The modern chip designer is required to make many complex and mutually dependent trade-offs. Power integrity (PI) is among these. Generally, one must find the most economical way to meet certain functionality and performance requirements: The design’s PI must be adequate to meet these goals. The responsibility for this is shared between various designers, including the chip architect, the logic designer, the physical layout designer, and the package designer, as well as the system team for the greater product. The integrated circuit (IC) design team focuses on areas impacting PI such as average and peak power management as well as chip floorplanning considerations. These include robust power distribution architectures, the arrangement of active power-consuming circuits, and the interspersing of passive components such as power supply decoupling capacitors.

8.1 Design for Power Integrity: Nanometer Era Considerations

While many aspects of good PI remain constant, there is a seemingly never-ending list of additional complexities that must be addressed at each new silicon technology node. As semiconductor device features continue to shrink, a growing set of design techniques is needed to achieve sufficient power scaling to enable the level of integration possible in these processes. Such power management techniques increase the complexity of meeting chip PI requirements. And as the limitations of these techniques often result in increased power consumption density across the die, even the most basic considerations become harder to address as described ahead.
8.1.1 System Requirements

System constraints play a fundamental role in shaping the floor plan of an integrated circuit design. These constraints include the PCB, the package, and the required connectivity across functional areas of the silicon itself. An example floor plan for a modern system-on-chip (SoC) design is shown in Figure 8-1.

System-driven constraints on ICs are of critical importance to a chip’s PI. These constraints have a direct impact on the integrity of the system board’s power supply to the chip package, as well as the PI of the package itself (as the cost and therefore the complexity of a package has a first-order influence on its PI). The type of package used affects the power supply impedance of the package traces, the power and ground BGA counts, and the availability of supply planes.

Figure 8-1  System-on-chip floor plan example.

1. Printed Circuit Board.
2. Ball-grid array, an arrangement of solder balls in a 2D grid for attaching a package to a board.
The extent of these package features determines the number of power and ground pads—or bumps—on the die, as well as the variations of system and package voltage at these bumps. Ultimately, these trade-offs determine how much of the system’s power integrity budget is available for use by the chip designer.

The chip floor plan also strongly depends on PCB constraints, which drive the chip’s package BGA assignments and therefore the placement of chip input/output (I/O) drive circuits on the die. I/O placement in turn affects the floor plan of the chip’s major functional blocks. The need for high performance I/Os and data flow between critical functional blocks may conflict with functional block arrangements determined by PCB constraints. This may drive further changes to the package BGA assignment and PCB routing.

8.1.2 Die Cost

Minimizing the cost of the chip is always one of the design team’s highest priorities. While there are many trade-offs that can minimize the overall cost of the die and the package, the first consideration is often to reduce die size and the number of required mask layers.

The size of contemporary chip designs is often limited by the availability of metal resources (or routing capacity), and a chip’s performance (in the maximum attained operating frequency) is often limited more by routing than by its circuits. Metal resources are therefore at a premium in chip layout. While a robust power grid is essential to PI, it must be efficiently designed to minimize chip area and maximize performance. This helps explain the increasing popularity of flip-chip-style packages, which effectively enhance the power grid of the chip itself.

It is becoming more common to see chip cores shrink (based on Moore’s Law\(^3\) technology scaling) as the rate of SoC integration struggles to keep pace. In such cases, chip sizes are often limited by the required I/O ring perimeter, since I/O devices and operating voltages do not scale as quickly as core devices and voltages. This leads to floorplanning techniques that array I/Os throughout the core to minimize overall chip size. However, this adversely affects both power and I/O signal integrity since it fractures package power planes (and similarly impacts PCB routing). In chips that use less expensive packages, and perhaps wire-bonded chips, this constraint also encourages a minimization of power and ground pads, which are critical to PI.

\(^3\) Discussed in Chapter 2.


8.1.3 Performance

As mentioned earlier, the chip design process is often one of meeting a given functionality and performance specification while minimizing the overall silicon and package cost. While some chips are designed to prioritize maximum performance over cost, most chip designs place a very high priority on performance closure, optimizing performance within cost constraints.

Automated chip “place-and-route” tools will typically place and route active circuits and then place decoupling capacitors in unused areas of the floor plan. In such a design flow, critical high-speed circuits tend to be tightly clustered, along with their clock distribution circuits. This commonly results in a very dense arrangement of cells with relatively high power demand. It correspondingly results in decoupling capacitors placed too far away to ensure a steady power supply to these critical circuits. This obviously degrades the performance required of these circuits and must be addressed, but often at the expense of complicating the overall performance-closure process.

For a chip design to meet performance requirements, it must often allocate more metal resources to signal routing in order to optimize the parasitic impedances of these connections. It must then minimize the allocation of metal resources to the power network. This is typically offset by the inclusion of one or more thick metal layers at the top of the chip stack (which are primarily used for power routing but may also be used for long-range signal buses). It is also offset by the placement of power and ground supply sources throughout the die, as in flip-chip packages, rather than just on its periphery.

8.1.4 Power Minimization

Traditionally, area and performance were the only primary constraints on the design process. In modern nanometer scale designs, the power reduction requirement relating to Moore’s Law can no longer be met solely through enhancements to the fabrication process. Scaling of chip power now requires co-optimizations in the manufacturing process, design, software, and so on. As such, chip design power management techniques are no longer exceptional. The additional overhead associated with these techniques exacerbates existing issues. For example, metal resource allocation is further strained by the additional power supplies corresponding to voltage domains or islands that typically need to be routed. These additional power supplies, as well as the impact of routing on package power planes, often further degrade package IR drop and signal integrity.
8.1.5 Other Considerations

As modern IC design is now firmly in the SoC era, bus architectures for switching signals must often span extremely long distances across the chip to connect major functional blocks with common logic and memory. As such, signal integrity design is another critical consideration. While good PI design is certainly beneficial to signal integrity, designing for the latter typically requires trade-offs related to distances between repeaters as well as the metal layers used and their corresponding width and spacing. This typically vies for metal resources with the chip power grid and distribution.

Other than power supply decoupling capacitors, unused device space is often claimed by components such as spare gates for design changes and bug (design error) fixes, electrostatic discharge protection (ESD) cells, and test circuits. These additional constraints are typically given far less priority than performance, area, and power minimization. As much as we may wish to advocate the importance of PI closure in the overall design process, the driving goal is still to meet a performance specification while minimizing system cost. An IC designer must hence make the best of limited resources to meet power integrity goals at minimal cost.

8.2 Design for Power Integrity: Techniques

A variety of techniques can be employed to ensure high PI within ICs. The choice of techniques is determined by the requirements of a given chip design. The most demanding applications are generally, though not always, high-performance chips built in the most advanced manufacturing process.

8.2.1 Power Consumption Management

From the perspective of PI, the most demanding chip designs are those with the greatest power “density” or power per area (relative to the placement of power and ground sources). It is therefore critical to minimize the power consumption of such designs. Even more importantly, power reduction can help reduce overall system complexity and cost.

Historically, it was expected that new manufacturing processes would resolve this power problem. This expectation can be traced back as far as Gordon Moore’s paper titled “Cramming More Components onto Integrated Circuits” in the April 19, 1965, edition of Electronics magazine [1]. This is the origin of “Moore’s Law,” which observed initially that “The complexity for minimum component costs has increased at a rate of roughly a factor of two per year.” In the same article, Moore pondered the problem of removing the heat generated at the levels of integration that might be permitted by this scaling, and stated that the
“the amount of capacitance which must be driven is distinctly limited. In fact, shrinking dimensions of an integrated structure makes it possible to operate the structure at higher speed for the same power per unit area.”

In modern complementary metal-oxide semiconductor (CMOS) ICs, the power that Moore refers to is the switching power dissipation, which is given as:

\[ P = \alpha CV^2f \]

where \( f \) is the switching frequency of the circuit, \( \alpha \) a circuit activity factor, and \( C \) represents the capacitive load (both device and wiring capacitances) that is being driven by the circuit. Moore’s observation was that this load capacitance would be limited by the area into which those devices would be “crammed.” It does not, however, account for the constantly increasing circuit frequencies of IC designs, the possibility of thinner dielectrics (which increase the capacitance per unit area), or nonswitching currents found in nanometer-scale ICs. Additionally, at the time of Moore’s writing, the parasitic resistance and capacitance of wiring between devices were mostly negligible, which is no longer the case.

Since the switching power of a CMOS circuit is exponentially a function of voltage, reduced operating voltages have been used to allow higher levels of device integration at increasing frequencies while keeping power dissipation density roughly constant. Transistor switching threshold voltages must be reduced correspondingly. The combined effect of this reduces the “head room” available to manage the voltage drop (i.e., the total collapse of the potential difference between \( V_{dd} \) (power) and \( V_{ss} \) (ground)) across the design. These tightened tolerances increase both the complexity and the cost of good PI.

To minimize the power consumption of CMOS circuits, we must also minimize “crowbar” currents caused by slow input transitions, which are in turn caused by the relatively large wiring resistance and capacitance of modern circuits. By minimizing these parasitic loads, crowbar currents are typically brought significantly lower than circuit switching currents. This is accomplished by inserting circuit repeaters in the wiring (which add to the overall switching power of the design), as well as by reducing resistances and capacitances per unit length through, for example, wider signal wires at greater spacing, preferably in the upper metal layers. Again, this detracts from the wiring resources available to the power distribution network.

In modern semiconductor devices, nonswitching currents represent a significant portion of total chip power. Commonly referred to as leakage power, detailed in Chapter 2, this has become a major concern for nanometer-scale designs. It is now on roughly the same order of magnitude as a device’s switching power—it may even exceed switching power—and constantly drains batteries in portable
applications. This standby power consumption arises from the subthreshold leakage current of transistors. These are a result of device scaling and the reduced transistor switching thresholds required by lower operating voltages. Many different techniques are used to minimize leakage power, including embedded power switches, active device body biasing, clock gating, the usage of nonminimal transistor gate lengths, on-chip voltage generators (e.g., low-dropout regulators or LDOs), and actively managed device operating frequencies and voltages. These techniques impact power integrity in many ways, including tighter operating voltage tolerance requirements. The increased number of power and ground nets required for many of these methods significantly diverts the wiring resources available to the chip power grid. An example of chip floor plan with multiple power domains and flip-chip bump placements is shown in Figure 8-2.

Another power minimization issue in modern circuits is related to the ability of nanometer-scale processes to integrated hundreds of millions or even billions of...
transistors on a chip. The sheer number of devices (and the enormous number of possible operational scenarios for those circuits) along with the even larger number of parasitic capacitors, resistors, and even inductors imply that it is impractical to accurately simulate the power consumed by these circuits. While the necessary abstractions are becoming more accurate, they are not entirely so. There is another cost trade-off here—the level of abstraction can be reduced by applying greater computing power and allocating more time to these simulations. There will usually still be a significant level of inaccuracy, and therefore uncertainty, in the resulting estimates of power consumption.

8.2.2 Power Grid Design

Along with the practical application of power minimization techniques, the design of the power grid is another important mechanism in maintaining good PI in the design. As discussed previously, there are many factors that limit the wiring resources available to the power grid designer, including constraints applied to signal wiring on the die, as well as modern design techniques that dilute resources across numerous unique power and ground nets. The constant need to reduce cost typically forces the use of cheaper packaging options, which are generally less conducive to good (and simple design for) PI.

The first step toward a robust power grid design is to limit the resistive voltage drop based on DC-average power estimates. An initial estimate can be reached using Ohm’s Law \( V = I \cdot R \) and budgets that account for system losses. An example follows for a nominal 1 V, 0.50 W design, with a goal of 5% maximum static voltage drop (i.e., 2.5% for \( V_{dd} \) and 2.5% for \( V_{ss} \) or 25 mV for each at 1.0 V), and using the following assumptions:

- Board + Package Power Delivery Effective Resistance = 0.60 \( \Omega \) (per pad-ball connection)
- Chip I/O Power Feed-through Resistance = 0.240 \( \Omega \)
- Chip I/O Power Pad Count = 24 (presumed to be equally distributed around die perimeter)

The characteristic resistance of the chip power grid \( (R_{grid}) \) needed to satisfy the constraints above can be easily calculated:

\[
0.025 \text{ V} = (0.5 \text{ W/1.0 V}) \cdot \left( \frac{(0.60 \text{ \Omega} + 0.240 \text{ \Omega})}{24} + R_{grid} \right)
\]

\[
= 0.5 \text{ A} \cdot (0.035 + R_{grid})
\]

\[
R_{grid} = (0.025 \text{ V/0.5 A}) - 0.035 \text{ \Omega} = 0.015 \text{ \Omega}
\]
This is validated using an electronic design automation (EDA) tool typically called a static IR (or voltage) drop checker. Such tools are commonly required to validate the power integrity of contemporary chip designs. They are increasingly used not only on the chip, but also the chip’s package and even the system board routing from the voltage regulators. An example of the results of such a simulation is shown in Figure 8-3.

Whether or not the design is assisted by such tools, the required characteristic power grid resistance is achieved through the prioritized usage of chip wiring layers. Most contemporary chips are manufactured in processes that have at least one or two thick interconnect layers (more in high-performance applications) to

**Figure 8-3** Example of static voltage drop analysis from the power supply regulator on a PCB, through the package, and to the circuits on the silicon die itself. In this example, the PCB and package routing account for no more than about one quarter of the total static IR drop for the system.
provide low-resistance routing options for power meshes (as well as long-distance signal wiring). It is also common to leave two or more layers for intermediate wire lengths, as well as one or two layers of local (short) interconnect. It is common practice to use a small amount of one or two of the intermediate layers as part of the power mesh, and to use the remaining lower layers as vias from the mesh to the underlying circuits. In designs using embedded power switches, these vias typically connect to the switches themselves. Additional layers must be used for localized distribution of the switched power supplies to the functional circuits.

While the simplistic computations described above are a useful part of the design process, the final power grid must be created in the context of the floor plan of the chip’s functional blocks and power and ground sources (I/O power and ground cells and/or power and ground bumps for devices using flip-chip packaging). These blocks may create gaps in the chip’s power mesh and also represent singular points of power dissipation on the die. Static IR drop validation tools are being adapted to provide power grid synthesis tools, which recursively execute calculations similar to those above, validate the design, and modify it until the requirements are met. The relative weighting of chip routing layers is typically provided as a constraint to these algorithms. Such a power grid synthesis can be executed with a chip block floor plan or without one (e.g., in order to gauge the metal resources that need to be reserved for power distribution to facilitate early chip signal routing experiments). Using these tools lets us resolve discontinuities in the basic chip power mesh, in the distribution of power and ground sources, and in circuit blocks, which dissipate power. These characteristics can be used to provide denser power routing where needed or sparser routing in areas of relatively low power density. This permits more efficient area optimization and routing resource utilization, which can in turn help reduce device cost.

The long-term reliability of ICs can be compromised by excessively high static currents due to a phenomenon known as electromigration. This is caused by the electrostatic forces created by the electric field, producing an effect sometimes referred to as “electron wind.” This force is capable of moving portions of the conductor material, which, over time, gradually increases the resistivity of the conductor and can cause intermittent circuit failures. Over a longer period of time, the conductor may become an open circuit (as shown in Figure 8-4). In addition, the conductor material that has migrated may collect in other parts of the conductor, thereby widening it and possibly shorting to other wires.

This issue is avoided in design by analyzing the currents in each wire and via against a maximum average current density for each layer. This is performed in a manner similar to static voltage drop analysis, although the worst case conditions
may be different for each, and therefore may require power estimation and extraction of the electrical model of the wires to be done twice.

### 8.2.3 Chip Floorplanning and Decoupling Capacitance

Validating a design’s static PI may be sufficient for many applications, but a dynamic voltage drop analysis is increasingly required to validate the effect of switching power on circuit performance and noise margins. To enable this analysis for the silicon die itself, circuit macros are typically modeled as time-variant current sources and attached to the chip’s power grid. The grid is typically extracted as a resistive-capacitive (RC) network, including any existing placements of

---

**Figure 8-4** Scanning electron microscopy image of a failure caused by electromigration in a copper interconnect. The dark areas along the copper line in the middle of the picture are voids where the copper has been physically moved over time by excessive static electrical current. (Source: Patrick-Emil Zörner, Wikimedia Commons)
dedicated power supply decoupling capacitors. The remaining switching network is typically still very complex and requires a significant amount of time and computing power to simulate. A key simplification is to select a set of input switching patterns, or vectors, that simulate the worst-case peak current situations. There may be multiple situations that must be analyzed to completely validate a design, including the various operation modes of devices with high integration levels. We must also analyze device start-up and situations related to leakage power management, such as clock gating and exiting sleep modes. This level of validation involves a significant investment in simulations, which must be treated as a cost and time trade-off in the design process.

The die, package, and system board combine to form a complex RLC network, which must be accounted for. Dynamic voltage drop analysis is not only an IR drop simulation, but must also include the effect of inductance and \( \frac{d}{dt} \left( L \cdot \frac{di}{dt} \right) \) on supply voltage. As this additional complexity increases simulation cost, simplifications are commonly employed to diminish the problem. For example, the die itself is typically reduced to a relatively small number of lumped switching current sources and RC networks, which can be coupled with a detailed electrical model of the chip’s package. Likewise, a simplified package model using lumped parasitic elements can be used as part of a detailed die analysis. Used together, these simplifications can often provide a reasonably accurate analysis at reduced simulation cost.

The results of such analysis can be used to evolve the power grid design as well as the chip floor plan itself. Modern EDA tools allow the designer to explore optimizations to correct weaknesses in the chip power grid, improve the placement of decoupling capacitors or power and ground sources, and refine chip package design. The designer can also use these results to improve the floor plan of the chip’s active circuits. Specifically, areas of high power dissipation density can be modified to reduce concentrations of power demand. Such changes are typically required for circuit paths with challenging timing constraints, which cell placement tools cluster tightly together. While this is useful for meeting the circuit timing requirements, such clustering often results in little or no available area for local decoupling capacitor placements relative to the increased power density. The clustering can be relaxed to improve the dynamic PI of the region. An example of this situation is shown in Figures 8-5 and 8-6.

Another technique used to manage PI involves a subtle characteristic of the power mesh itself. Placing the power straps of the mesh as close to the ground
straps as possible—versus placing one precisely one half of the repeat pitch of the other, for example—allows the parasitic lateral coupling self-capacitance of that interconnect layer to be utilized as “free” decoupling capacitance in the grid. Circuits close to these pairs of straps will experience relatively low voltage drops. Other circuits placed between the sets of pairs will, however, experience a significant drop in $V_{dd}$ delivery as well as a rise in $V_{ss}$, in terms of static and dynamic PI. This is due in part to the increased distance from the extra capacitance provided by the paired straps. The alternative of staggering the power and ground straps tends to smooth the difference between the best and worst drops (e.g., a circuit experiencing a worst-case $V_{dd}$ drop would typically be close to the $V_{ss}$ straps and vice versa, thereby leveling total supply collapse for most circuits). In a third approach, paired straps are incorporated on the uppermost metal layers in the mesh while utilizing a uniformly staggered placement on the lowermost layers of the mesh.

Figure 8-5  Example of a dynamic voltage drop failure. The picture on the left highlights the maximum dynamic voltage drop. The picture on the right shows the floor plan for the same block.
8.3 Power Management and Power Integrity

The drive toward low-power devices comes from two distinct fronts. On one side, there are SoCs for handheld devices running on batteries, which need aggressive power management techniques to minimize energy consumption and increase battery life. On the other side are devices such as high-performance microprocessors and DSPs, where power must be managed in response to the system’s thermal constraints and a need to keep chip junction temperatures low.

The simplest way to minimize dynamic power is by lowering the supply voltage. Over the past decades, very-large-scale integration (VLSI) technology has moved into smaller geometries and lower voltages, allowing low-power devices with more functionality and higher performance. Most of the power benefit came from process entitlement: power could be scaled as we moved to smaller geometries. This is because supply voltages were scaled to keep the electric fields