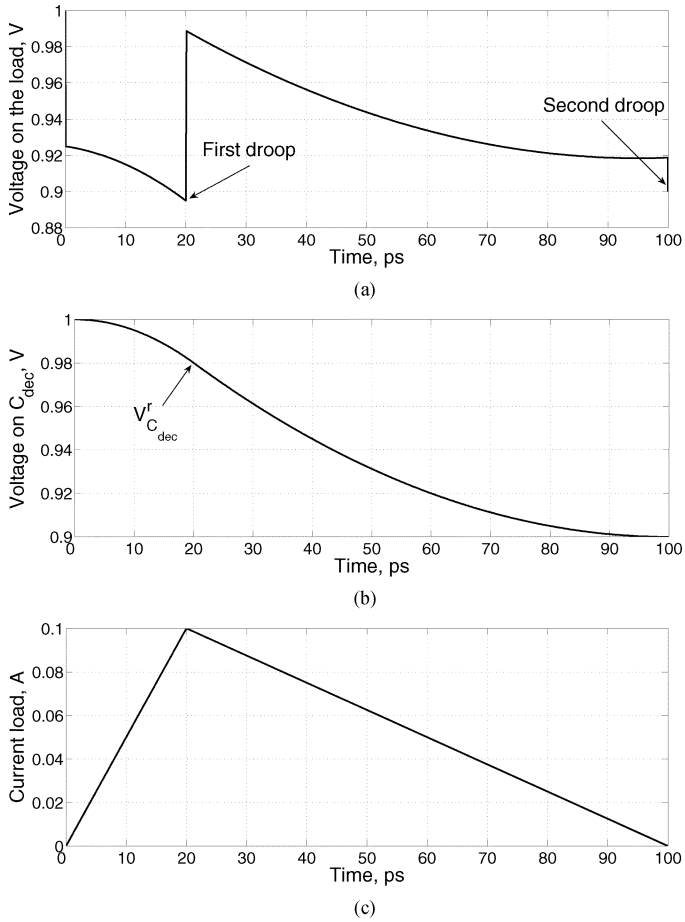


Fig. 4. Power distribution noise during discharge of an on-chip decoupling capacitor: $I_{\max} = 100$ mA, $V_{\rm dd} = 1$ V, $t_r = 20$ ps, $t_f = 80$ ps, $R = 100$ mΩ, $L = 15$ pH, and $C_{\rm dec} = 50$ pF. (a) Voltage across the terminals of the current load. (b) Voltage across the decoupling capacitor. (c) Current load modeled as a triangular current source. For these parameters, the parasitic impedance of the metal lines connecting the decoupling capacitor to the current load is larger than the critical impedance. The first droop is therefore larger than the second droop and (5) underestimates the required decoupling capacitance. The resulting voltage drop on the power terminal of a current load is therefore larger than the maximum tolerable noise.

B. First Dominant Droop

Note that (5) is applicable only to the case where the voltage drop at the end of the switching event is larger than the voltage drop at the peak time t_p. This phenomenon can be explained as follows. The voltage drop as seen at the current load is caused by current flowing through the parasitic resistance and inductance of the on-chip power distribution system. The resulting voltage fluctuations are the sum of the ohmic IR voltage drop, inductive $L(dI/dt)$ voltage drop, and the voltage drop across the decoupling capacitor at t_p. A critical parasitic *RL* impedance, therefore, exists for a specific set of rise and fall times. Beyond this critical impedance, the first droop becomes larger than the second droop, as shown in Fig. 4. The decoupling capacitor should therefore be increased to reduce the voltage drop across the capacitor during the rise time V^r_C, lowering the magnitude of the power noise.

The charge $Q^r_{\rm dis}$ required to support the current demand during the rise time of the current load is equal to the area of the triangle formed by $I_{\max}$ and t_r. The required charge is provided by the on-chip decoupling capacitor. The voltage

across the decoupling capacitor drops below the power supply level by ΔV_C^r. The required charge during t_r is[2]

$$Q_{\rm dis}^r = \frac{I_{\max} \times t_r}{2} = C_{\rm dec} \times \Delta V_C^r \tag{6}$$

where $Q_{\rm dis}^r$ is the charge drawn by the current load during t_r and ΔV_C^r is the voltage drop across the decoupling capacitor at t_p. From (6)

$$\Delta V_C^r = \frac{I_{\max} \times t_r}{2\, C_{\rm dec}}. \tag{7}$$

By time t_p, the voltage drop as seen from the current load is the sum of the ohmic IR drop, the inductive $L(dI/dt)$ drop, and the voltage drop across the decoupling capacitor. Alternatively, the power noise is further increased by the voltage drop ΔV_C^r. In this case, the voltage at the current load is

$$V_{\rm load}^r = V_{\rm dd} - I \times R - L\,\frac{dI}{dt} - \Delta V_C^r \tag{8}$$

where R and L are the parasitic resistance and inductance of the P/G lines, respectively. Linearly approximating the current load, dI is assumed equal to $I_{\max}$ and dt to t_r.

Assuming that $V_{\rm load}^r \geq 0.9\, V_{\rm dd}$, substituting (7) into (8), and solving for $C_{\rm dec}$, the minimum on-chip decoupling capacitance to support the current demand during t_r is

$$C_{\rm dec}^r \geq \frac{I_{\max} \times t_r}{2\,\left(0.1\, V_{\rm dd} - I \times R - L\,\frac{dI}{dt}\right)}. \tag{9}$$

Note that if the first droop dominates, $C_{\rm dec}$ is excessively large. The voltage drop at the end of the switching event is hence always smaller than the maximum tolerable noise.

Also note that, as opposed to (5), (9) depends upon the parasitic impedance of the on-chip power distribution system. Alternatively, in the case of the first dominant droop, the required charge released by the decoupling capacitor is determined by the parasitic resistance and inductance of the P/G lines connecting the decoupling capacitor to the current load.

C. Critical Line Length

Assuming the impedance of a single line, the critical line length $d_{\rm crit}$ can be determined by setting $C_{\rm dec}^r$ equal to $C_{\rm dec}^f$

$$\frac{I_{\max} \times t_r}{\left(0.1\, V_{\rm dd} - I\, r\, d_{\rm crit} - l\, d_{\rm crit}\,\frac{dI}{dt}\right)} = \frac{I_{\max} \times (t_r + t_f)}{0.1\, V_{\rm dd}}. \tag{10}$$

Solving (10) for $d_{\rm crit}$

$$d_{\rm crit} = \frac{0.1\, V_{\rm dd}\,\left(1 - \frac{t_r}{t_r + t_f}\right)}{I\, r + l\,\frac{dI}{dt}}. \tag{11}$$

For a single line connecting a current load to a decoupling capacitor, the minimum required on-chip decoupling capacitor is determined by (5) for lines shorter than $d_{\rm crit}$ and by (9) for lines longer than $d_{\rm crit}$, as illustrated in Fig. 5. Note that for a line length equal to $d_{\rm crit}$, (5) and (9) result in the same required capacitance. Also note that the maximum length of a single line is determined by (2). A closed-form solution for the critical line length has not been developed for the case of multiple current paths existing between the current load and a decoupling capacitor. In this case, the impedance of the power grid connecting a decoupling capacitor to a current load is extracted and compared to the critical impedance. Either (5) or (9) is utilized to estimate the required on-chip decoupling capacitance.

[2]In the general case with a given current profile, the required charge can be estimated as the integral of $I_{\rm load}(t)$ from 0 to t_r.

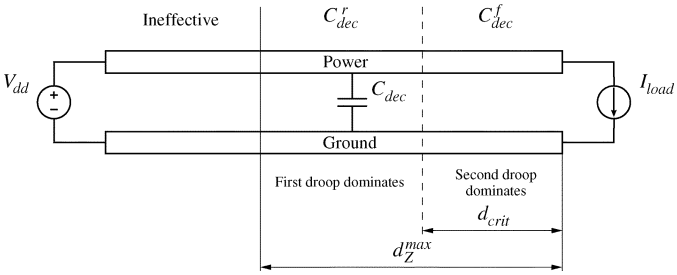


Fig. 5. Critical line length of an interconnect between a decoupling capacitor and a current load. The minimum required on-chip decoupling capacitance is determined by (5) for lines shorter than $d_{\rm crit}$ and by (9) for lines longer than $d_{\rm crit}$. The decoupling capacitor is ineffective beyond the maximum effective radius as determined by the target impedance $d_Z^{\max}$.

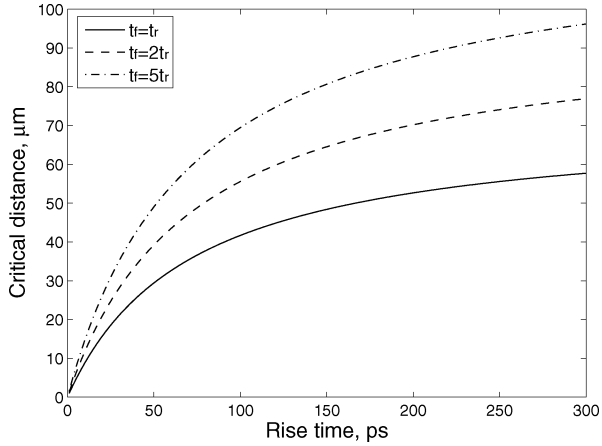


Fig. 6. Dependence of the critical line length $d_{\rm crit}$ on the rise time of the current load: $I_{\max} = 0.1$ A, $V_{\rm dd} = 1$ V, $r = 0.007\ \Omega/\mu$m, and $l = 0.5$ pH/μm. Note that $d_{\rm crit}$ is determined by t_r/t_f, increasing with larger t_f. The critical line length will shrink in future nanometer technologies as transition times become shorter.

The dependence of the critical line length $d_{\rm crit}$ on the rise time t_r of the current load as determined by (11) is depicted in Fig. 6. From Fig. 6, the critical line length decreases sublinearly with shorter rise times. Hence, the critical line length will decrease in future nanometer technologies as transition times become shorter, significantly increasing the required on-chip decoupling capacitance. Also note that $d_{\rm crit}$ is determined by t_r/t_f, increasing with larger fall times.

Observe in Fig. 5 that the design space for determining the required on-chip decoupling capacitance is broken into two regions by the critical line length. The design space for determining the required on-chip decoupling capacitance ($C_{\rm dec}^r$ and $C_{\rm dec}^f$) is depicted in Fig. 7. For the example parameters shown in Fig. 7, the critical line length is 125 μm. Note that the required on-chip decoupling capacitance $C_{\rm dec}^r$ depends upon the

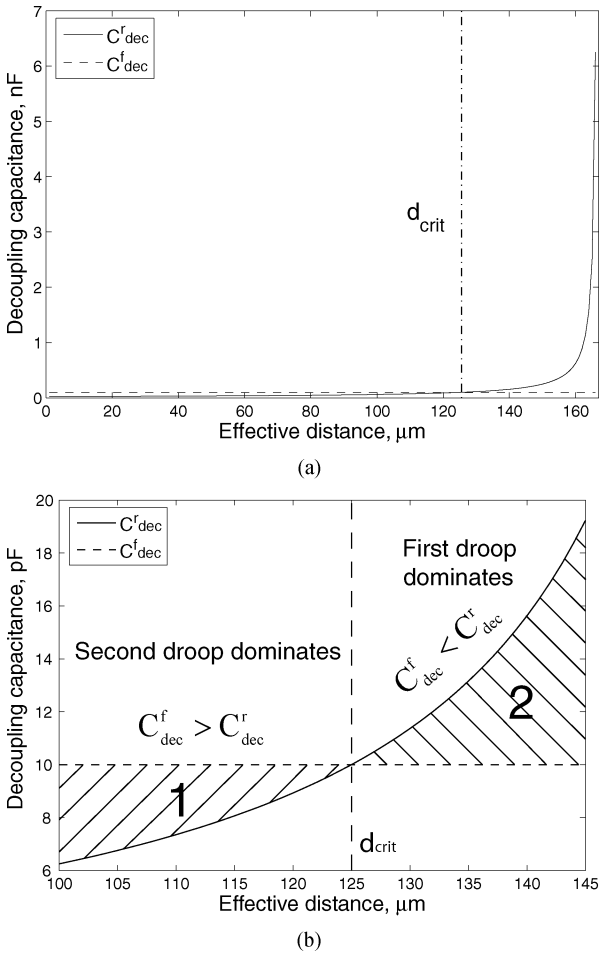


Fig. 7. Design space for determining the minimum required on-chip decoupling capacitance: $I_{\max} = 50$ mA, $V_{\rm dd} = 1$ V, $r = 0.007\ \Omega/\mu$m, $l = 0.5$ pH/μm, $t_r = 100$ ps, and $t_f = 300$ ps. (a) The design space for determining the minimum required on-chip decoupling capacitance is broken into two regions by $d_{\rm crit}$. (b) The design space around $d_{\rm crit}$. For the example parameters, the critical line length is 125 μm. In region 1, $C^f_{\rm dec}$ is greater than $C^r_{\rm dec}$ and does not depend upon the parasitic impedance. In region 2, however, $C^r_{\rm dec}$ dominates, increasing rapidly with distance between the decoupling capacitor and the current load.

parasitic impedance of the metal lines connecting the decoupling capacitor to the current load. Thus, for lines longer than $d_{\rm crit}$, $C^r_{\rm dec}$ increases exponentially as the separation between the decoupling capacitor and the current load increases, as shown in Fig. 7(a). Also note that for lines shorter than $d_{\rm crit}$, the required on-chip decoupling capacitance does not depend upon the parasitic impedance of the power distribution grid. Alternatively, in the case of the second dominant droop, the required on-chip decoupling capacitance $C^f_{\rm dec}$ is constant and greater than $C^r_{\rm dec}$ [see region 1 in Fig. 7(b)]. If the first droop dominates (the line length is greater than $d_{\rm crit}$), the required on-chip decoupling capacitance $C^r_{\rm dec}$ increases substantially with line length and is greater than $C^f_{\rm dec}$ [see region 2 in Fig. 7(b)]. Conventional techniques therefore significantly underestimate the required decoupling capacitance in the case of the first dominant droop. Note that in region 1, the parasitic impedance of the metal lines connecting a decoupling capacitor to the current load is not important. In region 2, however, the parasitic impedance of the P/G

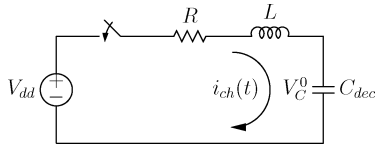


Fig. 8. Circuit charging an on-chip decoupling capacitor. The parasitic impedance of the power distribution system connecting the decoupling capacitor to the power supply is modeled by a series *RL* circuit.

lines should be considered. A tradeoff therefore exists between the size of $C^r_{\rm dec}$ and the distance between the decoupling capacitor and the current load. As $C^r_{\rm dec}$ is placed closer to the current load, the required capacitance can be significantly reduced.

V. Effective Radius as Determined by Charge Time

Once discharged, a decoupling capacitor must be fully charged to support the current demands during the following switching event. If charge on the capacitor is not fully restored during the relaxation time between two consecutive switching events (the charge time), the decoupling capacitor will be gradually depleted, becoming ineffective after several clock cycles. A maximum effective radius from the power supply, therefore, exists for an on-chip decoupling capacitor as determined during the charging phase for a target charge time. Similar to the effective radius based on the target impedance presented in Section III, an on-chip decoupling capacitor should be placed in close proximity to the power supply (the power pins) to be effective.

To determine the current flowing through a decoupling capacitor during the charging phase, the parasitic impedance of a power distribution system is modeled as a series *RL* circuit between the decoupling capacitor and the power supply, as shown in Fig. 8. When the discharge is completed, the switch is closed and the charge is restored on the decoupling capacitor. The initial voltage V^0_C across the decoupling capacitor is determined by the maximum voltage drop during discharge.

Assuming an overdamped power distribution system with an on-chip decoupling capacitor (which is typical for lines longer than several micrometers in a 90-nm CMOS technology), applying the initial conditions to the Kirchhoff's voltage law (KVL) equation [16] for the current in the charging circuit (see Fig. 8), and integrating the current supplied to a decoupling capacitor during the charging phase from zero to the charge time, the voltage across the decoupling capacitor during the charging phase is determined by (12)

$$V_{C_{\rm dec}}(t_{\rm ch}) = \frac{I_{\max}(t_r+t_f)}{4C^2_{\rm dec}L\sqrt{\left(\frac{R}{2L}\right)^2-\frac{1}{LC_{\rm dec}}}} \times \left\{ \frac{\exp\left[\left(-\frac{R}{2L}+\sqrt{\left(\frac{R}{2L}\right)^2-\frac{1}{LC_{\rm dec}}}\right)t_{\rm ch}\right]-1}{-\frac{R}{2L}+\sqrt{\left(\frac{R}{2L}\right)^2-\frac{1}{LC_{\rm dec}}}} + \frac{1-\exp\left[\left(-\frac{R}{2L}-\sqrt{\left(\frac{R}{2L}\right)^2-\frac{1}{LC_{\rm dec}}}\right)t_{\rm ch}\right]}{-\frac{R}{2L}-\sqrt{\left(\frac{R}{2L}\right)^2-\frac{1}{LC_{\rm dec}}}} \right\}. \quad (12)$$

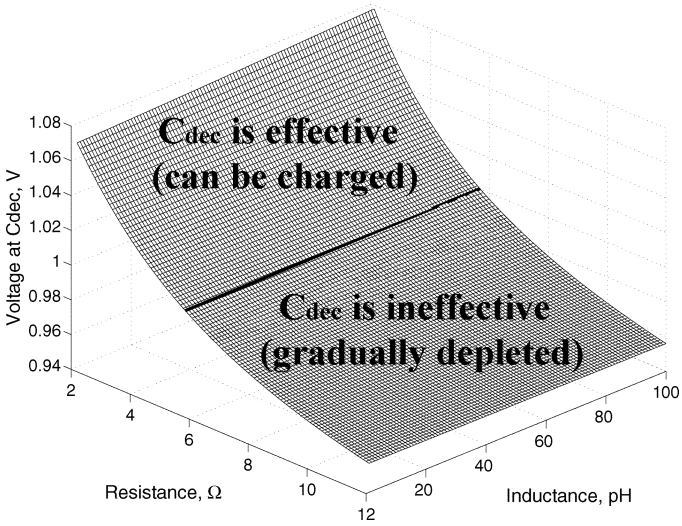


Fig. 9. Design space for determining the maximum tolerable parasitic resistance and inductance of a power distribution grid: $I_{\max} = 100$ mA, $t_r = 100$ ps, $t_f = 300$ ps, $C_{\rm dec} = 100$ pF, $V_{\rm dd} = 1$ V, and $t_{\rm ch} = 400$ ps. For a target charge time, the maximum resistance and inductance produce a voltage across the decoupling capacitor that is greater or equal to the power supply voltage (region above the dark line). Note that the maximum voltage across the decoupling capacitor is the power supply voltage. A design space that produces a voltage greater than the power supply means that the charge on the decoupling capacitor can be restored within $t_{\rm ch}$.

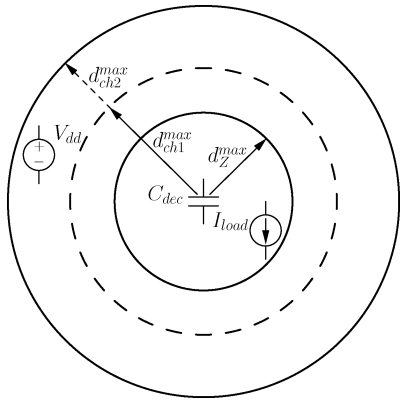


Fig. 10. Effective radii of an on-chip decoupling capacitor. The on-chip decoupling capacitor is placed to ensure that both the current load and the power supply are located inside the effective radius. The maximum effective radius as determined by the target impedance $d_Z^{\max}$ does not depend on the decoupling capacitance. The maximum effective radius as determined by the charge time is inversely proportional to $C_{\rm dec}^2$. If the power supply is located outside the effective radius $d_{\rm ch1}^{\max}$, the current load should be partitioned, resulting in a smaller decoupling capacitor and, therefore, an increased effective distance $d_{\rm ch2}^{\max}$.

Observe that the criterion for estimating the maximum effective radius of an on-chip decoupling capacitor as determined by the charge time is transcendental. A closed-form expression is therefore not available for determining the maximum effective radius of an on-chip decoupling capacitor during the charging phase. Thus, from (12), a design space can be graphically described in order to determine the maximum tolerable resistance and inductance that permit the decoupling capacitor to be recharged within a given $t_{\rm ch}$, as shown in Fig. 9. The parasitic resistance and inductance should be maintained below the maximum tolerable values, permitting the decoupling capacitor to be charged during the relaxation time.

Note that as the parasitic resistance of the power delivery network decreases, the voltage across the decoupling capacitor increases exponentially. In contrast, the voltage across the decoupling capacitor during the charging phase is almost independent of the parasitic inductance, slightly increasing with inductance. This phenomenon is due to the behavior that an inductor resists sudden changes in the current. Alternatively, an inductor maintains the charging current at a particular level for a longer time. Thus, the decoupling capacitor is charged faster.

A design methodology for placing on-chip decoupling capacitors is as follows. The maximum effective radius based on the target impedance is determined from (2) for a particular current load (circuit block), power supply voltage, and allowable ripple. The minimum required on-chip decoupling capacitance is estimated to support the required current demand. If the second droop dominates (the minimum supply voltage is achieved at the end of the switching event), (5) is used to determine the required on-chip decoupling capacitance. If the first droop dominates (the minimum supply voltage is achieved at the rise time), the on-chip decoupling capacitance is determined by (9). In the case of a single line connecting a decoupling capacitor to a current load, the critical wire length is determined by (11).

The maximum effective distance based on the charge time is determined from (12). Note that (12) results in a range of tolerable parasitic resistance and inductance of the metal lines connecting the decoupling capacitor to the power supply. Also note that the on-chip decoupling capacitor should be placed to ensure that both the power supply and the current load are located within the respective effective radius, as shown in Fig. 10. If this allocation is not possible, the current load (circuit block) should be partitioned into several blocks and on-chip decoupling capacitors should be allocated for each block, satisfying both effective radii requirements. The effective radius as determined by the target impedance does not depend upon the decoupling capacitance. In contrast, the effective radius as determined by the charge time is inversely proportional to $C_{\rm dec}^2$. The on-chip decoupling capacitors should therefore be distributed across a circuit to provide sufficient charge for each functional unit.

VI. Activity Radius of On-Chip Decoupling Capacitor

To characterize the activity radius of an on-chip decoupling capacitor in the frequency domain, the on-chip decoupling capacitance is assumed to be uniformly distributed across the die with a density $C_{\rm die}$ per unit area. The power grid connects the decoupling capacitors to the current loads, adding a local impedance in series with remote portions of $C_{\rm die}$, as illustrated in Fig. 11. This additional impedance reduces the efficiency of the distant on-chip decoupling capacitors, making the charge supplied to the switching circuit from the distant decoupling capacitors negligible.

The activity radius of the distributed on-chip decoupling capacitors in the frequency domain is discussed in this section. An analysis of the on-chip power distribution grid is presented in Section VI-A. The activity radius of the on-chip decoupling capacitors is determined in Section VI-B. Design techniques for estimating the effective on-chip decoupling capacitance available to the switching circuits are described in Section VI-C.

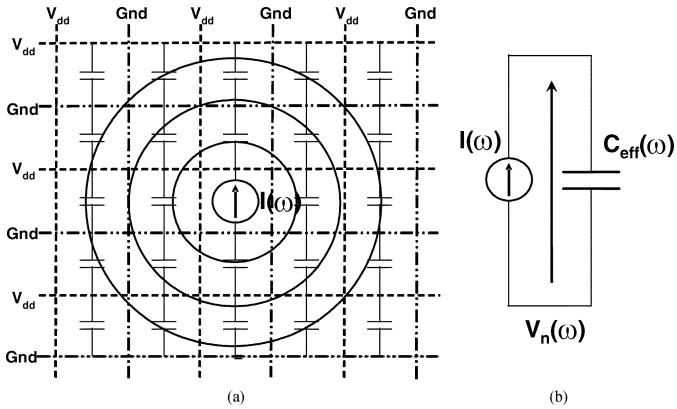


Fig. 11. Simplified representation of an on-chip decoupling capacitance. (a) The power distribution network and on-chip decoupling capacitors distributed around a single switching circuit modeled by a current source. (b) A model of a frequency-dependent effective capacitance.

A. Power Grid Analysis

An on-chip power distribution grid in high performance digital ICs is commonly structured as a multilayer grid [2]. In such a grid, parallel P/G lines in each metalization layer span the entire die (or large functional unit) and are orthogonal to the lines in adjacent layers. The power and ground lines typically alternate in each layer [17]. Vias connect a power (ground) line to another power (ground) line at overlap sites. For simplicity, an on-chip power distribution grid composed of only two metal layers is considered in this subsection. Neglecting the parasitic capacitance, the on-chip power distribution grid is modeled as an *RL* mesh, as shown in Fig. 12.

All of the circuit elements including the current loads and decoupling capacitors are connected between the power and ground paths. A simplified model of a single grid segment is illustrated in Fig. 12(b). The size of each segment is determined by the interconnect section between two adjacent vias. The power distribution grid in typical high performance ICs includes multiple vias. The number of grid cells, therefore, can be extremely large.

A dense multi-segment power distribution grid can be mathematically approximated, behaving as a continuous structure, as depicted in Fig. 12(c). Observe that all of the points connected to $V_{\rm dd}$ and $\rm Gnd$ are assumed to reside in virtually continuous planes. Note that such points [numbered in Fig. 12(b)] reside in both metal layers of the original grid [see Fig. 12(a)]. The power and ground lines are modeled as continuous planes with a sheet impedance equivalent to the metal lines in the original grid. The same approach is applicable to the on-chip decoupling capacitors. In contrast to an *RL* mesh, a continuous structure can be characterized by a continuous current density at each point rather than by a number of discrete branch currents.

Since the proposed model is linear, the system can be analyzed in the frequency domain. The planes are therefore modeled by an effective sheet impedance ($\Omega/\square$) at a particular frequency

$$\rho_{S_{\rm grid}} = R_{\rm grid} + j\omega L_{\rm grid} \tag{13}$$

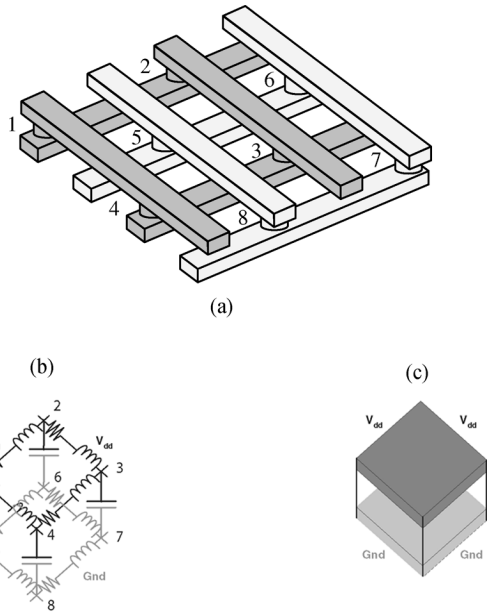


Fig. 12. On-chip power distribution grid. (a) Physical structure. Two metal layers are shown. The power lines are dark grey and the ground lines are light grey. (b) Circuit model of the grid (a single segment of a dense *RL* mesh is depicted). (c) Approximation of a power distribution grid by a continuous structure.

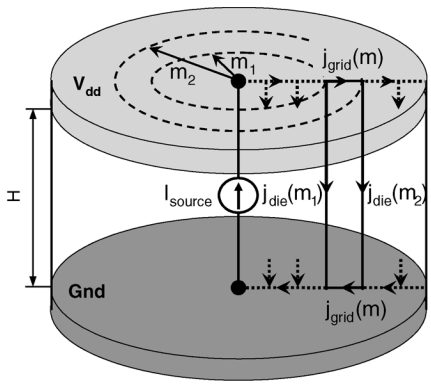


Fig. 13. Equivalent model of a power distribution grid with a uniformly distributed decoupling capacitance.

where $R_{\rm grid}$ and $L_{\rm grid}$ are the characteristic resistance and inductance of the power and ground grid segments, respectively, and ω is the switching frequency. Similar to the power grid, the impedance of the decoupling capacitance is modeled as a layer of material with an equivalent resistivity $\rho_{\rm die}$ and length H. Assuming $H = 1$, as shown in Fig. 13

$$\rho_{\rm die} = R_{\rm die} + \frac{1}{j\omega C_{\rm die}} \tag{14}$$

where $R_{\rm die}$ is the ESR of an on-chip decoupling capacitor. Note that the decoupling capacitance is characterized by the density per unit area, where $\rho_{\rm die}$ is characterized in units of Ohms by micrometers squared.

The effective activity radius is typically much smaller than the die size in modern high performance ICs [18]. The area of

an IC is therefore modeled as infinite, permitting boundary effects to be neglected. A rectangular structure is therefore modeled as a cylindric structure, as illustrated in Fig. 13. The currents are determined at each point from the current densities $j_{\text{grid}}(m)$ and $j_{\text{die}}(m)$. The current through the ring with radius m is composed of current $i_{\text{grid}}(m)$ in the horizontal direction and $i_{\text{die}}(m)$ in the vertical direction (see Fig. 13). The current passing through the ring with radius m is therefore

$$i_{\text{grid}}(m) = 2\pi m j_{\text{grid}}(m). \tag{15}$$

From the principle of current conservation, the charge of current in each ring is due to the current flowing to the bottom layer with a density $j_{\text{die}}(m)$. Thus

$$\frac{di_{\text{grid}}(m)}{dm} = -2\pi m j_{\text{die}}(m). \tag{16}$$

From KVL for two adjacent rings with radii m_1 and m_2, respectively

$$j_{\text{die}}(m_1)\rho_{\text{die}} = j_{\text{grid}}(m_1)\rho_{S_{\text{grid}}}(m_2 - m_1) + j_{\text{die}}(m_2)\rho_{\text{die}} \tag{17}$$

where $\rho_{S_{\text{grid}}}$ accounts for the impedance of both the power and ground layers. Simplifying (17) with $(m_2 - m_1)$ approaching zero and substituting (15) into (16)

$$j''_{\text{die}}(m) + \frac{j'_{\text{die}}(m)}{m} = a^2 j_{\text{die}}(m) \tag{18}$$

where $a^2 = \rho_{S_{\text{grid}}}/\rho_{\text{die}}$ (μm^{-1}). Note that am is dimensionless. Also note that since $\rho_{S_{\text{grid}}}$ and ρ_{die} are frequency dependent, a is also frequency dependent.

Depending on whether a^2 is real, imaginary, or complex, (18) results in three different solutions. At low frequencies, where the impedance of the power grid is resistive [19] (below 1 GHz), $a = j\omega C_{\text{die}} R_{\text{grid}}$ is imaginary. Equation (18) becomes a zero order Kelvin's equation with a solution

$$j_{\text{die}}(m) = j_0 \left[Ker_0(am) + jKei_0(am)\right] \tag{19}$$

where Ker_0 and Kei_0 are Bessel Kelvin functions and j_0 is determined from the boundary conditions. Note that the absolute value of the current density exhibits the expected localization for small radii [20].

At high frequency, where the impedance of a power grid is dominated by the inductance [19], $a^2 = -\omega C_{\text{die}} L_{\text{grid}}$ is a negative real number. Expression (18) becomes a Bessel equation of zero order. Applying boundary conditions, the solution of (18) is

$$j_{\text{die}}(m) = j_0 H_0^{(2)}(am) \tag{20}$$

where $H_0^{(2)}$ is a second type zero order Hankel function, also exhibiting localization for small radii. An activity radius can be determined from the current localization as described in the following subsection.

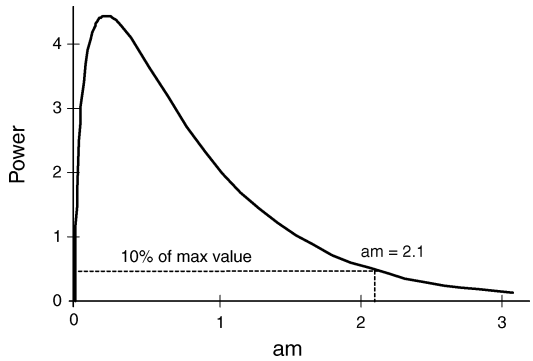


Fig. 14. Average power as a function of the effective activity radius in a resistive power distribution grid (operating at low frequency).

B. Activity Radius

In a resistive grid (operating at low frequency), the average power at radius m is

$$P(m) \propto 2\pi m \left|j_{\text{die}}(m)\right|^2. \tag{21}$$

The average power in a resistive grid, as determined by (21), is plotted in Fig. 14. If the effective activity radius is determined to ensure that 90% of the total power is collected within a circle of radius $m_{\text{eff}_{\text{res}}}$, the value of am is about 2.1. Thus, the effective activity radius is

$$m_{\text{eff}_{\text{res}}} = \frac{2.1}{a} = 2.1\sqrt{\frac{\rho_{\text{die}}}{\rho_{s_{\text{grid}}}}}. \tag{22}$$

Substituting (13) and (14) into (22) and neglecting R_{die}

$$m_{\text{eff}_{\text{res}}} = 2.1\frac{1}{\sqrt{\omega C_{\text{die}} R_{\text{grid}}}}. \tag{23}$$

Note that the effective activity radius $m_{\text{eff}_{\text{res}}}$ diminishes with larger power grid impedance and a higher density of decoupling capacitance C_{die}.

The average power in an inductive grid (operating at high frequency) is plotted in Fig. 15. Similar to the resistive grid, the effective activity radius is

$$m_{\text{eff}_{\text{ind}}} = 0.75\frac{1}{\sqrt{\omega C_{\text{die}} \omega L_{\text{grid}}}} = \frac{0.75}{\omega}\frac{1}{\sqrt{C_{\text{die}} L_{\text{grid}}}}. \tag{24}$$

Note that the effective activity radius decreases with frequency in both resistive and inductive grids.

C. Effective On-Chip Decoupling Capacitance

At low frequency (resistive power grid), the effective decoupling capacitance $C_{\text{eff}_{\text{res}}}$ within a radius $m_{\text{eff}_{\text{res}}}$ around the current load is

$$\begin{aligned} C_{\text{eff}_{\text{res}}} &= \pi m_{\text{eff}_{\text{res}}}^2 C_{\text{die}} \\ &= \pi\left(2.1\frac{1}{\sqrt{\omega C_{\text{die}} R_{\text{grid}}}}\right)^2 C_{\text{die}} \propto \frac{1}{\omega R_{\text{grid}}}. \end{aligned} \tag{25}$$

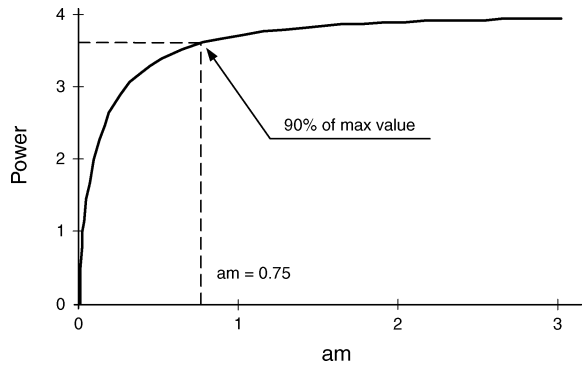


Fig. 15. Average power as a function of the effective activity radius in an inductive power distribution grid (operating at high frequency).

At high frequency, where $\omega L_{\rm grid} \gg R_{\rm grid}$, the effective decoupling capacitance is

$$C_{\rm eff_{ind}} = \pi m^2_{\rm eff_{ind}} C_{\rm die} = \pi \left(\frac{0.75}{\omega} \frac{1}{\sqrt{C_{\rm die} L_{\rm grid}}} \right)^2 C_{\rm die} \propto \frac{1}{\omega^2} \frac{1}{L_{\rm grid}}. \quad (26)$$

Note that $C_{\rm eff_{res}}$ decreases linearly with ω, whereas $C_{\rm eff_{ind}}$ exhibits an inverse dependence on ω^2. Also note from (25) and (26) that $C_{\rm eff}$ does not depend upon $C_{\rm die}$. Intuitively, a larger $C_{\rm die}$ should increase $C_{\rm eff}$. The greater $C_{\rm die}$, however, reduces the effective radius and results in a lower $C_{\rm eff}$. Note that the effective activity radius is infinite in power distribution grids without decoupling capacitors. This scenario is not physically possible, however, as some intrinsic decoupling capacitance always exists [8], [21].

VII. Case Study

The dependence of the effective radii of an on-chip decoupling capacitor on a power distribution system is described in this section to quantitatively illustrate these concepts. The load is modeled as a triangular current source with a 100-ps rise time and 300-ps fall time. The maximum tolerable ripple at the load is 10% of the power supply voltage. The relaxation time between two consecutive switching events (the charge time) is 400 ps. Two scenarios are considered for determining the effective radii of an on-chip decoupling capacitor. In the first scenario, an on-chip decoupling capacitor is connected to the current load by a single line (local connectivity). In the second scenario, the on-chip decoupling capacitors are connected to the current loads by an on-chip power distribution grid (global connectivity). A flip-chip package is assumed. For high performance ICs with die sizes of 1.5 in $\times$ 1.5 in inside a flip-chip package, the distance between two adjacent power *or* ground pads is about 1300 μm [1]. An on-chip power distribution system with a flip-chip pitch (the area formed by the four closest pins) is modeled as an *RL* distributed mesh of 40 $\times$ 40 equal segments to accurately determine the maximum effective distance of an on-chip decoupling capacitor. The resulting pitch size is therefore 32.5 μm, as listed in Table II. The parasitic resistance and inductance of the package (four closest pins of a flip-chip package) are also included in the model. The proposed methodology for placing on-chip decoupling capacitors provides a highly accurate estimate of the magnitude and location of the on-chip decoupling capacitors. The resulting voltage drop as determined from SPICE is equal to the target value of 0.9 V.

TABLE I
Maximum Effective Radii of an On-Chip Decoupling Capacitor For a Single Line Connecting a Decoupling Capacitor to a Current Load

Metal Layer	Res. (Ω/μm)	Ind. (pH/μm)	I_{load} (A)	C_{dec} (pF)	d_{max} (μm) Z	d_{max} (μm) t_{ch}
Top	0.001	1	0.01	20	159	2740
	0.001	1	0.05	100	32	540
	0.001	1	0.1	200	16	270
Intermediate	0.04	0.3	0.01	182	226	521
	0.04	0.3	0.05	907	45	16
	0.04	0.3	0.1	1829	23	2
Bottom	0.01	0.1	0.01	73	847	130
	0.01	0.1	0.05	365	169	25
	0.01	0.1	0.1	731	85	13

$V_{dd} = 1$ V, $V_{ripple} = 100$ mV, $t_r = 100$ ps, $t_f = 300$ ps, and $t_{ch} = 400$ ps

TABLE II
Maximum Effective Radii of an On-Chip Decoupling Capacitor For an On-Chip Power Distribution Grid Modeled as a Distributed *RL* Mesh

Metal Layer	Res. (Ω/μm)	Ind. (pH/μm)	I_{load} (A)	C_{dec} (pF)	d_{max} (cells) Z	d_{max} (cells) t_{ch}
Top	0.007	0.5	0.01	20	>40	>40
	0.007	0.5	0.1	357	2	>40
	0.007	0.5	1	–	<1	–
Intermediate	0.04	0.3	0.01	20	>40	>40
	0.04	0.3	0.1	227	1	<1
	0.04	0.3	1	–	<1	–
Bottom	0.1	0.1	0.01	20	>40	>40
	0.1	0.1	0.1	–	<1	–
	0.1	0.1	1	–	<1	–

$V_{dd} = 1$ V, $V_{ripple} = 100$ mV, $t_r = 100$ ps, $t_f = 300$ ps, $t_{ch} = 400$ ps, and cell size is 32.5 μm $\times$ 32.5 μm
The "–" symbol means that the maximum effective radii cannot be determined based on the RL mesh (the mesh is too coarse). The maximum effective radii are smaller than the cell size. A finer mesh is therefore required to accurately determine the maximum effective radii.

For a single line, the maximum effective radii as determined by the target impedance and charge time for three sets of on-chip parasitic resistances and inductances are listed in Table I. These three scenarios listed in Table I represent typical values of the parasitic resistance and inductance for top, intermediate, and bottom layers of on-chip interconnects in a 90-nm CMOS technology [1]. In the case of the top metal layer, the maximum effective distance as determined by the target impedance is smaller than the critical distance as determined by (11). Hence, the second droop dominates, and the required on-chip decoupling capacitance is determined by (5). Note that the decoupling capacitance increases linearly with the current load. For a typical parasitic resistance and inductance of the intermediate and bottom layers of the on-chip interconnects, the effective radius as determined by the target impedance is longer than the critical distance $d_{\rm crit}$. In this case, the overall voltage drop at the current load is determined by the first droop.

The on-chip decoupling capacitance can therefore be estimated by (9).

In the case of an *RL* mesh, the maximum effective radii as determined by the target impedance and charge time for three sets of on-chip parasitic resistances and inductances are listed in Table II. From (11), for the parameters listed in Table II, the critical voltage drop is 75 mV. If the voltage fluctuations at the current load do not exceed the critical voltage, the second droop dominates and the required on-chip decoupling capacitance is determined by (5). Note that for the aforementioned three interconnect scenarios assuming a 10-mA current load, the maximum effective radii of the on-chip decoupling capacitor based on the target impedance and charge time are larger than forty cells (the longest distance within the mesh from the center of the mesh to the corner). The maximum effective radii of the on-chip decoupling capacitor is therefore larger than the pitch size. The decoupling capacitor can therefore be placed anywhere inside the pitch. For a 100-mA current load, the voltage fluctuations at the current load exceed the critical voltage drop. The first droop dominates and the required on-chip decoupling capacitance is determined by (9).

Note that in both cases, $C_{\rm dec}^{r}$ as determined by (9) increases rapidly with the effective radius based on the target impedance, becoming infinite at $d_Z^{\max}$. In this case study, the decoupling capacitor is allocated almost at the maximum effective distance $d_Z^{\max}$, simulating the worst case scenario. The resulting $C_{\rm dec}$ is therefore significantly large. As the decoupling capacitor is placed closer to the current load, the required on-chip decoupling capacitance as estimated by (9) can be reduced. A tradeoff therefore exists between the maximum effective distance as determined by the target impedance and the size of the minimum required on-chip decoupling capacitance (if the overall voltage drop at the current load is primarily caused by the first droop). Alternatively, the current load can be partitioned into several blocks, lowering the required on-chip decoupling capacitance. The parasitic impedance between the decoupling capacitor and the current load and power supply should also be reduced, increasing the maximum effective radii of the on-chip decoupling capacitors.

Observe that the maximum effective radius as determined by the charge time decreases quadratically with the decoupling capacitance. The maximum effective distance as determined by the charge time becomes impractically short for large decoupling capacitances. Note that the maximum effective radius during the charging phase has been evaluated for the case where the decoupling capacitor is charged to the power supply voltage. In practical applications, this constraint can be relaxed, assuming the voltage across the decoupling capacitor is several millivolts smaller than the power supply. In this case, the effective radius of the on-chip decoupling capacitor as determined by the charge time can be significantly increased.

An activity radius has been compared to simulation results to quantitatively analyze analytic criteria developed in Section VI. A power distribution grid is modeled as an *RL* mesh. A two metal layer 600 μm $\times$ 600 μm power grid is assumed. The width of the lines is 0.6 μm and the line spacing is 3 μm (for a 65-nm CMOS technology), resulting in a characteristic resistance $R_{\rm grid} = 1\ \Omega$ and inductance $L_{\rm grid} = 10$ pH. Note that

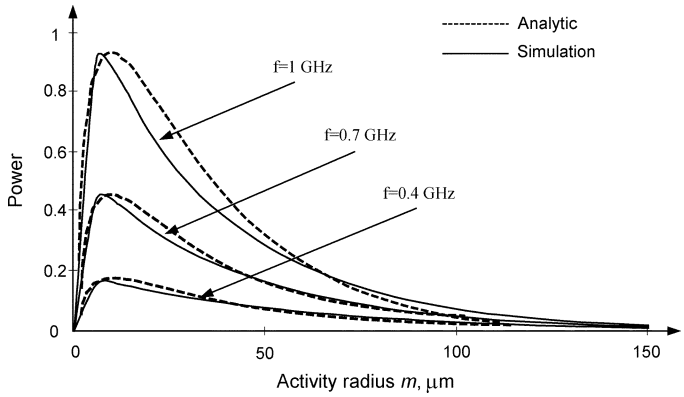


Fig. 16. Average power as a function of the activity radius of an example power distribution grid. The power distribution grid behaves as a resistive grid at relatively low frequencies.

the power distribution grid is resistive in the frequency range of interest, agreeing with the results presented in [22]. Decoupling capacitors are placed uniformly across the circuit. The dependence of the power of an example power distribution grid on the activity radius is illustrated in Fig. 16. Note that the power concentrates around the current source as the frequency increases. Also note that the proposed analytic model exhibits high accuracy as compared to simulations of the average power of a multi-segment network. Similar to the maximum effective radii as determined in the time domain, the activity radius as determined in the frequency domain shrinks as the frequency increases (the rise time decreases).

VIII. Design Implications

Typical effective radii of an on-chip decoupling capacitor is in the range of several hundreds micrometers. In order to determine the location of an on-chip decoupling capacitor, the size of each *RL* mesh segment should be much smaller than the effective radii. In modern high performance ICs such as microprocessors with die sizes approaching 1.5 in $\times$ 1.5 in, a fine mesh is infeasible to simulate. In the case of a coarse mesh, the effective radius is smaller than the size of each segment. The location of each on-chip decoupling capacitor, therefore, cannot be accurately determined. To resolve this dilemma, the accuracy of the capacitor location can be traded off with the complexity of the power distribution network. A "hot" spot (an area where the power supply voltage drops below the minimum tolerable level) is first determined based on a coarse mesh. A finer mesh is used next within each "hot" spot to accurately estimate the effective radius of the on-chip decoupling capacitor. Note that in a mesh structure, the maximum effective radius is the Manhattan distance between two points. In disagreement with Fig. 10, the overall effective radius is actually shaped more like a diamond, as illustrated in Fig. 17.

Typically, multiple current loads exist in an IC. An on-chip decoupling capacitor is placed in the vicinity of the current load to ensure that both the current load and the power supply are within the maximum effective radius. Assuming a uniform distribution of the current loads, a schematic example placement of the on-chip decoupling capacitors is shown in Fig. 18. Each decoupling capacitor provides sufficient charge to the

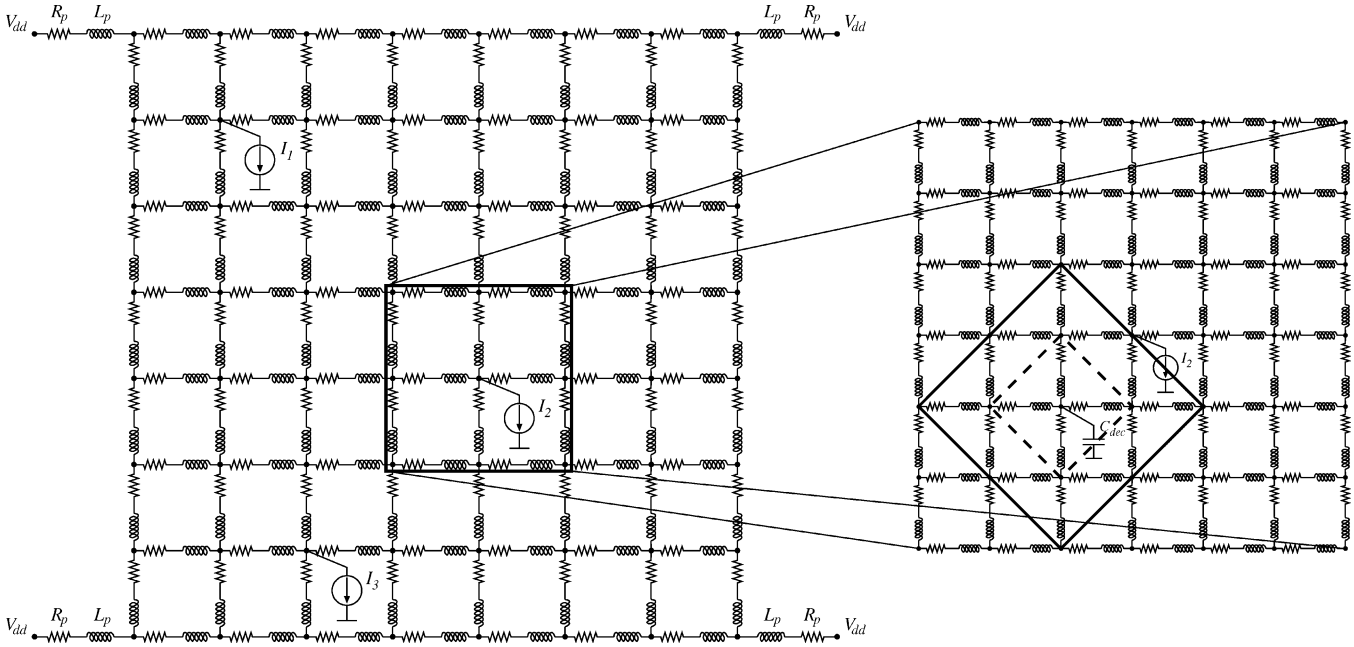


Fig. 17. Effective radii of an on-chip decoupling capacitor. The on-chip power delivery system is modeled as a distributed *RL* mesh with seven by seven equal segments. R_p and L_p denote the parasitic resistance and inductance of the package, respectively. A finer mesh with six by six segments is utilized within the "hot" spot – the area where the power supply voltage drops below the minimum tolerable level (represented by the thick rectangle). For a power distribution system modeled as a distributed *RL* mesh, the maximum effective radius is the Manhattan distance between two points. The overall effective radius is therefore shaped like a diamond.

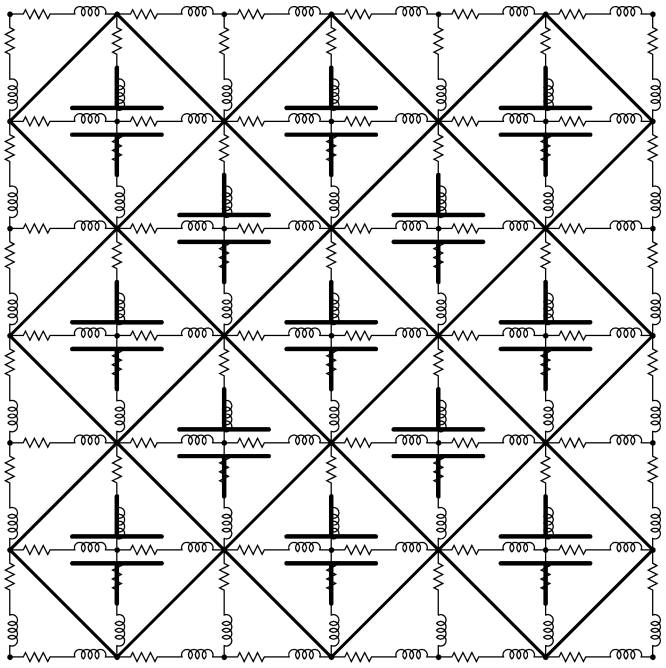


Fig. 18. Schematic example allocation of on-chip decoupling capacitors across an IC. Similar current loads are assumed to be uniformly distributed on the die. Each on-chip decoupling capacitor provides sufficient charge to the current load(s) within the maximum effective radius.

current load(s) within the maximum effective radius. Multiple on-chip decoupling capacitors are placed to provide charge to each of the circuit blocks. In general, the size and location of an on-chip decoupling capacitor are determined by the required charge (drawn by the local transient current loads) and certain system parameters (such as the per length resistance and inductance, power supply voltage, maximum tolerable ripple, and the switching characteristics of the current load).

IX. Conclusion

On-chip decoupling capacitors have traditionally been allocated into the available white space on a die based on an unsystematic or *ad hoc* approach. On-chip decoupling capacitors, however, behave locally and should therefore be treated as a local phenomenon. The efficiency of on-chip decoupling capacitors depends upon the impedance of the P/G lines connecting the capacitors to the current loads and power supplies. A design methodology for placing on-chip decoupling capacitors based on the maximum effective radii is presented in this paper. A maximum effective distance between the current load or power supply and the decoupling capacitor is shown to exist. Beyond this distance, the decoupling capacitor becomes ineffective.

The maximum effective radii of an on-chip decoupling capacitor based on a target impedance (discharge) and charge time in the time domain have been determined. An effective activity radius in the frequency domain is also developed assuming the on-chip decoupling capacitance is uniformly distributed. The effective decoupling capacitance as determined by the activity radius is shown to not depend upon the on-chip capacitance density. Depending upon the parasitic impedance of the P/G lines, the maximum voltage drop is achieved either at the end of the switching activity (the second dominant droop) or during the rise time (the first dominant droop). Design expressions to estimate the minimum on-chip decoupling capacitance required to support expected current demands based on the dominant

voltage drop are provided. The critical length of the interconnect between the decoupling capacitor and the current load is also determined.

To be effective, an on-chip decoupling capacitor should be placed to ensure that both the power supply and the current load are located inside the appropriate effective radius. If this allocation is not feasible, the current load should be partitioned into several circuit blocks, reducing and distributing the localized current demands. The on-chip decoupling capacitors should be allocated to each block while satisfying both effective radii criteria. Summarizing, on-chip decoupling capacitors should be allocated within appropriate effective radii across an IC to satisfy local transient current demands.

REFERENCES

[1] "International technology roadmap for semiconductors, 2006 update," Semiconductor Industry Association, 2006.

[2] M. Popovich, A. V. Mezhiba, and E. G. Friedman, *Power Distribution Networks with On-Chip Decoupling Capacitors*. New York: Springer, 2008.

[3] M. Popovich and E. G. Friedman, "Decoupling capacitors for multi-voltage power distribution systems," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 3, pp. 217–228, Mar. 2006.

[4] M. Popovich and E. G. Friedman, "Noise aware decoupling capacitors for multi-voltage power distribution systems," in *Proc. ACM/IEEE Int. Symp. Quality Electron. Des.*, Mar. 2005, pp. 334–339.

[5] L. D. Smith, "Decoupling capacitor calculations for CMOS circuits," in *Proc. IEEE Conf. Electr. Perform. Electron. Packag.*, Nov. 1994, pp. 101–105.

[6] L. D. Smith, R. E. Anderson, D. W. Forehand, T. J. Pelc, and T. Roy, "Power distribution system design methodology and capacitor selection for modern CMOS technology," *IEEE Trans. Adv. Packag.*, vol. 22, no. 8, pp. 284–291, Aug. 1999.

[7] L. D. Smith and D. Hockanson, "Distributed SPICE circuit model for ceramic capacitors," in *Proc. IEEE Electron. Components Technol. Conf.*, May/Jun. 2001, pp. 523–528.

[8] M. Sotman, M. Popovich, A. Kolodny, and E. G. Friedman, "Leveraging symbiotic on-die decoupling capacitance," in *Proc. IEEE Conf. Electr. Perform. Electron. Packag.*, Oct. 2005, pp. 111–114.

[9] H. H. Chen and S. E. Schuster, "On-chip decoupling capacitor optimization for high-performance VLSI design," in *Proc. IEEE Int. Symp. VLSI Technol., Syst., Appl.*, May 1995, pp. 99–103.

[10] M. D. Pant, P. Pant, and D. S. Wills, "On-chip decoupling capacitor optimization using architectural level prediction," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 10, no. 6, pp. 319–326, Jun. 2002.

[11] H. Su, S. S. Sapatnekar, and S. R. Nassif, "Optimal decoupling capacitor sizing and placement for standard-cell layout designs," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 22, no. 4, pp. 428–436, Apr. 2003.

[12] S. Zhao, K. Roy, and C.-K. Koh, "Decoupling capacitance allocation and its application to power-supply noise-aware floorplanning," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 21, no. 1, pp. 81–92, Jan. 2002.

[13] M. Popovich, E. G. Friedman, M. Sotman, A. Kolodny, and R. M. Secareanu, "Maximum effective distance of on-chip decoupling capacitors in power distribution grids," in *Proc. ACM/IEEE Great Lakes Symp. VLSI*, Apr. 2006, pp. 173–179.

[14] F. Moll and M. Roca, *Interconnection Noise in VLSI Circuits*. New York: Kluwer, 2003.

[15] A. Mezhiba and E. G. Friedman, "Inductive properties of high performance power distribution grids," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 10, no. 12, pp. 762–776, Dec. 2002.

[16] M. E. Van Valkenburg, *Network Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[17] M. Popovich, E. G. Friedman, M. Sotman, and A. Kolodny, "On-chip power distribution grids with multiple supply voltages for high performance integrated circuits," in *Proc. ACM Great Lakes Symp. VLSI*, Apr. 2005, pp. 2–7.

[18] E. Chiprout, "Fast flip-chip power grid analysis via locality and grid shells," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2004, pp. 485–488.

[19] L. Zlydina and Y. Yagil, "3D power grid modeling," in *Proc. IEEE Int. Conf. Electron., Circuits Syst.*, Dec. 2004, pp. 129–132.

[20] M. Sotman, A. Kolodny, M. Popovich, and E. G. Friedman, "On-die decoupling capacitance: Activity radius and effective value," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 489–492.

[21] S. R. Nassif, K. Agarwal, and E. Acar, "Methods for estimating decoupling capacitance of non-switching circuit blocks," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 3998–4001.

[22] S. Pant and E. Chiprout, "Power grid physics and implications for CAD," in *Proc. ACM/IEEE Des. Autom. Conf.*, Jul. 2006, pp. 199–204.

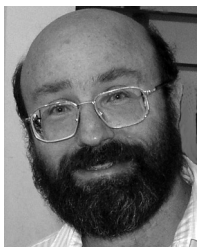
Mikhail Popovich (M'08) received the B.S. degree in electrical engineering from Izhevsk State Technical University, Izhevsk, Russia, in 1998, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Rochester, Rochester, NY, in 2002 and 2007, respectively.

He was an intern at Freescale Semiconductor, Inc., Tempe, AZ, in summer 2005, where he worked on signal integrity in RF and mixed-signal ICs and developed design techniques for placing distributed on-chip decoupling capacitors. In 2007, he joined Qualcomm Corportation as a Senior Engineer. His research interests include the areas of noise, signal integrity, and interconnect design including on-chip inductive effects, optimization of power distribution networks, and the design of on-chip decoupling capacitors.

Dr. Popovich was a recipient of the Best Student Paper Award at the ACM Great Lake Symposium on VLSI in 2005 and the GRC Inventor Recognition Award from the Semiconductor Research Corporation in 2007.

Michael Sotman received the B.S. and M.S. degrees in electrical engineering from Technion—Israel Institute of Technology, Haifa, Israel, in 1996 and 2007, respectively.

In 1995, he joined Intel Corporation, Israel Development Center, Haifa, Israel, where he is a Validation Engineer. His research interests include power delivery and signal integrity.



Avinoam Kolodny (M'81) received the D.Sc. in microelectronics from Technion—Israel Institute of Technology, Haifa, Israel, in 1980.

In 2000, he joined the Electrical Engineering Department, Technion. He was with Intel Corporation, where he was engaged in diverse research and development activities related to device physics, VLSI circuits, and electronic design automation. His research interests include VLSI systems and R&D methodologies.

Eby G. Friedman (F'00) received the B.S. degree from Lafayette College, Easton, PA, in 1979, and the M.S. and Ph.D. degrees from the University of California, Irvine, in 1981 and 1989, respectively, all in electrical engineering.

He has been with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, since 1991, where he is a Distinguished Professor, the Director of the High Performance VLSI/IC Design and Analysis Laboratory, and the Director of the Center for Electronic Imaging Systems. He is also a Visiting Professor at the Technion – Israel Institute of Technology. From 1979 to 1991, he was with Hughes Aircraft Company, rising to the position of Manager of the Signal Processing Design and Test Department, responsible for the design and test of high performance digital and analog ICs. His current research and teaching interests include high performance synchronous digital and mixed-signal microelectronic design and

analysis with application to high speed portable processors and low power wireless communications. He is the author of more than 300 papers and book chapters, numerous patents, and the author or editor of nine books in the fields of high speed and low power CMOS design techniques, high speed interconnect, and the theory and application of synchronous clock and power distribution networks.

Dr. Friedman is the Regional Editor of the *Journal of Circuits, Systems and Computers*, a Member of the editorial boards of the *Analog Integrated Circuits and Signal Processing*, *Microelectronics Journal*, *Journal of Low Power Electronics*, and *Journal of VLSI Signal Processing*, Chair of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS steering committee, and a Member of the technical program committee of a number of conferences. He previously was the Editor-in-Chief of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, a Member of the editorial board of the PROCEEDINGS OF THE IEEE and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: ANALOG AND DIGITAL SIGNAL PROCESSING, a Member of the Circuits and Systems (CAS) Society Board of Governors, Program and Technical chair of several IEEE conferences. He was a recipient of the University of Rochester Graduate Teaching Award and College of Engineering Teaching Excellence Award. Dr. Friedman is a Senior Fulbright Fellow.