

Chapter 12

Effective Radii of On-Chip Decoupling Capacitors

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 [29]. Since the inductance scales slowly [129], the location of the decoupling capacitors significantly affects the design of the 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 power supply noise [279].

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 resistances and inductances of the power delivery network. The resulting voltage drop at the current load is therefore always greater than the voltage drop at the decoupling capacitor. Thus, a maximum parasitic impedance between the decoupling capacitor and the current load exists at which the decoupling capacitor is effective. Alternatively, to be effective, a decoupling capacitor should be placed close to a current load during discharge (within the maximum effective distance d_Z^{\max}), as shown in Fig. 12.1.

Once the switching event is completed, a decoupling capacitor has to be fully charged before the next clock cycle begins. During the charging phase, the voltage across the decoupling capacitor rises exponentially. The charge time of a capacitor is determined by the parasitic resistance and inductance of the interconnect between the capacitor and the power supply. A design space for a tolerable interconnect resistance and inductance exists, permitting the charge on the decoupling capacitor to be restored within a target charge time. 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

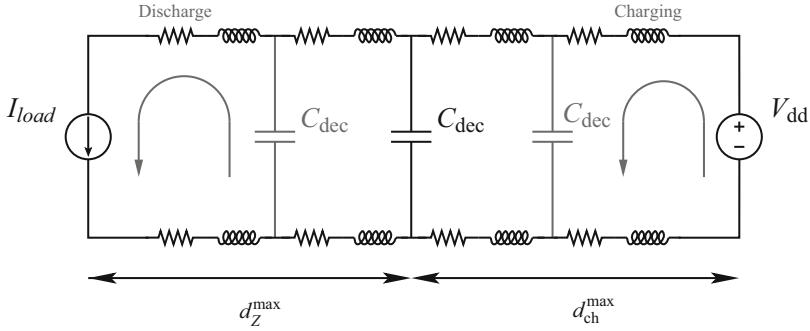


Fig. 12.1 Placement of an on-chip decoupling capacitor based on the maximum effective distance. To be effective, a decoupling capacitor should be placed close to the current load during discharge. During the charging phase, however, the decoupling capacitor should be placed close to the power supply 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_{ch}^{\max} during charging

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 supply during the charging phase (within the maximum effective distance d_{ch}^{\max} , see Fig. 12.1). 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^{\max} and d_{ch}^{\max} . This location is characterized by the effective radii of the on-chip decoupling capacitors and is the primary subject of this chapter. A design methodology to estimate the minimum required on-chip decoupling capacitance is also presented.

This chapter is organized as follows. Existing work on placing on-chip decoupling capacitors is reviewed in Sect. 12.1. The effective radius of an on-chip decoupling capacitor as determined by the target impedance is presented in Sect. 12.2. Design techniques to estimate the minimum magnitude of the required on-chip decoupling capacitance are discussed in Sect. 12.3. The effective radius of an on-chip decoupling capacitor based on the charge time is determined in Sect. 12.4. A design methodology for placing on-chip decoupling capacitors based on the maximum effective radii is presented in Sect. 12.5. A model of an on-chip power distribution network is developed in Sect. 12.6. Simulation results for typical values of on-chip parasitic resistances and inductances are presented in Sect. 12.7. Some circuit design implications are discussed in Sect. 12.8. Finally, some specific conclusions are summarized in Sect. 12.9.

12.1 Background

Decoupling capacitors have traditionally been allocated on a circuit board to control the impedance of a power distribution system and suppress 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 [280], 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 [136]. The authors of [136] show that each decoupling capacitor is effective only within a narrow frequency range. Larger decoupling capacitors have a greater form factor (physical dimensions), resulting in higher parasitic impedances [229]. The concept of an effective series resistance and an effective series inductance of each decoupling capacitor is also described. 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 10–20 MHz. Package decoupling capacitors should therefore be utilized in the frequency range from several megahertz to several hundred megahertz [136]. In modern high performance ICs operating at several gigahertz, only those decoupling capacitors placed on-chip are effective at these frequencies.

The optimal placement of on-chip decoupling capacitors has been discussed in [281]. The power noise is analyzed assuming an *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 described in [281]. The resulting budget of on-chip decoupling capacitance is therefore significantly overestimated. Another technique for placing on-chip decoupling capacitors has been described in [282]. The decoupling capacitors are placed based on activity signatures determined from microarchitectural simulations. This 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 described in [283]. The problem is formulated as a nonlinear optimization and solved using a sensitivity-based quadratic programming solver. The 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 [284], the problem of on-chip decoupling capacitor allocation is evaluated. This 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 [284] 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 [277], 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 chapter, on-chip decoupling capacitors are only effective in close vicinity to the switching circuit. The maximum effective distance for both the discharge and charging phase 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 presented in this chapter.

12.2 Effective Radius of On-Chip Decoupling Capacitor Based on Target Impedance

Neglecting the parasitic capacitance [285], 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 effective 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 [73]. The target impedance of the metal line of a particular length is therefore

$$Z(\omega) = Z'(\omega) \times d, \quad (12.1)$$

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_{target} presented in [28] into (12.1), the maximum effective radius d_Z^{max} between the decoupling capacitor and the current load is

$$d_Z^{max} = \frac{Z_{target}}{Z'(\omega)} = \frac{V_{dd} \times Ripple}{I \times \sqrt{r^2 + \omega^2 l^2}}, \quad (12.2)$$

where $\sqrt{r^2 + \omega^2 l^2}$ denotes the magnitude of the impedance of a unit length wire, Z_{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. 12.2 [286]. Also note that in a meshed structure, multiple paths between any two points are added in parallel. The maximum effective distance corresponding to Z_{target} is, therefore, larger than the maximum effective distance of a single line, as discussed in Sect. 12.7. The maximum effective radius is defined in this chapter 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.

12.3 Estimation of Required On-Chip Decoupling Capacitance

Once the specific location of an on-chip decoupling capacitor is determined as described in Sect. 12.2, 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

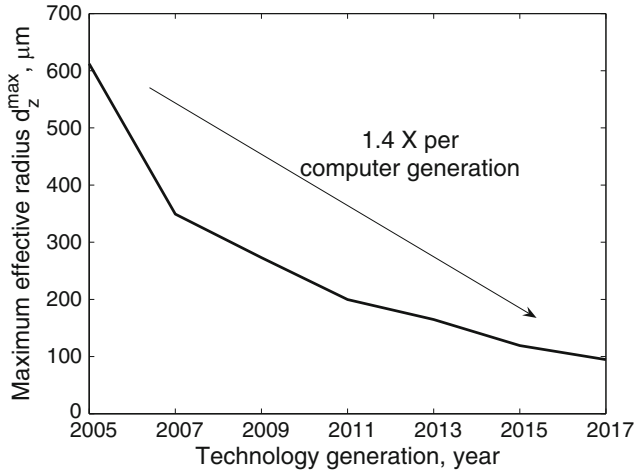


Fig. 12.2 Projection of the maximum effective radius as determined by the target impedance d_Z^{\max} for future technology generations: $I_{\max} = 10 \text{ mA}$, $V_{\text{dd}} = 1 \text{ V}$, and $\text{Ripple} = 0.1$. Global on-chip interconnects are assumed, modeling the highly optimistic scenario. 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)

capacitors based on the dominant power noise are presented in this section. A conventional approach with dominant resistive noise is described in Sect. 12.3.1. Techniques for determining the magnitude of on-chip decoupling capacitors in the case of dominant inductive noise are developed in Sect. 12.3.2. The critical length of the P/G paths connecting the decoupling capacitor to the current load is presented in Sect. 12.3.3.

12.3.1 Dominant Resistive Noise

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. 12.3. The on-chip power distribution network is modeled as a series RL circuit. To qualitatively illustrate the 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_{dis} required to satisfy the current demand during a switching event is modeled as the sum of the area of two triangles (see Fig. 12.3). Since the required charge is provided by an on-chip decoupling capacitor, the voltage across

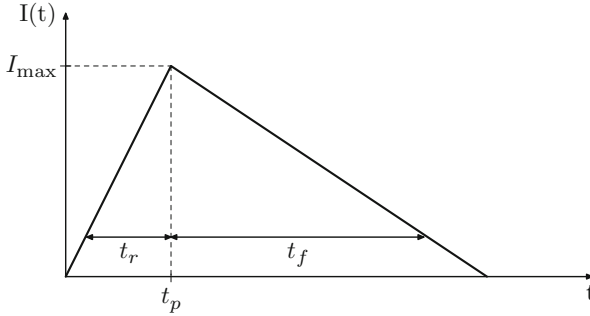


Fig. 12.3 Linear approximation of the current demand of a power distribution network by a current source. The magnitude of the current source reaches the maximum current I_{\max} at peak time t_p . Transition times t_r and t_f denote, respectively, the rise and fall time of the current load

the capacitor during discharge drops below the initial power supply voltage. The required charge during the entire switching event is thus¹

$$Q_{\text{dis}}^f = \frac{I_{\max} \times (t_r + t_f)}{2} = C_{\text{dec}} \times (V_{\text{dd}} - V_C^f), \quad (12.3)$$

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_{dec} is the decoupling capacitance, V_{dd} is the power supply voltage, and V_C^f 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 is equal to 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 [153]) to guarantee fault-free operation. Thus,

$$V_C^f \equiv V_{\text{load}}^f \geq 0.9 V_{\text{dd}}. \quad (12.4)$$

Substituting (12.4) into (12.3) and solving for C_{dec} , the minimum on-chip decoupling capacitance required to support the current demand during a switching event is

$$C_{\text{dec}}^f \geq \frac{I_{\max} \times (t_r + t_f)}{0.2 V_{\text{dd}}}, \quad (12.5)$$

where C_{dec}^f is the decoupling capacitance required to support the current demand during the entire switching event.

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

12.3.2 Dominant Inductive Noise

Note that (12.5) 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 ($IR \gg L di/dt$). Alternatively, the minimum voltage at the load is determined by the resistive drop and the parasitic inductance can be neglected. 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 any given set of rise and fall times. Beyond this critical impedance, the voltage drop at the load is primarily caused by the inductive noise ($L di/dt \gg IR$), as shown in Fig. 12.4. The decoupling capacitor should therefore be increased in the case of dominant inductive noise to reduce the voltage drop across the capacitor during the rise time V_C^r , lowering the magnitude of the power noise.

The charge Q_{dis}^r 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 is²

$$Q_{\text{dis}}^r = \frac{I_{\text{max}} \times t_r}{2} = C_{\text{dec}} \times \Delta V_C^r, \quad (12.6)$$

where Q_{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 (12.6),

$$\Delta V_C^r = \frac{I_{\text{max}} \times t_r}{2 C_{\text{dec}}}. \quad (12.7)$$

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

$$V_{\text{load}}^r = V_{\text{dd}} - I \times R - L \frac{dI}{dt} - \Delta V_C^r, \quad (12.8)$$

where R and L are, respectively, the parasitic resistance and inductance of the P/G lines. Linearly approximating the current load, dI is assumed equal to I_{max} and dt to t_r . Note that the last term in (12.8) accounts for the voltage drop ΔV_C^r across the decoupling capacitor during the rise time of the current at the load.

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

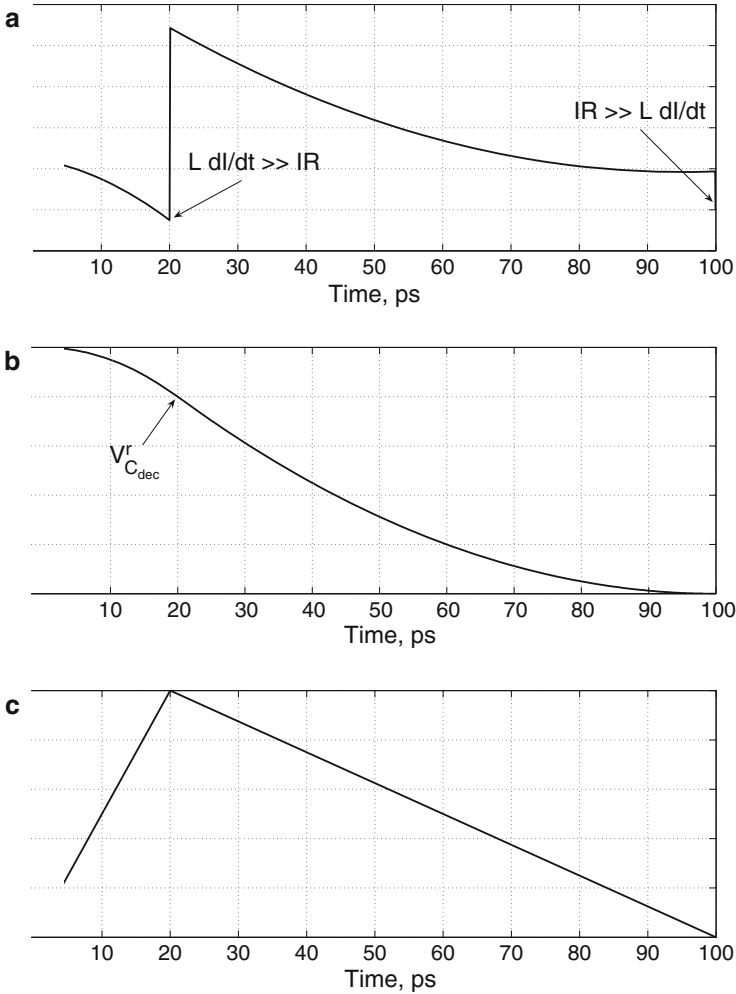


Fig. 12.4 Power distribution noise during discharge of an on-chip decoupling capacitor: $I_{\max} = 100$ mA, $V_{dd} = 1$ V, $t_r = 20$ ps, $t_f = 80$ ps, $R = 100$ m Ω , $L = 15$ pH, and $C_{dec} = 50$ pF; (a) voltage across the terminals of the current load, (b) voltage across the decoupling capacitor, (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 inductive noise therefore dominates the resistive noise and (12.5) 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

Assuming that $V_{load}^r \geq 0.9 V_{dd}$, substituting (12.7) into (12.8), and solving for C_{dec} , the minimum on-chip decoupling capacitance to support the current demand during t_r is

$$C_{\text{dec}}^r \geq \frac{I_{\text{max}} \times t_r}{2 \left(0.1 V_{\text{dd}} - I \times R - L \frac{dI}{dt} \right)}. \quad (12.9)$$

Note that if $L dI/dt \gg IR$, C_{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 (12.5), (12.9) depends upon the parasitic impedance of the on-chip power distribution system. Alternatively, in the case of the dominant inductive noise, 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.

12.3.3 Critical Line Length

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

$$\frac{I_{\text{max}} \times t_r}{\left(0.1 V_{\text{dd}} - I r d_{\text{crit}} - l d_{\text{crit}} \frac{dI}{dt} \right)} = \frac{I_{\text{max}} \times (t_r + t_f)}{0.1 V_{\text{dd}}}. \quad (12.10)$$

Solving (12.10) for d_{crit} ,

$$d_{\text{crit}} = \frac{0.1 V_{\text{dd}} \left(1 - \frac{t_r}{t_r + t_f} \right)}{I r + l \frac{dI}{dt}}. \quad (12.11)$$

For a single line connecting a current load to a decoupling capacitor, the minimum required on-chip decoupling capacitor is determined by (12.5) for lines shorter than d_{crit} and by (12.9) for lines longer than d_{crit} , as illustrated in Fig. 12.5. Note that for a line length equal to d_{crit} , (12.5) and (12.9) result in the same required capacitance. Also note that the maximum length of a single line is determined by (12.2). 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 (12.5) or (12.9) is utilized to estimate the required on-chip decoupling capacitance.

The dependence of the critical line length d_{crit} on the rise time t_r of the current load as determined by (12.11) is depicted in Fig. 12.6. From Fig. 12.6, 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_{crit} is determined by $\frac{t_r}{t_f}$, increasing with larger fall times.

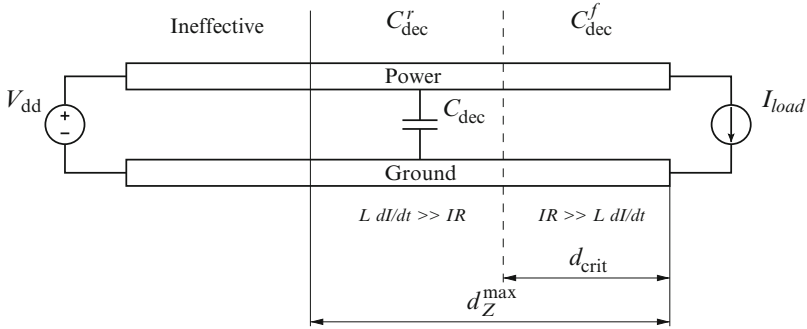


Fig. 12.5 Critical line length of an interconnect between a decoupling capacitor and a current load. The minimum required on-chip decoupling capacitance is determined by (12.5) for lines shorter than d_{crit} and by (12.9) for lines longer than d_{crit} . The decoupling capacitor is ineffective beyond the maximum effective radius as determined by the target impedance d_Z^{max}

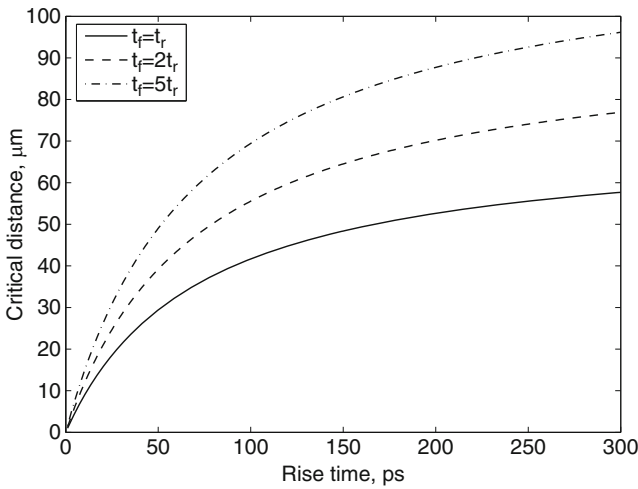


Fig. 12.6 Dependence of the critical line length d_{crit} on the rise time of the current load: $I_{max} = 0.1 \text{ A}$, $V_{dd} = 1 \text{ V}$, $r = 0.007 \Omega/\mu\text{m}$, and $l = 0.5 \text{ pH}/\mu\text{m}$. Note that d_{crit} is determined by $\frac{l}{t_r}$, increasing with larger t_r . The critical line length will shrink in future nanometer technologies as transition times become shorter

Observe in Fig. 12.5 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_{dec}^r and C_{dec}^f) is depicted in Fig. 12.7. For the example parameters shown in Fig. 12.7, the critical line length is $125 \mu\text{m}$. Note that the required on-chip decoupling capacitance C_{dec}^r depends upon the parasitic impedance of the metal lines connecting the decoupling capacitor to the current load. Thus, for lines longer than d_{crit} , C_{dec}^r increases exponentially as the separation between the decoupling capacitor and the

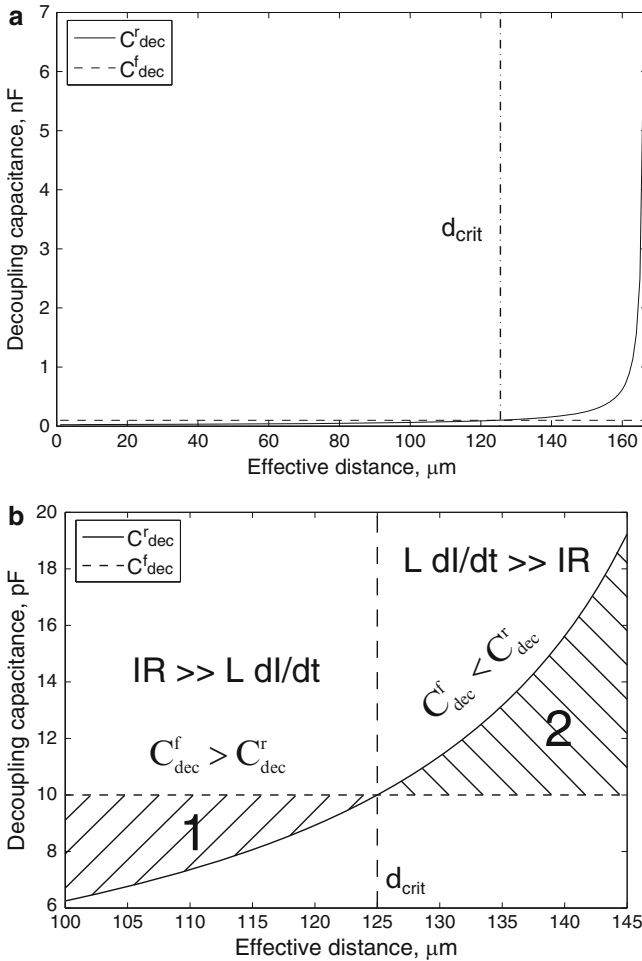


Fig. 12.7 Design space for determining minimum required on-chip decoupling capacitance: $I_{max} = 50$ mA, $V_{dd} = 1$ V, $r = 0.007 \Omega/\mu\text{m}$, $l = 0.5$ pH/ μm , $t_r = 100$ ps, and $t_f = 300$ ps; (a) design space for determining the minimum required on-chip decoupling capacitance is broken into two regions by d_{crit} , (b) design space around d_{crit} . For the example parameters, the critical line length is $125 \mu\text{m}$. In region 1, C_{dec}^f is greater than C_{dec}^r and does not depend upon the parasitic impedance. In region 2, however, C_{dec}^r dominates, increasing rapidly with distance between the decoupling capacitor and the current load

current load increases, as shown in Fig. 12.7a. Also note that for lines shorter than d_{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 dominant resistive drop, the required on-chip decoupling capacitance C_{dec}^f is constant and greater than C_{dec}^r (see region 1 in Fig. 12.7b). If $L dl/dt$ noise dominates the IR noise (the line length is greater than d_{crit}), the required on-chip decoupling capacitance

C_{dec}^r increases substantially with line length and is greater than C_{dec}^f (see region 2 in Fig. 12.7b). Conventional techniques therefore significantly underestimate the required decoupling capacitance in the case of the dominant inductive noise. 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 lines should be considered. A tradeoff therefore exists between the size of C_{dec}^r and the distance between the decoupling capacitor and the current load. As C_{dec}^r is placed closer to the current load, the required capacitance can be significantly reduced.

12.4 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 the 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, 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 Sect. 12.2, 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. 12.8. When the discharge is completed, the switch is closed and the charge is restored on the decoupling capacitor. The initial voltage V_C^0 across the decoupling capacitor is determined by the maximum voltage drop during discharge.

For the circuit shown in Fig. 12.8, the KVL equation for the current in the circuit is [287]

$$L \frac{di_{\text{ch}}}{dt} + R i_{\text{ch}} + \frac{1}{C_{\text{dec}}} \int i_{\text{ch}} dt = V_{\text{dd}}. \quad (12.12)$$

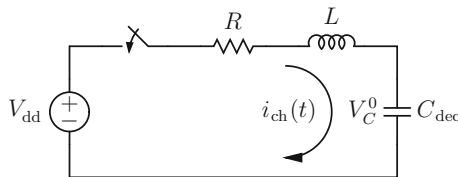


Fig. 12.8 Circuit charging an on-chip decoupling capacitor. The parasitic impedance of the power distribution system connecting the decoupling capacitor to the power supply is modeled by a series RL circuit

Differentiating (12.12),

$$L \frac{d^2 i_{\text{ch}}}{dt^2} + R \frac{di_{\text{ch}}}{dt} + \frac{1}{C_{\text{dec}}} i_{\text{ch}} = 0. \quad (12.13)$$

Equation (12.13) is a second order linear differential equation with the characteristic equation,

$$s^2 + \frac{R}{L} s + \frac{1}{LC_{\text{dec}}} = 0. \quad (12.14)$$

The general solution of (12.13) is

$$i_{\text{ch}}(t) = K_1 e^{s_1 t} + K_2 e^{s_2 t}, \quad (12.15)$$

where s_1 and s_2 are the roots of (12.14),

$$s_{1,2} = -\frac{R}{2L} \pm \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}}. \quad (12.16)$$

Note that (12.15) represents the solution of (12.13) as long as the system is overdamped. The damping factor is therefore greater than one, i.e.,

$$\left(\frac{R}{L}\right)^2 > \frac{4}{LC}. \quad (12.17)$$

For a single line, from (12.17), the critical line length resulting in an overdamped system is

$$D > \frac{4l}{r^2 C_{\text{dec}}}, \quad (12.18)$$

where C_{dec} is the on-chip decoupling capacitance, and l and r are, respectively, the per length inductance and resistance. Inequality (12.18) determines the critical length of a line resulting in an overdamped system. Note that for typical values of r and l in a 90 nm CMOS technology, a power distribution system with a decoupling capacitor is overdamped for on-chip interconnects longer than several micrometers. Equation (12.15) is therefore a general solution of (12.13) for a scaled CMOS technology.

Initial conditions are applied to determine the arbitrary constants K_1 and K_2 in (12.15). The current charging the decoupling capacitor during the charging phase is

$$\begin{aligned}
 i_{\text{ch}}(t) = & \frac{I_{\text{max}}(t_r + t_f)}{4LC_{\text{dec}} \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}}} \\
 & \times \left\{ \exp \left[\left(-\frac{R}{2L} + \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}} \right) t \right] \right. \\
 & \left. - \exp \left[\left(-\frac{R}{2L} - \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}} \right) t \right] \right\}. \quad (12.19)
 \end{aligned}$$

The voltage across the decoupling capacitor during the charging phase can be determined by integrating (12.19) from zero to the charge time,

$$V_C(t) = \frac{1}{C_{\text{dec}}} \int_0^{t_{\text{ch}}} i_{\text{ch}}(t) dt, \quad (12.20)$$

where t_{ch} is the charge time, and $V_C(t)$ and $i_{\text{ch}}(t)$ are, respectively, the voltage across the decoupling capacitor and the current flowing through the decoupling capacitor during the charging phase. Substituting (12.19) into (12.20) and integrating from zero to t_{ch} , the voltage across the decoupling capacitor during the charging phase is

$$\begin{aligned}
 V_{C_{\text{dec}}}(t_{\text{ch}}) = & \frac{I_{\text{max}}(t_r + t_f)}{4C_{\text{dec}}^2 L \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}}} \\
 & \times \left\{ \frac{\exp \left[\left(-\frac{R}{2L} + \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}} \right) t_{\text{ch}} \right] - 1}{-\frac{R}{2L} + \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}}} \right. \\
 & \left. + \frac{1 - \exp \left[\left(-\frac{R}{2L} - \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}} \right) t_{\text{ch}} \right]}{-\frac{R}{2L} - \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_{\text{dec}}}}} \right\}. \quad (12.21)
 \end{aligned}$$

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

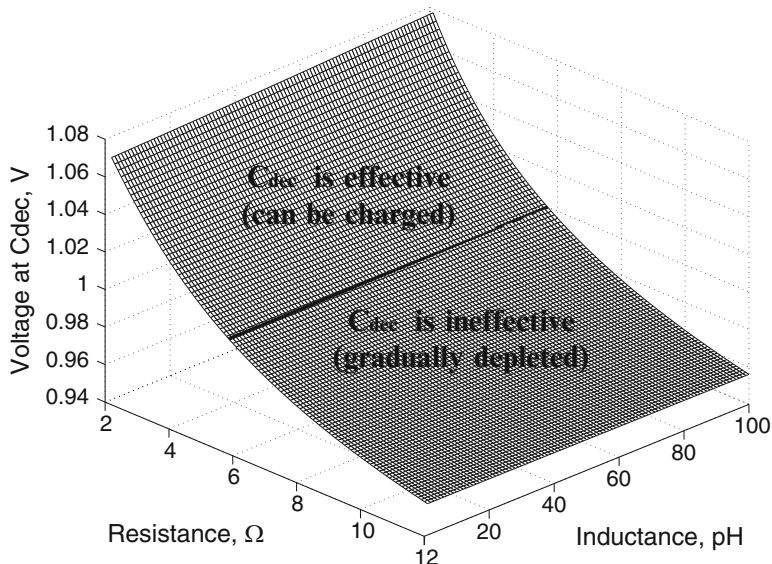


Fig. 12.9 Design space for determining the maximum tolerable parasitic resistance and inductance of a power distribution grid: $I_{\max} = 100$ mA, $t_r = 100$ ps, $t_f = 300$ ps, $C_{\text{dec}} = 100$ pF, $V_{\text{dd}} = 1$ V, and $t_{\text{ch}} = 400$ ps. For a target charge time, the maximum resistance 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 voltage means that the charge on the decoupling capacitor can be restored within t_{ch} .

effective radius of an on-chip decoupling capacitor during the charging phase. Thus from (12.21), 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_{ch} , as shown in Fig. 12.9. 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.

12.5 Design Methodology for Placing On-Chip Decoupling Capacitors

A design methodology for placing on-chip decoupling capacitors based on the maximum effective radii is illustrated in Fig. 12.10. The maximum effective radius based on the target impedance is determined from (12.2) 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 resistive drop is larger than the inductive drop, (12.5) is used to determine the required on-chip decoupling capacitance. If $L di/dt$ noise dominates, the on-chip decoupling capacitance is determined by (12.9). In the case of a single line connecting a decoupling capacitor to a current load, the critical wire length is determined by (12.11).

The maximum effective distance based on the charge time is determined from (12.21). Note that (12.21) 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 such that both the power supply and the current load are located inside the effective radius, as shown in Fig. 12.11. If this allocation is not possible, the current load (circuit block) should be partitioned into several blocks and the 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_{dec}^2 . The on-chip decoupling capacitors should be distributed across the circuit to provide sufficient charge for each functional unit.

12.6 Model of On-Chip Power Distribution Network

In order to determine the effective radii of an on-chip decoupling capacitor and the effect on the noise distribution, a model of a power distribution network is required. On-chip power distribution networks in high performance ICs are commonly modeled as a mesh. Early in the design process, minimal physical information characterizing the P/G structure is available. A simplified model of a power distribution system is therefore appropriate. For simplicity, equal segments within a mesh structure are assumed. The current demands of a particular module are modeled as current sources with equivalent magnitude and switching activities. The current load is located at the center of a circuit module which determines the connection point of the circuit module to the power grid. The parasitic resistance and inductance of the package are also included in the model as an equivalent series resistance R_p and inductance L_p . Note that the parasitic capacitance of the power distribution grid provides a portion of the decoupling capacitance, providing additional charge to the current loads. The on-chip decoupling capacitance intentionally added to the IC is

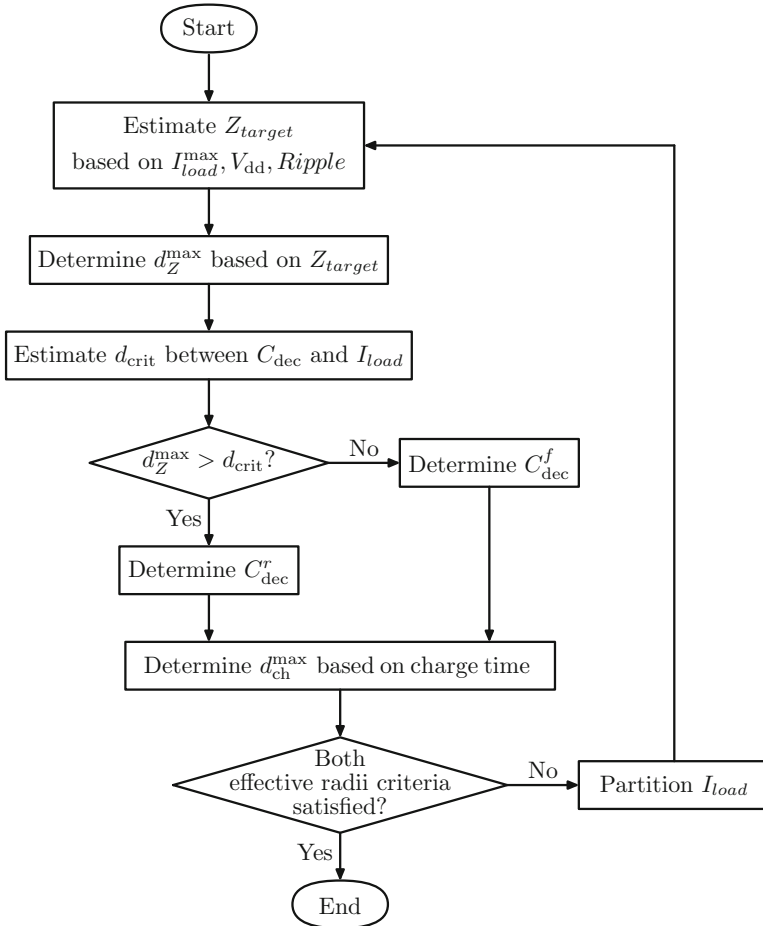


Fig. 12.10 Design flow for placing on-chip decoupling capacitors based on the maximum effective radii

typically more than an order of magnitude greater than the parasitic capacitance of the on-chip power grid. The parasitic capacitance of the power delivery network is, therefore, neglected.

Typical effective radii of an on-chip decoupling capacitor is in the range of several hundred 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. by 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

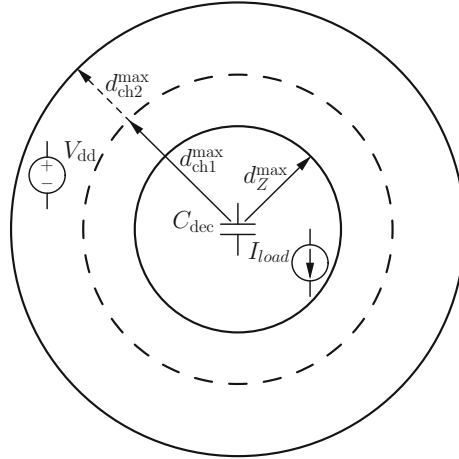


Fig. 12.11 The effective radii of an on-chip decoupling capacitor. The on-chip decoupling capacitor is placed such that both the current load and the power supply are located inside the effective radius. The maximum effective radius as determined by the target impedance d_Z^{max} does not depend on the decoupling capacitance. The maximum effective radius as determined by the charge time is inversely proportional to C_{dec}^2 . If the power supply is located outside the effective radius d_{ch1}^{max} , the current load should be partitioned, resulting in a smaller decoupling capacitor and, therefore, an increased effective distance d_{ch2}^{max} .

spot (an area where the power supply voltage drops below the minimum tolerable level) is first determined based on a coarse mesh, as shown in Fig. 12.12. 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. 12.11, the overall effective radius is actually shaped more like a diamond, as illustrated in Fig. 12.13.

In modern high performance ICs, up to 3000 I/O pins can be necessary [286]. Only half of the I/O pads are typically used to distribute power. The other half is dedicated to signaling. Assuming an equal distribution of power and ground pads, a quarter of the total number of pads is typically available for power or ground delivery. For high performance ICs with die sizes of 1.5 in. by 1.5 in. inside a flip-chip package, the distance between two adjacent power or ground pads is about 1300 μm . By modeling the flip chip area array by a six by six distributed RL mesh, the accuracy in determining the effective radii of an on-chip decoupling capacitor is traded off with the computational complexity required to analyze the power delivery network. In this chapter, an on-chip power distribution system composed of the four closest power pins is modeled as an RL mesh of forty by forty equal segments to accurately determine the maximum effective distance of an on-chip decoupling capacitor. Note that this approach of modeling a power distribution system is applicable to ICs with both conventional low cost and advanced high performance packaging.

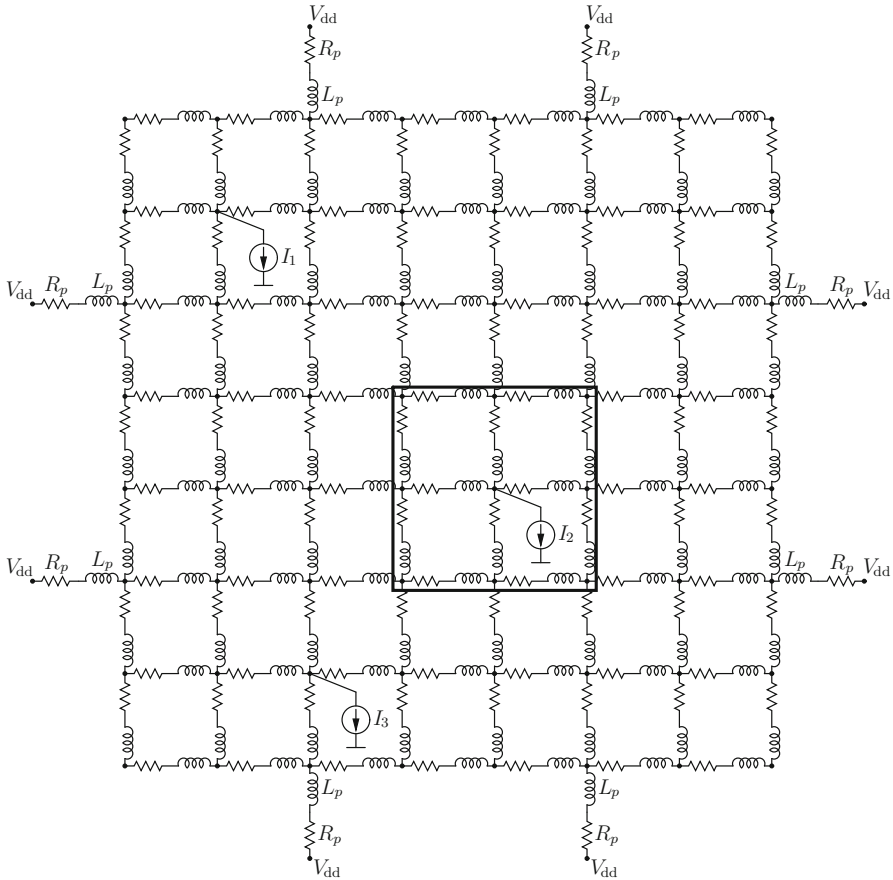


Fig. 12.12 Model of a power distribution network. The on-chip power delivery system is modeled as a distributed RL mesh with seven by seven equal segments. The current loads are modeled as current sources with equivalent magnitude and switching activities. Impedances R_p and L_p denote, respectively, the parasitic resistance and inductance of the package. The *rectangle* denotes a “hot” spot—the area where the power supply voltage drops below the minimum tolerable level

12.7 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 (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).

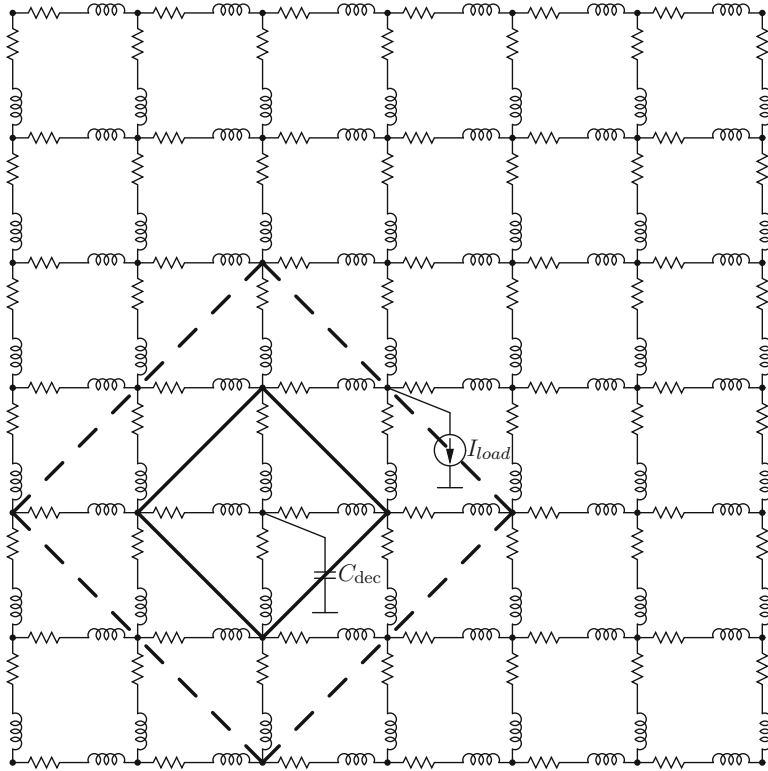


Fig. 12.13 Effective radii of an on-chip decoupling capacitor. For a power distribution system modeled as a distributed RL mesh, the maximum effective radius is the Manhattan distance between two points. The overall effective radius is therefore shaped like a *diamond*

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. 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 forty by forty equal segments to accurately determine the maximum effective distance of an on-chip decoupling capacitor. The parasitic resistance and inductance of the package (the four closest pins of a flip-chip package) are also included in the model. The methodology for placing on-chip decoupling capacitors provides a highly accurate estimate of the magnitude and location of the on-chip decoupling capacitors. The maximum error of the resulting power noise is less than 0.1 % as compared to SPICE.

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 12.1. These three scenarios listed in Table 12.1 represent typical values of the parasitic resistance and inductance of the top, intermediate, and bottom

Table 12.1 Maximum effective radii of an on-chip decoupling capacitor for a single line connecting a decoupling capacitor to a current load

Metal layer	Resistance ($\Omega/\mu\text{m}$)	Inductance ($\text{pH}/\mu\text{m}$)	I_{load} (A)	C_{dec} (pF)	d_{max} (μm)	
					Z	t_{ch}
Top	0.007	0.5	0.01	20	310.8	1166
	0.007	0.5	0.1	200	31.1	116
	0.007	0.5	1	2000	3.1	11.6
Intermediate	0.04	0.3	0.01	183	226.2	24.2
	0.04	0.3	0.1	1773	22.6	2.4
	0.04	0.3	1	45,454	2.3	0.2
Bottom	0.1	0.1	0.01	50,000	99.8	0
	0.1	0.1	0.1	∞	0	0
	0.1	0.1	1	∞	0	0

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

layers of on-chip interconnects in a 90 nm CMOS technology [286]. 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 (12.11). Hence, $IR \gg L di/dt$, and the required on-chip decoupling capacitance is determined by (12.5). 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_{crit} . In this case, the overall voltage drop at the current load is determined by the inductive noise. The on-chip decoupling capacitance can therefore be estimated by (12.9).

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 12.2. From (12.11), for the parameters listed in Table 12.2, the critical voltage drop is 75 mV. If the voltage fluctuations at the current load do not exceed the critical voltage, $IR \gg L di/dt$ and the required on-chip decoupling capacitance is determined by (12.5). 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 $L di/dt$ noise dominates and the required on-chip decoupling capacitance is determined by (12.9).

The effective radii of an on-chip decoupling capacitor decreases linearly with current load. The optimal size of an RL distributed mesh should therefore be determined for a particular current demand. If the magnitude of the current requirements is low, the mesh can be coarser, significantly decreasing the simulation time.

Table 12.2 Maximum effective radii of an on-chip decoupling capacitor for an on-chip power distribution grid modeled as a distributed RL mesh

Metal layer	Resistance ($\Omega/\mu\text{m}$)	Inductance ($\text{pH}/\mu\text{m}$)	I_{load} (A)	C_{dec} (pF)	d_{max} (cells)	
					Z	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	–

$V_{dd} = 1\text{ V}$, $V_{ripple} = 100\text{ mV}$, $t_r = 100\text{ ps}$, $t_f = 300\text{ ps}$, $t_{ch} = 400\text{ ps}$, cell size is $32.5 \times 32.5\ \mu\text{m}$

For a 10 mA current load, the effective radii as determined from both the target impedance and charge time are longer than the pitch size. Thus, the distributed mesh is overly fine. For a current load of 1 A, the effective radii are shorter than one cell, meaning that the distributed RL mesh is overly coarse. A finer mesh should therefore be used to accurately estimate the maximum effective radii of the on-chip decoupling capacitor. In general, the cells within the mesh should be sized based on the current demand and the acceptable computational complexity (or simulation budget). As a rule of thumb, a coarser mesh should be used on the perimeter of each grid pitch. A finer mesh should be utilized around the current loads.

Note that in both cases, C_{dec}^* as determined by (12.9) 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 at almost the maximum effective distance d_Z^{max} , simulating the worst case scenario. The resulting C_{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 (12.9) 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 inductive $L\ di/dt$ drop).

The effective radii listed in Table 12.1 are determined for a single line between the current load or power supply and the decoupling capacitor. In the case of a power distribution grid modeled as a distributed RL mesh, multiple paths are connected in parallel, increasing the effective radii. For instance, comparing Table 12.1 to Table 12.2, note that the maximum effective radii as determined by the target impedance are increased about three times and two times for the top metal layers with, respectively, a 10 and 100 mA current load. Note also that for typical values of the parasitic resistance and inductance of a power distribution grid, the effective

radius as determined by the target impedance is longer than the radius based on the charge time for intermediate and bottom metal layers. For top metal layers, however, the effective radius as determined by the target impedance is typically shorter than the effective radius based on the charge time.

Also note 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. For the bottom metal layer, the maximum effective radius based on the charge time approaches zero. 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.

The maximum effective radius as determined by the charge time becomes impractically short for large decoupling capacitors, making the capacitors ineffective. In this case, the decoupling capacitor should be placed closer to the current load, permitting the decoupling capacitance to be decreased. Alternatively, the current load can be partitioned into several blocks, lowering the requirements on a specific local on-chip decoupling capacitance. The parasitic impedance between the decoupling capacitor and the current load and power supply should also be reduced, if possible, increasing the maximum effective radii of the on-chip decoupling capacitors.

12.8 Design Implications

A larger on-chip decoupling capacitance is required to support increasing current demands. The maximum available on-chip decoupling capacitance, which can be placed in the vicinity of a particular circuit block, is limited however by the maximum capacitance density of a given technology, as described in Chap. 13. Large functional units (current loads) should therefore be partitioned into smaller blocks with local on-chip decoupling capacitors to enhance the likelihood of fault-free operation of the entire system. An important concept described in this chapter is that on-chip decoupling capacitors are a *local* phenomenon. Thus, the methodology for placing and sizing on-chip decoupling capacitors results in a greatly reduced budgeted on-chip decoupling capacitance as compared to a uniform (or blind) placement of on-chip decoupling capacitors into any available white space [284].

Typically, multiple current loads exist in an IC. An on-chip decoupling capacitor is placed in the vicinity of the current load such 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. 12.14. Each decoupling capacitor provides

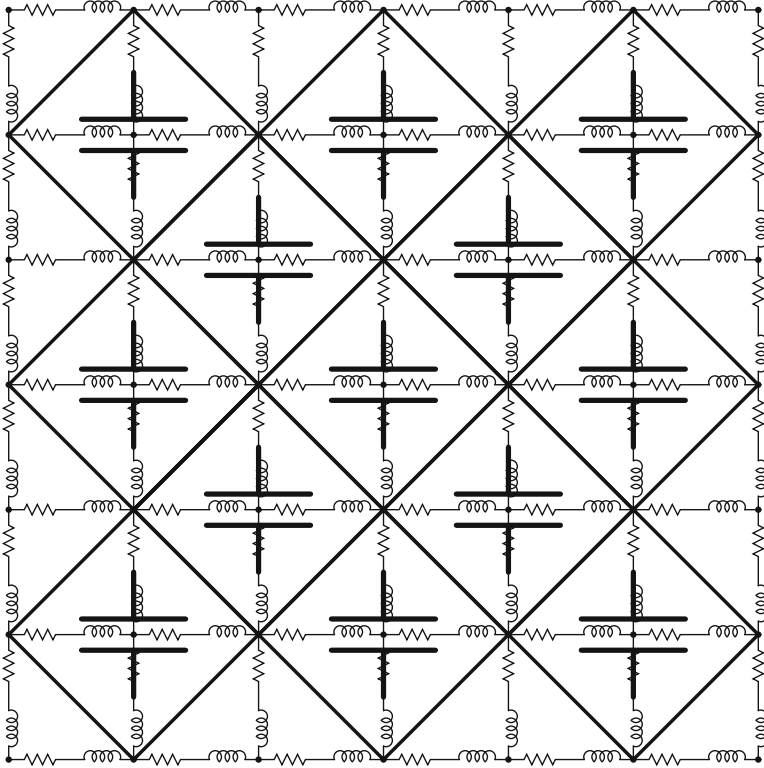


Fig. 12.14 A schematic example allocation of on-chip decoupling capacitors across an IC. Similar current loads are assumed to be uniformly distributed on the die. Each on-chip decoupling capacitor provides sufficient charge to the current load(s) within the maximum effective radius

sufficient charge to the current load(s) within the maximum effective radius. Multiple on-chip decoupling capacitors are placed to provide charge to all 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).

12.9 Summary

A design methodology for placing and sizing on-chip decoupling capacitors based on effective radii is presented in this chapter and can be summarized as follows.

- On-chip decoupling capacitors have traditionally been allocated into the available white space on a die, i.e., using an unsystematic or ad hoc approach

- On-chip decoupling capacitors behave locally and should therefore be treated as a local phenomenon. The efficiency of on-chip decoupling capacitors depends upon the impedance of the power/ground lines connecting the capacitors to the current loads and power supplies
- Closed-form expressions for the maximum effective radii of an on-chip decoupling capacitor based on a target impedance (during discharge) and charge time (during charging phase) are described
- Depending upon the parasitic impedance of the power/ground lines, the maximum voltage drop is caused either by the dominant inductive $L di/dt$ noise or by the dominant resistive IR noise
- Design expressions to estimate the minimum on-chip decoupling capacitance required to support expected current demands based on the dominant voltage drop are provided
- An expression for the critical length of the interconnect between the decoupling capacitor and the current load is described
- To be effective, an on-chip decoupling capacitor should be placed such that both the power supply and the current load are located inside the appropriate effective radius
- On-chip decoupling capacitors should be allocated within appropriate effective radii across an IC to satisfy local transient current demands