Foundations and Trends® in
Electronic Design Automation
Vol. 4, No. 4 (2010) 223–306

DOI: 10.1561/1000000014

now
the essence of knowledge

System-in-Package: Electrical and Layout Perspectives

By Lei He, Shauki Elassaad, Yiyu Shi,
Yu Hu and Wei Yao

Contents

Foundations and Trends® in
Electronic Design Automation
Vol. 4, No. 4 (2010) 223–306

DOI: 10.1561/1000000014

now
the essence of knowledge

System-in-Package: Electrical and Layout Perspectives

Lei He[1], Shauki Elassaad[2], Yiyu Shi[3], Yu Hu[4] and Wei Yao[5]

[1] *University of California, Los Angeles, CA 90095, USA, lhe@ee.ucla.edu*
[2] *Stanford University, Stanford, CA 94305, USA, shauki@stanford.edu*
[3] *Missouri University of Science and Technology, Rolla, MO 65409, USA, yshi@mst.edu*
[4] *University of Alberta, Edmonton, Alberta T6G 2R3, CANADA, bryanhu@ece.ualberta.ca*
[5] *University of California, Los Angeles, CA 90095, USA, weiyao@ee.ucla.edu*

Abstract

The unquenched thirst for higher levels of electronic systems integration and higher performance goals has produced a plethora of design and business challenges that are threatening the success enjoyed so far as modeled by Moore's law. To tackle these challenges and meet the design needs of consumer electronics products such as those of cell phones, audio/video players, digital cameras that are composed of a number of different technologies, vertical system integration has emerged as a required technology to reduce the system board space and height in addition to the overall time-to-market and design cost. System-in-package (SiP) is a system integration technology that achieves the aforementioned needs in a scalable and cost-effective way, where

multiple dies, passive components, and discrete devices are assembled, often vertically, in a package. This paper surveys the electrical and layout perspectives of SiP. It first introduces package technologies, and then presents SiP design flow and design exploration. Finally, the paper discusses details of beyond-die signal and power integrity and physical implementation such as I/O (input/output cell) placement and routing for redistribution layer, escape, and substrate.

1

Introduction

Since birth of the integrated circuit (IC), the ever-increasing integration level has been enabling more functions at reduced cost. This has been primarily driven by Moore's Law, which dictates the scaling of a single chip in the past half-century. On top of this, at the system integration level, technologies such as wafer-scale integration and multi-chip modules (MCM) have been explored to further increase the design size and reduce the cost. Today, with the growing scalability of semiconductor processes, the higher level of functional integration at the die level, and the system integration of different technologies needed for consumer electronics, system-in-package (SiP) is the new advanced system integration technology, which integrates (or vertically stacks) within a single package multiple components such as CPU, digital logic, analog/mixed signal, memory, and passive and discrete components in a single system.

SiP reduces the form factor of a system. Compared with system-on-a-chip (SoC), SiP decreases the cost due to the following reasons. First, different components may be fabricated in different generations or different types of technologies, without complications and high cost associated with integrating heterogeneous technologies in one process.

Second, the same component can be fabricated in a large volume and used for different systems, amortizing the ever-increasing non-recurring engineering expenses such as those for designing and mask. Finally, the size of each individual die of the SiP is much smaller than the size of the chip if SoC is used for the same system. Smaller size improves yield rate and reduces production cost. It also makes design easier and reduces time-to-market.

While SiP clearly has advantages, the design complexities and costs associated with designing the package and integrating the different components in a system may eclipse the design challenges of the stand-alone dies. Packaging has evolved over the years from the point where chips had few pins to designs that have thousands of pins. Traversing the evolution of the electronic packaging, different technologies have been designed and adopted to solve the design and cost problems associated with the ever-increasing number of I/Os. Electronic packaging has started with dual-in-line package (DIP), and evolved to include a variety of technologies such as tape-automated bonding (TAB), pin grid array (PIG), ball grid array (BGA), and many other forms of system outline packages (SOP) and chip-scale packages (CSP). SiP with multiple dies and passive components in one package introduces more design challenges than CSP.

This survey focuses on electrical and layout perspectives of SiP, without discussing thermal and mechanic characteristics of SiP. In addition, this survey does not consider three-dimensional (3D) integration using through-silicon vias (TSVs). The remainder of the survey is organized as follows. Section 2 presents a tutorial on IC package, and Section 3 introduces overall design challenges and design exploration of SiP with consideration of beyond-die power and signal integrity, and Section 4 presents placement and routing for SiP.

2

IC Package Tutorial

ICs are created to integrate an increasing number of devices in a given area of silicon chip using technologies such as metal-oxide semiconductor (MOS), bipolar, bipolar-complementary MOS, and gallium arsenide technologies [69]. These silicon chips have to be protected from the environment; electrical connections have to be created to the external world; and the generated heat must be effectively dissipated. In other words, the IC must be packaged for use in an electronic system [69].

IC packaging supplies the chips with wires to distribute signals and power, typically providing a transposition from a tighter I/O pitch at the die to a wider pitch at the next level of packaging, removes the heat generated by the circuits, and provides chips with physical support and environmental protection [33]. All functions must occur in the most cost-effective way without significant performance reduction. As a result, the best IC package contains the chip and does not draw attention to itself. To achieve this, the IC package should be compact, the wiring on the package should be very dense, and the extra interconnections should not disrupt high-speed signal transmission. The package should provide a stable power supply level and should not cause the die temperature to exceed the performance and/or reliability threshold.

It should protect the chip and avoid stress-induced cracks and failures. The package should cost much less than the chip it carries [5].

With developments in the area of electronic equipment, more devices are accommodated within a given chip and the number of functions a chip can perform is enhanced. This increase in functional complexity leads to an increase in the number of a chip's inputs and outputs as well as an increase in the amount of power that is dissipated by the device as heat. Meanwhile, the additional materials and structures used in packaging increase the thermal resistance from the chip to the ambient, increase the electrical delay, and reduce the reliability of the device due to material incompatibility [33]. The requirements at the product level, however, are continuously increasing in terms of performance, size, weight, and operating conditions. Any one type of packaging cannot possibly meet the present day range of product requirements. Consequently, a large variety of chip-level package configurations and technologies have been created and new ones are constantly introduced. This section presents a very brief overview of semiconductor packaging technologies and packages.

In Section 2.1, a general packaging hierarchy is first introduced and then, according to this hierarchy, different die-to-package interconnect methods are presented in Section 2.2. Several package substrate materials are compared in Section 2.3. In Section 2.4, different types of packages are introduced according to different package-to-board interconnections. Finally, the multi-chip module, SiP, and some future trends are discussed in Section 2.5.

2.1 Packaging Hierarchy

After fabrication, semiconductor wafers are diced and chips are mounted on the carrier. Chip carriers can be made of many different materials, including organic or ceramic materials, or even silicon. They can also have from as few as a dozen pins to thousands of pins. The carriers may also be composed of multiple levels of materials as well such as package on package. As a result, depending on cost and performance requirements, the chip carrier can be in many different forms. The chip is mounted on the package on a substrate or metal lead frame by a

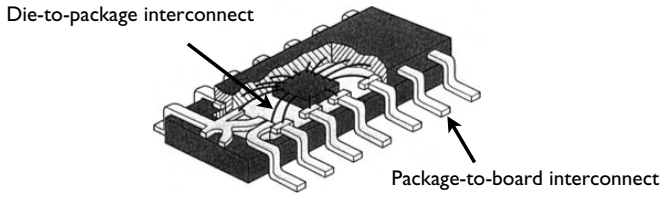


Fig. 2.1 Elements of a semiconductor package [5].

die attach material, which will permit heat conduction while assuring mechanical stability. The three primary types of die attach materials used are soldering, metal-filled polymers, and metal-filled glass [10].

The packaged ICs are then placed on the printed circuit boards (PCBs), which connect to other boards by connectors and cables [33]. The lead, pin, or pad in a semiconductor package connects a conductor on a PCB to the body of the package (*package-to-board interconnect*) and another connects to a bond site on the chip (*die-to-package interconnect*), as shown in Figure 2.1. The details of package interconnects are discussed in the following sections.

2.2 Die-to-package Interconnect

The die-to-package interconnection refers to the technology required to get electrical signals into and out of the IC [5]. In other words, it connects the bonding pads on the IC and the pins of the package. This is generally accomplished by wire bonding, flip chip bonding, and TAB. The structure of these interconnections is illustrated in Figure 2.2.

2.2.1 Wire Bonding

Wire bonding is the oldest interconnect approach and is still the most widely used method today, particularly for chips with moderate lead counts (i.e., <200) [10]. The process uses gold or aluminum wires to connect between I/O pads located around the periphery of the silicon die and its associated package pin, shown in Figure 2.3. These interconnections are created one at a time and the process is time consuming because each wire, requiring two bonding operations, must be attached individually [10].

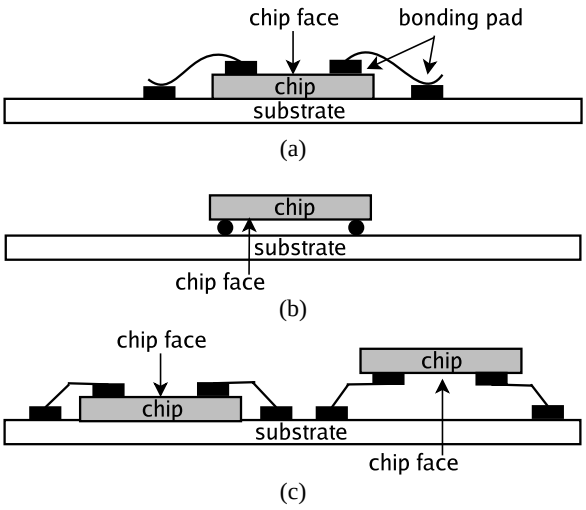


Fig. 2.2 Illustrations of (a) wire, (b) flip chip, and (c) tape-automated bonding [5].

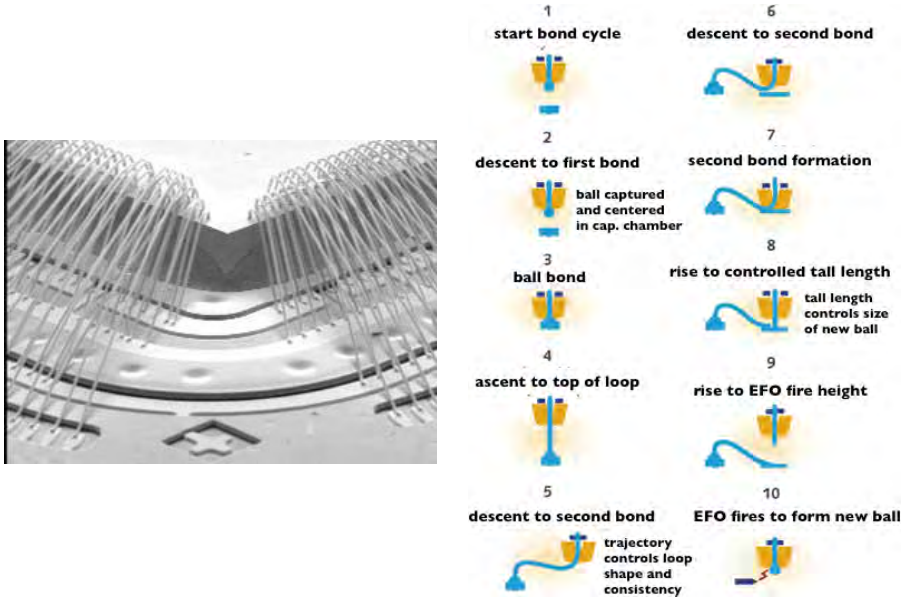


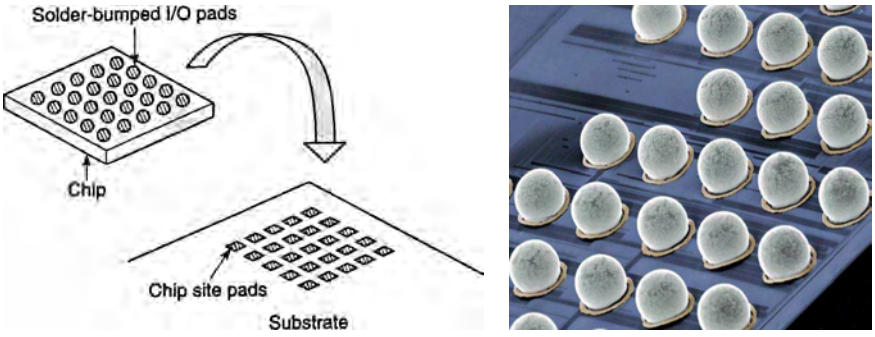
Fig. 2.3 (a) Example of wire bonding and (b) wire bonding (ball bond) steps [68].

In the gold wire bonding process, the tip of the wire is melted to form a ball [68]. A detailed procedure for wire bonding (ball bond) is shown in Figure 2.3. The bonding tool is first in position over the bond pad of the chip and attaches the chip through a spherical ball using thermal or ultrasonic energy. Then, the tool is moved to the package bond lead. Once over the bond lead, the tool comes back down and repeats the process to complete the interconnection [68]. On the other hand, the process used with aluminum wire is known as "wedge bonding" and the bond head is actually shaped as a wedge. The wire is fed through a wire guide and extends beneath the wedge. Wedge bonding is usually used for fine-pitch applications [33].

The major advantages for wire bonding are its highly automated assembly infrastructure. Bell Labs introduced wire bonding process in 1957 and since then, the reliability of the process and the speed of the bonding equipment have significantly improved. Wiring changes can also be accomplished very quickly and easily, without special tooling and material changes [10]. Limitations of wire bonding, however, include the requirement for minimum spacing between adjacent bonding leads and the limited bonding space around the periphery of the chip. As the number of I/O and power/ground (P/G) pads increases, wire bonding these pads without causing shorts between wires becomes challenging [5]. This leads to the so-called "*pad limited design*" and forces the designer to either increase the chip size or reduce the wire bond pitch, which may lead to yield and reliability problems [33]. In addition, the cost of wire bonding increases proportionally to the lead count because of the peripheral bonding and sequential process. The large parasitic inductance of long looping bonding wires also causes severe signal integrity issues. Signal distribution is also constrained because the I/O drivers must be located near die periphery. Power distributed only around the periphery also causes significant voltage drops across the entire chip and leads to a power integrity issue [33].

2.2.2 Flip Chip Bonding

The first flip chip bonding, also known as *Controlled Collapse Chip Connection* (C4), was introduced by IBM in 1964 [69]. In flip chip



(a) (b)

Fig. 2.4 (a) Flip chip bonding and (b) a fabrication example [33].

bonding, the chip is mounted upside down onto the carrier, as shown in Figures 2.2(b) and 2.4(a). The connection is made through solder bumps so that the pads and bumps on the chip directly align with the bond pads on the package, as shown in Figure 2.4(a). In Figure 2.4(b), wafers are "*bumped*" with interconnect metallurgy across the active side of the die. Once the chip is flipped and mounted into position, a reflow-soldering technique forms all the joints between the chip and package simultaneously.

The solder bumps are located all over the surface of the chip in the form of an array or random pattern and, compared with the periphery limitation in wire bonding, do not constrain the I/O capability. All bumps are formed to the IC at the same time by a process called *gang bonding* [69]. As a result, the lead count is proportional to the chip area rather than the die periphery and the bonding cost is essentially independent of the lead count. Moreover, the reflow process is self-aligning in nature because of the shape and the surface tension of the molten solder.

The major advantages of flip chip bonding are as follows. The I/O density is only limited by the minimum distance between adjacent bonding pads and the entire chip area can be used for I/O [10]. Signal bumps can also be addressed at locations other than the periphery of the chip. In addition, the interconnect distance between chip and package is minimized, which results in smaller parasitic inductances

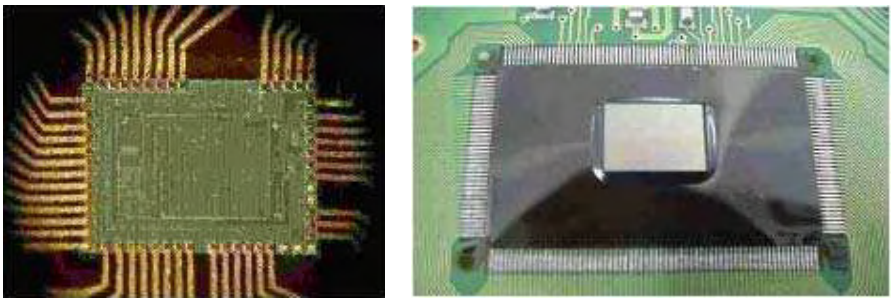
and resistances that improve signal integrity as compared with wire bond. The power and ground can also be distributed on the package rather than the chip and can be accessed anywhere on the chip. On the other hand, reliability concerns are the major manufacturing disadvantages of flip chip bonding [33], especially for larger dies or dies with low K dielectrics. Most flip chip interconnects require underfill to meet reliability requirements. Non-conductive underfill is used to fill the space between the chip and the package and mechanically adhere the entire chip surface to the substrate. The system depends on underfill to ensure the intimate attachment of hundreds or thousands of interconnects between dissimilar materials with vastly different coefficients of thermal expansion (CTE) and underfill typically precludes assembly rework after test [33].

2.2.3 Tape-Automated Bonding

TAB technology is the process of mounting a die on a flexible tape made of polymer material [10], such as polymide, as shown in Figure 2.2(c). The mounting is done such that the bonds of the die, usually in the form of bumps or balls made of gold or solder, are connected to metal stripes on the tape, which allows the die to connect to the package or directly to external circuits [10]. The TAB bonds connecting the die and the tape are known as inner lead bonds (ILB), whereas those that connect the tape to the package or to external circuits are known as outer lead bonds, as shown in Figure 2.5.

For ILB, the attachment to the tape is affected by thermocompression bonding and all bonds to the IC are formed via the gang bonding process. Copper, a commonly-used metal in tapes, can be electro-deposited onto the tape or simply attached to the tape using adhesives [10]. The metal patterns of the circuit are imaged onto the tape by photo-lithography. Sometimes the tape on which the die is bonded already contains the actual application circuit of the die [89], and the IC can be tested and burned-in, allowing the elimination of defective chips from further packaging process.

The main advantage of TAB, which is similar to flip chip bonding, is its large number of I/O that can be supported by attaching the TAB



(a) (b)

Fig. 2.5 (a) Inner lead bonds and (b) outer lead bonds [89].

tape to the entire surface of the chip rather than just the periphery. Moreover, the shape of the interconnect and the use of copper in tapes provide low-inductance and low-resistance interconnection, which minimizes signal distortion [10]. However, TAB requires the use of complex metallurgy, multi-layer solder bumps in order to affect a bond on either the tape or IC. A TAB tape can only be used for a chip and package, which matches its interconnection pattern. As a result, each TAB tape needs to be custom-designed.

2.3 Package Substrate

2.3.1 Lead Frames

The lead frame is made of copper alloy and plated with gold and silver or palladium, either completely or in selected areas over nickel or nickel/cobalt. The silicon chip is usually attached to the lead frame with an organic conductive formulation of epoxy [10]. Gold or aluminum wires are bonded to the aluminum bonding pads on the chips and to the fingers of the lead frame, as shown in Figure 2.6. The peripheral leads of the package are then attached to the board.

The main advantage of using a lead frame is its ubiquitous assembly manufacturing capability available worldwide [10]. However, it suffers from thermal, electrical, and mechanical performance limitations as well as lead count restrictions. Typical examples of lead frame substrates are quad flat package (QFP) and DIP.

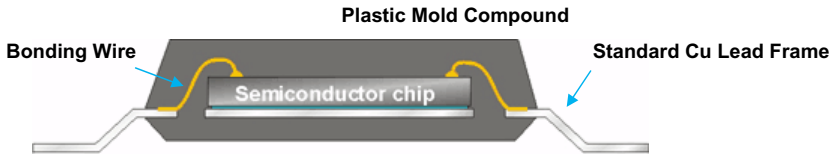
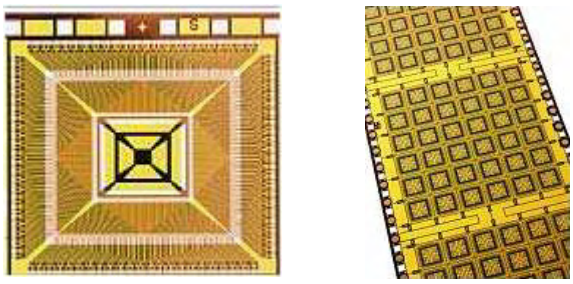


Fig. 2.6 Illustration of lead-frame-based plastic package [10].



(a) (b)

Fig. 2.7 Examples of flex substrate [33].

2.3.2 Flex

Flex is typically a single-sided-polyimide-based circuit tape, although two-metal tape is available. Either copper is electro-deposited to the tape or rolled copper is attached to the tape using an adhesive. The circuitry is then imaged using a photo-lithography process. Some examples of the flex substrate are shown in Figure 2.7. The principal advantage of the flex substrate is the tight pitch of the circuit, which can provide high wiring density suitable for high pin count devices. Moreover, flex substrates can be bent or twisted and this ability to conform is required for some 3D applications [33].

The main disadvantage, however, is that the flex must be rigidified for most mainstream IC package applications and there are a limited number of metal layers in conventional flex.

2.3.3 Printed Circuit Board

The fine-pitch PCB is a thin core substrate with minimal fiberglass reinforcement and can be compatible with CO_2 laser drilling [33]. It also

provides very fine-pitch plated-through-holes (PTH) and usually contains 2 or 4 metal wiring layers. Most importantly, the PCB substrate has a very competitive cost and further cost reductions are expected in the future as drilling, handling, and testing of thin core substrates are refined. PCB substrate is typically used with a wire-bonding, overmolding, BGA attachment and can also be used with a low-power, low-lead count flip chip interconnect and underfill [33].

The main advantage of this substrate is its competitive cost and compatibility with both flip chip and wire bonding interconnect. The disadvantage is, however, that the PCB substrate has limitations on wireability for high lead count devices and limitations on thermal dissipation in *cavity up* configurations [33].

Figure 2.8 shows the two different configurations for wire bond substrates: *cavity up* and *cavity down*. Traditional *cavity up* substrate does not allow for good thermal dissipation due to high thermal resistance. Devices are limited to $2-3W$, depending on die temperature tolerances. Cavity down substrate construction was invented to address this limitation and greatly enhances thermal dissipation capability by using a metallic lid as heat spreader, as shown in Figure 2.8(b). Moreover, cavity down substrates can also facilitate multi-tier wire bonding for high lead counts and provides better signal fidelity [33]. A typical example of a cavity down wire bond is a graphics processor. Note that for cavity down substrates, while power dissipation is excellent, power distribution is still marginal due to peripheral wire bonding [33].

2.3.4 Organic Build-up Substrates

Organic build-up package substrates were developed in response to the need for high-speed, large-scale integrated chips. Sequential

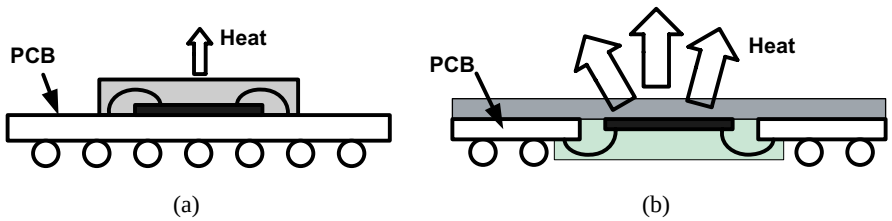


Fig. 2.8 (a) Cavity up and (b) Cavity down wire bond substrate [33].

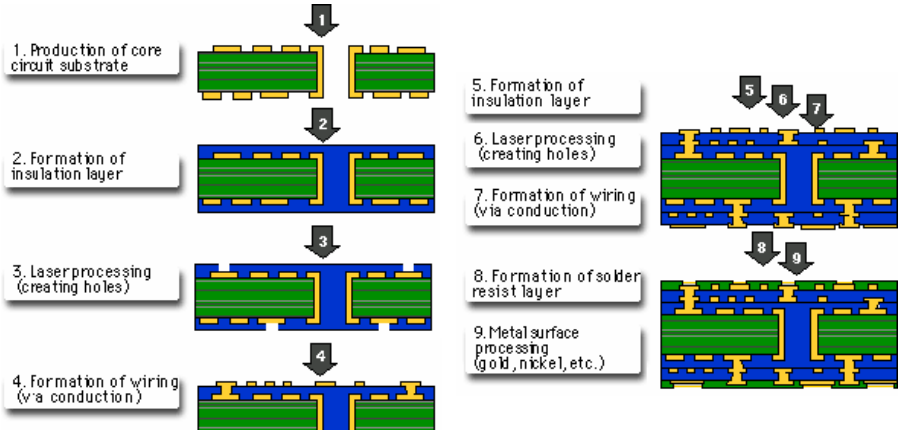


Fig. 2.9 Manufacturing process for multi-layer organic build-up substrate [88].

high-density layers with fine-pitch wiring and laser etched vias in unfilled dielectrics were built around a 2 or 4 metal layer drilled PCB core, as shown in Figure 2.9 [88]. Note that all signals must flow from the top side (chip attach) to the bottom side (balls) through the PTH in the core. As circuits are formed on pure copper conductors, resistance is low. Additionally, as an insulation layer is formed by organic resin, permittivity is low and the substrate is very lightweight. Furthermore, it has very fine-pitch wiring capability on a per-layer basis.

Lamination in the organic build-up substrates, however, must be balanced or it is subject to significant warpage. High-density wiring layers on the bottom side of the package substrates are usually not fully utilized in the design [33]. Also, as mentioned, all signals must traverse the PTHs in the core, which can represent a large impedance mismatch. Finally, organic substrates are sensitive to moisture which can impact the performance of the underfill.

2.3.5 Ceramic Substrates

A ceramic substrate is fabricated with screened metal paste on punched ceramic dielectric sheets that are aligned and fired [33], as shown in Figure 2.10. Ceramic provides a rigid, stable surface with superb dielectric strength and very good high-frequency performance. Moreover, ceramic substrate has both a low CTE when compared with organic

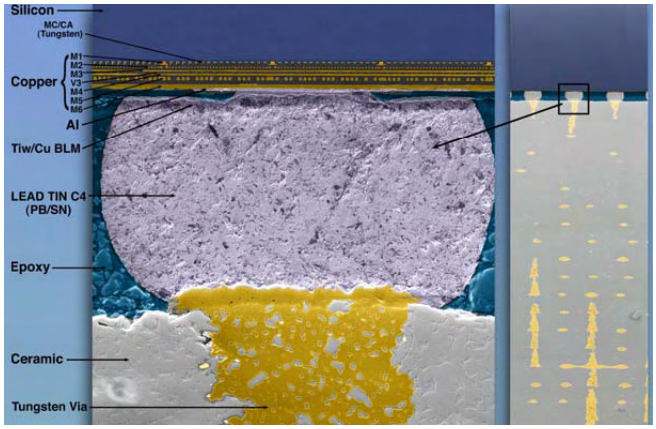


Fig. 2.10 Cross-section of ceramic substrate [33].

substrates and a lower thermo-mechanical stress with large die [33]. It also has higher substrate thermal conductivity, which results in better intrinsic thermal performance and an insensitivity to moisture. Ceramic substrate, however, at the same time creates CTE mismatches with board for large BGA modules because of its low CTE property [33]. Taking into consideration the limited supply base, a higher infrastructure cost is also one of the major disadvantages for ceramic substrates [33].

Ceramic substrates have advantages that include low-incremental cost-per-wiring-layer, a stable dielectric constant, and a low dissipation, which is an important factor for analog and RF designs [33]. Additional wiring layers can also be added easily to the design. On the other hand, ceramic substrates have a lower wiring density on a per-layer basis compared with organic build-up substrates and may create a special problem in flip chip fanout wiring on large lead count devices. Furthermore, screened and fired metal is not as conductive as plated metal on organic substrates, which causes inferior lateral power distribution within the module. However, there are ceramic packages using copper instead of the less conductive tungsten.

2.4 Package-to-board Interconnect

The package-to-board interconnect refers to the electrical connection of an IC to a circuit board, such as the conventional Printed Wiring

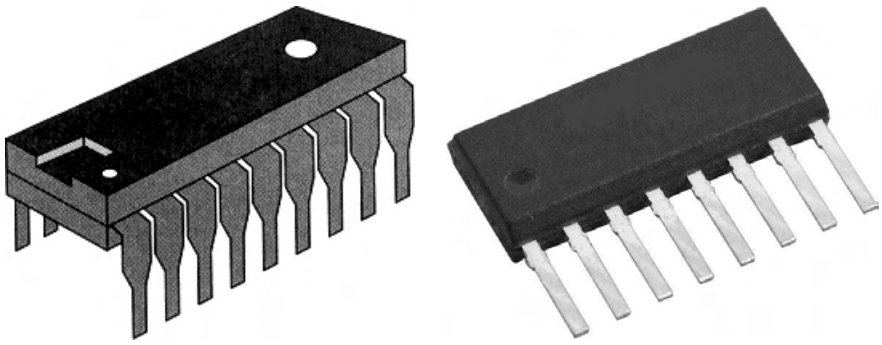
Board (PWB). According to the mounting method, the packages can generally be divided into three categories: *through-hole mounting*, *surface mounting*, and *direct die mounting*. Further categorization can be achieved according to available I/O locations, I/O density, and different substrate types.

2.4.1 Through-hole Mounting

Through-the-board mounting technology has been the traditional low-cost packaging choice and is still used nowadays. The PC board is manufactured by stacked layers of thin copper sheets and epoxy fiberglass insulating layers on top of each other [5]. These copper sheets are patterned and etched to form the interconnection and holes are drilled through the board and plated with copper. This copper-plated holes serve as the connections between different layers and also connect with the packages with their through-the-hole metal pins. Wave soldering is used to adhere the metal pins with the plated holes and their surrounding contact area.

2.4.1.1 Dual-in-line Package

The DIP, as shown in Figure 2.11(a), is one of the earliest packaging styles introduced. It is available in low-cost plastic and hermetic ceramic types, with through-hole mounting I/O terminals. DIP was



(a) (b)

Fig. 2.11 (a) Dual-in-line package and (b) Single-in-line package [5].

extremely popular in the early era of ICs because it was cost-effective and reliably satisfied the electrical and mechanical requirements for many years [5, 69]. With its 8 through 40 I/O terminals, it satisfied the needs of a wide range of IC devices and also the assembly requirements on PCBs either through manual soldering or using automatic termination techniques.

Similarly, single-in-line packages, as shown in Figure 2.11(b), are rectangular with leads on one of the long sides. This type of packaging is mainly used to provide a means for packaging memory chips in a high-density format [5]. When inserted into a PCB, many of these can be placed side by side and stacked closely together, separated by just the thickness of the package, because the leads or contacts are located along just one side of the package [5].

2.4.1.2 Pin Grid Array

PGA is an array package available with ceramic and organic substrates [5]. The pins are arranged across the bottom of the package and used for through-hole mounting, as shown in Figure 2.12. The ceramic PGA package body is typically constructed as a cofired multilayered ceramic and wire bonding is usually used for chip to package interconnect, although TAB and C4 interconnects can also be used [5]. PGA packages played a dominant role in high-density packaging for many years because of their universal assembly compatibility and easy implementation as a Pb-free interconnect. However, PGA packages had drawbacks such as high cost, through-holes on the PCB for mounting, as well as the high inductance and resistance associated with pins and

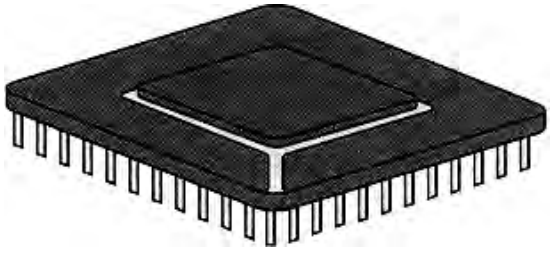


Fig. 2.12 Pin grid array [5].

socket housing [5]. The expensive pins of PGA packages were replaced by cheaper solder balls in BGA packages, as discussed later.

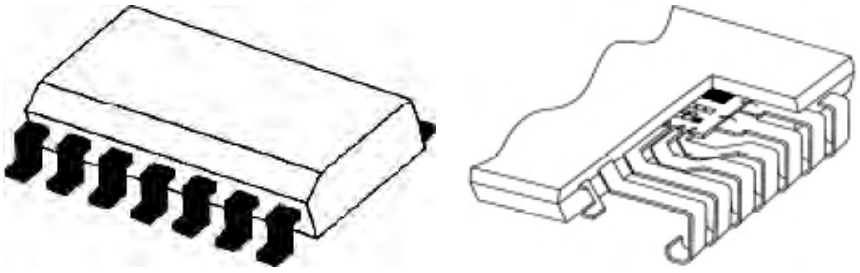
2.4.2 Surface Mounting

In surface mounting, a chip carrier is soldered to the pads on the surface of a board without requiring any through-holes [5]. The smaller component sizes, lack of through-holes, and possibility of mounting chips on both sides of the PC board improve the board density. In addition, the manufacturing process is easier since placing chips is much simpler than putting pins through-holes.

Surface mounting, however, has disadvantages including thermal expansion mismatches due to lack of support pins. It is also difficult to test for faulty mounting connection due to electrical and mechanical defects.

2.4.2.1 Small Outline Package

A small package body characterizes the small outline peripheral leaded packages [69]. These packages have surface mount leads on two sides. The primary advantages of a small outline package are its small size and its suitability for surface mounting. The gull wing package, shown in Figure 2.13(a), is commonly known as SOP. The SOP is a plastic molded lead-frame-based package [69]. The lead is formed away from the body in the shape of a gull wing, for surface mounting. On the



(a) (b)

Fig. 2.13 Small outline package with (a) gull wing leads and (b) J-bend leads [5].

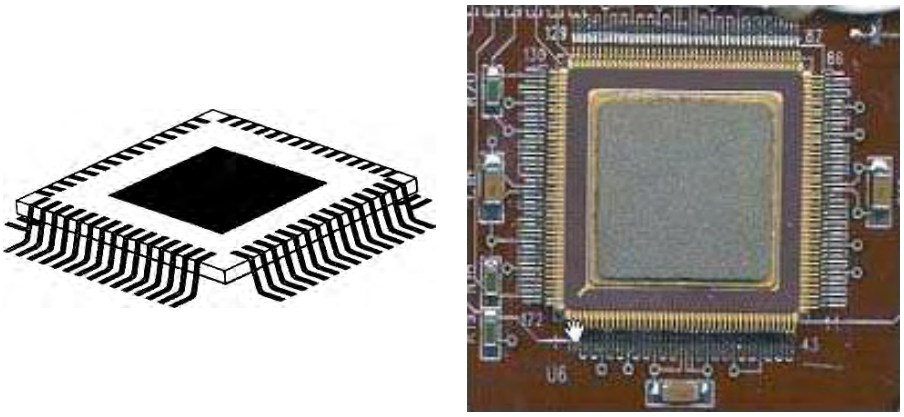
other hand, Figure 2.13(b) shows the J-bend SOP. It is also a lead-frame-based package with a molded plastic body. The *J* in the package name refers to the lead configuration. The leads extend out from the side of the package and wrap underneath the body, forming the shape of the letter J [69].

2.4.2.2 Quad Flat Package

The QFP is a surface mounting package characterized by a larger body and higher lead counts than an SOP package, with soldered leads on all four sides (Figure 2.14) [5]. The QFP is compatible with lead frame substrates and popular because of its low cost and low power property. It has very high manufacturing volumes and standardized assembly can be done anywhere in the world at a very low cost. However, the lead counts are still limited by the peripheral and the QFP needs non-standard thermal enhancements for higher power devices, which can be expensive.

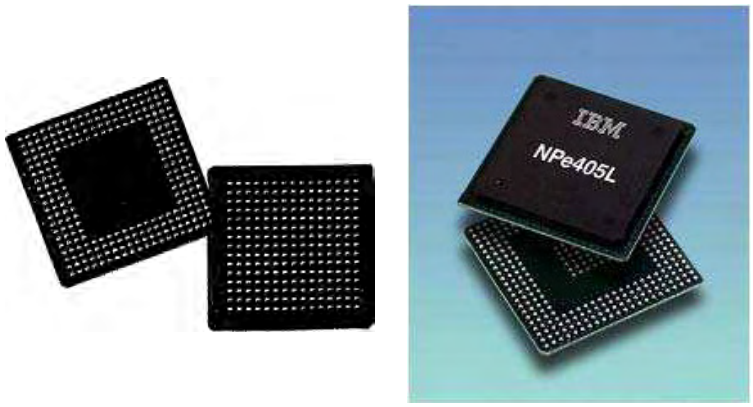
2.4.2.3 Ball Grid Array

BGA has an area array of solder balls and, as a result, can support high lead count devices (Figure 2.15) [33]. Unlike PGA that also has area



(a) (b)

Fig. 2.14 Quad flat package [5].



(a) (b)

Fig. 2.15 Ball grid array [5].

array contact, BGA is compatible with the surface mounting technology (SMT) board assembly. The package-to-board connection is done directly through the reflow process [33]. Moreover, the SMT attachment improves board-level wireability with improved power and signal distribution on the board. Compared with the QFP that also supports the SMT, the most evident advantage is the package area: BGAs typically are 20–25% smaller than QFPs, given the same lead count [5]. In addition to the mechanical considerations, BGA packages offer better thermal and electrical characteristics than QFPs [69]. The most obvious drawback to using a BGA package is its significant higher cost. Similar to other SMT packages, BGA suffers from thermo-mechanical stress on the joints due to warpage and CTE expansion of dissimilar materials. Also, BGA interconnects cannot be reliably socketed due to solder creep.

2.4.3 Direct Die Mounting

Direct die mounting, also known as chip on board, eliminates the chip carrier and places die directly on a PWB. The die can be attached to the substrate using wire bonding, TAB, or flip chips with solder bump connections.

2.5 Multi-chip Modules and SiP

SiP and MCM in recent years have seen expanded applications [46]. MCM interconnect and package more than one bare IC chip (die), enable the designer to drastically reduce the interchip distance (delay), and lead to significant size reduction of the product [10]. On the other hand, the SiP consists of not only multiple dies, but also passive components and discrete devices designed and assembled into a standard or custom package to achieve a modular function previously only accomplished by using several separated single chip packages [46]. The SiP forms a functional block, or module, that can be used for board-level manufacturing.

The traditional two-dimensional (2D) MCMs are divided broadly into three types depending upon the substrate dielectric construction, as shown in Figure 2.16 [10]. MCM-L technology uses laminated dielectric-like PWB substrates. MCM-C technology uses ceramic dielectric material and MCM-D technology uses deposited dielectric material.

MCM-L packaging is constructed with PCB laminates and the interconnections on the laminate are almost always copper, and created by photo imaging [10]. Vias are created through drilling and electroless plating followed by electroplating. On the other hand, MCM-Cs are constructed on cofired ceramic or glass-ceramic substrates using thick film technologies to form the conductor patterns with fireable metals [10]. Vias are formed during the conductor screen printing operation and are of the same material as the conductors. MCM-D packaging is based on depositing a dielectric over a substrate, which is usually ceramic, silicon, copper, other metals, or metal composites [10]. The dielectrics used can be broadly classified into two categories: polymeric

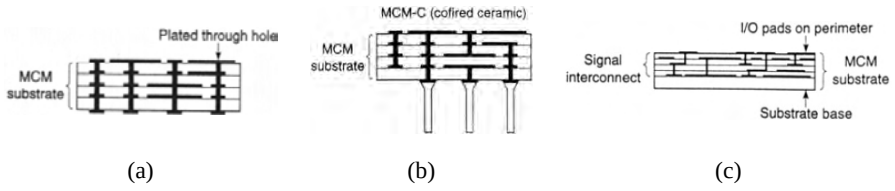


Fig. 2.16 (a) MCM-L, (b) MCM-C, and (c) MCM-D [10].

(a) (b)

Fig. 2.17 (a) System in package example and (b) 3D stacking [5].

and inorganic. Aluminum, copper, and gold are used for conductor metallization. Copper and nickel are the most common via fill materials.

The SiP performs all or most of the functions of an electronic system, and is typically used inside a mobile phone, digital music player, etc. [10]. Dies containing ICs may be stacked vertically on a substrate, as shown in Figure 2.17. This kind of 3D package or Chip Stack MCM contains two or more chips stacked vertically and takes IC industry a step forward in the Moore curve. Chips are internally connected by fine bonding wires or solder bumps and usually require an extra "interposer" layer between chips. As shown in Figure 2.17(b), the 3D stacking technique has been rapidly adopted because 3D stacking integration provides more functionality in less space and it is easier to integrate die from multiple suppliers through 3D stacking [10].

TSVs can be used to replace edge wiring by creating vertical connections through the body of silicon chip in some new 3D packages. With the introduction of TSV, a 3D package is slowly becoming a 3D IC technology in an attempt to increase the number of feasible interconnects and enhance electrical performance.

SiP or 3D IC simplifies product system board design and assembly and provides increased functionality per unit area/volume [10] by heterogeneous integration of materials, devices, and signals. Owing to reduced interchip distance, system electrical performance improves while manufacturing costs decrease and at the same time have better reliability. However, there are a few challenges restraining the industry from large volume of commercializing 3D ICs. For example, when two dies are 3D stacked for lower RC delay, these two high power

density regions reside on top of each other dramatically increasing power density of the entire stack. And some 3D IC technologies, such as TSV, are not evolved to be cost-efficient and bring up numerous manufacturing yield challenges, such as chip alignment, accurate-controlled bonding, and mechanical stress issues.

3

System-in-Package Design Exploration

3.1 Introduction

For high-speed nanometer designs, it is imperative to design the chip and package in a concurrent fashion to successfully meet the design and market constraints. The nature and speed of these designs make the traditional separated package-chip design a non-feasible design paradigm. For these designs, the package plays a critical role in meeting the tight performance and reliability constraints as well as the cost and time-to-market constraints. The tight performance constraints coupled with the tight time-to-market constraints necessitate a holistic approach to the design of the components that make up the SiP. For successful design and implementation of the system, characterization and optimization of the entire system are needed in order to understand its electrical, thermal, and mechanical behavior. When the number of I/Os in designs was relatively small (few hundreds) and switching at low to medium frequencies along with core devices switching at a modest speed (hundreds of MHZ), package design was a task left till the end of the design cycle when the die is completely designed and taped out. In that paradigm, the design of the die dominated the design cycle

and the on-chip constraints in terms of timing, power, and reliability were given the highest priority without regards to the performance of the package nor its impact on the performance of the die. Standard packages in terms of the number of layers and number of pins are commonly used for designs that do not push the envelop in terms of performance constraints. However, as eluded to earlier, these standard packages would not fit the designs that have a large number of I/Os or that push the envelop in terms of performance and reliabiliy. For such designs, custom packages are designed and optimized as an integral part of the overall design cycle of the system in order to achieve a first-pass success. These are the applications that require SiPs and require a change in the design paradigm.

Furthermore, a great interest in SiPs stems from the need to extend Moore's law and continue the aggressive design integration trend. In addition to satisfying the performance and cost constraints of high-end designs, SiPs also provide a medium to integrate multiple technologies and different process nodes in order to realize an efficient, cost-effective, and reliable hybrid system. The integrated components could be a combination of digital, RF, MEMS, passive elements, and optical modules. To be able to analyze the different tiers (layers) of this hybrid system, electrical, mechanical, and thermal modelings of the various components are necessary.

Design exploration for SiPs tries to balance the needs and constraints of the various components that make up the SiP as well as the packaging technology that supports the various components. In addition to meeting the performance, power, and reliability constraints for all the components that make up the system, it is equivalentaly important to choose a packaging substrate that satisfies the design constraints while still minimizing the overall cost and complexity of the design cycle.

There are a number of design parameters that need to be explored and optimized during the package exploration phase. The system design specifies budgets for performance, power, and reliability that the overall system must satisfy. Satisfying these system constraints in turn produces a set of constraints that need to be met if the system goals and constraints are to be met.

For high-end applications, the cost of the package exceeds the manufacturing cost. Thus, during package exploration in general, and for high-volume designs in particular, it is very important to choose a package that supports the design constraints while minimizing the total cost. This can be attained by optimizing the number of layers that a package has as well as optimizing the design rules and the materials used to meet the design targets.

3.2 Overview

SiPs serve as a design solution that bridges the continuously widening gap between the I/O needs of the silicon and support provided by the packaging technology. There are many design metrics and constraints that dictate the selection of a packaging technology. For high-end applications, the package plays a major role in attaining the performance and reliability goals of the design. These goals constrain the selection of the packaging technology in terms of the number, material, thickness of the routing and the plane layers in the package. In addition, the size of the package is constrained by the size of the design components and the number of I/O pins that need to be fanned-out and shielded to attain high performance and reduce switching noise. In addition, the heat flux constraints as well as the performance and reliability constraints drive the push for new technologies and higher level of integration such as those provided by 3D ICs, TSV, and nanotechnologies to circumvent the deficiencies in the cooling technologies and the density gap between silicon and packaging. Today, the technology allows us to integrate more transistors on a single die than we can use and power on. This is due to the thermal constraints associated with powering such a large number of transistors (10^9 s). SiPs alleviate some of these issues by distributing the components on different dies and allow for a more efficient heat flow. SiPs provide many benefits including:

- Higher level of integration by placing a number of chips along with discrete components on a single packaging substrate.
- Higher performance, better reliability, and reduction in time-to-market as compared with the on-chip silicon integration.

- Integration of different interconnection styles such as flip chip, SMT, and other technologies on one substrate.
- Integration of different process technologies and the ability to migrate to new process technologies of selected components without redesigning the entire SiP.

In this section, we will highlight the design process as well as the design factors associated with SiPs. These design goals and constraints need to be addressed in order to realize a cost-effective system, which meets the performance and reliability constraints. We start by highlighting the design process and decisions that come with it, and then we discuss how to tackle the problems associated with the chip-package co-design early in the SiP design cycle. In particular, we will focus the discussion on the two most important design problems that need to be addressed: the design of the channel (signal interfaces) and the design of the power delivery network. In this section, the emphasis is on the design exploration of SiPs using efficient and reasonably accurate models. The objective is to concurrently design the chip(s) and package to meet the design constraints in a fast time-to-market fashion. The emphasis here is on reasonably-accurate models to assess whether the technology and design choices that have been made in the design exploration stage are sufficient. At this stage in the design exploration, neither the package has been implemented nor the design or floorplanning of the associated silicon components is necessarily done. Thus, it is wise neither from a performance point of view nor from a design-effort point of view to use detailed analysis models and incur costly design efforts for a virtual system that is under design. In a later section, detailed analysis and simulation methods and algorithms are described that are needed to analyze the implementation of the chip-package system.

Also, we will address the constraints associated with the physical design and implementation of SiPs and their impact on technology selection and exploration. In particular, we will discuss the issues related to timing, noise, cooling that are associated with the floorplanning, placement, and routing of the dies and the I/Os. All these constraints impact the package technology selection and the design exploration of the overall system. It is necessary to employ a design

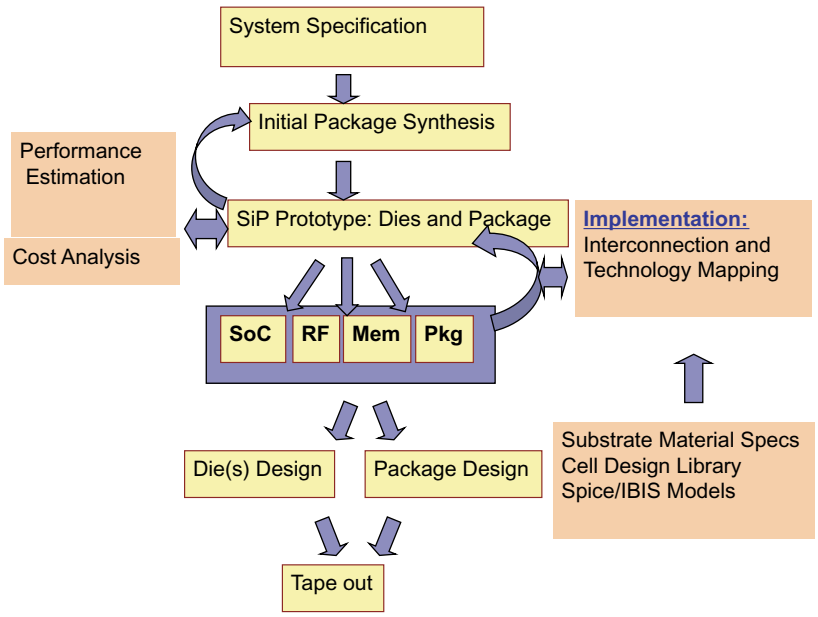


Fig. 3.1 A coherent SiP design methodology.

methodology that addresses all these constraints simultaneously in order to successfully implement the design and meet its constraints.

Figure 3.1 shows the system design flow of SiPs. The design process starts with a system specification that outlines the design objectives of power, performance, noise tolerance amongst others. In addition, the system design provides specifications of the design interfaces needed for the SiP to communicate with the other modules in the system. Armed with the system specification, the chip and package designers work in tandem to meet the design objectives in a cost-efficient manner. The chip-package designers build and employ efficient models to study the characteristics of the package as given by the substrate material specifications and the given I/O buffers library and models (IBIS/Spice). If the models are in the ballpark of the design objectives, a prototype of the chip(s) and package is built. Further refinement and optimization to the chip design and implementation as well as the package design proceed. A detailed look at the extraction and simulation steps follows to validate the entire system. Although the chip and package design

steps are coupled and some take place concurrently, we expound them here separately to make the design flow clearer.

From the point of view of the chip, the designer requires the entire system and the package in particular, to provide a stable and reliable power supply, a noise-free (if possible) communication medium, all in a cost-efficient and productive manner.

From the point of view of the package, the designer requires the chips' I/Os to be planned and routed in a fashion that is amenable to efficient implementation of the package with regards to the number of package layers needed and the design of the power delivery system (PDS). The management of the I/O noise in the system depends greatly on the type, size, and placement of the I/O buffers. The package designer also desires the package technology to provide a bump pitch that results in adequate power sources as needed by the SiP without degrading the routability of the package.

In the design exploration stage, decisions on the type of I/O interfaces as well as the bandwidth needed are driven by the receiving components and the performance targets for the system as a whole. The models needed to concurrently carry out the design and planning of the core(s), I/Os, and package should be efficient yet accurate to provide reasonable metrics in terms of noise, and timing to be able to estimate the operating frequency range of the interconnect.

3.3 On-chip Design Decisions

During the early stages of SiP planning, the goal of the chip designer is to meet a certain channel bandwidth that is set by the system design specification. To meet such a bandwidth, the system design specifies the channel characteristics in terms of characteristic impedance, bit error rate (BER), timing, timing skew, crosstalk budget among others. The goal of the chip I/O designer is to optimize the I/O circuitry and place and route the I/Os on the chip to the package bumps (redistribution layer (RDL) routing). To be able to design the channel and meet the system-level specifications, the chip designer has to work closely with the package designer to make sure that any decision taken accounts for the substrate characteristics.

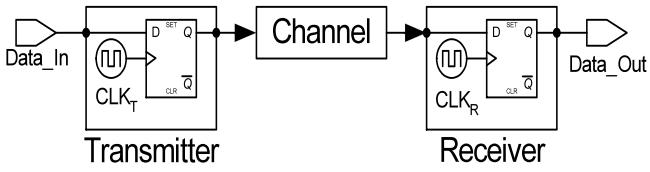


Fig. 3.2 Block diagram of a chip-to-chip communication channel. The main components are the transmitter, the receiver, and the lossy channel that is composed of the RDL routes, the package traces, and the PCB traces.

3.3.1 I/O Circuit Design

The proper design of the interface channel (Figure 3.2) must adhere to a complex set of specifications and constraints. These include voltage levels and noise, BER, signal jitter, and slew rate. Proper design and optimization of these circuits are needed to meet the design specifications at an acceptable cost in terms of area and power [6, 21, 34, 53, 66].

To meet the performance and signal integrity constraints for the I/O drivers, good models of the package and board are needed to account for the capacitive loading and coupling as well as the inductive coupling in the system. Early in the design cycle where the package is not routed yet or where no board models exist, good estimates are needed. In high-end designs, it is no longer sufficient to use *ad hoc* metrics such as rules of thumb to make decisions on the type and size of the I/Os as well as on the capacitive and inductive values of the load that the driver sees. Good virtual models that are able to capture the package and board effects are needed.

Figure 3.3 shows the modeling components needed to estimate the performance and noise metrics of the signal interface in SiPs. Guided by the channel interface spec, the I/O circuits are partitioned into power domains, each with a given voltage level. Each of the power domains needs to be characterized to study the impact of the package on the channel bandwidth. Also, the model studies the impact of the noise injected by the I/O drivers on the power delivery network and the other drivers on the chip.

To control the power consumption of the I/O buffers, the circuit impedance should be reduced. However, good design of the circuit should take into account the impedance of the lossy transmission

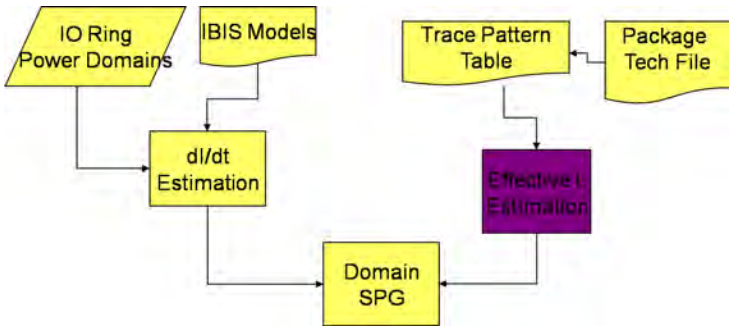


Fig. 3.3 Quick simultaneous switching noise estimation model that uses the IBIS models of the drivers, the package layer parameters, and the trace model of the signal routes from the driver to the package pins.

medium in the package and PCB. The impedance should be as close as possible to that of the channel or proper termination is needed to reduce any reflections that result from the impedance mismatch. For high-bandwidth SOCs, there is a number of high-speed I/O circuits in use today that employ different signaling techniques such as LVDS to optimize the delay, power, and noise [40]. There is a set of standard I/O buffer circuits that are commonly used such as high-speed transfer logic (HSTL), series-stub transfer logic (SSTL), and PECL. Differential signaling is also very common to reduce the noise associated with the high-switching rate and improve the resultant bandwidth. HSTL and SSTL are two very popular design choices for high-bandwidth applications. For proper functionality of these circuits, tight control on the driver and receiver's impedance is needed. For example, SSTL requires its output impedance to match that of the transmission line impedance seen in the package along with maintaining low capacitive loading at the receiver's end. This results in a critically damped waveform that aids in reducing noise and crosstalk as well as keeping the clock and data jitter under control [19].

3.3.2 I/O Buffer Physical Planning

There are physical and electrical constraints that need to be met when planning and placing the I/O circuits on the die [12, 41, 44, 60, 91]. The floorplan of the die imposes physical and timing constraints on the

I/O circuits. There are timing constraints that need to be honored on the path from the I/O circuits to the latches connected to them. There are timing skew constraints associated with the placement and routing of differential pairs [52, 95]. There are power and signal integrity constraints that need to be honored, chief amongst those is the simultaneous switching noise (SSN) constraint [18]. For correct modeling of the SSN effect, the entire circuit loop from the I/O output pin to the respective power pin needs to be modeled. This necessitates modeling the virtual trace routes in the package. It is not acceptable anymore to wait until after routing the package to do this analysis and verify that the signal and power integrity constraints are met [11, 47].

3.3.3 On-chip Power Design and Planning

One of the primary objectives in the design of a power network is the delivery of a low-impedance stable power supply level from DC to the transition frequency of the signals. The success of this mission requires a close interaction between the chip and package. On the chip, a reliable power grid with an acceptable dc voltage drop is of paramount interest. If the on-chip *IR drop* is not controlled, the performance of the switching devices is directly impacted. The on-chip power network is affected by on-chip noise sources such as the simultaneous switching of the on-chip devices as well as perturbations due to capacitive and inductive coupling in the package. To achieve a low-impedance stable PDS for the whole frequency spectrum of interest, the entire power delivery network needs to be taken into account. The design should account for low-, medium-, and high-frequency sources of noise. The PCB planes and traces are the source of low-frequency noise while the package with its planes, traces, and vias play the major part in the mid-frequency noise injected into the power delivery network. The high-frequency noise is due to switching rates in the chip near to the circuit.

3.4 Package Design and Exploration

Meeting the performance target of the system at an acceptable price is the most important criterion for a successful design. If cost is not

accounted for, an overdesigned system in terms of the number of layers in the package and the spacing on each layer will meet the performance target. Such a system is not a viable solution due to the exorbitant cost. Owing to the difficulty of taping out high-speed package designs, it is likely that a package solution can be more expensive than the chip. For high-volume designs, such a solution is not acceptable. Early in the chip-package design stage, not much is known about the switching profile of the circuits as well as the size and placement of the I/O buffers. Therefore, constraints must be defined for co-design. This is the primary reason making package design difficult.

3.4.1 Package Stack-up Order

Once a package substrate is chosen, one of the most important decisions in deciding on a package design is the number of layers (substrate build-up) and the determination of optimal design rules. Often, when a package has more than two layers, a core layer (very thick) is designed and a number of layer pairs are added by mirroring the layers around the core layer. For example, for a package that has three layers, the substrate build-up would be 111. For a package that has five layers, the substrate build-up would have 212, which means a core layer, and two layers above it and two layers below it. Each additional layer pair adds to the cost of the package. For an efficient package design, the number of layers should be minimized while meeting the design constraints and performance targets [22, 61].

3.4.2 Substrate Layer Assignment

The decision of layer assignment is a very crucial one with respect to performance, reliability, and cost of a package. For high-speed designs, more layers are assigned to power and ground in order to ensure a steady low-impedance power supply as well as reduce the inductive and capacitive coupling in the package to shield the signal traces and improve the performance. The need for more power planes increases the cost of the package and hence the overall system. On the other hand, lower speed designs require fewer layers and thus have a lower cost [3, 25, 28, 57, 58].

When characterizing the package either through measurements or modeling, the computed parameters such as SSN or package resonance impact the decision of how many package layers to be dedicated for power and ground.

To study and optimize the variance in high-speed signaling characteristics such as Z0 (characteristic impedance), insertion loss, and crosstalk as functions of key substrate design dimensions, design of experiment (DOE) methods are employed by high-end design houses. As for Z0, target center value and tolerance are given by chip and system designers. The design rules have to be determined to meet these requirements. In addition, they have to be optimized for low-cost and high-volume manufacturing. The primary constraint is trace width. A narrower trace results in larger variance in characteristic impednace and a wider trace inordinately constraints wiring density [26]. Most often, the design rules are overdesigned and are provided by the substrate manufacturers as templates to be followed by the package design teams.

3.5 Voltage Domain Planning

The different technologies in SiPs are referenced to a number of voltage domains, each with a voltage level decided by the circuitry in the module, the process node, or a performance constraint for that tier or component. In addition, each voltage domain is composed of one or more power domains. These power domains constitute logical partitioning of the different signals and buses in the system (Figure 3.4).

3.6 Modeling and Analysis Decisions

For pre-layout characterization of the package, lumped models or transmission line models are utilized. Package designers are often familiar with transmission line models, and there are mature tools that can be employed to calculate S-parameters to ascertain the package behavior and the signal characteristics.

Package substrates should be designed to meet the time-domain specifications as well as frequency-domain specifications such as

insertion loss or crosstalk. Eye opening and jitter define time-domain constraints that the system must meet. These specifications themselves cannot be guaranteed by characterizing only the package substrate because the eye diagram is determined by the characteristics of the total channel, including PCB and the package, on the other end (Figure 3.5).

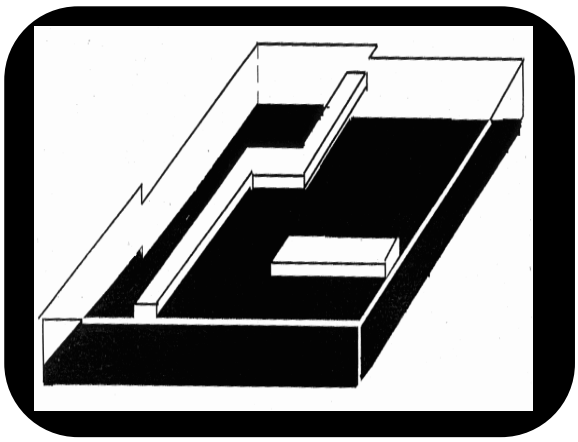


Fig. 3.4 This picture shows a solid plane that provides a stable reference for the traces above and below it. This design provides a tighter return path loop, which renders the effective loop inductance smaller.

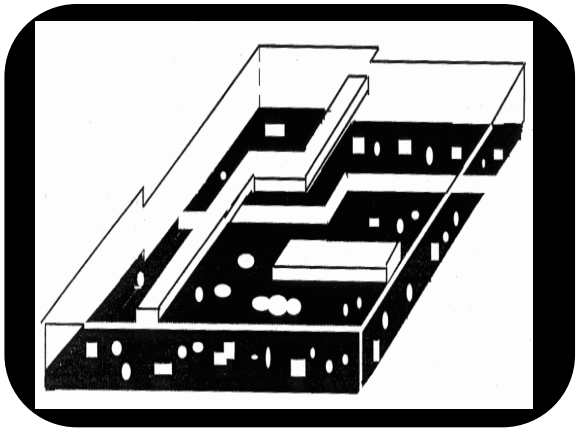


Fig. 3.5 The picture on the right shows a poor voltage domain partitioning in the package. The perforated plane does not provide a stable reference and causes a large inductance loop and noise problems in the package.

To validate the channel, worst-case time-domain simulations are usually carried out [26, 27].

For systems where building a detailed model of the chip-package at very high frequencies (millimeter-wave designs), measurements are utilized to design and verify systems. Also, for measurement-based modeling of the system, extracted models can be used to design and optimize systems. This method is feasible for small systems with a small number of I/Os. However, for systems with high number of I/Os, measurement-based models will be prohibitively expensive, and more emphasis is placed on modeling and simulation for the design and optimization of chip-package systems.

For simulation methodologies, the type of models used becomes a critical one. In general, the more detailed the model, the more it can capture the intricacies of the system under design. However, such detailed models if exist come at a cost in terms of simulation and design. In the early phases of the design where neither the I/Os are placed nor the package is routed, detailed models are not warranted.

In order to specify the major factors that affect the critical parameters of the package design, DOE technique and statistical analysis are used. Both 2D and 3D electromagnetic analyses and circuit model simulations are carried out in design optimization process. Passive characterizations are also carried out to get the basic parameters used in the package design.

3.7 SiP Design Problems

In this section, we will expound the following two most important design problems that need to be addressed early in the design cycle:

(1) power integrity modeling,
(2) timing and signal integrity modeling.

These problems dictate and define the type of I/O circuits needed as well as how they are placed and routed. Also, these decisions define the package characteristics in terms of the number of layers, the design rules followed in the package implementation, and the layer assignment in the package.

3.7.1 Design and Planning of PDS

The design of the PDS requires time- and frequency-domain modeling and analyses of the chip and package. To control the size and complexity of the RLC model of the power network, the modeling and validation of the chip and package power networks are done in isolation. During the design planning stage of the chip and package, RLC models of the various I/O power domains along with the package voltage domains can be built and analyzed. As for the chip power network, the package power network plays more of a prominent role in wire bonding designs as opposed to flip chip ones. For flip chip packaging, there is abundance of C4 power bumps feeding the core logic as opposed to the wire-bonded designs where the power is constrained to be delivered by the power pads on the periphery of the chip. Insertion of decoupling capacitors (decaps) is an integral part of the power design of SiPs. A good understanding of the behavioral model of the capacitors and the dynamic power fluctuation is necessary in order to accurately predict the noise on the voltage rails [83].

3.7.2 Design and Planning of Signal Interface

It is imperative to model the entire channel that signals go through as they make their way from the transmitting chip side to the receiving side. This channel consists of the on-chip transmitters, the different substrate layers in the SiP including the package layers, and onto the receiving end. If the transmitter and the receiver are not in the same SiP, then the PCB traces should be modeled as well as the receiving SiP. These stages could be a number of PCBs and intermediaries whose electrical characteristics affect the performance and reliability of the signal. Modeling the signal transmission into and out of the chip requires understanding the behavior of the signal transmission in the package. Also, in order to study the power needs of the chip, we need to understand the power network in the package, which carries power from the voltage regulator module (VRM) down through PCB and up to the chip. At the early stages of chip planning, many of the important components of this signal transmission system are not defined.

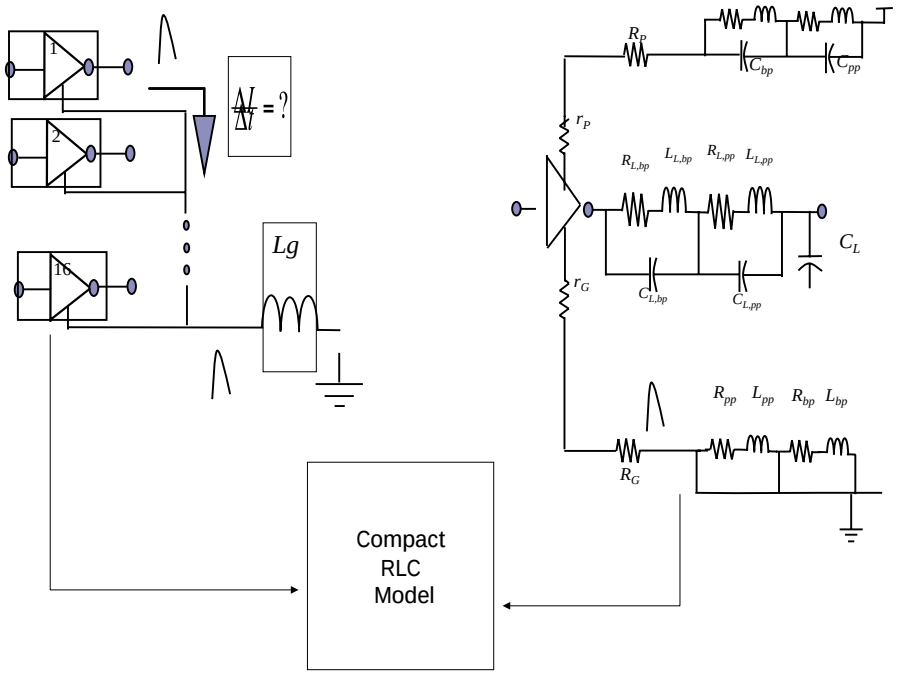


Fig. 3.6 RLC model for estimating the noise and timing of the signal interface early in the SiP design stage.

3.7.2.1 Signal Interface Characterization

Signal interface characterization in terms of bandwidth, jitter, and skew has to be carried out before an SiP is taped out. For design and modeling of the channel, Figure 3.6 shows a basic model that can be constructed as soon as the chip and package decisions outlined above have been implemented.

Figure 3.6 shows a detailed model for estimating the channel noise as well as the signal timing, skew, and jitter. The model's accuracy is reasonable in the design planning stage before the chip is fully implemented and the package is routed. The parasitics of the RDL routes and those in the package reflect virtual routes of the signals based on the power and voltage domain partitioning on the chip and in the package. The RLC model of the plane is shown in reduced form, and it does not have to account for any vias at this stage in the design.

The subject of channel design and characterization is treated in more depth in [6, 21, 34, 53, 66].

It is often the case that both the chip and package share some layers or voltage domains. In such cases, some of the package integrity problems could creep into the chip and *vice versa*. A simple example of such a scenario is when both the I/O signals and the on-chip signals share the same ground plane. Any ground bounce phenomenon or a general power integrity problem could travel in the ground plane and reach the on-chip devices. To remedy such problems, careful analysis and characterization of the I/O signals and the power planes in the package must be done. In addition, for enhanced reliability of the system, the chip and package should be isolated from each other as much as possible for power noise.

The following section expands on detailed modeling and analysis techniques for sign-off and validation of the SiP design based on the decisions outlined above.

3.8 Parasitic Modeling for Design

As technology scales down towards nano-regime and the trend of SiP continues, the industry is moving toward higher pin-count, greater complexity, higher-density packaging with increased functionality, higher operating frequency, lower operating voltages, and reduced package size. This trend has also brought about the increased less ideal wires and traces, as shown in Figure 3.7. The non-ideal effects can generally be categorized into resistive, capacitive, or inductive parasitics, impacting the power and signal integrity of die. To capture those non-ideal effects and to characterize the electrical performance of the package, in this section we will discuss how to extract the full parasitics of a package and form circuit models to facilitate the design for power and signal integrity.

3.8.1 Definition and Effects of Inductance

Among all the parasitics, it is the inductive effect that makes the extraction and modeling of the package differ from that of the chip. In this section, we will introduce the concept of inductance.

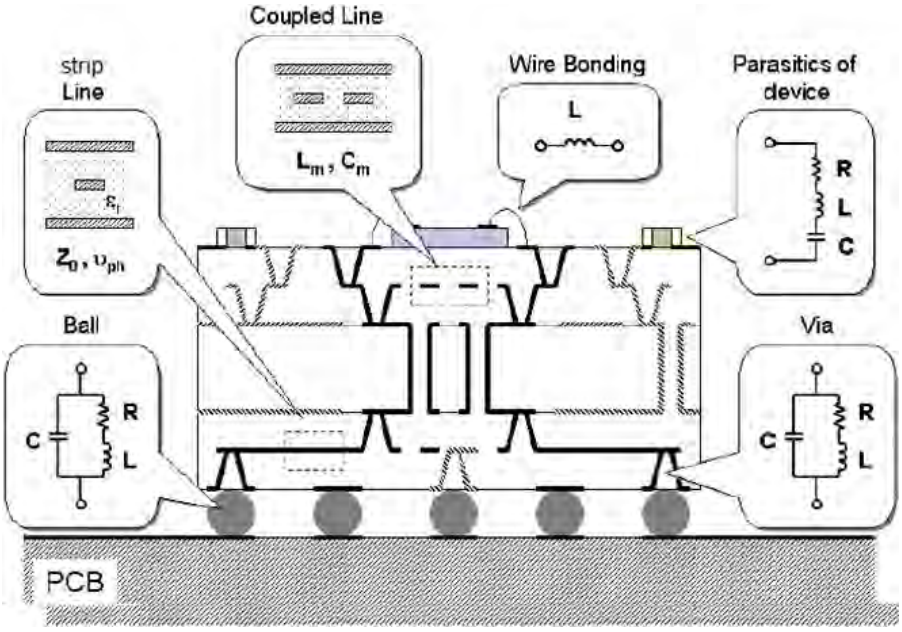


Fig. 3.7 Resistive, capacitive, and inductive parasitics of a package [87].

Inductance is the property within an electrical circuit where a change in the current flowing through that circuit induces an electromotive force that opposes the current change. All inductors have inductance associated with them. The inductance of a loop increases as the area of a loop is increased. Such a loop can be formed starting from the driver, through power and ground lines (both are AC ground), substrate or other signal lines, then returning to the driver. As the return current always chooses the paths with the smallest impedance $Z = R + j\omega L$ that varies with frequency ω, the inductance is also frequency-dependent.

At low frequencies, the resistive part R dominates and accordingly, the return current spreads with a relatively large inductance. With the increase of frequency, $j\omega L$ becomes dominant and the return current concentrates closer to the conductors. Hence, the inductance decreases and the resistance increases somewhat with the increase of frequency. At very high frequencies, the skin effect becomes important and the current concentrates to the surface and the resistance increases

proportional to the square-root of the frequency. As an example, the frequency dependence of both the resistance and the inductance of an interconnect is shown in Figure 3.8.

It is clear from this that inductance can have significant impact on the power and signal integrity, causing overshoot, undershoot and increased propagation delay as shown in Figure 3.9. The voltage overshoot or undershoot can exceed VDD or GND and may threaten the

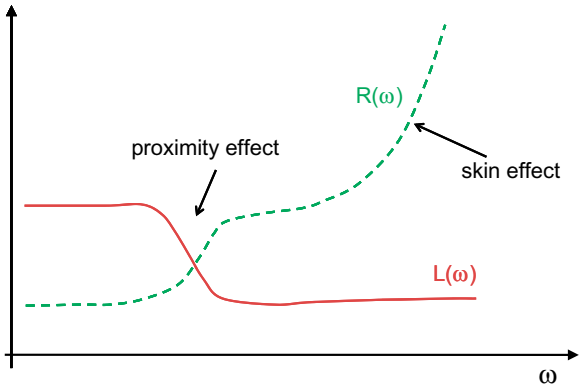


Fig. 3.8 Frequency dependence of package resistance and inductance.

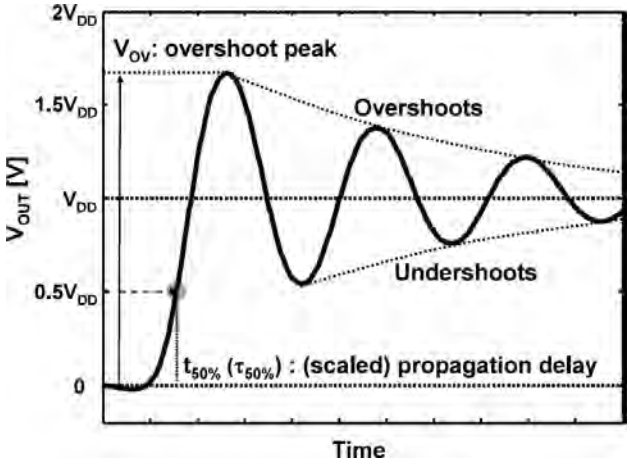


Fig. 3.9 Overshoot, ring back, and propagation delay caused by inductive effect [2].

reliability of the circuits, while the increased propagation delay may cause timing violations.

Numerous techniques have been proposed to effectively extract the parasitic elements of the package, especially the inductive parasitics. Below we will briefly review three types of methods: the partial element equivalent circuit (PEEC) method that derives the model directly from the Maxwell equations, a loop inductance method that is known for its efficiency and can be used in design iterations, and the measurement/simulation-based S-parameter methods.

3.8.2 PEEC Models

As a brief review,[1] the full-wave PEEC model starts by cutting the package into cells, where for each cell the following equation is solved, which is the solution to Maxwell's equations:

$$\hat{n} \times \vec{E}_0(\vec{r},t) = \hat{n} \times \left(\frac{\vec{J}(\vec{r},t)}{\sigma} + \frac{\partial \vec{A}(\vec{r},t)}{\partial t} + \bigtriangledown \Phi(\vec{r},t) \right). \tag{3.1}$$

In the above equation, $\vec{E}_0$ is an external electric field, $\vec{J}$ is the current density in the conductor, $\vec{A}$ is the magnetic vector potential, and Φ is the electric scalar potential. The magnetic vector potential $\vec{A}$ and scalar potential Φ at $\vec{r} = (x,y,z)$ are

$$\vec{A}(\vec{r},t) = \mu \int_{V'} G(\vec{r},\vec{r'}) \vec{J}(\vec{r'},t_d) dV', \tag{3.2}$$

$$\Phi(\vec{r},t) = \frac{1}{\epsilon} \int_{S'} G(\vec{r},\vec{r'}) \rho(\vec{r'},t_d) dS', \tag{3.3}$$

with t_d being the retarded time

$$t_d = t - \frac{|\vec{r} - \vec{r'}|}{c}, \tag{3.4}$$

where c is the speed of light in the medium, ρ is the surface charge density, and G is the Green function for homogeneous media:

$$G(\vec{r},\vec{r'}) = \frac{1}{4\pi} \frac{1}{|\vec{r} - \vec{r'}|}. \tag{3.5}$$

[1] For a complete theoretical derivation of the PEEC models, interested readers are referred to [71].

Substituting the vector (3.2) and scalar potentials (3.3) into (3.1) yields the following integral equation:

$$\hat{n} \times \left(-\vec{E}_0(\vec{r},t) + \frac{\vec{J}(\vec{r},t)}{\sigma} + \mu \int_{V'} G(\vec{r},\vec{r'}) \frac{\partial \vec{J}(\vec{r'},t)}{\partial t} dV' \right.$$
$$\left. + \frac{\bigtriangledown}{\epsilon} \int_{S'} G(\vec{r},\vec{r'}) \rho(\vec{r'},t_d)\, dS' \right) = 0 \tag{3.6}$$

To solve (3.6), current and charge densities are discretized into volume and surface cells, respectively. Applying the KVL and the Galerkin method [29] leads to the full-wave PEEC models and circuit matrices can be obtained accordingly.

The PEEC model has several advantages: first, without prior knowledge of return paths, current loops can be reconstructed by assuming that current returns at the infinity. Second, it can exactly evaluate the skin-depth effects by discretizing the conductors as a combination of many thin-wires (filaments). This model is called the volume filament method.

Recently, a library-based PEEC extraction was presented in [37] and can be used inside iterative design and verification. However, as the PEEC model includes mutual inductances between every pair of conductors, the resulting circuit matrix is dense [32]. Accordingly, sparsification or acceleration methods are required for large models.

The most direct and perhaps the simplest way to sparsify the matrix is by truncation. The idea of truncation is to remove all the coupling terms that are below a user-specified threshold. However, a major problem with this approach is that the resulting matrix can become non-positive definite. As a result, the sparsified system is no longer passive. To address the passivity problem, block-diagonal sparsification has been proposed. It uses circuit-topology-based partition techniques, which can guarantee the passivity of the sparsified model.

For specific problems of long on-chip wires, alternative approaches have been proposed. As an alternative, one approach associates each segment with a distributed current return path out to a shell of certain radius [49]. The inductive coupling between segments with spacing greater than the specified radius is neglected. To compensate for the

truncation, the inductance values of the segments within the radius are shifted properly. Such a shift-truncate method can guarantee the generation of positive definite sparse approximations of the original matrix. However, it is not straightforward to determine the value of the shell radius. An extension of this work [38] uses a moment-based algorithm to compute the shell radius. Another approach for limiting the inductive interaction is proposed in [72], which introduces return-limited inductances for sparsification and the use of "Halos" to limit the number of mutual inductances.

Another alternative method is discovered from the fact that the inverse of the partial inductance matrix usually has a higher degree of locality and sparsity [24], which is very similar to the capacitance matrix. It is very easy to do the truncation and sparsification on the inversion of the partial inductance matrix. However, such a method requires inversion of the partial inductance matrix, which is very expensive. In addition, a special circuit simulator is required to handle the inversion of the inductance matrix.

Finally, model order reduction for the linear portion of the circuit can be combined with the gate models and simulated in SPICE [62]. The linear circuit can also be parameterized [76] or symbolized [78]. The reduced order models have a much reduced size and yet serve as good approximations of the input–output behaviors of the original circuits. In turn, they are very efficient in terms of simulation time. One of the problems with the model order reduction techniques is that they cannot handle circuits with a large number of ports well.

3.8.3 Loop Inductance Models

Though accurate, the high complexity of not reduced PEEC models prohibits them from being used in design iterations. As discussed previously, such complexity comes from the fact that for the simple problem with equal return conductors, the inductance is formed by loop currents, and partial mutual inductance exists between all pairs of conductors. Instead of using the partial inductance, a simplified approach using the loop inductance model is commonly used with less accuracy but more efficiency.

For 2D structures, the inductance model [86] is developed in a three-step process: to start with, individual conductor cross-sections are sub-divided into sections smaller than the skin depth at the maximum frequency of interest, as is done for the volume filament model in PEEC. With this discretization, uniform current density can be assumed in each section. Next, resistance and inductance values are computed for each conductor cross-section assuming uniform current density in each cross-section. This allows the well-developed inductance formulas to be employed for computing the L matrix. Finally, the coupled RL skin-effect macromodel depicted in Figure 3.10(a) is analyzed at several points in the frequency domain. Such a method still suffers from the complexity problem as the obtained linear equation needs to be solved at multiple frequency points, and at each point a matrix factorization or inversion is required. This method is the foundation of frequency-dependent extraction tools such as FastHenry [43].

To reduce the number of solves required and to reduce the size of the circuit synthesized, Kim et al. [45] propose to use smaller numbers of the π-segments to model the frequency-dependent behavior. Moreover, only the frequency characteristics at maximum frequency and dc $(R_{\max}, L_{\max}, R_{\mathrm{dc}}, L_{\mathrm{dc}})$ are required to synthesize such a circuit. Accordingly, expensive methods such as [86] only need to be applied at those two frequency points, and the characteristics at any frequency points between can be efficiently obtained by solving the compact model instead. However, the efficiency of those two methods are still limited

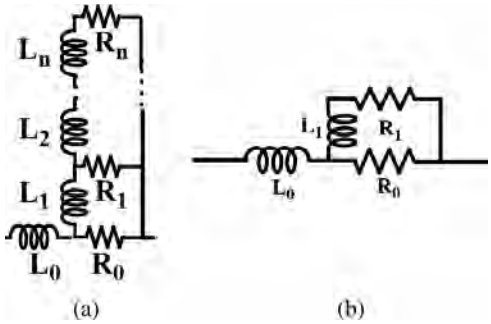


Fig. 3.10 RL ladder circuit to model frequency dependence.

because finding $R_{\max}$ is costly and the synthesis procedure is iterative. Most recently, Krauter and Mehrotra [48] propose the construction of a compact ladder circuit as shown in Figure 3.10(b) without the necessity of calculating $R_{\max}$ or employing iterative synthesis.

3.8.4 S-parameter Methods

Scattering matrices or S-parameters can also be used to quantify how energy propagates through the package: it describes the electrical properties of the package by counting all transmitted and reflected power from a given incident signal. For the simplicity of presentation, we will limit our discussion to the 2-port case. Multi-port S-parameters can be established in a similar manner. Figure 3.11 shows the schematic representation of a two-port network. The port voltages and currents are shown along with the "power waves" for the respective ports. The power waves are defined so that $|a|^2$ is equal to the power incident on the port, whereas $|b|^2$ is equal to the reflected power.

The S-parameters quantify the relationship between the incident and reflected power waves as follows:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \tag{3.7}$$

In practice, S-parameters can be obtained either through measurement directly, or through electromagnetic simulation such as Ansofts HFSS or Agilents Momentum, without the requirement of extensive computation as in the PEEC-based methods.

The S-parameters are frequency-dependent and consequently provide a broadband description of the package. The power waves a_n and b_n depend on the voltage difference across the port terminals via

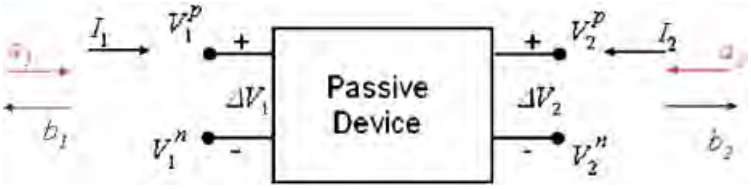


Fig. 3.11 Illustration of voltages and power waves for a two-port system.

the relationship

$$\Delta V_n = (a_n + b_n)\sqrt{Z_{on}} \quad , \tag{3.8}$$

$$I_n = \frac{a_n - b_n}{\sqrt{Z_{on}}}, \quad n = 1, 2, \tag{3.9}$$

where n is the port number and Z_{on} is the characteristic impedance of the port.

In order to use the S-parameters for time-domain simulation, we need to extract a circuit model from the measured S-parameters. This is often a tedious task and it even requires a heuristic guess to construct the topology of equivalent circuit, as different topology would give different values of the element for the same S-parameters. The problem is even more complicated when we also need to make sure that the extracted circuit is passive and causal. Recently, Lee et al. [54] proposed an efficient way by first converting the S-parameter to the T-parameter (transmission matrix), which can be further represented by a cascaded network of series RL and shunt GC ladder circuit.

3.9 In-package Power Integrity

With the technology scaling down to 45 nm and beyond, power integrity has become the major bottleneck for the reliability of high-performance system in SiP integration. The reduced supply voltage and increased clock frequency and chip density have made the circuits more vulnerable to power supply noise than ever before.

SSN, also referred to in the literature as ΔI noise, is considered to be the major threat to the power integrity. Accordingly, we will focus on SSN in this paper. It primarily occurs due to a very large amount of instantaneous P/G current from the simultaneous switching gates, which is quite common in clock synchronized circuits. SSN is mainly an inductive noise and can be generally characterized by the equation $V_n = L(dI/dt)$, where V_n is the magnitude of SSN. L is the parasitic inductance of the chip and the package, and I is the total switching current. In other words, the magnitude of SSN is proportional to the total parasitic inductance and to the rate of change of the switching current. SSN noise causes supply voltage fluctuation; it reduces noise

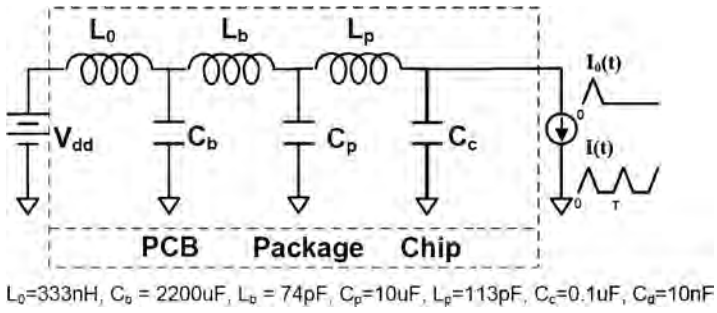


Fig. 3.12 Illustration of power delivery system.

margin of digital circuits; it shifts the operating point of analog circuits, it decreases the effective driving strength of the gates; and it causes output signal distortion (e.g., jitters) impairing signal integrity.

Figure 3.12 illustrates a lumped model for PDS for board, package, and chip, where L_0, L_b, and L_p are the inductance of power regulator, board, and package, respectively. C_b, C_p, and C_c are the capacitance of the board, package, and chip, respectively, V_{dd} is the power supply voltage, and $I(t)$ is the on-chip switching current. A set of their typical values is also provided [75].

Such a PDS structure leads to three distinct impedance peaks, if looking from the chip into the PDS, as shown in Figure 3.13 [7]. The first and smallest peak is in the kHz range, mainly caused by the coupling between the power regulator and the board. The second in the MHz range, mainly caused by the coupling between the package and the board. The third in the 100 MHz range, mainly caused by the coupling between the chip and the package. As discussed in the previous section, the SSN from I/O buffers is first injected into PDS via such coupling.

SSN is most significantly observed around the output pads of the chip. The main reasons are in three folds: First, in order to drive large off-chip loads, the I/O buffers are usually very large in size, drawing significant amount of instantaneous currents when they switch to change the output binary value, as shown in Figure 3.14. Second, in clock synchronized chips multiple I/O buffers tend to switch simultaneously to create a large surge current with a sharp slope. Third, the parasitic inductance of the power distribution network of the package, including

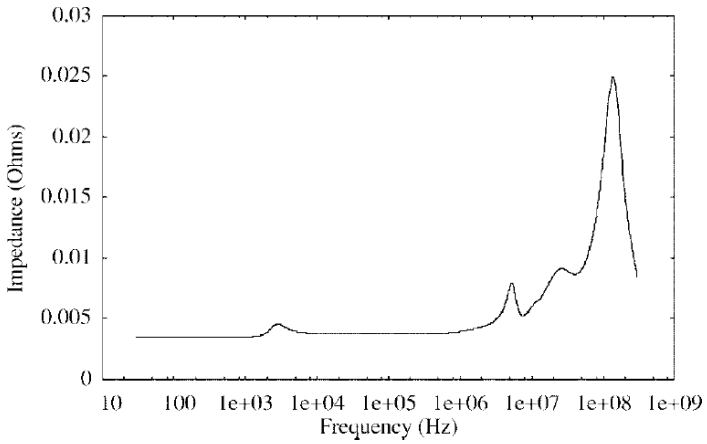


Fig. 3.13 PDS impedance seen from the chip [7].

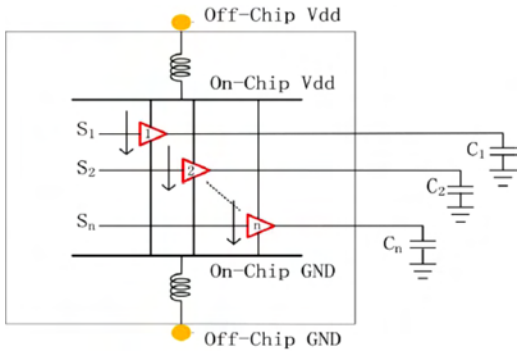


Fig. 3.14 SSN caused by the simultaneous switching of I/O buffers.

the interconnections to both the chip and the board, is usually in the range of a few hundred pico-Henries. Such large inductance has made package a major contributor to the SSN.

In order to reduce SSN, many design techniques have been proposed in literature. They target at various design freedoms in different design stages, trying to reduce the parasitic inductance, or trying to reduce the impedance between the power and ground plane. In this section, we will briefly discuss a few of these techniques and point out the critical issues in each of them. A more thorough review of these techniques can be found in [77].

3.9.1 I/O Planning and Placement

I/O placement plays a key role as the interface between chip and package designs in a co-design flow. Most of today's high performance ICs are designed with flip chip technology, which eliminates wires for chip-package bonding. The bonding is achieved through bumps via the SMT. As discussed previously, I/O cells are first connected to bumps on the die via RDL routing, then the die is "flipped" and mounted on the surface of the substrate, where bumps are connected to bump pads on the substrate. Finally, package trace routing is performed to furnish the connection between bump pads to package pins.

For flip chip designs, there are two types of I/O cell placement schemes: peripheral I/O and area I/O. Peripheral I/O scheme restricts the placement of I/O cells at the chip boundaries, and it is a cost-effective way to transform traditional wire bonding chip designs to flip chip designs. Area I/O scheme allows I/O cells to be placed anywhere within the die area and it is inherently suitable for flip chip packaging. In [23], the peripheral and area I/O schemes are also called extrinsic and intrinsic flip chip designs, respectively. Interested readers are referred to [90] for more details.

For I/O planning and placement, we need to assign the pins and pads to different signals and P/G supply. Different assignments can significantly impact the system performance, including the signal and power integrity.

While so many factors can affect the I/O placement, in this section we will focus on the issues that need to be considered for SSN suppression. There are mainly two criteria.

First, as illustrated in Figure 3.15, the power and ground pins and pads for analog and digital circuits should be separated whenever possible. As such, the switching noise generated by the digital portion of the system will not affect the operation of the analog portion of the systems, which is generally more sensitive to power supply noise.

Second, the pads and pins for power and ground should be made as many as possible. With the increased P/G pin/pad number, the inductance will become smaller (parallel connection) which leads to lower SSN. In addition, the slew of the relationship between SSN and

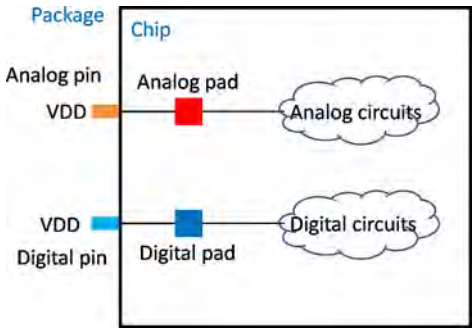


Fig. 3.15 Separation of digital and analog power/ground pads.

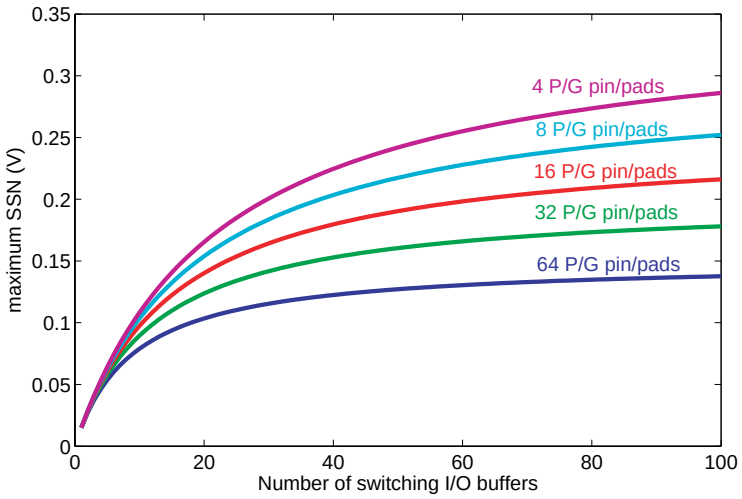


Fig. 3.16 Relationship between SSN and the number of switching I/O buffers for different numbers of power/ground pins/pads.

the number of switching I/O buffers becomes smaller with the increase in power and ground pin/pad number, as shown in Figure 3.16.

The robustness of the power delivery in package mainly depends on the resistance, inductance, and capacitance of the delivery path and the mutual inductance and coupling capacitance formed between the path and any other surrounding metals including traces and P/G planes. With the low inductance (pico-Henries) and the large capacitance (micro-Farads), metal planes provide much lower impedance for

return current paths than traces or point-to-point wires. Accordingly, the plane structures are much more superior in delivering power than other traces.

Accordingly, the layer stackup and assignment problem can be formulated to assign signal nets and P/G planes to different layers in the package so as to minimize the number of layers required while maintaining power integrity. At this point, we will look from the power integrity point of view, and during this process the following guidelines should be followed [4]:

(1) Solid reference planes must be parallel to the routed signals.
(2) Consider return current amplitude when determining power plane thickness.
(3) Pair up power and ground layers to provide additional "free" capacitance, reducing power supply noise.
(4) Use extra ground planes instead of power planes to isolate sets of routing layers.
(5) Use many vias to connect multiple ground layers.
(6) Minimize plane splits. Split planes create return path discontinuity and cause additional crosstalk problems. When signals travel over the split, this causes additional forward and reverse crosstalk. Vias can also create the same effect as a split plane because the reference plane may change.

3.9.2 Decoupling Capacitance Insertion

Decaps, which act as charge storage devices and low pass filters for ac signals, serve as another technique to enhance the power integrity by reducing voltage fluctuation in the PDS. An illustration of the decap insertion in a multi-layer package is shown in Figure 3.17.

For package decoupling purposes, discrete decaps ranging from a few pico-Farads to sub-micro-Farads are used. These decaps are not perfect and their frequency responses can be modeled with an equivalent series capacitance (ESC), an equivalent series inductance (ESL), and an equivalent series resistance (ESR), with different effective ranges. Generally, for the same amount of ESC, a decap with a lower ESL is more expensive [99]. The expensive decaps may not always be the

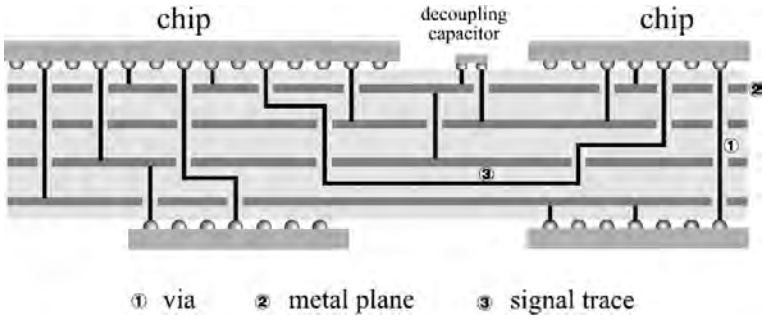


Fig. 3.17 Resistive, capacitive, and inductive parasitics of a package [17].

best choice for electrical performance. Moreover, the effectiveness of the decaps depends on its connected electrical environment and therefore varies with locations. Low-impedance decaps can be put at predefined locations in the package. Therefore, the sizes and locations of the decaps need to be optimized for the most effective design with the minimal cost.

In the literature, there exist two different types of metrics used to evaluate the SSN: the frequency domain impedance metric and the time-domain noise metric. As a frequency domain impedance metric, the impedance at given ports is required to be lower than a computed target impedance in the entire frequency bandwidth of interest. According to Smith et al. [80], the target impedance can be computed as follows:

$$Z_t = \frac{\delta V_{dd}}{I}, \tag{3.10}$$

where δV_{dd} is tolerable variation of V_{dd} and I is the switching current at the given ports. In [99], Zheng et al. also proposed a weighted combined impedance to consider the coupling between ports. Despite its ease of computation, the impedance metric may lead to significant overdesign as the impedance is a pessimistic estimation of the noise rather than a direct proportion to the noise [15]. Physically, the metric assumes that all the frequency components have the same impedance with the same phase and add up to the total noise. In fact, the current is not uniformly distributed in the entire frequency band, and impedance can be

different at different frequencies. Also, different frequency components have different amplitude and phase and may cancel each other. The impedance also varies with the frequency and does not need to be very small in the entire frequency band.

As an improvement over the frequency domain impedance metric, a time-domain noise metric can be used. Chen and He [15] propose to use the maximum voltage drop at all ports of interest as the metric to evaluate the SSN. Shi et al. [79], Yu et al. [96], and several others integrate the time-domain waveform over the time interval where the voltage drops below the tolerance bound, as shown in Figure 3.18, and take the sum over all the ports as a metric. Such a metric is more accurate in the sense that it also takes the time duration of the noise violation into consideration.

With the cost and noise computation methods as discussed above, the decap insertion problem can be formulated as to minimize the cost subject to a given noise constraint [15, 96, 99], which is a discrete and non-convex optimization problem.

Chen et al. [17] and Yang et al. [94] propose to use trial-and-error methods to manually optimize the problem. As an attempt to automatically optimize the decap insertion problem, Kamo et al. [42] propose to use the PEEC model and model order reduction techniques to compute the input impedance and then search for the optimal locations to minimize the impedance by gradient-based search. Hattori et al. [35] use FDTD and FFT to obtain the frequency-dependent Poynting vector

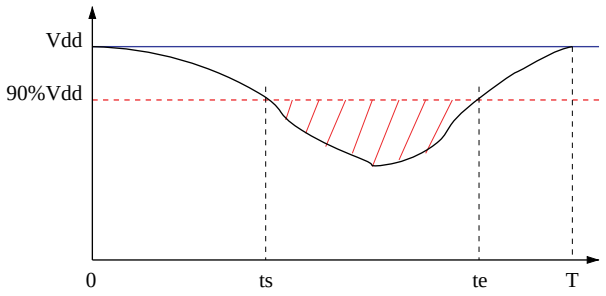


Fig. 3.18 Illustration of voltage drop with the shaded area showing the metric (tolerance bound = 90% V_{dd}).

and decaps are iteratively put at the port with maximum Poynting vector. However, in both papers the decap value is fixed and ESL or ESR is not considered, rendering the methods less practical.

Zheng et al. [99] model the inductive effect of packages with susceptance (the inverse of inductance) instead of inductance, and extract a resistance–capacitance–susceptance model of the package. Based on this model, a macromodel is built with a model order reduction technique. Then, based on the macromodel, a simulated annealing (SA) algorithm is developed to search for the optimal types of decaps at selected locations to minimize the cost under the constraint of a target impedance at chip I/O ports. Different types of decaps with different ESC, ESL, and ESR are considered. However, this method uses the impedance as the target objective function, and accordingly suffers from the overdesign problem as previously discussed. Chen and He [15] address the problem by using the time-domain noise metric in the constraint. SA is still used to find the optimal solution. However, both of the papers use the SA as an optimization approach, thus only capable of dealing with a pre-designed package with a limited legal positions for decap insertion. Therefore, they cannot accommodate the ever-increasing complexity of the package design.

Yu et al. [96] propose a fast decaps allocation method for a large number of legal positions. By applying a spectral clustering, a small amount of principal I/Os are found. Accordingly, the large power supply network is partitioned into several blocks with only one principal I/O in each. Then, a localized macromodeling for each block is performed by a triangular-structured reduction. In addition, to systemically consider a large legal position map in a manageable fashion, the map of legal positions is decomposed into multiple rings, which are further parameterized in each block. The decaps are then allocated according to the sensitivity obtained from the parameterized macromodel for each block. This method finishes large industrial designs with a quick runtime.

3.10 Signal Integrity for Off-chip Signaling

The electrical performance of the package substrate cannot be ignored in today's super-GHz design space. At these frequencies, the physical

dimensions in the package represent substantial fractions of a wavelength. As a result, electromagnetic phenomena such as transmission line effects, high-frequency propagation losses, and electromagnetic interference become prominent and affect the signal integrity of the communication between the different chips within the package. In this section, we will first review the SiP interconnects and then discuss how to analyze the signal integrity efficiently.

3.10.1 Overview of SiP Interconnects

SiP interconnects can be grouped in two sets: on-chip interconnects and off-chip interconnects. These two types of communication channels have very different physical dimensions and electrical characteristics, thus exhibiting different kinds of signal integrity problems [70]. Signal integrity problems of on-chip interconnects are mainly due to capacitive crosstalk, which has been widely described and characterized. On the other hand, off-chip interconnects may exhibit signal integrity problems mainly due to inter-symbol interference (ISI) and inductive crosstalk [55], which needs to be carefully characterized in order to take the proper countermeasures to guarantee a reliable communication. In this section, we will concentrate on the off-chip interconnects, as its signal integrity is becoming increasingly critical with continuous decreasing of the power supply and escalating of the operating frequencies [81, 84].

In order to evaluate the signal integrity of the considered system, either single transition analysis or eye diagram analysis can be carried out. Single transition analysis switches all signals simultaneously in the same direction, thus generating the maximum amount of SSN. Eye diagram analysis [9, 36], on the other hand, plots the transition of a single (victim) line for different transitions of other lines (aggressors). This procedure has to be accomplished for each signal. Different from the single transition analysis, the eye diagram analysis allows for the evaluation of the effects of crosstalk (both capacitive and inductive) among the lines and ISI, thus allowing a more realistic evaluation. Accordingly, below we will limit our discussion to the eye diagram analysis.

As shown in Figure 3.19, the eye diagram is defined as the synchronized superposition of all possible realizations of the signal viewed

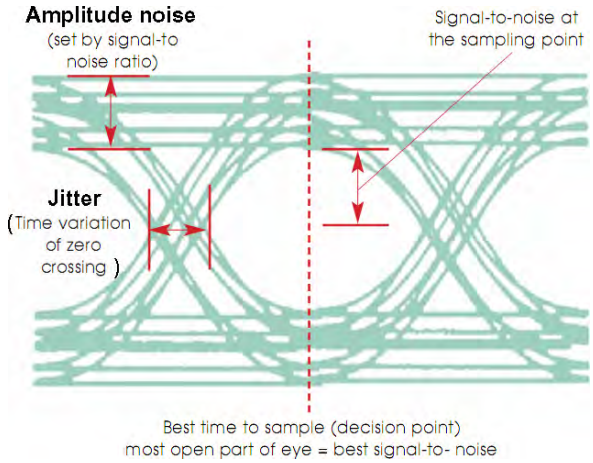


Fig. 3.19 Eye diagram.

within a particular signal interval. It provides a fast evaluation of system performance. The width of the eye opening defines the time interval over which the received signal can be sampled without error. The height of the eye opening with the amount of amplitude noise at a specified sampling time defines the signal-to-noise ratio (SNR) of the received signal [36].

To evaluate the signal integrity, we can measure the jitter and amplitude noise from the eye diagram. Jitter is defined as the deviation of the zero-crossing from its ideal occurrence time and decreases the eye's width [51]. Amplitude noise, on the other hand, decreases the SNR and, accordingly, the eye's vertical height.

In order to obtain the eye diagram for jitter and amplitude noise at the design stage, accurate modeling and time-domain simulation of the high-speed interconnect are required, which will be discussed in the following section.

3.10.1.1 Modeling and Simulation of High-speed Interconnects

A simplified model based on lumped resistance, capacitance, and inductance parameters can be adopted if the interconnect length is smaller than one-tenth of the wavelength of the signal [70]. However, with the

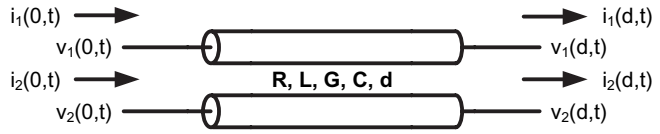


Fig. 3.20 Parallel transmission lines.

constant increases in clock frequency, such a condition can hardly hold and a transmission line model must be used in those cases. One way to characterize the transmission line is to use the well-known Telegrapher's equations based on the RLGC per-unit-length model [1]. In this section, we briefly discuss how we can obtain the time-domain waveform necessary for the eye diagram characterization starting from the Telegrapher equations. For the simplicity of presentation, we will use the transmission line differential pair shown in Figure 3.20 as an example. The derivation can be easily applied to more general cases and interested readers are referred to [1] for more details.

In a frequency domain, the voltage–current relationships between $x = 0$ and $x = d$ [1] from the Telegrapher equations can be expressed as

$$\frac{\partial}{\partial x}v(x,t) = -Ri(x,t) - L\frac{\partial}{\partial x}i(x,t), \tag{3.11}$$

$$\frac{\partial}{\partial x}i(x,t) = -Gv(x,t) - C\frac{\partial}{\partial x}v(x,t), \tag{3.12}$$

where v and i are voltage and current vectors, R, L, G, and C are the resistance, inductance, conductance, and capacitance matrices, which can be stamped using the per-unit-length resistance, inductance, conductance, and capacitance values, respectively.

Taking the Laplace transforms of (3.11) and (3.12), we can get

$$\frac{\partial}{\partial x}V(x,s) = -ZI(x,s), \tag{3.13}$$

$$\frac{\partial}{\partial x}I(x,s) = -YV(x,s), \tag{3.14}$$

where $Z = R + sL$ and $Y = G + sC$ are the impedance and admittance matrices. Differentiating the partial differential equations given in Equations (3.13) and (3.14) with respect to x, we get the following

two coupled equations:

$$\frac{\partial^2}{\partial x^2}V(x,s) = ZYV(x,s), \tag{3.15}$$

$$\frac{\partial^2}{\partial x^2}I(x,s) = YZI(x,s). \tag{3.16}$$

Decoupling of Equations (3.15) and (3.16) can be achieved by introducing a transformation matrix W relating to actual circuit voltage V and nodal voltage $\tilde{V}$ [1]:

$$V(x,s) = W\tilde{V}(x,s). \tag{3.17}$$

Using Equation (3.17), we can rewrite Equation (3.15) as

$$\frac{\partial^2}{\partial x^2}W\tilde{V}(x,s) = ZYW\tilde{V}(x,s), \tag{3.18}$$

or

$$\frac{\partial^2}{\partial x^2}\tilde{V}(x,s) = (W^{-1}ZYW)\tilde{V}(x,s). \tag{3.19}$$

To successfully couple the equations, the matrix product in parenthesis of Equation (3.19) must be a diagonal matrix:

$$W^{-1}ZYW = \begin{bmatrix} \gamma_1^2 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \gamma_N^2 \end{bmatrix}, \tag{3.20}$$

where the transformation matrix W corresponds to the eigenvectors of product ZY. The resulting diagonal matrix contains the eigenvalues of product ZY. The solution of Equation (3.19) can then be expressed as

$$\tilde{V}(x) = [E(x)]C_1 + [E(x)]^{-1}C_2 \tag{3.21}$$

and

$$V(x) = W[E(x)]C_1 + W[E(x)]^{-1}C_2, \tag{3.22}$$

where $E(x) = \text{diag}[e^{-\gamma_1 x} \cdots e^{-\gamma_N x}]$ and (C_1, C_2) are constants determined by boundary conditions.

Substituting Equation (3.22) in Equation (3.13), we have

$$I(x) = W_i[E(x)]C_1 + W_i[E(x)]^{-1}C_2, \tag{3.23}$$

where

$$W_i = Z^{-1}W\Gamma, \tag{3.24}$$

$$\Gamma = \begin{bmatrix} \gamma_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \gamma_N \end{bmatrix}. \tag{3.25}$$

Combining Equations (3.22) and (3.23) to eliminate constants (C_1, C_2) and taking into consideration the voltage–current relationships between $x = 0$ and $x = d$,

$$\begin{aligned} \begin{bmatrix} I(0) \\ -I(d) \end{bmatrix} &= \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V(0) \\ V(d) \end{bmatrix}, \\ &= \begin{bmatrix} W_i E_1 W^{-1} & W_i E_2 W^{-1} \\ W_i E_2 W^{-1} & W_i E_1 W^{-1} \end{bmatrix} \begin{bmatrix} V(0) \\ V(d) \end{bmatrix}, \end{aligned} \tag{3.26}$$

where

$$E_1 = \operatorname{diag}\left\{ \frac{1 + e^{-2\gamma_k d}}{1 - e^{-2\gamma_k d}} \right\}, \tag{3.27}$$

$$E_2 = \operatorname{diag}\left\{ \frac{-2e^{-2\gamma_k d}}{1 - e^{-2\gamma_k d}} \right\}, \quad k = 1, 2, \ldots, N, \tag{3.28}$$

$$W^{-1}ZYW = \begin{bmatrix} \gamma_1^2 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \gamma_N^2 \end{bmatrix} = \Gamma^2, \tag{3.29}$$

with $W_i = Z^{-1}W\Gamma$, $I(0)$, $I(d)$, $V(0)$, and $V(d)$ as the Laplace transforms of $i(0,t)$, $i(d,t)$, $v(0,t)$, and $v(d,t)$, respectively. $Z = R + sL$ and $Y = G + sC$ are the impedance and admittance matrices, respectively. The transformation matrix W corresponds to the eigenvectors of product ZY and the resulting diagonal matrix contains the corresponding eigenvalues.

Without loss of generality, we model the transmitter end as an independent voltage source V_s with matching conductance G_s and the receiver end with loading conductance G_L and capacitance C_L. Accordingly, the termination constraints become

$$V(0) = V_s - \frac{I(0)}{G_s}, \tag{3.30}$$

and

$$I(d) = (G_L + sC_L)V(d). \tag{3.31}$$

We can derive the frequency domain transfer function using (3.26), (3.30), and (3.31). The result is as follows:

$$V(d) = \tilde{H}V_s(s) = (Y_{12} + (Gs + Y_{11})\tilde{Y})^{-1}Gs \cdot V_s(s), \tag{3.32}$$

where

$$\tilde{Y} = Y_{21}^{-1}(-Y_{22} - G_L - sC_L). \tag{3.33}$$

Note that G_s, G_L, and C_L are all 2×2 diagonal matrices. $\tilde{H}$ describes the complete two-port relationship and includes the effect of ISI, crosstalk, reflection, and all other channel impairments. The frequency domain relationship between differential input and differential output now becomes

$$H(s) = [1 \quad -1]\tilde{H}\begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}. \tag{3.34}$$

In order to find the time-domain response, Equation (3.34) is approximated into the following pole-residue form:

$$H(s) = c + \sum_{i=1}^{q} \frac{k_i}{s - p_i}, \tag{3.35}$$

by using a least-square-approximation-based method [8]. In this manner, the time-domain step response can be obtained through the inverse Laplace transforms of $H(s)/s$ and we get

$$s(t) = c \cdot u(t) + \sum_{i=1}^{q} \frac{k_i}{p_i}(e^{p_i t} - 1)u(t). \tag{3.36}$$

With the step response, the eye diagram can be easily constructed by using the superposition theorem and waveform folding [36].

4

Placement and Routing for SiP

In this section, we discuss the placement and routing for SiP. We first use Figure 4.1 [92] to illustrate a viable interconnect model for SiP, where I/O cells are first connected to bumps on the die via RDL *routing*, then the die is flipped and mounted on the surface of the substrate, where bumps are connected to bump pads on the substrate. Finally, package trace routing is performed to furnish the connection between bump pads to balls (or package pins). Owing to the pitch mismatch between bumps (on the chip side) and balls (on the package side), package trace routing can be further divided into two parts. The first part is to route traces under the die, which is called *escape route* since its main goal is to escape traces from the die through an appropriate number of substrate layers. The second part is to route traces after escaping and we call this the *substrate route*. Package trace routing is preferred to be planar, as it not only reduces the number of high cost buried vias on the package, but also makes timing and SI analysis more predictable as transmission line modeling can be used for package traces.

In the remainder of this section, we will present the I/O planning and placement in Section 4.1, the RDL routing in Section 4.2,

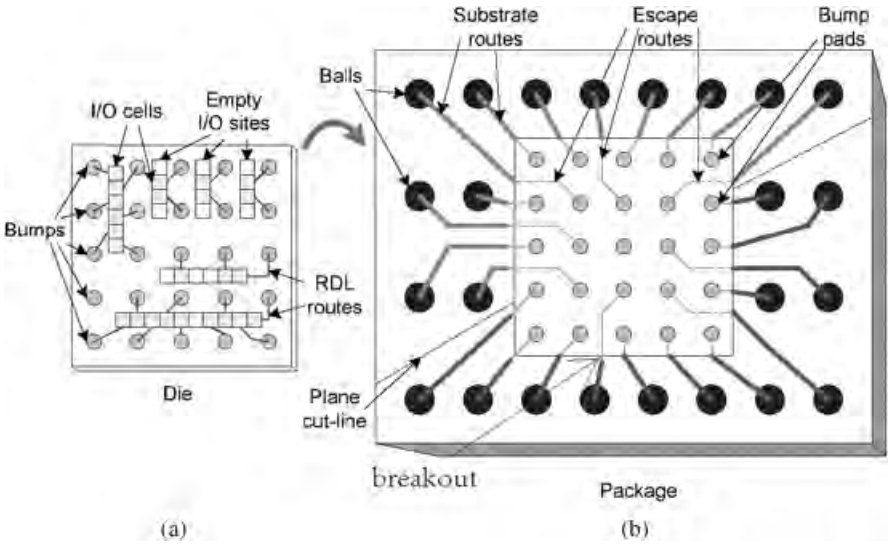


Fig. 4.1 I/O placement and package routing phases for flip chip design.

the escape routing in Section 4.3, and finally the substrate routing in Section 4.4.

4.1 I/O Placement

I/O placement plays a key role as the interface between chip and package designs in a co-design flow. I/O placement not only significantly affects chip performance, but also determines the feasibility of package designs. To tackle timing closure, signal integrity, and power integrity problems resulting from chip-package co-design, complicated design constraints are generated in practice to guide I/O placement. We first discuss some of the common design constraints that we encountered in a number of real industrial designs in this section.

- *Power integrity constraints*: All signal I/O cells that share common power and ground nets fall into one *power domain*, and they are expected to be physically placed close to each other. In addition, in order to provide good reference planes for signal I/O traces in the package, the P/G planes in the package have to be cut based upon the I/O cells' power

domain properties and their physical locations. A good I/O placement should find a solution that minimizes the number of cuts on a package's P/G planes. P/G nets that provide power supply to signal I/O cells also require a set of corresponding *P/G driver cells* to be connected with the package P/G planes. In practice, a *ratio* of the number of I/O cells to the number of neighboring P/G cells (the so-called *signal-power-ground ratio*, or *SPG ratio*) must be maintained such that the design can have a reliable power supply. Different SPG ratios may be derived for different groups of signal I/O cells.

- *Timing constraints*: Substrate routes in a package vary significantly. One of the impacts of such variation is on the timing measured from I/O cells to package pins. In other words, different substrate route lengths result in significantly different delays. If power supply variations and package stackup variations are taken into account, more significant delay variations would exhibit, making timing closure extremely difficult to attain. Therefore, for I/O cells that have critical timing relations such as differential pairs, we have to take this delay variation into account when we place these cells. A common practice is to place differential pairs close to each other so that the corresponding package routes will have a similar route length.
- *I/O standard related constraints*: Common I/O interfaces implemented in the same chip in today's high-speed IC designs (e.g., DDR2, SSTL, PCI-express, Serdes) are not unusual. Each I/O interface has its own specification on the relative timing requirements for signals within that interface (such as differential pairs). Moreover, because all signals belonging to the same I/O interface will be very likely routed to the same I/O interface in other chips on the PCB, it is desirable to have the I/O cells belonging to the same interface physically close to each other (or even in a preferred order) to reduce the delay and SI variations between signals of the same interface. In particular, differential signal pairs

are usually required to escape and to be routed together on the package. This imposes not only a *closeness* constraint but also *bump assignment feasibility* constraints (e.g., bumps escaped on the same layers).

- *Floorplan induced region constraints*: Some I/O cells may have a *region* preference or constraints that are imposed by either a chip floorplan or PCB floorplan. For example, in a top-down design hierarchy, the placement of I/O signals may have side preferences imposed by a board-level floorplan in a bottom-up design style, the placement of the core dictates that some I/O cells be placed within certain regions. Without respecting these region constraints, significant wire length increase and performance and routability degradation could occur.

Considering the timing constraints and circuit structure, Pedram et al. [67] proposed an algorithm for assigning off-chip I/O pads to a circuit. The technique treats I/Os as floating gates and uses a force-directed approach to assign positions to all gates. However, the proposed algorithm only considers the peripheral array I/O, which has been proven to be an inferior scheme to area-array I/O [13].

Caldwell et al. [13] presented an emperical study on the implications of area-array I/O for placement methodology by examining different I/O regimes (e.g., area-array *versus* peripheral pad locations), alternatives to different I/O and core placement methodologies, and different placement engines. The experimental data show that the area-array I/O regime can tolerate bad placement methodologies more than the peripheral I/O regime and the wrong methodology can entail substantial degradation of solution quality and efficiency.

To consider the complex I/O placement constraints, integer and linear programming (ILP) is an effective tool. Mak [60] presented an exact algorithm to solve the constrained I/O placement problem for FPGAs that support multiple I/O standards based on an ILP-based formulation. Using a similar approach, Xiong et al. [92] proposed an ILP-based algorithm that considers all practical I/O placement constraints as aforementioned.

4.2 Redistribution Layer Routing

As the first phase of the package routing, RDL routing, redistributes nets from I/O cells to bumps and then routes each of them. There are two types of RDL routing problems for the flip chip design [30]. The first one is the *free-assignment* routing problem, where an I/O cell is not assigned to any bump before routing, and therefore a router has the freedom to assign an I/O cell to any bump during the routing. The second type of RDL routing is the *pre-assignment* routing problem, where the mapping between I/O cells and bumps is specified. For those designs whose functions of I/O cells and bumps are pre-defined by the IC and packaging designers, such pre-assignment RDL in routing problems for each net is very important.

Algorithms have been proposed for both problems. To cope with the free-assignment RDL routing, Fang et al. [31] proposed a two-stage technique of global routing followed by detailed routing. In global routing, a network flow-based algorithm is used to solve the assignment problem from the wire bonding or flip pads to the bump pads, and a global routing path is created. The detailed routing consists of three stages: i.e., cross point assignment, net ordering determination, and track assignment, to complete the routing. To deal with the pre-assignment RDL routing, Fang et al. [30] presented an ILP-based global routing algorithm for RDL, followed by a detailed routing using X-based gridless routing to complete the routing.

In addition to the traditional routing cost metric of total wire length, the issue of signal skews is of significance in RDL routing because the signals are linked to off-chip high-speed communication channels. Fang et al. [30, 31] model the signal skew as the difference in wire length between two nets and take it into consideration in the routing by balancing the wire length among different nets.

4.3 Escape Routing

Escape routing is performed to connect the I/O signals in the area array underneath the die to the next level assembly, where wires are routed to I/O signal breakouts. The high I/O signals count and density require

an increase in the number of escape routing layers and increase in the severity of the signal integrity. Traditionally, the number of escape routing layers is reduced by dimensional changes, such as decreasing wire width and spacing. However, these changes increase design cost and cause yield and reliability issues. An efficient and effective escape routing strategy that achieves high performance with low cost would greatly benefit the electronic product. Owing to the regularity of the I/O pad distribution, pattern-based escape routing (Section 4.3.1) is an effective strategy for general packages in practice. In addition, to cope with special design considerations, such as MCM and dense PCBs, customized escape routing algorithms are needed (Section 4.3.2).

4.3.1 Escape Patterns

In the conventional escape routing manner for a typical square grid pad matrix, pads on an outermost row are routed out first, and then the pads on the next inner row are routed out through the *channel*. A channel is made up of a pair of pads rectilinearly adjacent to each other. Thus, the routing is performed in order from the outer side row and the number of rows routable on a layer are determined by the number of lines per channel plus one, as shown in Figure 4.2.

Horiuchi et al. [39] showed that wiring efficiency can be significantly improved by employing the *hybrid routing channel* in a preferential routing strategy. The hybrid channels consist of both rectilinear

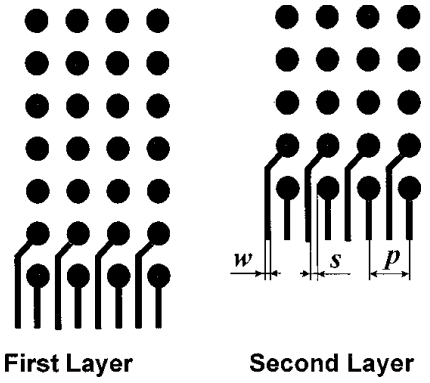


Fig. 4.2 Typical conventional routing structure on first and second layers.

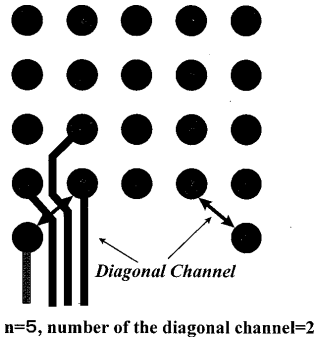


Fig. 4.3 Illustration of hybrid channels.

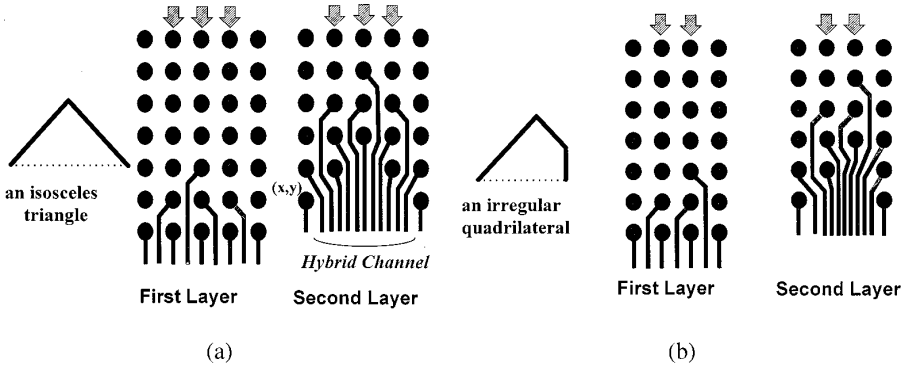


Fig. 4.4 Examples of preferential routing with hybrid channels: (a) $n = 5$ and (b) $n = 4$.

channels and diagonal channels (see Figure 4.3). The number of channels is determined by the number of pads (denoted by n) that form the peripheral of the escape area. For example, there are two diagonal channels with n=5 in Figure 4.3.

Figure 4.4 shows the preferential routing structure with the hybrid channels. Figure 4.4(a) includes the maximum number of diagonal channels for a given odd n and the number of diagonal channels available is $n - 1$. For an even n, the shape of the hybrid channels becomes an irregular quadrilateral with $n - 1$ diagonal channels as shown in Figure 4.4(b). To achieve the efficiency of the hybrid channel for all routing layers, n and the shape of the hybrid channel can vary from one layer to the other. Figure 4.5 shows an arrangement for the hybrid channels in three layers.

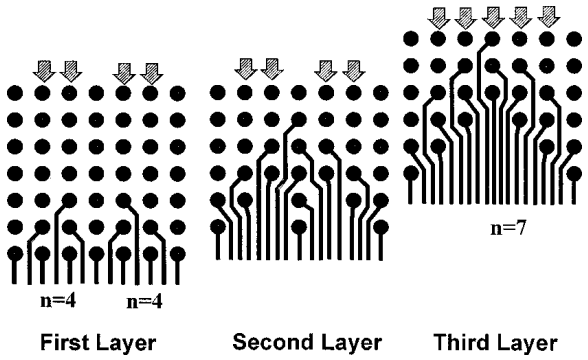


Fig. 4.5 An arrangement of the hybrid channels with $n = 7$.

Wang et al. [85] devised a central triangular sequence to minimize the escape routing layers. A network flow model is used to analyze the bottleneck of the routable pins. The triangular patterns are generated in a reverse order, from the last to the first layer. The proposed triangular pin sequence maximizes the sum of escape pins in the accumulated layers and thus minimizes the number of escape routing layers.

The preferential routing strategy proposed by Horiuchi et al. [39] and Wang et al [85] can only deal with the square grid array. Shi and Cheng [74] analyzed the properties of hexagonal array (see Figure 4.6), and showed that such a geometry increases the density of I/Os in the array remarkably. This work further proposed three escape patterns for the hexagonal array: column-by-column horizontal escape routing, two-sided horizontal/vertical escape routing, and multi-direction hybrid channel escape routing, originally presented by Shi et al. [73] for the square grid array. As a result, the hexagonal array can hold 15% more I/Os and implement the same array with a less or equal number of layers when compared with the square grid array.

In order to compare these three routing patterns, Figure 4.7 shows an example for a 10×11 hexagonal grid array. The column-by-column horizontal escape routing strategy is very straightforward for the hexagonal array. Using this strategy, I/Os in hexagonal array can escape in the same number of routing layers as the square grid array, which has the same area and same pitch, although more I/Os are packed inside. However, the vertical routing channels are wasted in this strategy.

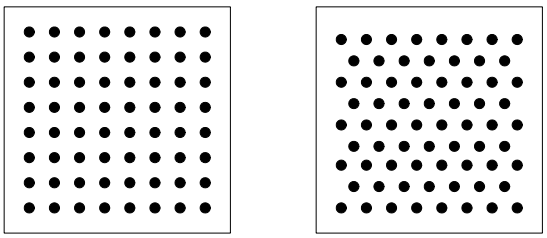
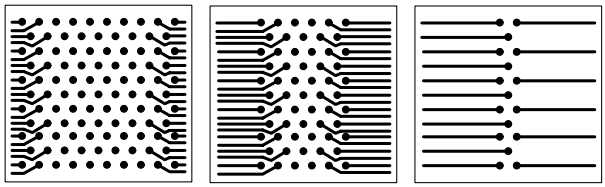
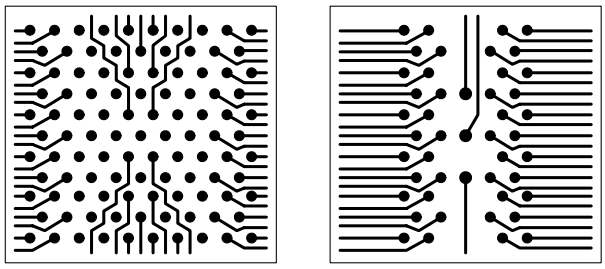


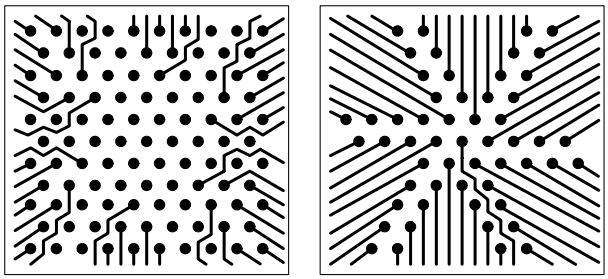
Fig. 4.6 Square grid array *versus* hexagonal grid array.



(a) Column-by-column horizontal escape routing



(b) Two-sided horizontal/vertical escape routing



(c) Multi-direction hybrid channel escape routing

Fig. 4.7 Escape patterns for a 10×11 hexagonal grid array.

In the two-sided horizontal/vertical escape routing strategy, I/Os in the outside zigzag columns escape through horizontal channels and simultaneously I/Os in the middle zigzag columns escape through vertical routing channels. All routing channels are utilized sufficiently and the number of escape routing layers are reduced. However, there is no simple routing rule for the wires breaking out I/Os in the middle zigzag columns; hence it is hard to implement this strategy in an automatic program and those wires need to go through many other I/Os, and thus, crosstalk is an important issue to be considered. The multi-direction hybrid channel escape routing strategy uses the symmetric property of hexagonal array to divide it into six partitions and exploits hybrid channels to increase escape efficiency. The hybrid channels increase the number of escape routing wires on every layer and consequentially decrease the number of layers. I/Os in each partition escape independently, which makes the problem simpler and the wires routing through hybrid channels are very ordered; hence, this strategy is easy to implement in an automatic program.

4.3.2 Escape Routing for Special Purposes

To cope with specific requirements in PCB routing, a customized escape routing algorithm is needed to enhance routability and performance. The PCB routing problem can be decomposed into two phases (as shown in Figure 4.8). Ozdal and Wong [63] and Ozdal et al. [64] proposed an effective algorithm to find the escape routing solution for multiple dies simultaneously such that the number of crossings in the intermediate area is minimized. Figure 4.8 illustrates a one-layer escape routing solution for two components. In this figure, nets have been routed from their terminal pins to the corresponding component boundaries. Here, only one net (net D) crosses with the others in the intermediate area. For this net, the area router will need to use a via to resolve the crossing. In [63, 64], the use of the number of crossings in the intermediate area is a good measure for the via requirements of an escape routing solution.

The above problem was formulated by Ozdal and Wong [63] as a *longest path with forbidden pairs* problem, which is then solved by an optimal polynomial algorithm and a random algorithm with high

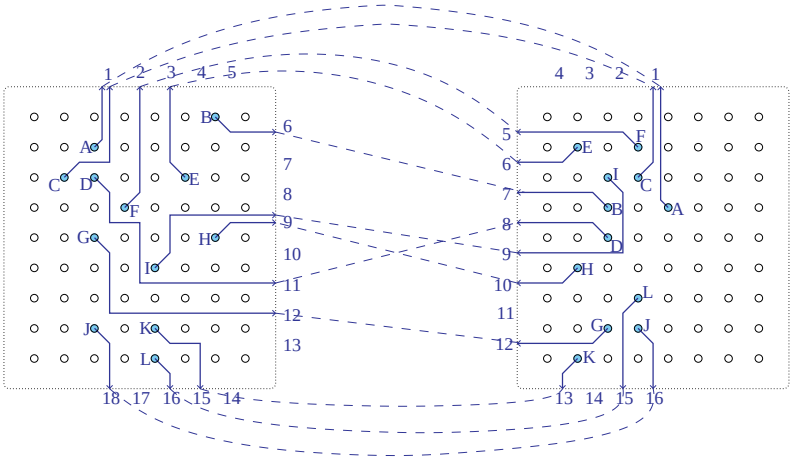


Fig. 4.8 An escape routing solution for 12 nets. The escape slots are identified on the boundaries of components. The connections in the intermediate area are shown by dashed lines.

scalability. Ozdal et al. [64] also proposed an improved algorithm to tackle the same problem by considering more general escape patterns instead of straight connections in [63]. As a result, the enhanced algorithm reduces the number of vias by 39%.

Ozdal et al. [65] studied a specialized escape routing for high-end MCM designs: *routing within dense pin clusters*. Pin clusters are often formed by pins that belong to the same functional unit or the same data bus and can become bottlenecks in terms of overall routability. Typically, these clusters have irregular shapes, which can be approximated with rectilinear convex boundaries. As such boundaries often have irregular shapes, a traditional escape routing algorithm may give unroutable solutions. Ozdal et al. [65] studied how the positions of escape terminals on a convex boundary affect the overall routability and proposed a set of necessary and sufficient conditions to model routability outside a rectilinear convex boundary. Given an escape routing solution, Ozdal et al. further proposed an optimal algorithm to select the maximal subset of nets routable outside the boundary.

Yan and Wong [93] specifically considered the modeling of the diagonal capacity in the escape routing and proposed an optimal algorithm based on the network flow. This algorithm also handles missing pins.

4.4 Substrate Routing

Substrate routing connects escape break points of flip chip dies or bond pads of wire bonding dies to solder balls (usually in the bottom layer) of a BGA package substrate. In general, substrate routing can be divided into two steps: *topological routing* and *detailed routing*.

Yu and Dai [97] proposed an algorithm for simultaneous pin assignment and topological routing for PGA packaging. The algorithm is to first find a monotonic pin assignment, i.e., the mapping from pins of a die to pads on the substrate follows a clockwise order. Given a monotonic pin assignment, it was proven that there must exist a monotonic routing with the shortest wire length. This work was generalized by Yu et al. [98] by considering multiple routing layers and other practical issues such as prerouted nets and all-angle wiring styles. A min-cost network flow-based heuristic was proposed [98] by representing the routing space with flows in a triangulated routing network. Figure 4.9 shows an example of the triangulation.

Assuming that the pad-to-pin assignment is given, Chen et al. [16] proposed a windows-based substrate routing algorithm, which clusters a set of grid array pins to form multiple rectangular rings and the substrate area. The algorithm tries to minimize the number of layers required to complete the routing. Three phases are employed, including

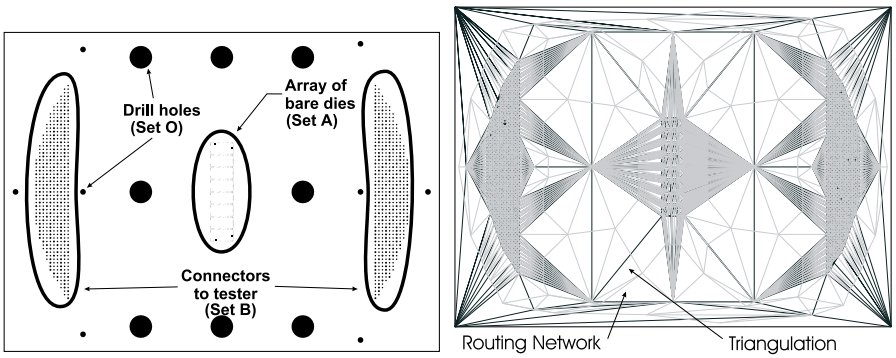


Fig. 4.9 (a) An unrouted example: each pad in the center must be connected to one pad on the edge connector, and (b) the triangulated routing network of the (a).

a layer assignment phase to assign 2-terminal nets to two or more distinct layers, a topological routing phase to realize the pad-to-pin interconnect on each layer in a type of planar sketch, and a geometrical routing phase to transform the planar sketch into a physical wiring layout. Taking inter-layer vias into consideration, Kubo and Takahashi [50] proposed a via assignment and topological routing method for two-layer BGA packages, which considers total wire length and wire congestion. The algorithm begins with an initial via assignment and incrementally improves the via assignment by local modification, including moving a via to the adjacent grid node one by one, exchanging two adjacent vias or rotating three adjacent vias. The modification with the maximum gain is applied.

When dropping signal vias, the vias need to be staggered and close to the locations above assigned destination balls (see Figure 4.10). In substrate routing, nets are connected to the solder balls in the bottom-layer by staggered vias, where vias crossing multiple layers cannot be stacked exactly one on top of the other due to the required offset called minimum and maximum staggered via pitches. The former is determined by via manufacturing technology. The latter is determined by the P/G network as the pitch should not impact the integrity of the P/G plane. This important practical constraint was largely ignored in existing substrate routing algorithms. Liu et al. [59] proposed a flexible via-staggering technique to improve routability, and developed a substrate routing algorithm, which applies dynamic pushing to tackle the net ordering problem as well as reordering and rerouting to further reduce wire length and congestion. The algorithm reduces unrouted nets by 4.5× compared with an industrial tool.

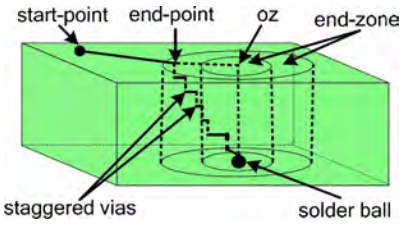


Fig. 4.10 Illustration of staggered vias.

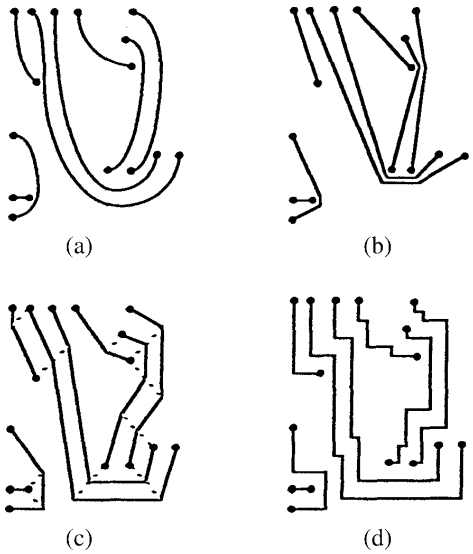


Fig. 4.11 Rubber-band sketch-based detailed routing.

After the topological routing, the resultant netlist in a single layer can be represented by the *rubber-band equivalent (RBE) of a sketch*, proposed by Leiserson and Maley [56]. A sketch consists of a finite set of rigid objects, called *features*, F, and a finite set of wires, W. Each wire in W may not intersect itself or any other wires, and each wire is defined to intersect exactly two features, thus, forming 2-terminal nets (see Figure 4.11(a)). As a rough routing can be represented by many different sketches, a canonical form is used, a *rubber-band sketch*. This rubber-band sketch parallels the RBE of a sketch [56]. A rubber-band sketch is a sketch where each wire in W is represented by a rubber-band and is the minimum length routing for that wire's topology (see Figure 4.11(b)). To hold wires away from each feature in F (because of design rules), *spokes*, open-ended segments, are used to maintain the minimum spacing between adjacent objects where the length of each spoke reflects the minimum separation between such objects (see Figure 4.11(c)). A scan-based algorithm can be used to embed the rubber-band sketch with spokes to certain geometries, such as a rectilinear shape or a octilinear shape (see Figure 4.11(d)).

Based on the above procedure, the SURF system developed by Dai et al. [20] presented a grid-less detailed package routing, which creates spokes explicitly, updating the given sketch dynamically as they are created. SURF then performs a routability test using geometric algorithms such as convex hull, range search, and the Delaunay triangulation. A detailed analysis of routability and geometric embedding based on the rubber-band representation can be found in [82]. Chen and Lee [14] showed that the dynamic updating of the RBE is computationally expensive and proposed an efficient and easy-to-implement algorithm for RBE transformation, resulting in the $O(|F| \cdot |W|)$ time algorithm for a topological sketch with $|F|$ features and $|W|$ non-crossing wire segments connecting n two-terminal nets.

References

[1] R. Achar and M. Nakhla, "Simulation of high-speed interconnects," in *Proceedings of the IEEE*, vol. 89, pp. 693–728, May 2001.

[2] D. D. Antono, K. Inagaki, H. Kawaguchi, and T. Sakurai, "Simple waveform model of inductive interconnects by delayed quadratic transfer function with application to scaling trend of inductive effects in VLSI's," *IEICE Transactions Fundamentals*, pp. 3569–3578, 2006.

[3] J. Audet, D. O'Connor, M. Grinberg, and J. Libous, "Effect of organic package core via pitch reduction on power distribution performance," in *Proceedings of the Electronic Components and Technology Conference*, pp. 1449–1453, 2004.

[4] S. Aziz, "It's still an analog world: Signal and power integrity," in *2005 MAPLD International Conference*, 2005.

[5] H. Bakoglu, *Circuits, Interconnections and Packaging for VLSI.* Addison-Wesley, 1990.

[6] G. Balamurugan, J. Kennedy, G. Banerjee, J. Jaussi, M. Mansuri, F. O'Mahony, B. Casper, and R. Mooney, "A scalable 5–15 Gbps, 14–75 mW low-power I/O transceiver in 65 nm CMOS," *IEEE Journal of Solid-State Circuits*, pp. 1010–1019, 2008.

[7] W. D. Becker and et al., "Modeling, simulation, and measurement of mid-frequency simultaneous switching noise in computer systems," in *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, pp. 157–163, 1998.

[8] W. Beyene and J. Schutt-Aine, "Efficient transient simulation of high-speed interconnects characterized by sampled data," in *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging*, vol. 21, pp. 105–114, February 1998.

[9] G. Breed, "Analyzing signals using the eye diagram," *High Frequency Electronics*, pp. 50–53, November 2005.

[10] W. Brown, *Advanced Electronic Packaging with Emphasis on Multichip Modules*. IEEE Press, 1999.

[11] P. Buffet, J. Natonio, R. Proctor, Y. Sun, and G. Yasar, "Methodology for I/O cell placement and checking in ASIC designs using area-array power grid," in *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 125–128, 2000.

[12] A. Caldwell, A. Kahng, S. Mantik, and I. Markov, "Implications of area-array I/O for row-based placement methodology," in *IEEE Symposium on IC/Package Design Integration*, pp. 93–98, 1998.

[13] A. Caldwell, A. B. Kahng, S. Mantik, and I. L. Markov, "Implications of area-array I/O for row-based placement methodology," in *IEEE Symposium on IC/Package Design Integration*, 1998.

[14] H. F. S. Chen and D. T. Lee, "A faster algorithm for rubber-band equivalent transformation for planar vlsi layouts," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 1996.

[15] J. Chen and L. He, "Efficient in-package decoupling capacitor optimization for I/O power integrity," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 734–748, 2007.

[16] S. S. Chen, J. J. Chen, S. J. Chen, and C. C. Tsai, "An automatic router for the pin grid array package," in *Proceedings Asia South Pacific Design Automation Conference*, pp. 275–281, 1999.

[17] Y. Chen, Z. Chen, and J. Fang, "Optimum placement of decoupling capacitors on packages and printed circuit boards under the guidance of electromagnetic field simulation," in *Electronic Components and Technology Conference*, pp. 756–760, 1996.

[18] M. Chrzanowska-Jeske and S.-K. Her, "Improved I/O pad assignment for sea-of-gates placement algorithm," in *Proceedings of the 35th Midwest Symposium on Circuits and Systems*, pp. 1396–1399, 1992.

[19] S. Dabral and T. Maloney, *Basic ESD and I/O Design*. John Wiley and Sons, 1998.

[20] W. W. Dai, R. Kong, and M. Sato, "Routability of a rubber-band sketch," in *Proceedings Design Automation Conference*, 1991.

[21] W. J. Dally and J. Poulton, *Digital Systems Engineering Cambridge*. Cambridge University Press, 1998.

[22] J. Darnauer, D. Chengson, B. Schmidt, E. Priest, D. Hanson, and W. Petefish, "Electrical evaluation of flip-chip package alternatives for next generation microprocessors," *IEEE Transactions on Advanced Packaging*, pp. 407–415, 1999.

[23] P. Dehkordi, C. Tan, and D. Bouldin, "Intrinsic area array ics: What, why, and how," in *Multi-Chip Module Conference*, pp. 120–124, 1997.

[24] A. Devgan, J. Hao, and W. Dai, "How to efficiently capture on-chip inductance effects: Introducing a new circuit element K," in *Proceedings International Conference on Computer Aided Design*, pp. 150–155, 2000.

[25] B. e. a. Garben, "Novel organic chip package technology and impacts on high speed interfaces," in *IEEE Topical Meeting on Electrical Performance of Electronic Packaging*, 2002.

[26] Y. e. a. Goto, "Electrical design optimization and characterization in cell broadband engine package," in *Proceedings of the Electronic Components and Technology Conference*, 2006.

[27] P. e. a. Harvey, "Chip/package design and technology trade-offs in the 65 nm cell broadband engine," in *Proceedings 57th of the Electronic Components and Technology Conference*, pp. 27–34, 2007.

[28] D. e. a. OConnor, "Electrical modeling and characterization of packaging solutions utilizing lead-free second level interconnects," in *Proceedings of the Electronic Components and Technology Conference*, pp. 450–455, 2003.

[29] A. Ern and J. L. Guermond, *Theory and Practice of Finite Elements*. Springer, 2004.

[30] J.-W. Fang, C.-H. Hsu, and Y.-W. Chang, "An integer linear programming based routing algorithm for flip-chip design," in *Proceedings Design Automation Conference*, pp. 606–611, 2007.

[31] J.-W. Fang, I.-J. Lin, P.-H. Yuh, Y.-W. Chang, and J.-H. Wang, "A routing algorithm for flip-chip design," in *Proceedings International Conference on Computer Aided Design*, pp. 753–758, 2005.

[32] K. Gala, D. Blaauw, J. Wang, V. Zolotov, and M. Zhao, "Inductance 101: Analysis and design issues," in *Proceedings Design Automation Conference*, pp. 329–334, 2001.

[33] P. Harvey, "IC packaging tutorial," in *Proceedings of Asia South Pacific Design Automation Conference*, 2006.

[34] H. Hatamkhani, F. Lambrecht, V. Stojanovic, and C.-K. K-Y, "Power-centric design of high-speed I/Os," in *Proceedings of the Design Automation Conference*, pp. 867–872, 2006.

[35] I. Hattori, A. Kamo, T. Watanabe, and H. Asai, "A searching method for optimal locations of decoupling capacitors based on electromagnetic field analysis by fdtd method," in *IEEE Conference on Electrical Performance of Electronic Packaging, 2000*, pp. 159–162, 2002.

[36] S. Haykin, *Communication Systems*. John Wiley and Sons, 2000.

[37] L. He, N. Chang, S. Lin, and O. S. Nakagawa, "An efficient inductance modeling for on-chip interconnects," in *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 457–460, 1999.

[38] Z. He, M. Celik, and L. Pileggi, "Spie: Sparse partial inductance extraction," *Proceedings Design Automation Conference*, pp. 137–140, 1997.

[39] M. Horiuchi, E. Yoda, and Y. Takeuchi, "Escape routing design to reduce the number of layers in area array packaging," in *IEEE Transaction on Advanced Packing*, 2000.

[40] M. Horowitz, C.-K. K-Y., and S. Sidiropoulos, "High-speed electrical signaling: Overview and limitations," *IEEE Micro*, pp. 12–24, 1998.

[41] H.-Y. Hsieh and T.-C. Wang, "Simple yet effective algorithms for block and I/O buffer placement in flip-chip design," in *Circuits and Systems, 2005. IEEE International Symposium on, ISCAS 2005*, pp. 1879–1882, 2005.

[42] A. Kamo, T. Watanabe, and H. Asai, "An optimization method for placement of decoupling capacitors onprinted circuit board," in *IEEE Conference on Electrical Performance of Electronic Packaging, 2000*, pp. 73–76, 2000.

[43] M. Kamon, M. J. Tsuk, and J. White, "FASTHENRY: A multipole-accelerated 3-D inductance extraction program," in *IEEE Transactions on MTT*, pp. 1750–1758, 1994.

[44] J. Kar, B. Shukla, and R. Bhattacharyya, "Optimizing C4 bump placements for a peripheral I/O design," in *Proceedings of the Electronic Components and Technology Conference*, pp. 250–254, 1999.

[45] S. Kim and D. P. Neikirk, "Compact equivalent circuit mode for the skin effect," in *IEEE-MTT-S International Microwave Symposium*, 1996.

[46] W. Koh, "System in package (SiP) technology applications," pp. 61–66, August–September, 2005.

[47] J. Kozhaya, S. Nassif, and F. Najm, "I/O buffer placement methodology for ASICs," in *International Conference on Electronics, Circuits and Systems*, pp. 245–248, 2001.

[48] B. Krauter and S. Mehrotra, "Layout based frequency depended inductance and resistance extraction for on-chip interconnect timing analysis," in *Proceedings Design Automation Conference*, pp. 303–308, 1998.

[49] B. Krauter and L. T. Pileggi, "Generating sparse partial inductance matrices with guaranteed stability," in *Proceedings International Conference on Computer Aided Design*, pp. 45–52, 1995.

[50] Y. Kubo and A. Takahashi, "A global routing method for 2-layer ball grid array packages," in *Proceedings International Symposium on Physical Design*, 2005.

[51] A. Kuo, R. Rosales, T. Farahmand, S. Tabatabaei, and A. Ivanov, "Crosstalk bounded uncorrelated jitter (BUJ) for high-speed interconnects," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, pp. 1800–1810, October 2005.

[52] M.-F. Lai and H.-M. Chen, "An implementation of performance-driven block and I/O placement for chip-package codesign," in *International Symposium on Quality Electronic Design*, pp. 604–607, 2008.

[53] Lee, "Low-power area-efficient high-speed I/O circuit techniques," *IEEE Journal of Solid-State Circuits*, 2000.

[54] H. Lee, K. Choi, K. Jang, T. Cho, and S. Oh, "A new efficient equivalent circuit extraction method for multi-port high speed package using multi-input multi-output transmission matrix and polynomial curve fitting," in *Electronic Components and Technology Conference*, pp. 1582–1588, 2003.

[55] S.-B. Lee, J. Park, S.-J. Hong, and H.-G. Jeon, "Electromagnetic interference (EMI) behavior of system in package (SIP)," in *IEEE International Symposium on Electromagnetic Compatibility*, pp. 876–888, 2004.

[56] C. Leiserson and F. Maley, "Algorithms for routing and testing routability of planar vlsi layouts," in *The 17th Annual ACM Symposium on Theory of Computing*, pp. 69–78, 1985.

[57] J. P. Libous, "Characterization of flip chip CMOS ASIC simultaneous switching noise on multilayer organic andceramic bga packages," in *IEEE 7th Topical Meeting on Electrical Performance of Electronic Packaging*, pp. 191–194, 1998.

[58] J. P. Libous and D. P. OConnor, "Measurement, modeling, and simulation of flip-chip CMOS ASIC simultaneous switching noise on a multilayer ceramic bga," in *IEEE Transacitions on Components, Packaging, and Manufacturing Technology*, pp. 266–271, 1997.

[59] S. Liu, G. Chen, T. T. Jing, L. He, T. Zhang, R. Dutta, and X.-L. Hong, "Topological routing to maximize routability for package substrate," in *Proceedings Design Automation Conference*, 2008.

[60] W.-K. Mak, "I/O placement for FPGAs with multiple I/O standards," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 315– 321, 2004.

[61] M. Mechaik, "Effects of package stackups on microprocessor performance," in *International Symposium on Quality Electronic Design*, pp. 475–481, 2000.

[62] A. Odabasioglu, C. M, and L. Pileggi, "PRIMA: Passive reduced-order interconnect macromodeling algorithm," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 1998.

[63] M. Ozdal and M. Wong, "Simultaneous escape routing and layer assignment for dense pcbs," in *Proceedings International Conference on Computer Aided Design*, pp. 822–829, 2004.

[64] M. Ozdal, M. Wong, and P. Honsinger, "An escape routing framework for dense boards with high-speed design constraints," in *Proceedings International Conference on Computer Aided Design*, pp. 759–766, 2005.

[65] M. Ozdal, M. Wong, and P. Honsinger, "Optimal routing algorithms for pin clusters in high-density multichip modules," in *Proceedings International Conference on Computer Aided Design*, pp. 767–774, 2005.

[66] R. Palmer, J. Poulton, W. Dally, J. Eyles, A. Fuller, T. Greer, M. Horowitz, M. Kellam, F. Quan, and F. Zarkeshvari, "A 14 mw 6.25 Gb/s transceiver in 90 nm CMOS for serial chip-to-chip communications," in *IEEE International Solid-State Circuits Conference*, pp. 440–614, 2007.

[67] M. Pedram, K. Chaudhary, and E. S. Kuh, "I/O pad assignment based on the circuit structure," in *IEEE International Conference on Computer Design on VLSI in Computer & Processors*, pp. 14–16, 1991.

[68] I. Qun, "Wire bonding tutorial," in *Advanced Packaging*, July, 2005.

[69] N. Rao and G. A. Rao, "Introduction to electronic packaging," http://www.prc.gatech.edu/.

[70] D. Rossi, P. Angelini, C. Metra, G. Campardo, and G. P. Vanalli, "Risks for signal integrity in system in package and possible remedies," in *European Test Symposium*, pp. 165–170, 2008.

[71] A. E. Ruehli, "Equivalent circuit models for three-dimensional multiconductor system," *IEEE Transactions on MTT*, pp. 216–221, 1974.

[72] K. L. Shepard and Z. Tian, "Return-limited inductances: A practical approach to on-chip inductance extraction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 425–438, 2000.

[73] R. Shi, H. Chen, C. Cheng, D. Beckman, and D. Huang, "Layer count reduction for area array escape routing," in *International Conference on Device Packaging*, 2005.

[74] R. Shi and C.-K. Cheng, "Efficient escape routing for hexagonal array of high density I/Os," in *Proceedings Design Automation Conference*, pp. 1003–1008, 2006.

[75] Y. Shi and et al., "Stochastic current prediction enabled frequency actuator for runtime resonance noise reduction," in *Proceedings Asia South Pacific Design Automation Conference*, pp. 373–378, 2009.

[76] Y. Shi and L. He, "Empire: An efficient and compact multiple-parameterized model-order reduction method for physical optimization," in *tvlsi*, 2009.

[77] Y. Shi and L. He, "Modeling and design for beyond-the-die power integrity," in *Proceedings International Conference on Computer Aided Design*, pp. 411–416, 2010.

[78] Y. Shi, L. He, and C.-J. R. Shi, "Scalable symbolic model order reduction," in *IEEE Behavioral Modeling and Simulation Conference*, pp. 112–117, 2008.

[79] Y. Shi, J. Xiong, C. Liu, and L. He, "Efficient decoupling capacitance budgeting considering operation and process variations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1253–1263, 2008.

[80] L. Smith, R. Anderson, D. Forehand, T. Pelc, and T. Roy, "Power distribution system design methodology and capacitor selection for modern cmos technology," *IEEE Transactions on Advanced Packaging*, pp. 284–291, 1993.

[81] Y. H. Song, S.-G. Kim, K.-J. Rhee, D.-S. Cho, and T.-S. Kim, "The reliability issues on ASIC/memory integration by SiP (system-in-package) technology," in *IEEE International SOC Conference*, pp. 7–10, 2003.

[82] D. Staepelaere, "Geometric transformations for a rubber-band sketch," M.S. Thesis, University of California Santa Cruz, 1992.

[83] M. Suryakumar, W. Cui, P. Parmar, C. Carlson, B. Fishbein, U. Sheth, and J. Morgan, "Power delivery validation methodology and analysis for network processors," in *Proceedings of the Electronic Components and Technology Conference*, pp. 589–592, 2004.

[84] C. Trigas, "Design challenge for system-in-package vs system-on-chip," in *Proceedings IEEE Custom Integrated Circuits Conference*, pp. 663–666, 2003.

[85] R. Wang, R. Shi, and C.-K. Cheng, "Layer minimization of escape routing in area array packaging," in *Proceedings International Conference on Computer Aided Design*, pp. 815–819, 2006.

[86] W. T. Weeks, L. L. Wu, M. F. McAllister, and A. Singh, "Resistive and inductive skin effect in rectangular conductors," *IBM Journal of Research and Development*, pp. 652–660, 1979.

[87] X. Wu, "Solving your ic package si and ssn problems: Icemax and it's applications," in *Icemax Tutorials*, 2005.

[88] www.dnp.co.jp.

[89] www.microbonding.com.

[90] J. Xiong and et al., "Constraint driven I/O planning and placement for chip-package codesign," in *Proceedings Asia South Pacific Design Automation Conference*, pp. 207–212, 2006.

[91] J. Xiong, Y.-C. Wong, E. Sarto, and L. He, "Constraint driven I/O planning and placement for chip-package co-design," in *Asia and South Pacific Conference on Design Automation*, 2006.

[92] J. Xiong, Y.-C. Wong, E. Sarto, and L. He, "Constraint driven I/O planning and placement for chip-package co-design," in *Proceedings Asia South Pacific Design Automation Conference*, 2006.

[93] T. Yan and M. Wong, "A correct network flow model for escape routing," in *Proceedings Design Automation Conference*, pp. 332–335, 2009.

[94] X. Yang, Q. lun Chen, and C. te Chen, "The optimal value selection of decoupling capacitors based on fdfd combined with optimization," in *IEEE Conference on Electrical Performance of Electronic Packaging, 2000*, pp. 194–197, 2002.

[95] G. Yasar, C. Chiu, R. Proctor, and J. Libous, "I/O cell placement and electrical checking methodology for ASICs with peripheral I/Os," in *International Symposium on Quality Electronic Design*, pp. 71–75, 2001.

[96] H. Yu, C. Chu, and L. He, "Off-chip decoupling capacitor allocation for chip package co-design," in *Proceedings Design Automation Conference*, pp. 618–621, 2007.

[97] M. F. Yu and W. W. Dai, "Pin assignment and routing on a single-layer pin grid array," in *Proceedings Design Automation Conference*, pp. 203–208, 1995.

[98] M. F. Yu, J. Darnauer, and W. W. Dai, "Interchangeable pin routing with application to package layout," in *Proceedings International Conference on Computer Aided Design*, pp. 668–673, 1996.

[99] H. Zheng, B. Krauter, and L. Pileggi, "On-package decoupling optimization with package macromodels," in *Proceedings IEEE Custom Integrated Circuits Conference*, pp. 723–726, 2003.