
Computer-Aided Design and Analysis

The process of computer-aided design and analysis of on-chip power distribution networks is discussed in this chapter. The necessity for designing and analyzing the integrity of the power supply arises at various stages of the integrated circuit design process as well as during the verification phase. The design and analysis of power distribution networks, however, poses unique challenges and requires different approaches as compared to the design and analysis of logic circuits.

The requirement for analyzing on-chip power distribution networks arises throughout the design process, from the onset of circuit specification to the final verification phase, as discussed in Section 8.1. The primary tasks and difficulties in analyzing the power supply vary at different phases of the design process. At the initial and intermediate design phases, the specification of the power distribution network is incomplete. The primary goal of the power supply analysis process is to guide the general design of the on-chip power distribution network based on information characterizing the power current requirements of the on-chip circuits. The information characterizing the power current requirements is limited, giving rise to the principal difficulty of the analysis process: producing efficient design guidance based on data of limited accuracy. The character of the analysis process gradually changes toward the final phases of the design process. The design of both the power distribution network and the on-chip logic circuits becomes more detailed, making a more accurate analysis possible. The principal goal of the analysis process shifts to verifying the design and identifying those locations where the target specifications are not satisfied. The dramatically increased complexity of the analysis process is

the primary difficulty, requiring utilization of specialized computational methods.

The chapter is organized as follows. A typical flow of the power distribution network design process is described in Section 8.1. An approach for reducing the analysis of power distribution system to a linear problem is presented in Section 8.2. The process of constructing circuit models that characterize a power distribution system is discussed in Section 8.3. Techniques for characterizing the power current requirements of the on-chip circuits are described in Section 8.4. Numerical techniques used in the analysis of power distribution networks are briefly described in Section 8.5. Three strategies for allocating on-chip decoupling capacitors are described in Section 8.6. The chapter is summarized in Section 8.7.

8.1 Design flow for on-chip power distribution networks

In high performance circuits, the high level design of the global power distribution network typically begins before the physical design of the circuit blocks. This approach ensures preferential allocation of sufficient metal resources, simplifying the design process. The principal decisions on the structure of the power distribution network are therefore made when little is known about the specific power requirements of the on-chip circuits. The early design of the power grid is therefore based on conservative design tradeoffs and is gradually refined in the subsequent phases of the design process.

The design flow for a power distribution grid is shown in Fig. 8.1. As the circuit design becomes better specified, a more accurate characterization of the power requirements is possible and, consequently, the design of the power distribution network becomes more precise. The design process can be roughly divided into three phases: preliminary pre-floorplan design, floorplan-based refinement, and layout-based verification [161], [211]. These phases are described in the rest of this section.

Preliminary pre-floorplan design

In the initial pre-floorplan phase, little is known about the power current requirements of the circuit. Preliminary estimates of the power current consumed by a circuit are typically made by scaling the power consumption of previously designed circuits considering the target die

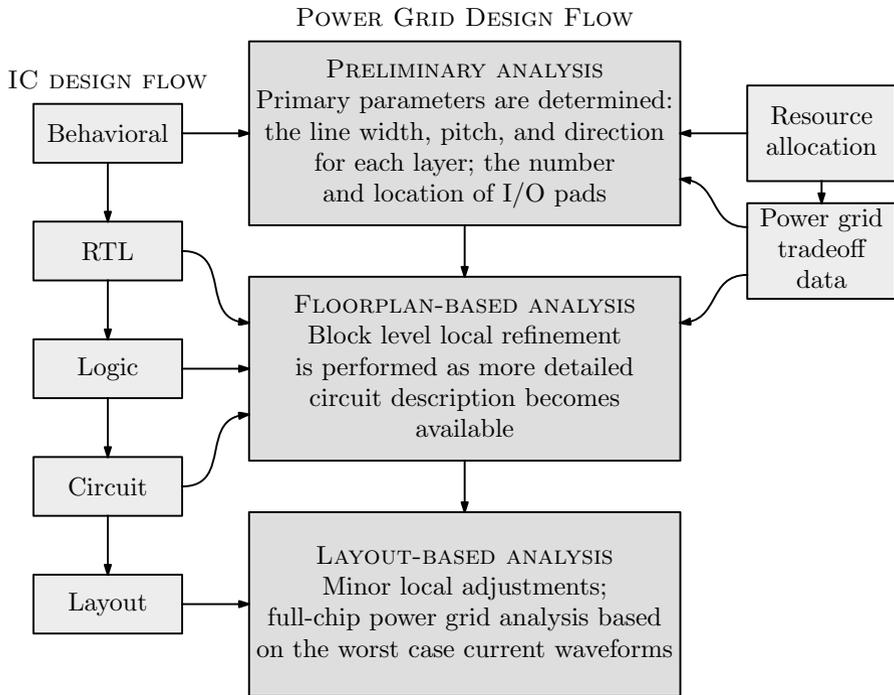


Fig. 8.1. Design flow for on-chip power distribution networks.

area, operating frequency, power supply voltage, and other circuit characteristics.

At this phase of the design process, the power distribution grid is often laid out as a regular periodic structure. A preliminary DC analysis of the IR drops within a power distribution grid is performed, assuming that the power current requirements are uniform across the die. The average power current requirements used in a DC analysis are increased by three to seven times to estimate the maximum power current. The power distribution network is assumed to be uniformly loaded with constant current sources. Basic parameters of the power distribution grid, such as the width and pitch of the power lines in each metal layer and the location of the power/ground pads, are determined based on a preliminary DC analysis of the network. This initial structure largely determines the tradeoff between robustness and the amount of metal resources used by an on-chip power distribution grid.

Floorplan-based refinement

Once the floorplan of a circuit is determined, the initial design of the power distribution grid is refined to better match the local current capacity of the power distribution grid to the power requirements of the individual circuit blocks [202]. The maximum and average power current of each circuit block is determined based on the function of an individual block (*e.g.*, the memory, floating point unit, register file), area, block architecture, and the specific circuit style (*e.g.*, static, dynamic, pass transistor [238], *etc.*). The current distribution is assumed uniform within the individual circuit blocks.

Block specific estimates of the power current provide an approximation of the non-uniform power requirements across the circuit die. The structure of the power distribution grid is tailored according to a DC analysis of a non-uniform power current distribution. Many of the primary problems in the design of power distribution networks are identified at this phase. Moderate computational requirements permit iterative application of a static analysis of the network. Large scale deficiencies in the coverage and capacity of the power distribution network are detected and repaired.

As the structure of the circuits blocks becomes better specified, the local power consumption of an integrated circuit can be characterized with more detail and accuracy. After the logic structure of the circuits is determined, the accuracy of the current requirements are enhanced based on the number of gates and clocking requirements of the circuit blocks. Gate level simulations provide a per cycle estimate of the DC power current for a chosen set of input vectors [211]. Cycle-to-cycle variations of the average power current provide an approximation of the temporal variations of the power current, permitting a preliminary dynamic AC analysis of the power distribution system. The accuracy of the dynamic analysis can be improved if more detailed current waveforms are obtained through gate level simulations. The worst case current waveform of each type of gate and circuit macro is precharacterized. The current waveforms of the constituent gates are arranged according to the timing information obtained in the simulations and are combined into an effective power current waveform for an entire circuit block. As the circuit structure and operating characteristics become better specified, the structure of the power distribution grid within each of the circuit blocks is refined to provide sufficient reliability and

integrity of the on-chip power supply while minimizing the required routing resources.

As the precise placement of the circuit gates is not known in the pre-layout phase, the spatial resolution of the floorplan-based models is relatively coarse. The die area is divided into a grid of $N \times M$ cells. The power and ground distribution networks within each cell are reduced to a simplified macromodel. These macromodels form a coarse RC/RLC grid model of the on-chip power distribution network, as shown in Fig. 8.2. The power current of the circuits located in each cell is combined and modeled by a current source connected to the appropriate node of the macromodel. The number of cells in each dimension of the circuit typically varies from several cells up to a hundred cells, depending on the size of the circuit and the accuracy of the power consumption estimate. The computational requirements of the analysis increase with the specificity of the circuit description. The total number of nodes, however, remains relatively small, permitting an analysis with conventional nonlinear circuit simulation tools such as SPICE.

Layout-based verification

When the physical design of a circuit is largely completed, a detailed analysis of the power distribution network is performed to verify that the target power supply noise margins are satisfied at the power/ground terminals of each on-chip circuit. A detailed analysis is first performed at the level of the individual circuit blocks. Those areas where the noise margins are violated are identified during this analysis phase. The current capacity of the power distribution grid is locally increased in these areas by widening the existing power lines, adding lines, and placing additional on-chip decoupling capacitance. The detailed verification process is repeated on the modified circuit. The iterative process of analysis and modification is continued until the design targets are satisfied. Finally, the verification process is performed for the entire circuit.

An analysis of an entire integrated circuit is necessary to verify the design of a power distribution network. Analyzing the integrity of the power supply at the circuit block level is insufficient as neighboring blocks affect the flow of the current through the power grid. For example, the design of a power grid within a circuit block drawing a relatively low power current (*e.g.*, a memory block) may appear satisfactory at a block-level analysis. However, it is likely to fail if the block is placed in

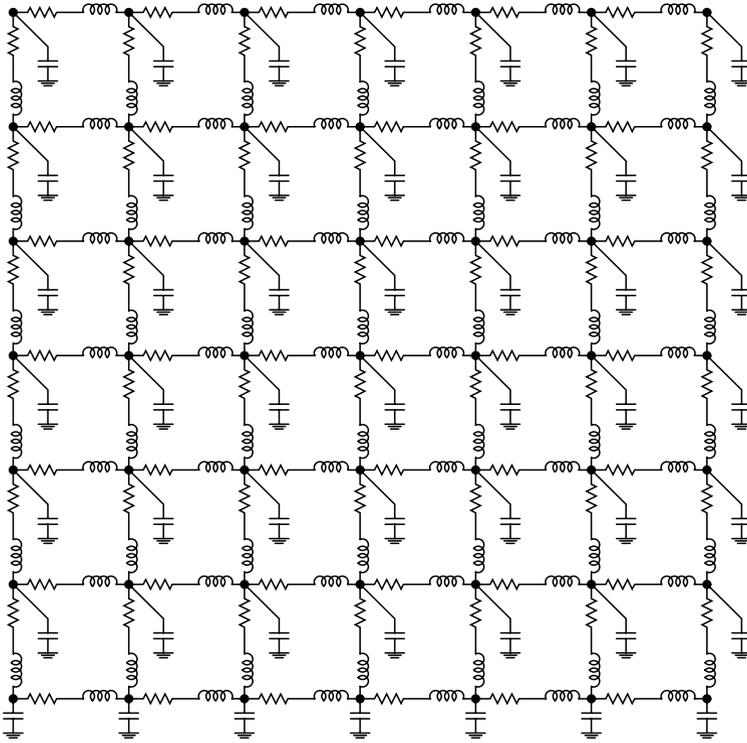


Fig. 8.2. An *RLC* model of an on-chip power distribution network [157].

close proximity to a block drawing high power current. The high power blocks can increase the current flowing through the power network of adjacent low power units [161]. It is therefore necessary to verify the design of the power distribution network at the entire circuit level.

The principal difficulty in verifying an entire power distribution network in a high complexity integrated circuit is the sheer magnitude of the problem. The on-chip power network of a modern high complexity integrated circuit often comprises tens of millions of interconnect line segments and circuit nodes forming a multi-layer power distribution grid, as described in Section 7.1. The circuits loading the power distribution network also consist of tens or hundreds of millions of interconnects and transistors. A transistor level circuit simulation of an entire circuit is therefore infeasible due to prohibitive memory and CPU time requirements. Final analysis and verification is therefore one of the most challenging tasks in the design of on-chip high complexity

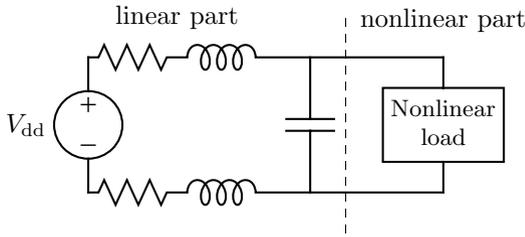
power distribution networks. The remainder of this chapter is largely focused on techniques and methodologies to manage the complexity of the analysis and verification process of power distribution networks.

This methodology is successful if the noise margin violations are local and can be corrected with available metal resources. However, if the necessary changes in the power grid require significant changes in the routing of the critical signal lines, the timing and noise performance characteristics of these critical signals can be significantly impaired. The laborious task of signal routing and timing verification of a circuit is repeated, drastically decreasing design productivity and increasing the time to market. This difficulty of making significant changes in the structure of the power distribution grid at late phases of the design process is the primary reason for using a highly conservative approach in the design of an on-chip power distribution network. Worst case scenarios are assumed throughout the design process. The resulting power distribution network is therefore typically oversized, significantly increasing the area of power distribution networks in modern interconnect-limited integrated circuits.

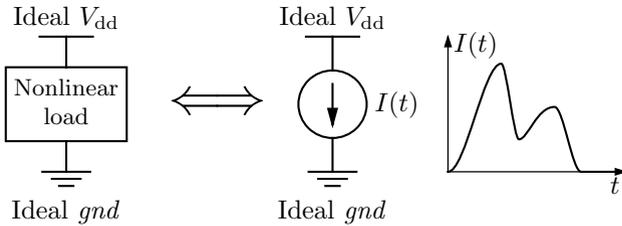
8.2 Linear analysis of power distribution networks

The process of analysis consists of building a circuit model of the power and ground networks including the circuits loading the networks. This step is followed by a numerical analysis of the resulting model. The problem is inherently nonlinear as the digital circuits loading the power distribution grid exhibit highly nonlinear behavior. The current drawn by the load circuits from the power distribution network varies nonlinearly with the voltage across the power terminals of the load. Analyzing a network with tens of millions of nodes is infeasible using a nonlinear circuit simulator such as SPICE due to the enormous computational and memory requirements. To permit the use of efficient numerical analysis techniques, the nonlinear part of the problem is separated from the linear part [161], [202], [211], as illustrated in Fig. 8.3. The current drawn from the power distribution network by nonlinear on-chip circuits is characterized assuming a nominal power and ground supply voltage. The load circuits are replaced by time dependent current sources emulating the original power current characteristics. The resulting network consists of power distribution conductors, decoupling

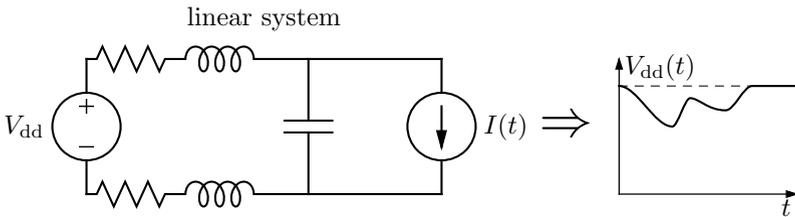
capacitors, and time dependent current sources. This network is linear, permitting the use of efficient numerical techniques.



(a) The original problem of analyzing a linear power distribution network with a nonlinear load



(b) The current requirements of the nonlinear load are characterized under an ideal supply voltage



(c) The nonlinear load is replaced with an AC current source. The resulting system can be analyzed with linear methods.

Fig. 8.3. Approximation for analyzing a power distribution network by replacing a nonlinear load with a time-dependent current source.

Partitioning the problem into a power current characterization part and a linear system analysis part ignores the negative feedback between

the power current and the power supply noise. The current flowing through the power and ground networks causes the power supply levels to deviate from the nominal voltage. In turn, the reduced voltage between the power and ground networks decreases the current drawn by the power load. This typical approach to the power noise analysis process is therefore conservative, overestimating the magnitude of the power supply noise. The relative decrease in the power current due to the reduced power supply voltage is comparable to the relative decrease in the power-to-ground voltage, typically maintained below 10%. The accuracy of these conservative estimates of the power noise is acceptable for most applications. To achieve greater accuracy, the analysis can be performed iteratively. The power current requirements are re-characterized at each iteration based upon the power supply voltage obtained in the previous step. Each iteration yields a more accurate approximation of the power supply voltage across the power distribution network.

The process of power distribution network analysis therefore proceeds in three phases: model construction, load current characterization, and numerical analysis. These tasks are described in greater detail in the following sections.

8.3 Modeling power distribution networks

It is essential that the on-chip power distribution network is considered in the context of the entire power distribution system, including the package and board power distribution networks. As discussed in Chapter 5, the package and board power distribution networks determine the impedance characteristics of the overall power distribution system at low and intermediate frequencies. It is therefore important to analyze the entire power distribution system, including the package and board power distribution networks and the decoupling capacitors, in order to obtain an accurate analysis of the on-chip power supply noise [157].

The complexity of a model depends upon the objectives of the analysis. Models for a DC analysis performed at the preliminary and floorplan-based design phases need to capture only the resistive characteristics of the interconnect structures. The inductive and capacitive circuit characteristics are unimportant in a DC analysis, greatly simplifying the model. As discussed in the previous section, the spatial

resolution of these models is typically limited, further simplifying the process of model characterization.

The reactive impedances of the system are, however, essential for an accurate AC analysis of a power distribution system. The capacitance of the board, package, and on-chip decoupling capacitors as well as the inductive properties of the network should be characterized with high accuracy.

The analysis and verification step towards the end of the design process requires highly detailed models that capture the smallest features of a power distribution system. These models are typically constructed through a back annotation process. The complexity of the board and package power distribution networks is relatively moderate, with the number of conductors ranging from hundreds to thousands. The moderate complexity supports the use of relatively sophisticated analysis tools, such as two- and three-dimensional quasi-static electromagnetic field analyzers [105], [111], [157], [239]. Characterizing the on-chip power distribution network is the most difficult part of the modeling process. The on-chip power distribution network comprises tens of million of nodes and interconnect elements. This level of complexity necessitates the use of highly efficient algorithms to extract the parasitic impedances of the on-chip circuit structures.

Resistance of the on-chip power distribution network

The resistance of on-chip interconnect can be efficiently characterized either with simple resistance formulas based on the sheet resistance of a metal layer [211], [239] or using well developed shape-based extraction algorithms [240], [241]. The temperature dependence of the interconnect resistance should also be included in the model. If R_{25} is the nominal metal resistance at a room temperature of 25°C , the metal resistance at the operating temperature of the circuit T_{op} is $R_{25}(1 + k_{\text{T}}(T_{\text{op}} - 25))$, where k_{T} is the temperature coefficient of the metal resistance. For a temperature coefficient of copper doped aluminum metalization of $0.003^{\circ}\text{C}^{-1}$ and an operating temperature of 85°C , the temperature induced per cent increase in the resistance is 18%, a significant change. Furthermore, the interconnect resistance increases over the circuit lifetime due to electromigration induced defects in the metal structure. This increase in resistance is typically considered in the design process by increasing the nominal metal resistance by a coefficient K_{em} , typically ranging from 10% to 20% [239]. The overall resistance of the on-chip metal R_{eff} can therefore be characterized

as [157]

$$R_{\text{eff}} = R_{25}(1 + (T_{\text{op}} - 25))(1 + K_{\text{em}}). \quad (8.1)$$

Characterization of the on-chip decoupling capacitance

Characterizing the capacitive impedances within the power distribution system is more difficult as compared to resistance characterization. The intrinsic capacitance of the power and ground lines is dominated by other sources of the decoupling capacitance, *i.e.*, the intrinsic circuit capacitance, well capacitance, and intentional capacitance, as discussed in Section 6.3. The capacitance of the power and ground lines can therefore be neglected in this analysis. The intentional and well diffusion decoupling capacitances can be readily characterized by shape-based extraction methods. The intrinsic decoupling capacitance of the on-chip circuits depends upon the state of the digital circuits, making this capacitance difficult to characterize.

The intrinsic circuit decoupling capacitance can be estimated based on the power consumption of the circuit, as described by Chen and Ling [239] and by Larsson [151]. Assuming that the total power P_0 is dominated by the dynamic switching power $P_{\text{switching}}$,

$$P_0 \approx P_{\text{switching}} = \alpha C_{\text{total}} f_{\text{clk}} V_{\text{dd}}^2, \quad (8.2)$$

where α is the switching factor of the circuit, C_{total} is the total intrinsic capacitance of the circuit, f_{clk} is the clock frequency, and V_{dd} is the supply voltage. The total capacitance C_{total} can therefore be determined from an estimate of the total circuit power: $C_{\text{total}} = \frac{P_0}{\alpha f V_{\text{dd}}^2}$. The fraction of the total capacitance being switched, *i.e.*, αC_{total} on average, is the load capacitance of the circuit. The rest of the total capacitance, $(1 - \alpha)C_{\text{total}}$, is quiescent and effectively serves as a decoupling capacitance. The intrinsic decoupling capacitance of the circuit is, therefore,

$$C_{\text{decap}}^{\text{ckt}} \approx \frac{P_0}{f V_{\text{dd}}^2} \frac{1 - \alpha}{\alpha}. \quad (8.3)$$

An estimate of the intrinsic decoupling capacitance represented by (8.3) is, however, strongly dependent on the switching factor α . The switching factor varies significantly depending upon the specific switching pattern and circuit type. The switching factor is therefore difficult to determine with sufficient accuracy in complex digital circuits.

Alternatively, the decoupling capacitance of quiescent circuits can be characterized by simulating a small number of representative circuit blocks, as described by Panda *et al.* [111]. A complete circuit model of each selected circuit block, including the parasitic impedances of the interconnect, is constructed through a back annotation process. The input terminals of a circuit block are randomly set to either the high or low state. The power terminals of the circuit are biased with the power supply voltage V_{dd} . A sinusoidal AC voltage of relatively small amplitude (5% to 15% of V_{dd}) is added to the power terminals of the circuit, modeling the power supply noise, as shown in Fig. 8.4(a). The current flowing through the power terminals is obtained and the small signal impedance of the circuit block *as seen from the power terminals* is determined for the specific frequency of the AC excitation. A series RC model is subsequently constructed, such that the model impedance approximates the impedance of the original circuit block, as shown in Fig. 8.4(b). The model capacitance is scaled by a factor $(1 - \alpha)$ to account for the switching of the circuit capacitance α which does not participate in the decoupling process. The resulting model is an equivalent circuit of the decoupling capacitance of the quiescent circuits, including the decoupling capacitance of both the transistors and interconnect structures. This estimate of the decoupling analysis is significantly less sensitive to the value of α , as compared to (8.3).

The elements R_{eff} and C_{eff} of the equivalent model depend on the state of the digital circuit and the frequency of the applied AC excitation. Nevertheless, these model parameters typically vary little with the input pattern and the excitation frequency in the range of 0.2 to 2 times the clock frequency [111]. For example, the model parameters exhibit less than a 3% variation over all of the input states of an example circuit block consisting of 240 transistors with ten primary inputs [111]. The decoupling characteristics of a larger circuit block are extrapolated from the characteristics of one or several of the precharacterized blocks, depending upon the circuit structure of the larger block. This technique allows for variations in the intrinsic decoupling capacitance for different circuit types.

In many circuits, however, the circuit decoupling capacitance is dominated by the well diffusion capacitance and the intentional capacitance [159]. The overall accuracy of the power supply noise analysis in these circuits is only moderately degraded by the inaccuracies in characterizing the circuit decoupling capacitance.

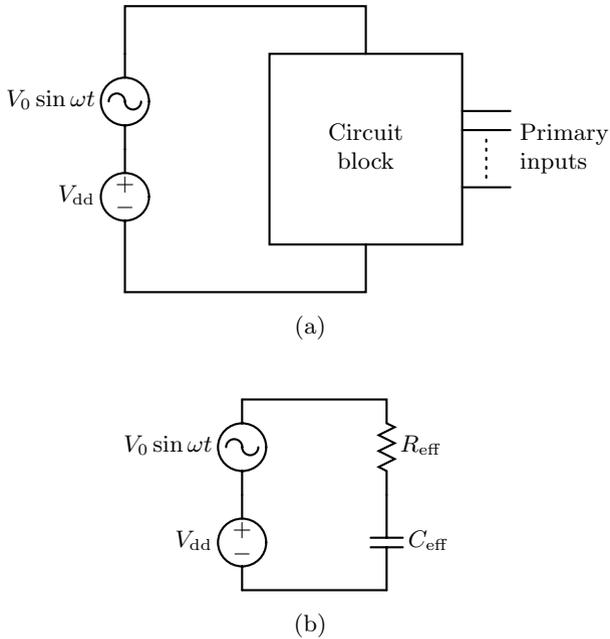


Fig. 8.4. Characterization of the intrinsic decoupling capacitance of the quiescent circuits; (a) circuit model to characterize the capacitance, (b) an equivalent circuit model of the intrinsic decoupling capacitance [111].

Inductance of the on-chip power distribution network

The inductance of the on-chip power and ground lines has historically been neglected [111]. The relatively high resistance R of the on-chip interconnect has dominated the inductive impedance ωL , suppressing the inductive behavior, such as signal reflections, oscillations, and overshoots. As the switching time of the on-chip circuits decreases with technology scaling, the spectral content of the on-chip signals has extended to higher frequencies, making on-chip inductive effects more pronounced. The significance of the on-chip inductance has been demonstrated by Chen and Schuster in an investigation of the sensitivity of the power supply noise to various electrical characteristics of the power distribution system [242]. Assuming the package leads provide an ideal nominal voltage of 2.5 volts, an RLC analysis of the on-chip power grid predicts a minimum on-chip voltage V_{dd} of 2.307 volts, 0.193 volts below the nominal level. If the inductance of the on-chip power grid is neglected, the analysis predicts a minimum on-chip power voltage V_{dd} of 2.396 volts, underestimating the on-chip power noise by 50% as

compared to a more complete *RLC* model. Including the package model in the analysis further reduces the on-chip power supply to 2.199 volts. Modeling the inductive properties of the on-chip power interconnect is therefore necessary to ensure an accurate analysis.

Incorporating the inductive properties of on-chip interconnect into the model of a power distribution network poses two challenges. First, existing techniques for characterizing the inductive properties of complex interconnect structures are computationally intensive, greatly reducing the efficiency of the back annotation process. This issue is discussed further below. Second, including inductance in the model precludes the use of highly efficient techniques for numerically analyzing complex power distribution networks, as discussed in Section 8.5.

The inductive properties of power and ground interconnect lines are difficult to characterize. Characterizing the inductance by a conventional method, *i.e.*, determining the loop inductance of the on-chip circuits based on the shape and size of the current loops is difficult as the current path consists of multiple conductors and the path of the current flow is, generally, not known *a priori*. The inductive properties of regular on-chip power distribution grids can be estimated based on an electromagnetic analysis of the grid structure [157], [239]. Alternatively, the inductive properties can be extracted in the form of a partial inductance matrix. While extracting the partial inductance matrix of an entire circuit is computationally efficient, this matrix is highly dense. The density of a complete partial inductance matrix drastically degrades the efficiency of the subsequent numerical analysis of the circuit model, as the computational efficiency of the most effective numerical methods is conditioned on the sparsity of the matrices characterizing the system. Techniques for sparsifying partial inductance matrices are an active area of research and several techniques has been proposed [63], [64], [65], [66]. The computational efficiency of these techniques is currently insufficient to make the analysis of multimillion conductor systems practical. A method to characterize the inductance of on-chip power distribution grids is described in Chapter 9.

Exploiting symmetry to reduce model complexity

The magnitude of the current flowing through the power and ground distribution networks is the same. The power and ground networks have the same electrical requirements and the structures of these networks are often (close to) symmetric, particularly at the initial

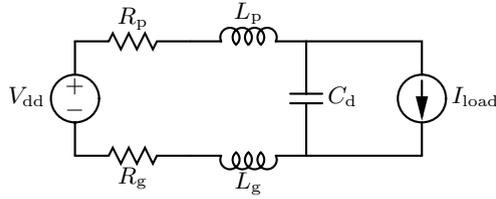
and intermediate phases of the design process. This symmetry can be exploited to reduce the complexity of the power distribution network model by half [111], as illustrated in Fig. 8.5. The model reduction is achieved by circuit “folding.” The original symmetric circuit, as shown in Fig. 8.5(a), is transformed into an equivalent circuit, where the sources, loads, and decoupling capacitors are replaced with equivalent symmetric networks, as shown in Fig. 8.5(b). The nodes on the axis of symmetry of the circuit (shown with a dashed line in Fig. 8.5(b)) are equipotential. It is convenient to use the potential of these nodes as a reference potential. These nodes are therefore referred to as a *virtual ground*. The original circuit is transformed into two independent circuits, as shown in Fig. 8.5(c). The independent circuits are symmetric; consequently, an analysis of only one circuit is necessary. The currents and voltages in one circuit have an opposite polarity as compared to the currents and voltages in the symmetric circuit. Where the impedances of the power and ground distribution networks are symmetric, the voltages in the power and ground networks (with reference to the virtual ground) are also symmetric. That is, wherever the power voltage is decreased by δV , the ground voltage is increased by δV (thereby decreasing the power rail-to-rail voltage by $2\delta V$).

8.4 Characterizing the power current requirements of on-chip circuits

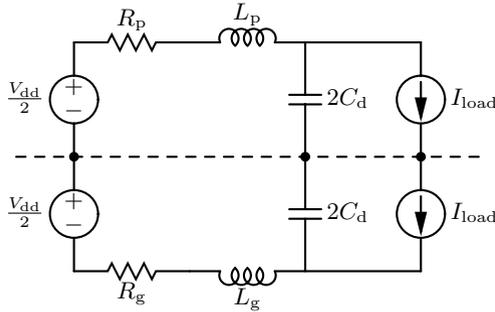
Accurate characterization of the power current requirements is an integral part of the power distribution analysis process, as discussed in Section 8.2. A brief overview of the methods for power current characterization is presented in this section. As the structure of an integrated circuit becomes specified in greater detail, the power current characteristics can be specified with greater accuracy. The complexity of the power current characterization process dramatically increases with the complexity of the circuit description.

Preliminary evaluation of power current requirements

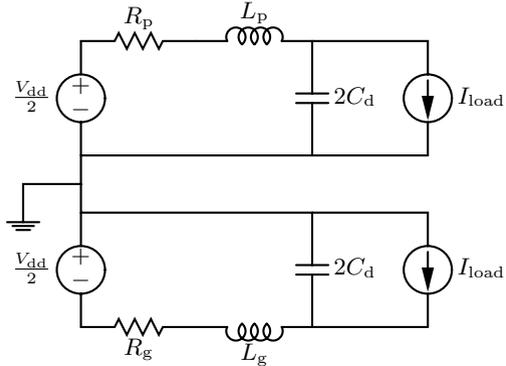
Early estimates of the power current are static. The temporal variation of the power current cannot be assessed in this phase as only the high level structure of the circuit has been developed. Static approaches are based on estimates of the average load currents drawn from the power network [161], [211], permitting static estimates of the IR



(a)



(b)



(c)

Fig. 8.5. Exploiting the symmetry of the power and ground distribution networks to reduce the model complexity by a factor of two. (a) The power and ground current paths in the original model are symmetric. (b) A virtual ground is introduced between the power and ground networks in the equivalent circuit as shown by the dashed line. (c) The resulting circuit model contains two independent symmetric circuits.

drops and electromigration reliability. Estimating average power current consumption is equivalent to estimating average circuit power, as the power supply voltage is maintained approximately constant. At the onset of the design process, the average power current per circuit area is estimated based on the function of a particular circuit block, circuit style used to implement a circuit block, and a scaling analysis of previously designed similar circuits. These estimates can be augmented by estimates of average circuit power based on a circuit description at the behavioral, register transfer, or microarchitectural levels [243], [244].

Gate level estimates of the power current requirements

Estimates of the power current requirements are refined once the logic structure of the circuit is determined. The primary difficulty in determining the power requirements of a CMOS circuit with sufficient accuracy is the dependency of the power current on the circuit input pattern [245]. The time of switching and the magnitude of the power current of a particular gate are determined by the temporal and functional relationships with other gates. The switching patterns that produce the greatest variation of the power supply voltage from a nominal specification (*i.e.*, the greatest power supply noise) are referred to as the worst case switching patterns. These worst case switching patterns are difficult to identify.

The worst case power current of small circuit structures, such as individual logic gates and circuit macrocells, are relatively easy to determine as the number of possible switching patterns is small and the patterns can be readily evaluated. The number of possible switching patterns increases exponentially with the number of inputs and internal state variables. The worst case switching patterns of relatively large circuit blocks comprising thousands of circuit gates and macrocells cannot be determined from an exhaustive analysis. Assuming that all of the gates draw the worst case power current at the same time is overly conservative. Incorporating the logical dependencies among the logic gates, however, greatly increases the complexity of the analysis. A tradeoff therefore exists between the accuracy and efficiency of the power current model.

Estimates of the average power current are typically based on a probabilistic or statistical analysis of the average switching activity and the output load (*i.e.*, the switched capacitance) of the gates [245]. Simple estimates of the average load currents can be obtained by determining the saturation current of each gate in a block and scaling this

current to account for the quiescent state of the majority of the gates at any particular time. More accurate estimates of the average current I_{avg} can be obtained through gate level simulations by evaluating the average switching activity P_s and capacitive load C_L of the gates. The average power current of the circuit is evaluated as $I_{\text{avg}} = \frac{1}{2}P_s f_{\text{clk}} C_L V_{\text{dd}}$, where f_{clk} is the clock frequency of the circuit. Several methods have been developed to determine the upper bound of the power current consumed by the circuit and the associated bound on the power supply noise in an input pattern independent manner [246], [247], [248], [249].

8.5 Numerical methods for analyzing power distribution networks

The circuit model of a power distribution network is combined with time dependent current sources emulating the worst case load to form a linear model of a power distribution network. The linear model is described by a system of linear differential equations. The system of differential equations is reduced to a system of linear equations, which can be numerically analyzed using a number of efficient linear system solution methods [250]. These linear solution methods are classified into direct and iterative methods [251]. The direct methods rely on factoring the coefficient matrix that characterizes the linear system. Once the matrix decomposition is performed, the system solution at each simulation time step is obtained by forward and backward substitution. Alternatively, iterative methods can be used to obtain the solution through a series of successive approximations. Assuming sufficient memory capacity to store the factorization matrices, the use of direct methods is preferable in analyses requiring a large number of time steps, as the solution at each step is obtained through an efficient substitution procedure. Iterative methods are more efficient in solving large systems with limited memory resources.

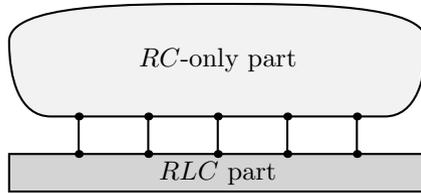
Numerical techniques exploiting special properties of the system are commonly employed to enhance the efficiency of the analysis process. The coefficient matrix of a linear system describing a power distribution network is highly sparse, with non-zero elements typically constituting only a 10^{-6} to 10^{-8} fraction of the total number of elements [202]. Furthermore, in a modified nodal analysis approach, the matrix is symmetric and, for an RC model of a power distribution network, positive

definite [202]. Of the direct methods, Cholesky factorization is particularly well suited, requiring moderate memory resources to store the factorization data. Of the iterative methods, the conjugate gradient method is more memory efficient for denser and larger systems [202]. Several techniques to further enhance the efficiency of analyzing power distribution networks are described in the remainder of this section.

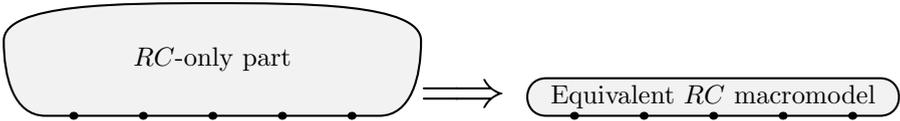
Model partitioning in RC and RLC parts

These numerical methods are modified if the mutual inductance among the interconnect segments are considered, as described by Panda *et al.* [111]. The matrix describing the system is no longer guaranteed to be positive definite, preventing the use of efficient methods based on this property and forcing the use of more general (and computationally expensive) methods. Including the mutual inductance elements is virtually always necessary to accurately describe the electrical properties of the power distribution networks of a printed circuit board and an integrated circuit package. An RC -only model is often adequate for describing the on-chip power distribution network. In many cases, therefore, only a relatively small part of the overall model (describing the package and board power interconnect) contains inductive elements. The computational complexity of the problem can be significantly reduced in these cases [111], as illustrated in Fig. 8.6. A comprehensive model of a power distribution system is partitioned into an RLC part containing all of the inductive elements (at the package and board level) and an RC -only part (the on-chip network). The RC -only part contains the vast majority of elements comprising the overall power distribution system. The complexity of the RC part of the system can be reduced by exploiting efficient techniques based on solving a symmetric positive definite system of equations. The RC part of the model is replaced with the equivalent admittance at the ports of the interface with the RLC part. The resulting system is significantly smaller than the original system and can be solved with general solution methods.

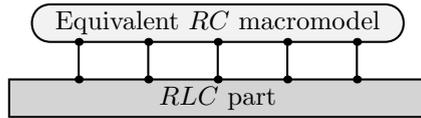
An approach to analyzing power distribution networks composed of RLC segments (with no mutual inductance terms) has been proposed by Zheng and Tenhunen [252], [253]. An enhanced matrix formulation of the power distribution network problem is proposed and numerical techniques to solve this formulation are described. As demonstrated on sample networks consisting of several hundred segments, this analysis approach is three to four hundred times faster as compared to SPICE simulations, while maintaining an accuracy within 5% of SPICE.



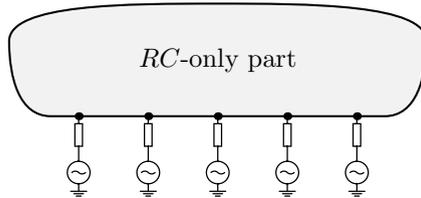
(a) A circuit model of a power distribution system can be partitioned into a relatively small *RLC* part and an *RC*-only part containing the vast majority of the circuit elements.



(b) An equivalent admittance macromodel of the *RC*-only part is constructed.



(c) The *RC* part is replaced with a reduced model and the system is analyzed using robust numerical methods. The voltages and equivalent admittances at the ports of the *RC*-only part are determined.



(d) The *RC*-only part is analyzed using efficient numerical methods. The *RLC* part of the model is replaced with equivalent circuits at the appropriate ports, as has been determined in the previous step.

Fig. 8.6. Reducing the computational complexity of the analysis process by separating the analysis of the *RLC* and *RC*-only parts of a power distribution system.

Improving the initial condition accuracy of the AC analysis

The efficiency of the transient analysis can be enhanced by accurate estimates of the steady state condition, *i.e.*, the currents passing through the network inductors and the voltages across the network capacitors. The steady state condition is not known before the analysis. The analysis starts with a rough estimate of the initial conditions, for example, with the voltages and currents determined in a DC analysis. In the beginning of the AC analysis, the initial excitation conditions are maintained and the system is allowed to relax to the AC steady state. After the steady state is reached, the switching pattern of interest can be applied to the circuit inputs, permitting a transient analysis to be initiated. No useful information is produced as the system settles to the AC steady state. In systems with a low damping factor, the time required to reach the steady state can be a substantial portion of the overall time span of the simulation, significantly increasing the computational overhead of the analysis [111].

An accurate estimate of the initial conditions is therefore desirable. This estimate can be efficiently obtained as follows [111]. A simplified circuit model of the power distribution network is constructed. Elements of the simplified model are determined based on the elements of the original network and the worst case voltage drop obtained by a DC analysis of the original network. The simplified circuit is simulated and the steady state inductor currents and capacitor voltages are determined. These currents and voltages are used as steady state values in the transient analysis of the original network. Using this technique to analyze the power distribution network of a 300 MHz PowerPC microprocessor, the initial conditions are estimated with an accuracy of 6.5% as compared to a 62% accuracy based on a DC analysis. The greater accuracy of the initial conditions shortens by a factor of three the time required to determine the AC steady state.

Global-local hierarchical analysis

A hierarchical approach to the electrical analysis of an on-chip power distribution network reduces the CPU time and memory requirements as compared to a flat (non-hierarchical) analysis, as described by Zhao *et al.* [254]. A power network is partitioned into a global grid and many local grids, as depicted in Fig. 8.7. A macromodel is built for every local partition. A macromodel is a linear multi-port network characterized by the same relationship between the port currents and voltages as the original local partition. The power network is simulated with each local

partition substituted by the respective macromodel. The problem size is thereby reduced from the total number of nodes in the original power distribution network to the number of nodes in the global partition plus the total sum of the local partition ports. Subsequently, to determine the voltage at the nodes of the local partitions, each local partition can be independently analyzed with the port currents determined during the analysis of the global grid with macromodels. The efficiency of the methodology therefore depends upon judicial partitioning, the computational cost of constructing a macromodel, and the complexity of the macromodel.

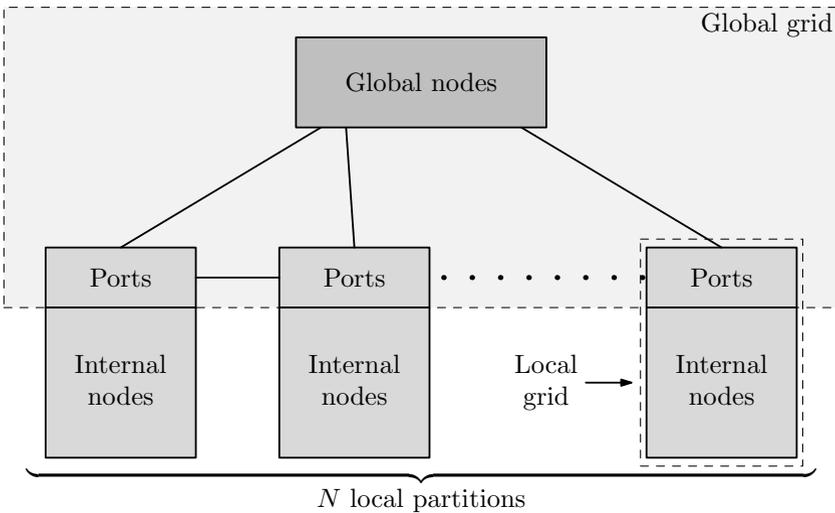


Fig. 8.7. A hierarchical model of a power distribution network. In a global analysis, the local grids are represented by multi-port linear macromodels.

The greatest reduction in the complexity of the analysis is achieved if the system is partitioned into subnetworks with the number of internal nodes much larger than the square of the respective number of ports [254]. The macromodel matrices tend to have a higher density than the matrix representation of the original power distribution network. The higher density can limit the choice of suitable numerical solution methods and, therefore, the efficiency of the analysis. Sparsification of the macromodels can be performed to address this problem [254].

The performance gains of the proposed hierarchical method over a conventional flat analysis have been assessed for power distribution networks of several industrial DSP and microprocessor circuits [254]. The memory requirements are reduced severalfold. The size of the linear system describing the hierarchical system is approximately ten times smaller than the size of the system in a flat (non-hierarchical) analysis. The memory requirement is reduced by ten to twenty times. The one-time overhead of the analysis setup (partitioning, macromodel generation, sparsification, *etc.*, in the case of the hierarchical methodology) is reduced by a factor of two to five. The run time of the subsequent time steps, however, is greater as compared to a flat analysis. The difference in the runtime decreases with the size of the system, becoming relatively small in networks with ten million nodes or more.

Since each local partition of the network is solved independently, a hierarchical analysis has two other desirable properties [254]. First, the hierarchical analysis is easily amenable to parallel computation, permitting additional speedup in the subsequent time steps. Being performed in parallel, hierarchical analysis is two to five times faster than a flat analysis. Second, the mutual independence of the macromodels renders the hierarchical analysis flexible. Local changes in the circuit structure necessitate regeneration of only a single macromodel. The rest of the setup can be reused, permitting an efficient incremental analysis. Alternatively, if a detailed power analysis of a specific block is only of interest, the local solution of other partitions can be omitted, accelerating the analysis while preserving the effect of these partitions on the partition of interest.

***Ad hoc* analysis techniques**

No assumptions regarding the topological structure of the on-chip power distribution network are made for the numerical analysis techniques described in the previous section. These techniques can therefore be applied to a network of general topology. A number of *ad hoc* techniques have also been developed. These techniques are either tailored to a specific network topology or exploit specific properties of the network. These techniques are briefly discussed below.

Multi-grid analysis

The efficient analysis of power distribution grids can be performed through the use of multi-grid methods, as described by Nassif and

Kozhaya [255], [256]. Power distribution grids are spatially and temporally well behaved (*i.e.*, smooth and damped) systems. General purpose robust techniques are unnecessary to achieve an accurate solution of such systems. A system of linear equations describing such well behaved systems is analogous to a finite element discretization of a two-dimensional parabolic partial differential equation [256]. Efficient numerical methods developed for parabolic partial differential equations can therefore be exploited to analyze power distribution grids. The multi-grid method is most commonly used for solving parabolic partial differential equations [257]. Using a fixed time step requires only a single inversion of a large and sparse matrix during the numerical analysis process.

Hierarchical analysis of networks with mesh-tree topology

The analysis and optimization of power distribution networks structured as a global mesh feeding local trees can be performed by a specially formulated hierarchical method, as proposed by Su, Gala, and Sapatnekar [258]. The process of hierarchical analysis proceeds in three stages. First, each tree is replaced with an equivalent circuit model obtained from the passive reduced-order interconnect macromodeling algorithm (PRIMA) [259]. The system is solved to determine all of the nodal voltages in step two. Each tree is analyzed independently based on the voltage at the root of the tree obtained in step two. The method produces results within 10% of SPICE with a greater than ten fold speedup.

Efficient analysis of RL trees

A worst case IR and ΔI noise analysis is efficiently performed in power and ground distribution networks structured as RL trees originating from a single I/O pad, as described by Zhao, Roy, and Koh [260]. The worst case power current requirements of each circuit attached to a power distribution tree are approximated by a trapezoidal waveform. The intrinsic and intentional decoupling capacitances are neglected in the analysis, allowing the power voltage to be efficiently calculated at each node of the tree.

A method for the frequency domain analysis of the noise in RL power distribution trees has also been developed [261]. A frequency domain noise spectrum is computed by analyzing the effective output impedance of the power distribution network at each current source and the spatial correlation among the trees. A time domain noise waveform is obtained by applying an inverse Fast Fourier transform to the

frequency domain spectrum. This approach is more than two orders of magnitude faster than HSPICE simulations, while maintaining an accuracy within 10% as compared to circuit simulation.

8.6 Allocation of on-chip decoupling capacitors

The allocation of on-chip decoupling capacitors is commonly performed iteratively. Each iteration of the allocation process consists of two steps, as shown in Fig. 8.8. In the power noise analysis phase, the magnitude of the power supply noise is determined throughout the circuit. The size and placement of the decoupling capacitors are then modified during the allocation phase based on the results of the noise analysis. This process continues until all of the target power noise constraints are satisfied. Occasionally, the power noise constraints cannot be satisfied for a specific circuit. In this case, the area dedicated to the on-chip decoupling capacitors should be increased. In some cases, large functional blocks should be partitioned, permitting the allocation of decoupling capacitors around the smaller circuit blocks.

Although a sufficiently large amount of on-chip decoupling capacitance distributed across an IC will ensure adequate power supply integrity, the on-chip decoupling capacitors consume considerable die area and leak significant amounts of current. Interconnect limited circuits typically contain a certain amount of white space (area not occupied by the circuit) where intentional decoupling capacitors can be placed without increasing the overall die size. After this area is utilized, accommodating additional decoupling capacitors increases the overall circuit area. The amount of intentional decoupling capacitance should therefore be minimized. A strategy guiding the capacitance allocation process is therefore required to achieve target specifications with fewer iterations while utilizing the minimum amount of on-chip decoupling capacitance.

Different allocation strategies are the focus of this section. A charge-based allocation methodology is presented in Section 8.6.1. An allocation strategy based on an excessive noise amplitude is described in Section 8.6.2. An allocation strategy based on excessive charge is discussed in Section 8.6.3.

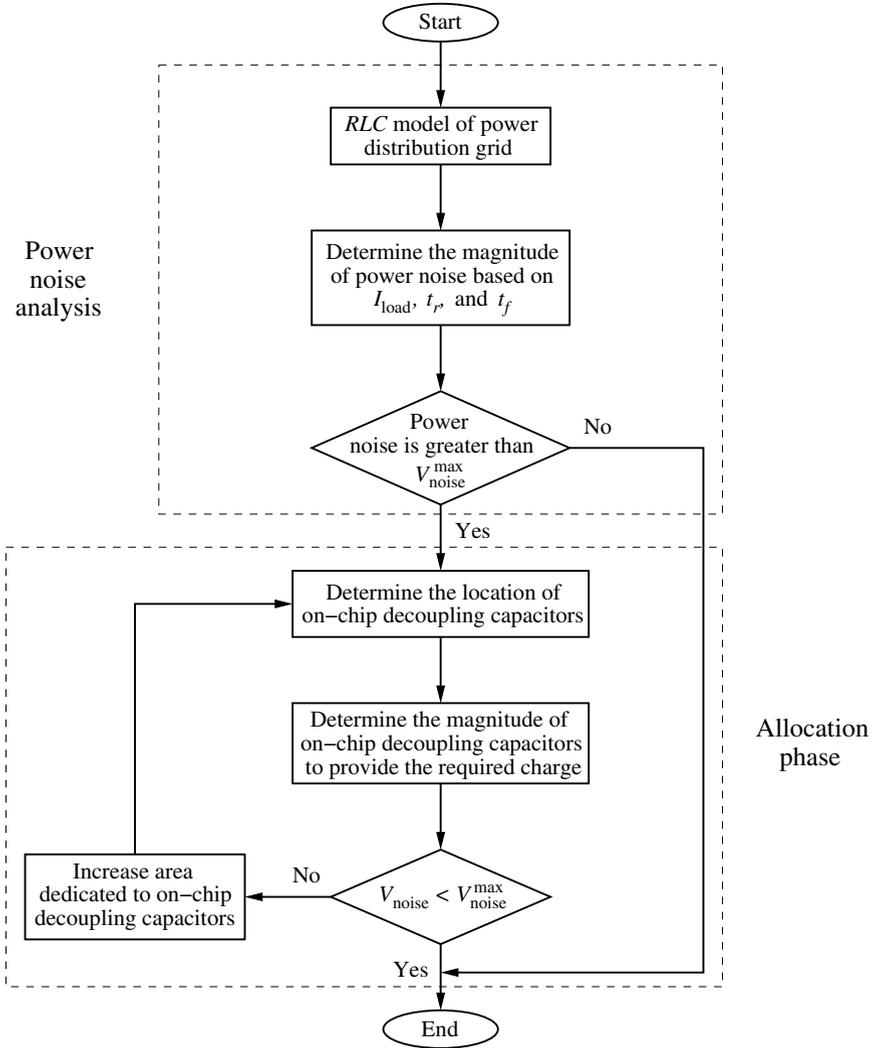


Fig. 8.8. Flow chart for allocating on-chip decoupling capacitors.

8.6.1 Charge-based allocation methodology

One of the first approaches is based on the average power current drawn by a circuit block [262]. The decoupling capacitance C_i^{dec} at node i is selected to be sufficiently large so as to supply an average power current I_i^{avg} drawn at node i for a duration of a single clock period, *i.e.*, to

release charge $\delta Q_i = \frac{I_i^{\text{avg}}}{f_{\text{clk}}}$ as the power voltage level varies by a noise margin δV_{dd} ,

$$C_i^{\text{dec}} = \frac{\delta Q_i}{\delta V_{\text{dd}}} = \frac{I_i^{\text{avg}}}{f_{\text{clk}} \delta V_{\text{dd}}}, \quad (8.4)$$

where f_{clk} is the clock frequency.

The rationale behind the approach represented by (8.4) is that the power current during a clock period is provided by the on-chip decoupling capacitors. This allocation methodology is based on two assumptions. First, at frequencies higher than the clock frequency, the on-chip decoupling capacitors are effectively disconnected from the package and board power delivery networks (*i.e.*, at these frequencies, the impedance of the current path to the off-chip decoupling capacitors is much greater than the impedance of the on-chip decoupling capacitors). Second, the on-chip decoupling capacitors are fully recharged to the nominal power supply voltage before the next clock cycle begins.

Both of these assumptions cannot be simultaneously satisfied with high accuracy. The required on-chip decoupling capacitance as determined by (8.4) is neither sufficient nor necessary to limit the power supply fluctuations within the target margin δV_{dd} . If the impedance of the package-to-die interface is sufficiently low, a significant share of the power current during a single clock period is provided by the decoupling capacitors of the package, overestimating the required on-chip decoupling capacitance as determined by (8.4). Conversely, if the impedance of the package-to-die interface is relatively high, the time required to recharge the on-chip decoupling capacitors is greater than the clock period, making the requirement represented by (8.4) insufficient. This inconsistency is largely responsible for the unrealistic dependence of the decoupling capacitance as determined by (8.4) on the circuit frequency, *i.e.*, the required decoupling capacitance decreases with frequency. Certain assumptions concerning the impedance characteristics of the power distribution network of the package and package-die interface should therefore be considered to accurately estimate the required on-chip decoupling capacitance.

The efficacy of the charge-based allocation strategy has been evaluated on the Pentium II and Alpha 21264 microprocessors using micro-architectural estimation of the average current drawn by a circuit block [263], [264], [265]. The characteristics of the power distribution network based on (8.4) are simulated and compared in both the frequency and time domains to three other cases: no decoupling

capacitance is added, decoupling capacitors are placed at the center of each functional unit, and a uniform distribution of the decoupling capacitors. The AC current requirements of the microprocessor functional units are estimated based on the average power current obtained with architectural simulations. The charge-based allocation strategy has been demonstrated to result in the lowest impedance power distribution system in the frequency domain and the smallest peak-to-peak magnitude of the power noise in the time domain.

8.6.2 Allocation strategy based on the excessive noise amplitude

More aggressive capacitance budgeting is proposed in [266], [267] to amend the allocation strategy described by (8.4). In this modified scheme, the circuit is first analyzed without an intentional on-chip decoupling capacitance and the worst case power noise inside each circuit block is determined. No decoupling capacitance is allocated to those blocks where the power noise target specifications have already been achieved. Alternatively, the intrinsic decoupling capacitance of these circuit blocks is sufficient. In those circuit blocks where the maximum power noise V_{noise} exceeds the target margin δV_{dd} , the amount of decoupling capacitance is

$$C_{\text{dec}} = \frac{V_{\text{noise}} - \delta V_{\text{dd}}}{V_{\text{noise}}} \frac{\delta Q}{\delta V_{\text{dd}}}, \quad (8.5)$$

where δQ is the charge drawn from the power distribution system by the current load during a single clock period.

The rationale behind (8.5) is that in order to reduce the power noise from V_{noise} to δV_{dd} (*i.e.*, by a factor of $\frac{V_{\text{noise}}}{\delta V_{\text{dd}}}$), the capacitance C_{dec} should supply a $1 - \frac{\delta V_{\text{dd}}}{V_{\text{noise}}}$ share of the total current. Consequently, the same share of charge as the power voltage is decreased by δV_{dd} , making $C_{\text{dec}} \delta V_{\text{dd}} = \frac{V_{\text{noise}} - \delta V_{\text{dd}}}{V_{\text{noise}}} \delta Q$. Adding a decoupling capacitance to only those circuit blocks with a noise margin violation, the allocation strategy based on the excessive noise amplitude implicitly considers the decoupling effect of the on-chip intrinsic decoupling capacitance and the off-chip decoupling capacitors [28].

The efficacy of a capacitance allocation methodology based on (8.5) has been tested on five MCNC benchmark circuits [268]. For a 0.25 μm CMOS technology, the proposed methodology requires, on average, 28%

lower overall decoupling capacitance as compared to the more conservative allocation methodology based on (8.4) [262]. A noise aware floorplanning methodology based on this allocation strategy has also been developed [268]. The noise aware floorplanning methodology results, on average, a 20% lower peak power noise and a 12% smaller decoupling capacitance as compared to a post-floorplanning approach. The smaller required decoupling capacitance occupies less area and produces, on average, a 1.2% smaller die size.

8.6.3 Allocation strategy based on excessive charge

The allocation strategy presented in Section 8.6.2 can be further refined. Note that (8.5) uses only the excess of the power voltage over the noise margin as a metric of the severity of the noise margin violation. This metric does not consider the duration of the voltage disturbance. Longer variations of the power supply voltage have a greater impact on signal timing and integrity. A time integral of the excess of the signal variation above the noise margin is proposed in [269], [270] as a more accurate metric characterizing the severity of the noise margin violation. According to this approach, a metric of the ground supply quality at node j is

$$M_j = \int_0^T \max \left[\left(V_j^{\text{gnd}}(t) - \delta V \right), 0 \right] dt, \quad (8.6)$$

or, assuming a single peak noise violates the noise margin between times t_1 and t_2 ,

$$M_j = \int_{t_1}^{t_2} \left(V_j^{\text{gnd}}(t) - \delta V \right) dt, \quad (8.7)$$

where $V_j^{\text{gnd}}(t)$ is the ground voltage at node j of the power distribution grid.

Worst case switching patterns are used to calculate (8.6) and (8.7). This metric is illustrated in Fig. 8.9. The value of the integral in (8.7) equals the area of the shaded region. Note that if the variation of the ground voltage does not exceed the noise margin at any point in time, the metric M_j is zero. The overall power supply quality M is calculated by summing the quality metrics of the individual nodes,

$$M = \sum_j M_j. \quad (8.8)$$

This metric becomes zero when the power noise margins are satisfied at all times throughout the circuit.

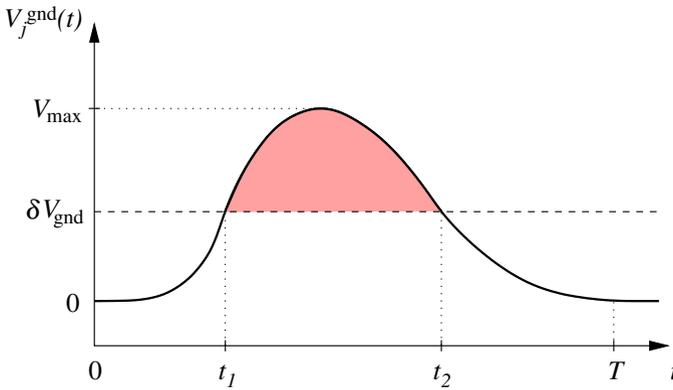


Fig. 8.9. Variation of ground supply voltage with time. The integral of the excess of the ground voltage deviation over the noise margin δV_{gnd} (the shaded area) is used as a quality metric to guide the process of allocating the decoupling capacitors.

Application of (8.6) and (8.8) to the decoupling capacitance allocation process requires a more complex procedure as compared to (8.4) and (8.5). Note that utilizing (8.6) requires detailed knowledge of the power voltage waveform V_j^{gnd} at each node of the power distribution grid rather than just the peak magnitude of the deviation from the nominal power supply voltage. Computationally expensive techniques are therefore necessary to obtain the power voltage waveform. Furthermore, the metric of power supply quality as expressed in (8.8) does not explicitly determine the distribution of the decoupling capacitance. A multi-variable optimization is required to determine the distribution of the decoupling capacitors that minimizes (8.8). The integral formulation expressed by (8.6) is, fortunately, amenable to efficient optimization algorithms. The primary motivation for the original integral formulation of the excessive charge metric is, in fact, to facilitate incorporating these noise effects into the circuit optimization process.

The efficacy of the allocation strategy represented by (8.8) in application-specific ICs has been demonstrated in [271], [272]. The distribution of the decoupling capacitance in standard-cell circuit blocks has been analyzed. The total decoupling capacitance within the circuit

is determined by the empty space between the standard cells within the rows of cells. The total budgeted decoupling capacitance (the amount of empty space) remains constant. As compared to a uniform distribution of the decoupling capacitance across the circuit area, the proposed methodology results in a significant reduction in the number of circuit nodes exhibiting noise margin violations and a significant reduction in the maximum power supply noise.

8.7 Summary

The process of designing and analyzing on-chip power distribution networks has been presented in this chapter. The primary conclusions of the chapter are summarized as follows.

- The design of on-chip power distribution networks typically begins prior to the physical design of the on-chip circuitry and is gradually refined as the structure of the on-chip circuits is developed
- The primary difficulty in the early stages of the design process is accurately assessing the on-chip power current requirements
- The primary challenge shifts to the efficient analysis of the on-chip power distribution network, once the circuit structure is specified in sufficient detail
- The complexity of analyzing an entire power distribution network loaded by millions of nonlinear transistors is well beyond the capacity of nonlinear circuit simulators
- Approximating nonlinear loads by time-varying current sources and thereby rendering the problem amenable to the methods of linear analysis is a common approach to manage the complexity of the power distribution network analysis process
- Several techniques have been developed to enhance the efficiency of the numerical analysis process
- The local impedance characteristics of an on-chip power distribution network depend upon the distribution of the decoupling capacitors
- Existing capacitance allocation methodologies place large decoupling capacitances near those on-chip circuits with the greater power requirements

- The time integral of the excess of the signal variation above the noise margin is a useful metric for characterizing the severity of a noise margin violation