Ho-Ming Tong · Yi-Shao Lai
C.P. Wong *Editors*

Advanced Flip Chip Packaging

Advanced Flip Chip Packaging

Ho-Ming Tong • Yi-Shao Lai • C.P. Wong
Editors

Advanced Flip Chip Packaging

Editors
Ho-Ming Tong
ASE Group
Kaohsiung
Taiwan, R.O.C

Yi-Shao Lai
ASE Group
Kaohsiung
Taiwan, R.O.C

C.P. Wong
Dean of Engineering
The Chinese University of Hong Kong
Shatin, NT, Hong Kong SAR, China

on leave from:
Georgia Institute of Technology
School of Materials Science & Engineering
Atlanta, GA, USA

ISBN 978-1-4419-5767-2 ISBN 978-1-4419-5768-9 (eBook)
DOI 10.1007/978-1-4419-5768-9
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012956538

© Springer Science+Business Media New York 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

To our knowledge, flip chip-centric reference books have for the most part been edited and published over a decade ago. A case in point includes those edited by John Lau. At that time, flip chip was still considered a luxury, reserved only for high-end, high-performance applications such as mainframes and workstations. Over the past decade, however, profound technological advancements have enabled the proliferation of low-cost highly reliable flip chip packages for applications covering all four Cs, namely, computing, communications, and consumer and car electronics. Going forward, the demand for flip chip will continue to grow to meet the insatiable appetite of consumers for performance, size, cost, and environmental compatibility as we enter into the consumer era.

In the past decade, many fundamental changes have taken place, rendering flip chip packages of today drastically different from their predecessors. As Moore's law progressed from non-low-k to low-K devices, chip-package compatibility needs to be assured prior to product realization due to the fragile nature of the low-K layers. The situation will become even more acute as the industry is migrating now from 45 to 32 nm and beyond. Also, effectively in 2006, RoHS initiative mandated by the European Community has driven the semiconductor industry, flip chip makers and providers included, to migrate from Pb-containing packages to Pb-free and halogen-free packages on a global scale. Unlike the previous decade where advanced technologies such as flip chip could fetch you a premium, flip chip of the current decade (the "consumer" era) has been replacing whatever packages it is intended for on a cost-competitive basis. This has fostered implementation of low-cost flip chip structures, processes, materials, as well as equipment across the entire flip chip industry. Moreover, as function integration accelerates at the system level, the industry has been witnessing accelerated migration to ever-finer pitches at all levels from IC to package to module to board/system, as well as ever-finer pitch and ever-larger flip chip packages for certain applications. In support of the growth of flip chip, companion technologies such as substrate, underfill, interconnect, design, simulation, and reliability have also continuously evolved. As a result of all the above, flip chip, being the densest packaging technologies, has continuously reinvented itself to cope with the

ever-more strict performance, cost, size, and environmental requirements of the past decade. Going forward, this trend is expected to continue at an accelerated pace as function integration accelerates at the system level, particularly for handheld electronics such as cell phones and possibly for portable PCs such as Netbooks and multimedia Internet devices.

At present, we strongly believe that a new reference book to capture the aforementioned advances in flip chip of the past decade is timely and would provide valuable insights to researchers working in this technology. Furthermore, this manuscript on "Advanced Flip Chip Packaging" covers the past, present, and future trends and technologies related to flip chip BGA and flip chip CSP.

Kaohsiung, Taiwan, R.O.C Ho-Ming Tong
Kaohsiung, Taiwan, R.O.C Yi-Shao Lai
Hong Kong SAR, China C.P. Wong

Contents

1 **Market Trends: Past, Present, and Future** 1
Robert Lanzone

2 **Technology Trends: Past, Present, and Future** 23
Eric Perfecto and Kamalesh Srivastava

3 **Bumping Technologies** 53
Michael Töpper and Daniel Lu

4 **Flip-Chip Interconnections: Past, Present, and Future** 85
Sung-Kwon Kang, Da-Yuan Shih, and William E. Bernier

5 **Flip Chip Underfill: Materials, Process, and Reliability** 155
Zhuqing Zhang, Shijian Luo, and C.P. Wong

6 **Conductive Adhesives for Flip-Chip Applications** 201
Daoqiang Daniel Lu and C.P. Wong

7 **Substrate Technology** 263
Yutaka Tsukada

8 **IC-Package-System Integration Design** 341
Yiyu Shi, Yang Shang, Hao Yu, and Shauki Elassaad

9 **Thermal Management of Flip Chip Packages** 413
Richard C. Chu, Robert E. Simons, Madhusudan Iyengar,
and Lian-Tuu Yeh

10 **Thermo-mechanical Reliability in Flip-Chip Packages** 471
Li Li and Hongtao Ma

11 **Interfacial Reactions and Electromigration**
in Flip-Chip Solder Joints 503
C.E. Ho, C.R. Kao, and K.N. Tu

Chapter 1
Market Trends: Past, Present, and Future

Robert Lanzone

Abstract Flip chip, as the name implies, is a means of chip assembly whereby the chip is attached face down (active Integrated Circuit (IC) side down) to the substrate. Flip chip technology was invented by IBM in 1961 and is now 51 years old. This development allowed IBM to be the leader in producing high performance circuits. The method for production that IBM employed was expensive and limited its widespread use. IBM kept this technology captive until the mid 1990s. As it is well known IBM created the technology for its advanced computing needs. This technology utilized ceramic carriers and was suitable for high power dissipation. Although the original "controlled collapse chip connection," also known as (C4), concept utilized solder coated copper balls to form the interconnect, high lead (97/3) solder evaporated through a molybdenum stencil became the staple of the technology for decades. IBM kept this technology captive, and the intellectual property (IP) protected and secret for many years past the original Patents useful lifetime.

1.1 Flip Chip Technology Overview and Early Beginnings

Flip chip, as the name implies, is a means of chip assembly whereby the chip is attached face down (active Integrated Circuit (IC) side down) to the substrate. Flip chip technology was invented by IBM in 1961 and is now 51 years old. This development allowed IBM to be the leader in producing high performance circuits. The method for production that IBM employed was expensive and limited its widespread use. IBM kept this technology captive until the mid 1990s. As it is well known IBM created the technology for its advanced computing needs. This technology

R. Lanzone (✉)
Advanced Packaging and Wafer Level Development at Amkor Technology Inc, 1900 South Price Road, Chandler, AZ 85286, USA
e-mail: robert.lanzone@amkor.com

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_1, © Springer Science+Business Media New York 2013

utilized ceramic carriers and was suitable for high power dissipation. Although the original "controlled collapse chip connection," also known as (C4), concept utilized solder coated copper balls to form the interconnect, high lead (97/3) solder evaporated through a molybdenum stencil became the staple of the technology for decades. IBM kept this technology captive, and the intellectual property (IP) protected and secret for many years past the original Patents useful lifetime.

As interconnect requirements advanced for other Integrated Device Manufacturers (IDM)'s, they too required the advantages for flip chip interconnect. Companies such as AMD, Digital Equipment Corporation (DEC) HP, Intel and Motorola licensed the technology from IBM in the mid 1990s and thus began the adoption of the technology for a broader but specific segment of the semiconductor market.

As referenced earlier, flip chip technology is an interconnect method where the semiconductor IC is manufactured such that the electrical terminals can be connected in a face down or flipped manner to the package substrate, also known as a carrier. These electrical terminals are made traditionally of solder and allow for connecting the IC die to the package substrate. As these electrical connections or "input/outputs" (I/O's) cover the entire active area, they could achieve unparalleled density using pitches that were course compared to wirebonding. Wirebonding only facilitates I/O's along the periphery of the edges of the IC. Therefore, even employing finer spacing (pitch), they could not match the density of I/O's distributed over the entire cross-sectional area of the IC. Bumping allows for direct attachment and utilization of the entire surface of the chip. Typical pitches for the IBM evaporative bumps were 250 μm. For instance, a chip with a 10 $\times$ 10 mm size could facilitate over 1500 I/O's. For large die such as 20 $\times$ 20 mm, this I/O count could be over 6,000 interconnections.

1.2 Wafer Bumping Technology Overview

The key constituent of wafer bumping is depositing the under-bump metallization (UBM). Note that the original term coined by IBM was Ball Limiting Metallurgy (BLM). The UBM serves several purposes, including the following:

- Providing a bonding layer for the interconnection structure to which the UBM mates.
- Providing a diffusion barrier to keep the mating material from penetrating down into the underlying metallurgy.
- Providing adhesion to the underlying dielectric and metallurgy and act as a barrier against horizontal migration of contaminants along the dielectric and into the underlying metallurgy.

Today, the majority of the UBM systems employed are sputtered. Sputtering is considered the most cost-effective methodology for depositing the UBM, especially when compared to evaporation, because it is the UBM that most directly affects the reliability of the solder bump structure.

Typically, the UBM must be able to survive many reflow cycles; often up to 20 times. As the UBM is the structure that adheres the solder bump to the pad metallization, it also must be able to endure both shear stress and tensile pull tests.

A common criterion for solder failure during mechanical destructive testing is for the failure to occur within the solder itself. Therefore, the UBM must be very robust and not degrade appreciably with exposure to time, temperature, moisture and electrical bias.

1.3 Evaporation (C4)

IBM invented the technology and has employed it for high-volume production in its own processes. Evaporative bump formation was also employed by the early licensees such as AMD and Motorola. A flow diagram for the evaporation process is illustrated in Fig. 1.1.

IBM employs an evaporated UBM with Cr, phased Cr/Cu, Cu, with a final flash of Au. This UBM structure, which is evaporated through a molybdenum shadow mask, has been demonstrated to be extremely reliable for high lead solders (97/3 and 95/5).

This soft solder interconnect was originally deployed on ceramic substrates. It has also been utilized to form a hybrid joint, uniting high lead with eutectic solder for joining die onto organic substrates.

Mask alignment is an important step in this process, and the mask must be designed with the TCE and the dynamics of changing temperature on the wafer in mind.

The solder is also evaporated through these same metal masks. Since most of the solder is evaporated on the masks and chamber walls, this process can be extremely costly. This is especially true for expensive low alpha solders.

The masks must be cleaned as the solder builds up on the metal mask. The cleaning process degrades the quality and opening of the mask, limiting its useful life. Multiple evaporators are required for the UBM and solder materials further increasing cost.

It is generally accepted that the area-array pitch limitation for evaporation is 225 μm. When application requirements approached 200 μm, practically all of the companies that originally deployed evaporation migrated to electroplating.

1.3.1 Stencil Printing

In the stencil printing process, the UBM is deposited over the entire face of the wafer using sputtering, including the metallized bonding pads. Typical UBM structures are Al/Ni-V/Cu, Ti/W/Cu, or equivalent. A generic process flow for sputtered/stencil printing is illustrated in Fig. 1.2. The solder is deposited using a stencil printing process. The stencil can be either metal masks or photo-defined.

Evaporation

- Vapor Phase
- UBM and solder are evaporated through a shadow mask

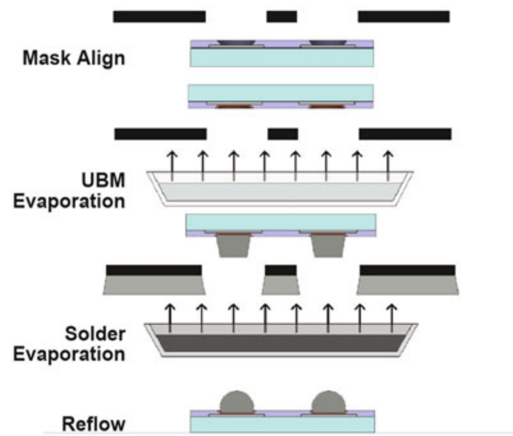


Fig. 1.1 Flow diagram of the evaporation process

Stencil Printing

- Solid Phase
- UBM is deposited and patterned
- Solder paste is screened through a stencil
- Low cost, coarse geometry

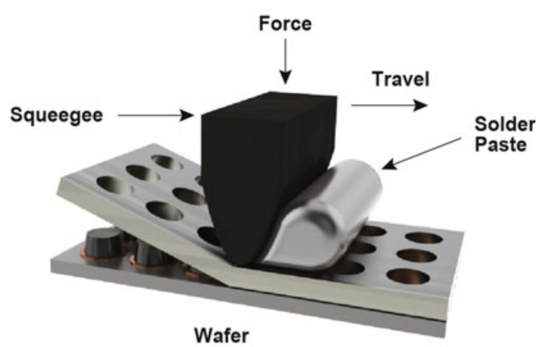


Fig. 1.2 The generic process flow for sputtered/stencil printing

The photo-defined methodology allows for tighter pitch, but still has limitations for very fine pitch. Yield issues are the predominant concern for fine pitch applications using screen printing. Stencil printing has other limitations for coarse pitch as well. Other issues occur during processing on thinned or brittle wafer materials such as Gallium arsenside (GaAs), Indium phosphide, or other III–V compounds.

Although stencil printing is generally considered to be the lower cost technology, due to its perceived ease of manufacturing for screen printing operations, this is not always the case. The stencils must be photo-defined as they are in electroplating. Screening and solder is not the most efficient use of materials.

Another concern for screen printing, and a limiting factor for smaller and tighter pitch bumps, is the formation of voids in the solder bump. Since flux is required to be mixed in with the paste, the formation of voids in the reflow process must be monitored and controlled.

1.3.2 Electroplating

Electroplating is the most cost-effective at the tightest pitch ranges since its yield is highest, small bump sizes are plated faster, and large bump densities amortize the cost over more bumps.

This process offers the widest set of material choices and compositions of any of the deposition processes used today. It also offers the broadest range of bump sizes, pitches, and patterns possible.

(A typical process flow for electroplating is illustrated in Fig. 1.3.)

The use of electro-deposition does require one modification to the list of functions that UBM must provide: Due to the nature of the process, it must also provide a layer for carrying the current necessary for material deposition. By virtue of this requirement, it is necessary that this layer be blanket deposited.

In the first step of the process, a blanket UBM is deposited. A popular UBM for electroplating is Ti/Cu/Ni. These components serve very different purposes. For example, the Ti is for adhesion and sealing. The Cu is the current carrying layer, and the Ni serves as the diffusion barrier and wettable layer.

1.3.3 Solder Dam

The lower cost UBM is sputtered. A unique process employed for electroplating is the deposition of a solder dam.

The next step is to deposit and image the plating template. The primary requirement in controlling the solder volume is to use a thick resist with a well-controlled sidewall slope. A cross section prior to plating is illustrated in Fig. 1.4.

Solder is plated within the entire volume of the defined structure by the template. This is generally referred to as standard plating, illustrated in Fig. 1.5, and tight bump uniformity can be achieved for very small pitches (A photo illustration is presented in Fig. 1.6). Solder pitches as low as 50 μm have been demonstrated with this methodology.

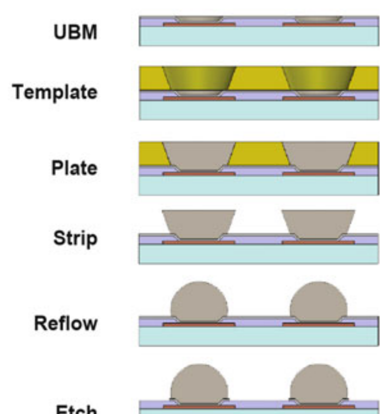


Fig. 1.3 The process flow utilized by the leading merchant provider of electroplating

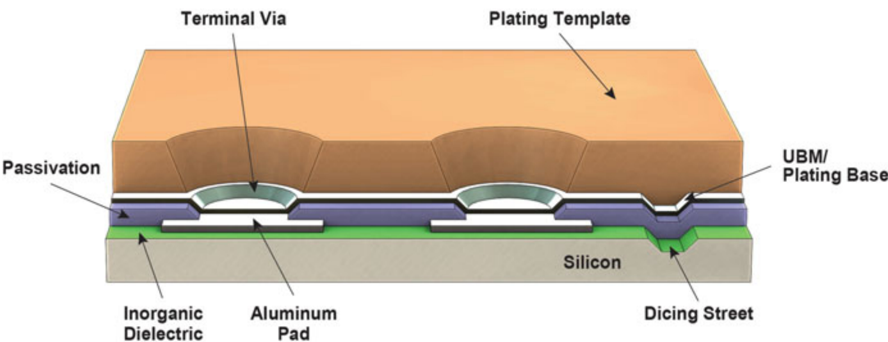


Fig. 1.4 Cross section of a solder dam prior to plating

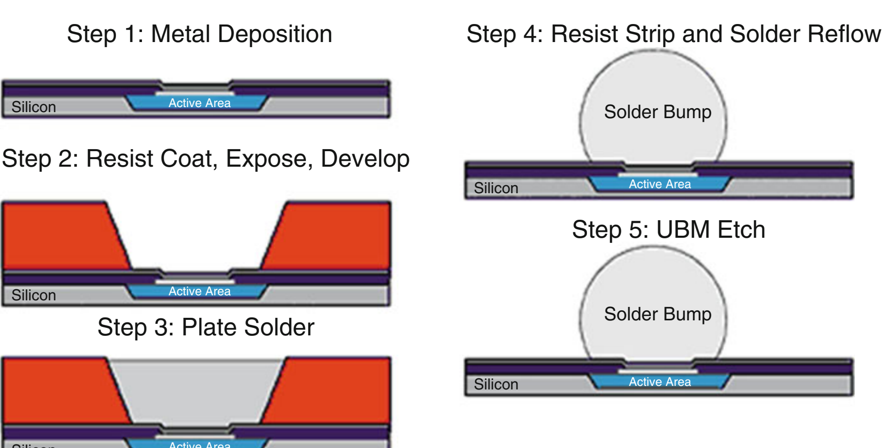


Fig. 1.5 Illustration of the standard plating process

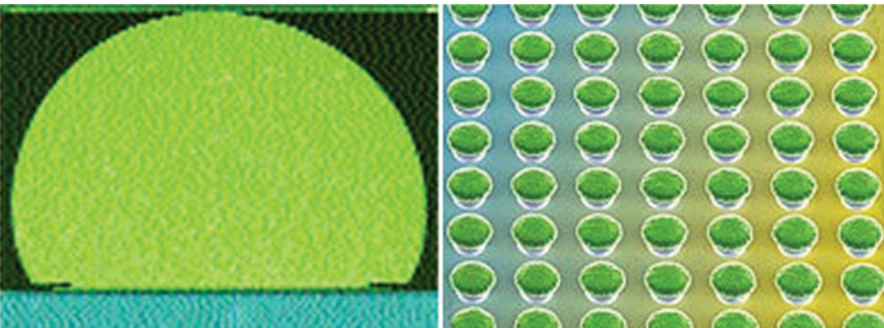


Fig. 1.6 Photo of eutectic bump post-reflow

1.3.4 Plating Outside a Defined Structure

“Mushroom” plating (which is shown in Fig. 1.7), or plating outside of a defined structure, is somewhat less controllable than plating within a defined structure. This form of plating, however, allows for the creation of much taller solder bumps than standard plating allows, because the volume of solder deposited is greater.

Once the material has been plated up in a uniform manner, the resist template may be stripped. Typically, the UBM will be etched at this point.

To avoid this issue is to first perform a solder reflow. This is the reason for depositing a solder dam in the UBM process, so that the solder does not flow outward during this operation.

The reason for reflowing at this step is to control (or eliminate) the undercutting that would normally occur with the use of the wet etching process. Undercutting can lead to loss of strength in the solder joint or may create variability of solder surface area, leading to reduced bump height coplanarity.

During the reflow process, the solder is able to force conversion of the underlying metal into an intermetallic form. This intermetallic is insensitive to the etchants in use for the field metals and hence etching through to complete removal of the field areas may continue.

A final reflow operation is performed post-etch to create a solder bump that is smooth and shiny in appearance. An example of a eutectic bump post-reflow is shown in Fig. 1.6.

Electroplating is the most scalable of all the technologies discussed to date. This is the predominant reason that companies such as major IDMs who originally instituted evaporation have migrated to electroplating (Fig. 1.8).

Essentially, if proper solder bath maintenance is performed, there is no waste, which is very important for higher cost, low alpha solder materials. Since no mechanical pressure is applied during the process, bumping on more brittle

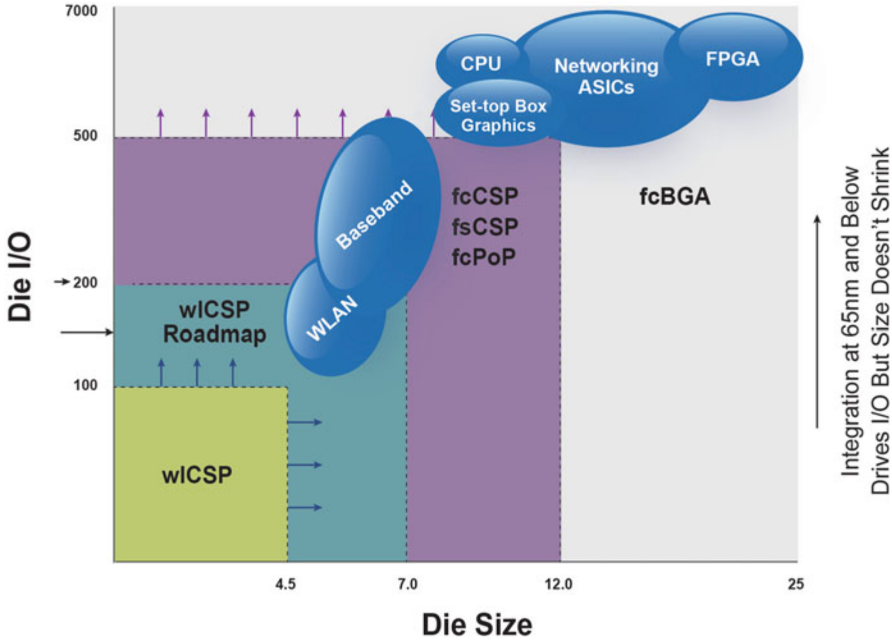


Fig. 1.7 "Application Space" for Flip Chip. *Source*: Amkor Technology Inc.

Structure	Max Die Size Qualified	Min Substrate Core Thickness
Bare Die (CUF)	Body Size minus 2.5 mm	0.20 mm (2L)
Overmold (CUF)	Body Size minus 2.0 mm	0.10 mm (2L) 0.06 mm (4L)
Moldable UF	Body Size minus 1.0 mm	0.10 mm (2L) 0.06 mm (4L)

Fig. 1.8 General Body Size for Flip Chip packaging. *Source*: Amkor Technology Inc.

materials such as GaAs or thinned silicon wafers is possible. A standard thickness specification for a 200 mm wafer is 18.0 mm.

Electroplating has been demonstrated to create large solder bumps (~175 μm in height) to very small solder bumps (~25 μm). This versatility is advantageous for a merchant wafer bumping provider (Fig. 1.9).

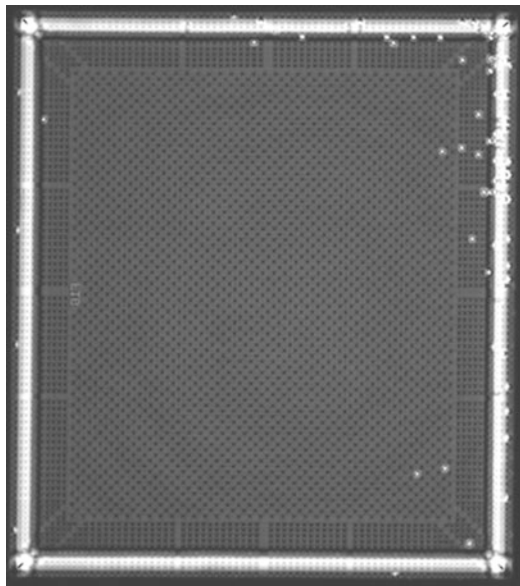


Fig. 1.9 Example of Sonocan white bump delamination. *Source*: Amkor Technology Inc.

1.4 Wafer Bump Summary

Electroplating has emerged the dominant wafer bumping technology although there still exists a large installed base of screen printed bumping. Electroplating has been proven to be the most cost-effective and versatile of the three technologies described here.

1.4.1 Substrate Technology

IBM originally developed carriers for flip chip and used either high temperature cofired ceramic (HTCC) or low temperature cofired ceramic (LTCC). These substrates could withstand the high reflow temperature, ~360 °C, reflow profiles. Ceramic materials are comparatively low Thermal Coefficient of Expansion (TCE) as compared to laminate carriers or printed wiring boards (PWB)'s. However, they are still mismatched to Silicon. Therefore, a material was needed to join these two dissimilar TCE materials. These are commonly referred to as underfills. The underfill adheres to the entire cross-sectional area of the die and substrate. The underfills' primary function is to distribute the stress and strain energy from the thermal mismatch of the die and substrate from the solder bumps themselves to the underfill material, extending solder joint life (Fig. 1.10).

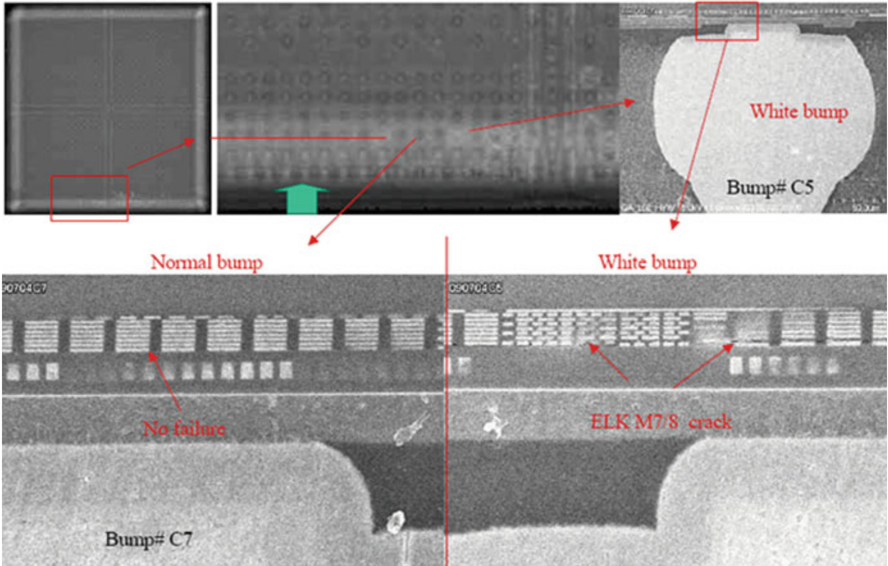


Fig. 1.10 Example of white bump ILD cracking. *Source*: Amkor Technology Inc.

1.5 Flip Chip Industry and Infrastructure Development

To summarize to date, flip chip in its early embodiment used high lead bumps evaporatively deposited at a 250 μm pitch on the wafer and reflowed onto ceramic carriers (substrates). This technology yielded high performance packaging that served IBM's advanced computing needs for decades. However, but due to its captive IP and high cost of manufacturing, widespread market adoption was limited. That said, as IDM's needed the performance advantages of flip chip, they sought out the technology from IBM. IBM licensed its flip chip technologies for fees in the tens of millions of dollars to these companies for their captive use. These were exciting times for the industry since the technology was poised for mainstream use, at least for advanced computing needs such as Microprocessors (MPUs) and Application Specific Integrated Circuits (ASICs). IBM had developed the technology which comprised the wafer bumping, ceramic carriers, and the flip chip assembly process. The flip chip assembly process had unique reflow furnaces and environments, flux and flux cleaning and the underfill technology (Fig. 1.11).

The challenge for the licensees was to adopt a new proprietary technology and create the required industry infrastructure independent from IBM to bring the technology to high volume manufacturing for the microprocessor industry. AMD, DEC, and Motorola basically brought in IBM's technology and practiced it much the same as IBM. Intel became the initial trail blazer to improve the technology's manufacturability and cost structure. To break down the infrastructure into more discrete pieces, there was the bumping technology, the substrate, and assembly

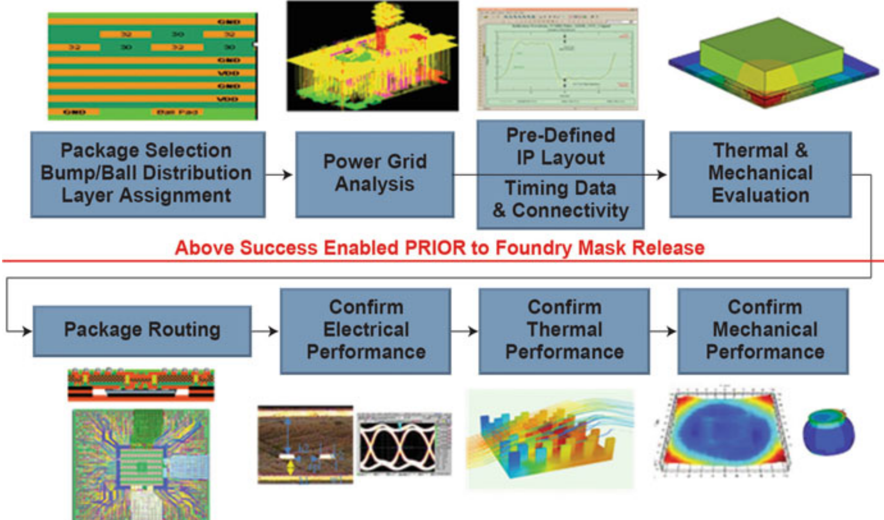


Fig. 1.11 Flowchart of co-design process. *Source*: Amkor Technology Inc.

process. The first licensees adopted IBMs bumping technology which, as discussed previously, was expensive. The first challenge they undertook was to create a global infrastructure for substrate suppliers. The author had worked at IBM for almost a decade in engineering and operations for flip chip packaging. Having sensed the first industry wide creation of the flip chip substrate infrastructure outside of IBM, the author joined Kyocera to work on flip chip ceramic substrate development in 1995. At the time, the big five MPU manufacturing companies; AMD, DEC, HP, Intel, and Motorola were all using ceramic substrates for wirebonding. Intel was the bold company to strike out and develop laminate substrates for flip chip packaging (Fig. 1.12).

The author worked with AMD, DEC, HP, LSI, and Motorola to develop the first Kyocera supplied flip chip ceramic packages. Intel had openly announced their intentions to their incumbent ceramic suppliers, Kyocera and NTK, and asked them to develop laminate flip chip substrates. At the time, the task to create a reliable laminate package with a much greater TCE mismatch to silicon seemed a huge development challenge. So much so that the incumbent suppliers decided to defer developing flip chip laminate technology. Instead they focused on the other previously named companies to create the first commercially available ceramic substrates aside from IBM. Even for ceramic technology this challenge was a significant undertaking. The author recalls the qualification of the AMD K6 MPU with a Kyocera substrate, where dozens of development engineers had been assigned to the project for well over a year. During the 2-year program, we were successful in launching flip chip ceramic substrate technology to the MPU and ASIC marketplace. The new business potential and revenue exceeded hundreds of millions of dollars.

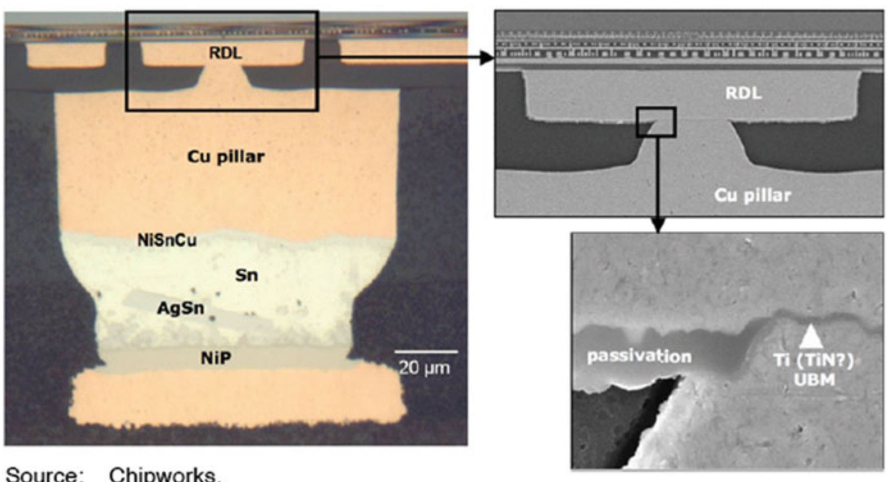


Fig. 1.12 Intel's Copper Pillar bump used in Atom processor

During this timeframe, Intel was successful bringing up laminate substrate suppliers such as Ibiden to produce the first laminate based flip chip packages. Once this had been demonstrated in the marketplace, the traditional ceramic suppliers also began development in advanced laminate build-up substrate technology and today are among the top advanced suppliers.

So now that the substrate infrastructure had been created among ceramic and laminate supply suppliers. The next challenge was to advance and reduce the cost of the bumping technology. IBM's evaporative bumping was expensive and due to using metal stencils limited in its ability to reduce pitch much below 250 μm. Intel once again among other independent companies developed electroplated bumping. Electroplated bumping abandoned evaporative deposition in favor of traditional wafer processing technologies. Electroplated bumping utilizes a sputtered "seed layer," previously discussed, to create a wafer coated electrode that can be subsequently used for electroplate a solder bump. Electroplating processes opened the possibility of other solder compositions. Previously there was only 97/3 Pb/Sn and now there was 95/5 Pb/Sn and 63/37 Sn/PB also available.

The bumping technology was still essentially captive to these initial IDM's, for flip chip to become more widespread other subcontract sources would be needed. This market need created opportunities for the first independent bump contractors such as Flip Chip Technology and Unitive Electronics. Unitive had created an advanced electroplated High Pb and Eutectic bump using licensed technology for the Microelectronics Center of North Carolina (MCNC). Flip Chip Technology licensed screen printing technology from Delphi Electronics for eutectic bumping. The author sensing once again the need for the subcontract bump supply chain joined Unitive Electronics in 1999. In many ways this was a do over for the business and technology development previously performed at Kyocera. The same

Microprocessor and ASIC companies wanted subcontract bumping services aside from their internal capability and thus created the wafer level bumping industry.

So now there were basically two bumping technologies competing in the marketplace; electroplating and screen printing. Evaporative bump technology limitations were well known by this time, so new players establishing internal capacity or subcontracting were turning to these two new offerings.

At first screen printing had the perceived advantage of being a low cost process and relatively easy to adopt. Screen printing had been well known in the packaging industry for many years and compared to complex electrochemistry seemed the logical first choice.

Circa 2000 the big industry question was whether bumping would be the domain of the semiconductor fabricator, the Semiconductor Assembly & Test (SATs) packaging manufacturer, the stand alone bumping manufacturer or even the Electronics Manufacturing Service (EMS) companies. Well the history is known and the packaging companies began licensing and investing in wafer bumping. Up to this time, they had been servicing the outsource wirebond packaging market. When these first IDMs brought flip chip packaging internally they were losing a segment of their packaging business. As first they focused on developing flip chip package assembly processes using wafer supplied with bumps from their customers. They also began collaboration with the upstart wafer bumping companies namely Flip Chip Technologies and Unitive Electronics. The author licensed Unitive's electroplating technology to Amkor Technology. Amkor also at the same time secured a license for screen printed bump technology from Flip Chip Technology. It was not Unitive's business model to become a licensing company so there were no other licenses granted other than Amkor. Unitive needed a second source in the marketplace for its customers to adopt its technology so that the overall market could be expanded and chose Amkor as that source. On the contrary, Flip Chip Technology adopted a licensing model and licensed all the major packaging subcontractors such as Advanced Semiconductor Engineering (ASE) and Silicon Precisionware (SPIL) in Taiwan.

In actuality Amkor's first bumping line in Korea was set up using screen printing. The author believes the earlier traction in broad licensing of the screen printing technology influenced the decision as to which technology to implement first among other considerations. Circa 2002, with the large SAT players investing heavily in flip chip packaging and bumping technology, this was the springboard for larger adoption of flip chip interconnect.

While IC and systems requirements are clearly driving a rapid conversion to flip chip interconnection, a number of ancillary capabilities emerged that supported rapid adoption:

- High-Density interconnect (HDI) or microvia organic substrates address the high pin count density requirements.
- Increased assembly outsourcing trends have shifted the capacity and industry infrastructure balance of power to the SATs. These SATs suppliers had embraced Ball Grid Array (BGA) and Chip Scale Packaging (CSP).

- Surface mount technologies such as BGA and CSP are easily suitable to flip chip adoption for the first level interconnection.
- Development of large industry database showing that high lead and eutectic solder bump first level interconnection is reliable in BGAs.

Another item that favored outsourcing, for the large vertically integrated IC companies that already had captive flip chip bumping, is first generation bumping technology obsolescence. As mentioned companies such as AMD, HP, Motorola, etc. which have licensed IBM's evaporative deposition C4 technology are now realizing this technology's limit for fine pitch bumping (<225 μm). These companies must now make the decision as to whether spend large sums of R&D and capital for electroplating technology, which will provide the solution for fine pitch (<200 μm), or outsource.

This trend of outsourcing to gain a time advantage for new technology adoption along with the benefits of savings on capital investment and the cost of manufactured products had become the norm for the previously large vertically integrated companies.

As the SATs companies invested in capacity and developed a lower cost high volume manufacturing infrastructure, this enabled flip chip packaging to be adopted by other market applications. Flip chip remained a comparatively high end performance package though the mid 2000s. By this time all MPUs, advanced ASICs and Field Programmable Gate Arrays (FPGAs) had adopted flip chip. Laminate substrates had also become the mainstream package. As a large percentage of flip chip packaging cost is the substrate itself, approximately 50 % or more, therefore lowering the cost laminate substrate was a necessity. Many laminate substrate suppliers from Japan, Korea and Taiwan have created the scale and competition to drive down the cost of the substrates.

1.6 Flip Chip Market Trends Covering C4

The proliferation of flip chip packaging as a mainstream packaging interconnect technology has occurred within the last 8 years or so. Up until this time flip chip was actually considered to be a package type versus an interconnect methodology. This were flip chip ball grid array (FCBGA) packages manufactured mostly using laminate technology in a singulated format using package boats for the assembly process. These were limited to the high performance IC segment.

The following graphic depicts the "Application Space" for Flip Chip:

The trends for traditional FCBGA packages are as follows:

- Bump Pitch
  - Reduction in bump pitch to enable increasing I/O densities
  - Pitch trend migration (250 ⇒ 225 ⇒ 200 ⇒ 180 ⇒ 150 ⇒ 140 ⇒ 125 μm)

- Solder Bump Deposition
  - Evaporation ⇒ Screen Printing ⇒ Electroplating
- Bump Solder Composition
  - High Lead ⇒ Eutectic ⇒ Lead Free (Sn.Ag) ⇒ Cu Pillar @ $<$125 μm
- Package Composition
  - Ceramic ⇒ HDI Laminate ⇒ Prepreg Laminate ⇒ Low CTE Laminate ⇒ Coreless
- Package Configuration
  - Hermetic Single Piece Lid (SPL), Non-hermetic SPL, Stiffener + Lid, Bare Die, Overmolded

As traditional wirebond chip scale packages (CABGA) were gaining popularity, they were driving higher density laminate package infrastructure in strip format. As the packages were being reduced in physical size to near chip size, via and line and space requirements and package layer counts were being reduced, these strip based laminate technology was becoming suitable for use with flip chip devices. By lowering the cost of the most expensive element of the flip chip package, the substrate, this allowed for wider adoption potential for other markets. These initial flip chip chip scale package (fcCSP) packages were developed for consumer applications for relatively higher performance applications such as baseband and application processors.

The market fit for fcCSP is as follows:

- Bump (I/O) density relative to die size
  - fcCSP for $>$200 I/O or $>$5.5 mm die size
  - Lower density applications—served better by wlCSP at lower cost
- Low power
  - Typical power $<$2 W (die size dependent)
  - Board level solution possible for high power bare die fcCSP ($>$2 W)
- Footprint
  - 40 nm/65 nm enables smaller die while integration drives more I/O resulting in insufficient area for peripheral I/O cells and substrate wire fan out footprint required for handheld devices
- Price
  - For high I/O small die, insufficient die periphery, Au wire cost, and larger substrate for wire fan-out Can drive price competitive solution in fcCSP

- Overmolding
  - Ease of test and handling, and common form (same as CABGA)

General values for Die Size, Body Size, and Substrate Core Thickness are shown in the following graphic:

- In general 14 mm body and smaller is assembled in strip format fcCSP.
- Today 15 mm and greater body size is assembled in single unit FCBGA.
- In some cases up to 17 mm body makes sense in strip format fcCSP.
- Trend for new devices are using a overmolded fcCSP package.
- Overmolding offers better assembly yields, improved warpage/BGA coplanarity, and ease of handling at board mount.

1.7 Flip Chip Market Drivers Including the Need for More Package Co-design and IC-Package-System Co-design Involving Cu/Low K Devices

As Moores Law continues with the ever advances is Silicon technology, Si Node reduction has been having a more profound impact on packaging technology. Traditionally Silicon development and advancements have been carried out with a disregard for the packaging technology. It has always been the packaging engineers job to make whatever was provided by the Silicon supplier work and doing so without impacting the Silicon design, process, or manufacturing. These times are coming to an end. For todays advanced Silicon nodes pretty much starting with 45 nm and below, chip to packaging interface (CPI) must be considered hopefully upfront. If not upfront, then some retrofits may be needed.

An issue still exists in that the Foundries do not want to lay open their process technology to the packaging community. This is resulting in costly development delays and a back and forth interaction where the packaging engineers point out a Silicon issue. The Silicon provider generally denies any issue, and then subsequently some unknown and unspecified improvement is made and the issue disappears. This had been occurring contemporaneously for Lead Free and Copper Pillar bump technologies for 45 nm and below nodes. As the need for finer pitch bump arrays and Lead Free solutions are occurring, these generally higher stress bump structures and encountering the more fragile and generally weaker Inner Level Dielectric (ILD) layers in the wafer fabrication process. Rarely does the packaging engineer gain access to the Silicon stack-up and related materials and thicknesses. Generally when encountering a new Silicon node, the packaging engineer attempts to apply the standard prior generation Bill of Material (BOM) and assembly processes. In very recent developments on 45 nm Silicon, it is common to see ILD cracking or white bumps. White bumps are the observed white bump shown in Sonoscan imaging highlighting the fact that there is some delamination occurring within the structure.

Through subsequent Failure Analysis (FA) is can be seen the source of the delamination which could be ILD cracking, UBM delamination, bump cracking, or other stress related issue.

An example of white bump delamination observed from Sonoscan is shown below:

An example of white bump related ILD cracking is shown in the following cross section:

Once ILD delamination is observed, the packaging engineer needs to isolate the root cause and determine the appropriate corrective action. Typically dialogue will commence with the Wafer manufacturer questioning the Silicon layer stack-up and related strength. To date the cooperativeness of these engagements has been challenging. What typically occurs is the Silicon provider works on strengthening the Silicon structure whereas the packaging engineer work on assembly processes that are gentler on the Silicon itself. Some adjustment measures for resolving ILD delamination or cracking as follows:

- Copper or Aluminum thickness on top layers
- Number of non-ELK capping layers
- Via density
- Silicon side bump pad opening
- Bump UBM diameter and height
- Addition of a passivation layer under the bump structure such as Polyimide or PBO
- UBM stack-up such as thicknesses and materials
- PCB pad size and Solder Mask Opening (SMO)
- UBM/SMO ratio
- UBM stress
- Die to package ratio
- Underfill material such as High Tg
- Reflow profile such as slow cool down
- Addition of a Lidded solution or other stiffener structural element
- Substrate core material and thickness
- Substrate CTE

The tools exist for package co-design and given a completely open situation as within an Integrated Device Manufacturer, package co-design is possible. A simple flowchart for the tools and processes for co-design is shown in the following illustration:

As traditionally package co-design was not required to find a solution, the need to change the methodology did not exist. Now that more and more companies are dealing contemporarily with these issues, hopefully the behaviors will change. There are no technical reasons why co-design cannot be performed; they are more business or managerial hierarchy related.

A way to address the managerial hierarchy issues is to create cross functional teams that include the packaging engineers upfront in the Silicon design phase. Additional characterization of the Silicon structural elements is also desired.

Although Finite Element Modeling (FEM) is often utilized for packaging analysis rarely if ever do these models include detailed Silicon Layer stack-ups. Many times obtaining this information is difficult as the Silicon manufacturer considers this information highly confidential and proprietary.

1.8 Migration of Flip Chip from IDM's to SAT's Providers

As mentioned earlier in this chapter, flip chip packaging was the domain of the IDM's up through the late 1990s. The SAT providers began their investments into creating flip chip packaging in the late 1990s with initial emphasis on packaging supplied bumped wafers. In the following years the SAT's began heavy investment in Bump and WLP technology.

In my own estimation, the Bump and WLP investment is a challenging but necessary investment for the SAT's. As compared to typical packaging CAPEX, bump is a large investment that exceeds hundreds of Millions of dollars for large scale capacity. On a historical basis for the last 5 years, flip chip CAPEX investment which includes bump can be as much as 50 % of the total SAT's CAPEX in any given year. Once the trend and continuity of these large investments was established it was logical that the IDM's no longer needed to invest to keep pace. In fact for an IDM making the foray into flip chip within the last few years, they did not need to make an investment at all. They could just access the existing and built up capabilities of the SAT's.

Intel continues to invest heavily in flip chip manufacturing and development, and even Intel has outsourced their Chipset business to the SAT's companies. IBM continues to have scale flip chip packaging capability in Bromont, Canada but also outsources a good portion of their business. The SAT's are a major player in flip chip packaging and the trend is to continue for them to grab more and more of the total available market.

Intel was the first high volume manufacturer for Copper Pillar bump based flip chip packaging. Intel's Copper Pillar bump technology used in the Atom processor is shown in the following micrograph:

Amkor was the first SAT to introduce Fine Pitch Copper Pillar packaging and ramp production in Q1 2010 in collaboration with Texas Instruments, see Fig. 1.13 for an example. This technology significantly advanced the pitch reduction of Flip Chip to 40 μm using a multiple row die perimeter design methodology with core bumps. This is an interesting example of advanced packaging that was designed to be low cost and was first deployed for the consumer wireless application space. Typically one might historically expect this to be deployed in the high-end Flip Chip market such as CPU's, Server's or FPGA's. Fine pitch flip chip packaging (FPFC) (i.e., less than 60 μm pitch) is an emerging technology for high speed portable devices, such as application processors in mobile phone, to meet the demands for both smaller form factors and lower cost products. A key advantage for FPFC devices is they do not need a

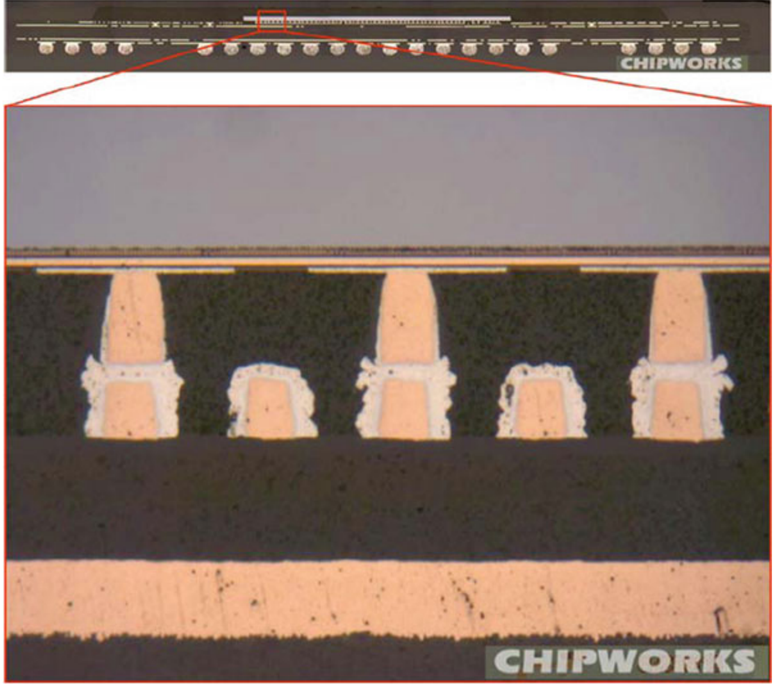


Fig. 1.13 Fine Pitch Copper Pillar in Texas Instrument's XAM3715 application processor

redistribution layer (RDL) on the wafer. Therefore, current die design and manufacturing infrastructures may be used. This enables peripheral inline pitch die to be applied directly to flip chip applications [5].

Rapid growth of copper pillar flip chip is expected as the FPFC technology Amkor had shipped over 100 million units in late 2011 and there are multiple customers adopting this technology. This success has created momentum for use of copper pillar in traditional area array flip chip packages in both BGA and CSP form. Fig. 1.14 below shows the rapid growth of copper pillar bumping demand forecast from 2009 upon its initial adoption through 2014, showing a sevenfold increase in that timeframe.

At the printing of this book, I believe it is accurate to characterize the SAT's as performing a significant proportion of flip chip development and production. This trend will continue as the SAT's investments have been made and most all future packaging technology advancement will utilize flip chip as an interconnect method. Whether it is any one of many 3D packages, Thru-Silicon-Via or Wafer Level Fan Out they all utilize the building blocks of the SAT's investment in flip chip or wafer level packaging.

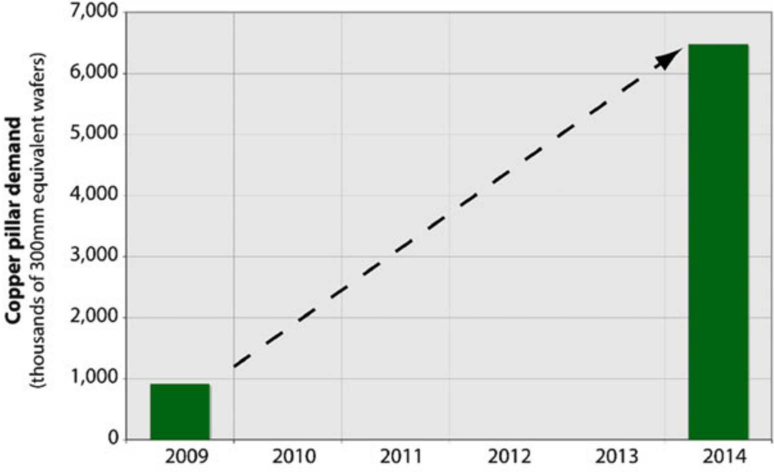


Fig. 1.14 2010 Tech Search Flip Chip Report Copper Pillar Bump Demand Forecast. *Source*: Tech Search International, Inc.

1.9 Implications of the Advent of Tighter Environmental Regulations Covering Underfill, Solder, Structural Design, etc.

Although Lead Free bumping and Lead Free packages have been available since 2004, widespread adoption has been fragmented by the market application. Generally consumer and handheld devices made the transition to Lead Free. Whereas large die singulated FCBGA’s have remained High Lead or Eutectic bump solutions. There have been many Lead Free mandate deadline which have come and gone. Extensions of flip chip packaging to remain leaded have pushed out the full migration to Lead Free.

1.10 Mounting Cost Pressures and Implications for Flip Chip, etc.

Flip chip has always been an exciting packaging technology, but its true growth to a mainstream packaging technology has been limited by its cost as compared to traditional wirebond packages. This limitation has been consistently been eroded away. Flip chip in strip format has significantly reduced the cost structure. As the Laminate substrate has been the largest part of the cost of goods sold, but reducing the Laminate substrate costs this is the most effective way to lower flip chip packaging comparative cost.

In addition, for FPFC design, Amkor has performed numerous studies on converting existing area array flip chip designs to fine pitch designs. In 80 % of the studies conducted, a fine pitch perimeter design resulted in a lower cost substrate due to metal layer count reduction and/or body size reduction. By lowering the cost of the most expensive element of the flip chip package (the substrate), this allowed for wider adoption potential for other markets [5].

References

1. Chip Scale Review (2001) Wafer bumping tutorial by Robert Lanzone
2. Amkor Technology Inc. various Company Technology presentations
3. Chipworks tear down reports
4. TechSearch International Inc (2010) Flip chip report
5. ECTC 2011—Next Generation Fine Pitch Cu Pillar Technology—Enabling Next Generation Silicon Nodes Mark Gerber, Craig Beddingfield, and Shawn O'Connor, Texas Instruments Inc., 13532 N. Central Expressway, Dallas, TX 75243, USA; Min Yoo, MinJae Lee, DaeByoung Kang, and SungSu Park, Amkor Technology Korea, 280-8, 2-ga, Sungsu-dong, Sundong-gu, Seoul 133-706, Korea; Curtis Zwenger, Robert Darveaux, Robert Lanzone, and KyungRok Park, Amkor Technology Inc., 1900 South Price Road, Chandler, AZ 85286, USA

Chapter 2
Technology Trends: Past, Present, and Future

Eric Perfecto and Kamalesh Srivastava

Abstract Since the invention of flip chip technology by IBM about 40 years ago, there has been a continuous need for increased I/O density. More recently fine pitch technology is being enabled in Pb-free through Cu pillar and Sn–Ag solders. Stiffer Pb-free interconnection coupled with fragile low-k dielectric materials imposes a significant challenge on first level packaging. In response to increased number of interconnections and higher performance needs, additional technologies are emerging, such as the following: fine pitch flip chip (<60 μm pitch) interconnections, 3D with and without TSV's, liquid phase connections, and bond-on line. This introductory chapter covers these technologies and sets the stage for current and future flip chip technologies discussed throughout the book.

2.1 Evolution of Flip Chip Technologies in Response to IC and System Technology Trends

Flip Chip through solder connection between the semiconductor chip and the first level package using controlled chip collapsed connection (C4) was invented by IBM and it has been practiced at for more than 40 years [1]. In 1970, IBM started evaporating the UBM and the C4 Solder through a metal mask (Fig. 2.1). This process has evolved to finer and finer pitch utilizing a variety of solder deposition methods. Figure 2.2 depicts the various elements for the traditional first level packages. The first level package function is to redistribute the wiring to the printed wiring board (PWB) pitch while maintaining electrical signal integrity, and to provide a cooling platform for high-end ASICs applications. Flip chips were

E. Perfecto (✉)
IBM Corporation
e-mail: perfecto@us.ibm.com

K. Srivastava
IBM Corporation, retired

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_2, © Springer Science+Business Media New York 2013

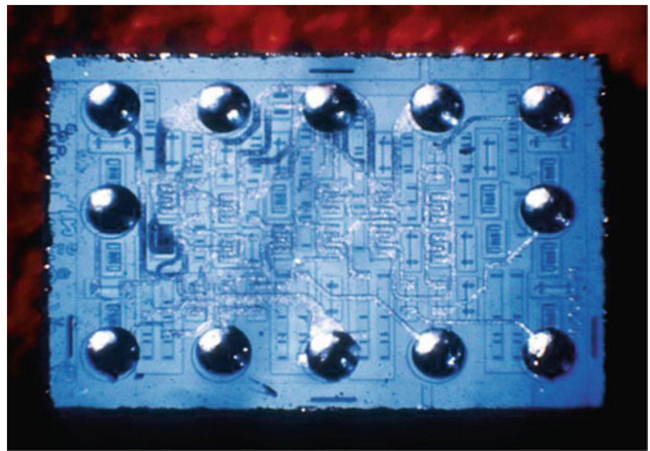


Fig. 2.1 High Lead solder on an IC (IBM, 1970)

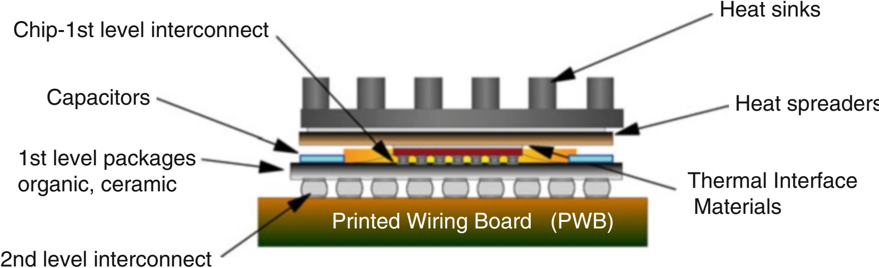


Fig. 2.2 First level package elements

initially joined mainly to ceramic substrates utilizing leaded solders. The introduction of surface laminar circuit (SLC) enabled the organic carriers to increase the wiring density and pitch making it the preferred chip carrier [2]. Low cost organic laminates are the main substrate base for single chip and dual chip modules.

A change from high Pb solders to eutectic solder (Pb63 %Sn) allowed the low temperature joining needed to implement flip chip on organic laminate carriers. However, to maintain mechanical and electromigration (EM) reliability, the usage of high Pb solders continues on the chip side while eutectic is selected for the laminate. Capacitors were subsequently added to either the top or bottom surface of the organic laminate. Underfills are applied after chip joining, and before module test.

The miniaturization in the chip ground rules has resulted in an increase of electrical resistance and the need for innovative thermal solutions to dissipate the heat. High-end processors utilize a thermal interface material (TIM), heat spreaders, and a heat sink. Junction temperatures of ~85 °C are now common.

The second level packaging refers to the module connecting to the PWB. It is here, at the board level, that all the peripheral are connected resulting in a final product.

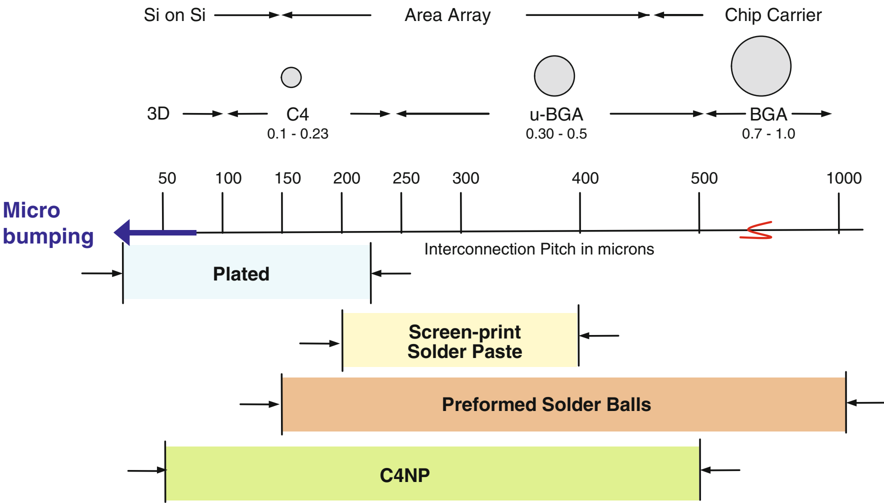


Fig. 2.3 Solder application for flip chip and module joining in first level packages

3D technology, as it will be discussed later, is enabling more and more peripherals to migrate to the chips and/or modules. Organic chip carriers are attached to the PWB through a solder connection, but at a much larger pitch, referred to as the Ball Grid array or BGA. The BGA pitch has been maintained around 1 mm but laminate technology improvements have driven the BGA pitch to 0.8 mm pitch and lower on advance products. A typical C4 will see at least four reflows: Initial solder attach, post test reflow (optional), C4 to laminate joining, BGA attach and BGA joining to PWB.

Wire bonding has been the lower cost technology for connecting the chips directly to the PWB. Its main limitation is that all the connections are through the chip perimeter, requiring a larger-than-chip area on the PWB surface. As the number of interconnection increased, more rows of wires are needed surrounding the chip.

Wafer level chip scale packages (WL-CSP) rearrange, through redistribution and solder connections, all the wire bond connections to an area array, thereby utilizing only the chip area on the PWB surface. The WL-CSP solder pitch is between 0.3 and 0.5 mm, with underfill as an option depending on the chip size and the component reliability requirements. Initially, eutectic Pb–Sn solder joints were used, but now Pb-free solder connections are the norm. There are multiple processes to deposit solders for area array connection. Figure 2.3 depicts solder deposition method base on the required interconnection pitch. Table 2.1 shows the four main technologies used for depositing solder. For completeness the table also include: anisotropic conductive adhesives (ACA), laser ball bump and Au Stud bumping.

The lowest cost deposition method is solder screening, but due to the large solder particle range, its coplanarity is poor, particularly at <200 μm pitch. Process improvements in solder screening volume control include changes in the screen mask (from a reusable metal mask to a disposable photo mask), and in a reduction of the solder particles grid size. Solder screening is the preferred method for

Table 2.1 Types and applications of Flip Chip joining-Test: Probers, testers, interfaces/hardware, etc.

	Applications	UBM required	Composition flexibility	High # of C4s	Volume control	Flux required	Propensity for voids	Underfill required
Ball drop	BGA and CSP	Y	Y	N	Y	Y	Y	Y
Laser ball bump	Sensors	Y	Y	N	Y	N	N	Y
Solder screening	PWB components	Y	Y	Y	N	N	Y	Y
C4NP	Flip chip, 3D	Y	Y	Y	Y	N	N	Y
Plating	Flip chip, 3D	Integrated	N	Y	Y	Y	N	Y
Au stud bump	Sensors	N	N	N	Y	N	N	N
ACA	Displays	Y	Y	Y	N	N	Y	N

depositing solder on to the laminate surface. When very large solder connections are needed, ball drop is the preferred method. As with Ball drop and solder paste, C4NP (developed by IBM in collaboration with Suss Microsystems) can deliver unique solder compositions with more than two metals [3].

2.2 Evolution of First Level Packaging

A series of technical challenges surfacing simultaneously are influencing the materials and process selection for the first level interconnection, i.e., increased power and thermal dissipation requirements, architectural shift to multi-cores giving increased I/O count and current density, and increased interconnection density.

2.2.1 Thermal Demands

Bipolar technology was the main semiconductor technology in the 1980s. In 1990, IBM ES9000 server system using bipolar technology required water cooling to support the ~13 W/cm^2 module heat flux (Fig. 2.4) [4]. With the movement to CMOS technology, the cooling requirements were reduced significantly. CMOS integration enabled full function integration into a single chip. However, as the semiconductor scaling continued, the current leakage generated by the increased wiring length and density (passive power) has increased exponentially and now, it has surpassed the active power in some high-end processors (Fig. 2.5) [5].

The increased power is affecting both thermal management and power delivery. The high junction temperatures coupled with the Joule heating in organic carriers is giving rise to EM concern for first and second interconnect levels, particularly in Sn-based Pb-free systems.

2.2.2 Increased Chip Size

The increase of bandwidth has resulted in an increase of I/O count requirements which in turn have decreased the solder pitch, from 200 to 150 μm. In addition, the chip size is increasing significantly to accommodate increase in functionality. A 20 × 20 mm chip size could be a common occurrence for high performance applications in the near future.

As chip size increased, underfill was introduced to compensate for the thermal mismatch between the Si die and organic carrier, enabling larger chips which otherwise would have failed thermal cycle testing. Organic laminates have a high coefficient of thermal expansion (CTE) of about 18 ppm compared to the 3 ppm of the Si chip. The underfill also protects the solder from corrosive environment.

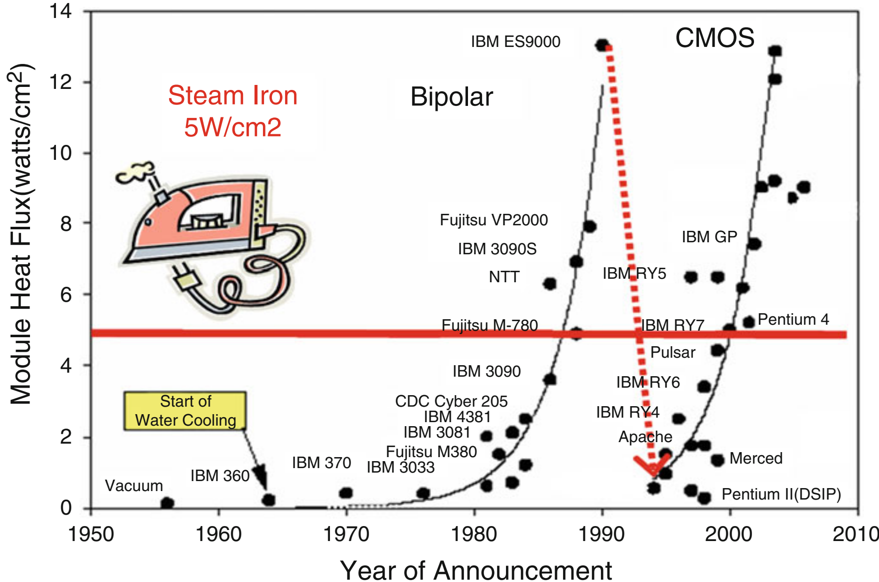


Fig. 2.4 Evolution of heat flux per module as a function of time

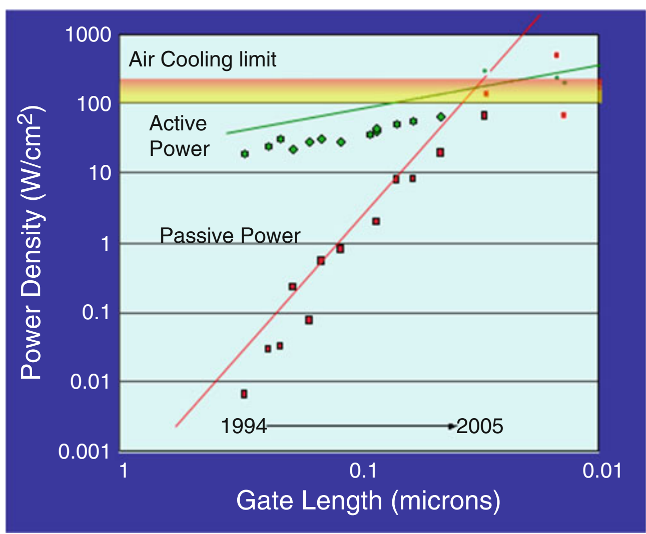


Fig. 2.5 Active and passive power effects as a consequence of scaling

A module sonoscan can confirm a void-free underfill, or crack-free chip corner. Underfills are available with a variety of CTE, Tg and modulus properties, which can be optimized for a particular application. These underfill properties are interrelated since CTE is controlled by the quantity of fillers which affects the material modulus.

2.2.3 Restriction of Hazardous Substances

The green compliance is managed by control of Hazardous Substances and Halogens in electrical or electronic products. RoHS [6], *Restriction of Hazardous Substances*, originated in the European Union in 2002, restricts the use of lead, mercury, cadmium, hexavalent chromium, and PBB's (polybrominated biphenyls) or PBDE's (polybrominated diphenyl ethers) as fire retardant, in new electrical or electronic equipments entered into force in February 2003. These directives came with provisions to exempt certain uses of otherwise restricted substances in specific cases where the negative effects were likely to outweigh the environmental, health, and/or consumer safety benefits of the substitution and requirements to review every 4 years to determine whether there is valid scientific data to support the continuation of each exemption. These restrictions and exemptions have been applicable since July 1, 2006. Following decision tree can be used determine if an electrical or electronic equipment is a subject of RoHS Directive (Fig. 2.6) [7].

Lead in solders for Flip Chip applications is the primary concern in electrical and electronic equipments since they are not expect to exceed the threshold of 1,000 ppm for mercury, hexavalent chromium, PBB, and PBDE and 100 ppm for cadmium.

Lead solder related exemptions include (4 year review cycle, next review in 2012)

- Spare parts for the repair of equipment that had been put on the market before 1 July 2006 and the equipment that was put on the market before the same date.
- Lead in high melting temperature solders (i.e., tin–lead solder alloys containing more than 85 % by weight or more lead)
- Lead in solders for server, storage and storage array systems, network infrastructure equipment for switching, signaling, transmission as well as network management for telecommunications
- Lead in solders to complete a viable electrical connection between semiconductor die and carrier within integrated circuit Flip Chip packages.
- Expiration dates for the above exemptions are expected to range from 2013 until July, 2014 with no grandfather clause, although "repair as produced" is expected to be approved for upgrades and field repairs.

High melting temperature solders containing >85 % lead are often used for high reliability flip chip bumps, some soldering processes for ceramic packages, and die attach for high power, wire bonded devices. Current flip chip interconnections using Pb–X% Sn ($X < 15$) with ceramic substrates and using Pb–3 %Sn with Pb–67 %Sn solder on organic laminate meet the criterion of exemption from RoHS [8]. However, all BGAs have transition to Pb-free solders.

Lead bearing solders have been the primary alloys for the assembly processes used within electronic equipment for decades. The quest to replace these solders with lead-free alloys is a tremendous task that has been undertaken by the electronics industry. It was recognized that to require this transition to lead-free solders prematurely in high performance, high reliability electronic equipment such as servers, storage, storage arrays, network infrastructure and telecommunications equipment that perform critical tasks could pose a health and safety risk to the public. These systems are

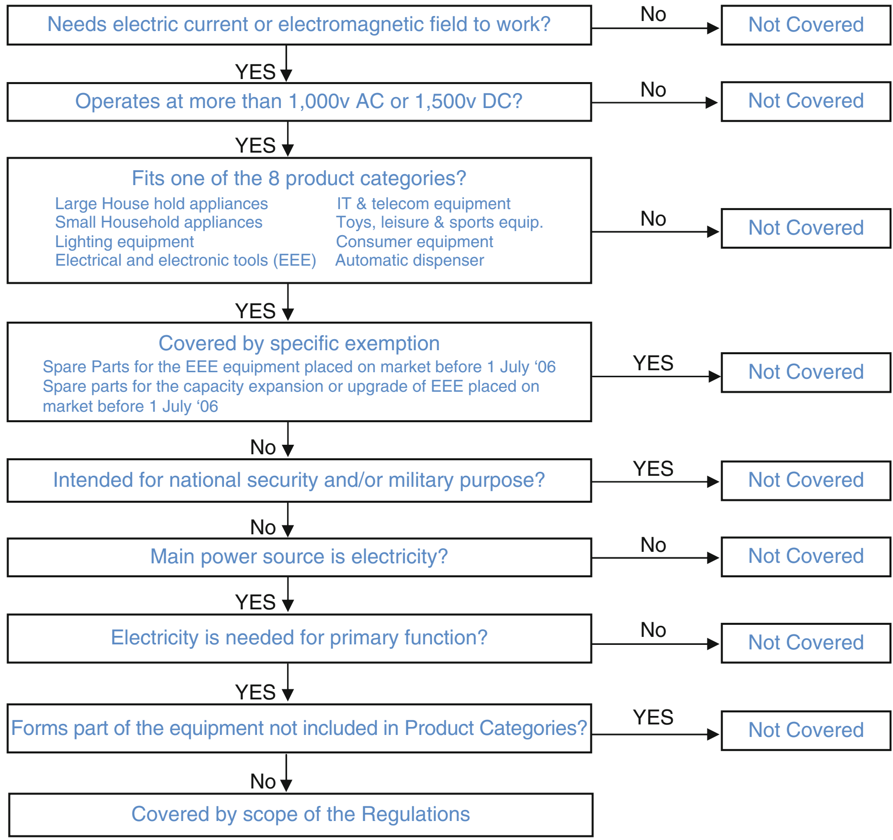


Fig. 2.6 Decision Tree for RoHS Pb-free regulations

expected to be in operation 24 hours a day, 7 days a week for at least 10 years with very little outage time during these years of operation. Until Pb-free solders could be proven as reliable for such high performance, high reliability applications, the need for the continued use of traditional lead–tin solder has been argued [8].

Two concerns have been identified as primary focus for development of Pb-free solder for flip chip interconnection; first, lack of adequate resistance to EM under severe operating conditions and second, compatibility with the fragile back end of line structure of the device. These two topics will be covered in detail in subsequent chapters.

2.2.4 Compliance Cost, and Future RoHS Directive

Initial compliance is estimated to have cost industry $32 billion, with a further $3 billion needed annually to maintain compliance, according to a study by

Technology Forecasters Inc (TFI) [9]. The Commissions own study puts the average overall cost related to RoHS at 1.9 % of the turnover. TFI survey also revealed 29 % companies lost sales due to RoHS (average loss US$1.84 million), 2/3 of these came from delayed new product or discontinued EU sales. In an attempt to reduce costs, resolve uncertainties increase market surveillance and enforcement, the European Commission revised the RoHS Directive in December 2008—with the objective "to develop a better regulatory environment, one that is simple, understandable, effective and enforceable."

The other component of green compliance comes from control of halogens, primarily bromine bearing compounds, such as PBB's, PBDE's, and TBBA's (Tetrabromobisphenol) used as fire retardants in a plastic package. In electronic products, thermoplastics, which include polycarbonate, acrylonitrilebutadien-styrene copolymer, and polystyrene, are mainly used in housings. Furthermore, thermosetting plastics—mainly epoxy resin compounds consisting of epoxyresin, hardener, and additives—are used as insulating materials for electronic parts.

In order to prevent fire from originating in electronic products, these plastics contain flame-retarding additives, most commonly in the form of organic halogen compounds such as brominated aromatic compounds. There is, however, a serious problem with such halogen compounds; as during burning they generate toxic substances that can injure people and contaminate the environment. In addition to the fire-related dangers, the treatment and recycling of the waste materials is also made extremely difficult.

The acute toxicity of the majority of the brominated flame retardants is low or very low. This is also the case for many of the breakdown products. The important risks of brominated flame retardants are therefore mainly connected to long-term effect; only relevant if the substance or its breakdown products may bioaccumulate. A substance that is fat soluble and stable, and that has a route of exposure to a specific organism may bioaccumulate. The third draft of the European Parliament and Council Directive on Waste Electrical and electronic Equipment, presented in 1999, required that member states phase out the use of PBBs and PBDEs by January 1, 2004.

Environmentally friendly flame-retardant plastics containing no toxic flame retarding additives such as halogen (bromine) compounds and phosphorus compounds have been developed for electronic products [10]. A polycarbonate resin containing a new silicone flame-retarding additive has been developed for use in housings. Furthermore, a self-extinguishing epoxy resin compound containing no flame-retarding additives was developed as a high-quality molding resin for electronic parts. These plastics show good general properties as well as high flame retardation [11].

2.2.5 Choice of Sn

Sn–Ag–Cu with a liquidus temperature of about 220 °C for the eutectic composition has been the solder composition of choice for most of Pb-free device applications.

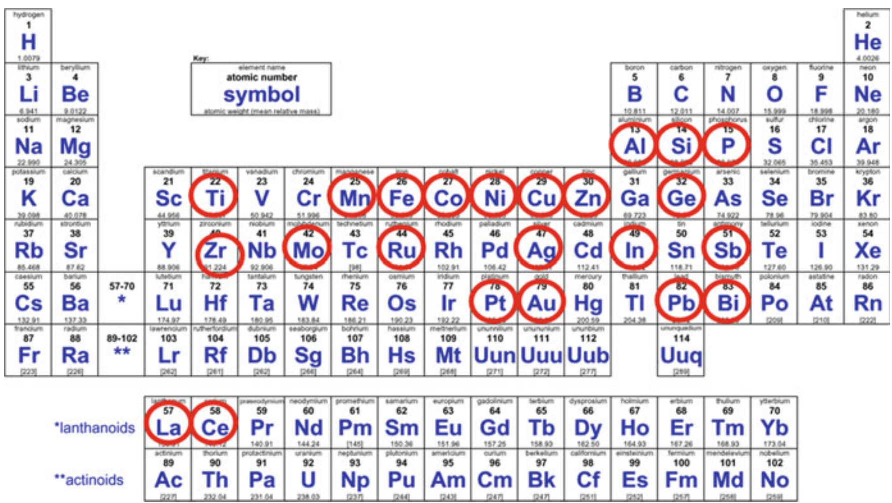


Fig. 2.7 Pb-free solder additive elements for specific applications

The individual composition is dependent on the design factors for a specific application. They include liquidus temperature, elimination of formation of rod-like Ag_3Sn precipitate, reduction in suppression of solidus, relatively low hardness, low undercooling and longer life expectation from the point of EM behavior. Pb-free solder composition effects on undercooling have been reported earlier for a BGA solder [12].

The precipitation of rod-like Ag_3Sn phase during chip-joining process has been observed in alloy containing as low as 2.3 %Ag. The ability of the solder to permit stress relaxation is measured in terms of its microhardness. J. Sylvestre of IBM studied various Sn <2.3 %Ag alloys joined to the laminates with Sn–0.7 %Cu or SAC 305, creating gradation of Ag compositions, and examined suppression of solidus and micro-hardness. Solder joint micro-hardness deceases as the percent of Ag in the interconnected bump deceases. This improvement on stress relaxation also resulted in a decrease in ultralow k dielectric delamination [13].

The quest for improved solder properties in a Sn-based system has resulted in an evaluation of many solder additives. Figure 2.7 shows many of the additives which have been reported in the literature. However, since pattern electroplating is the deposition method of choice for $\leq$150 μm pitch, solder additives are mainly of academic interest for fine pitch flip chip applications.

Finally, Sn Pest is another concern on Sn-based solders. It occurs when by β-Sn phase transforms to the α-phase, and has been reported at -18 and -40 °C requiring several years [14]. The α-Sn is readily crumbled and destroys the mechanical integrity of the solder joint. Sn pest has been observed in Sn–0.5Cu, Sn–3.5Ag, Sn–3.8Ag–0.7Cu, and Sn–3.0Ag–0.5Cu solders [14].

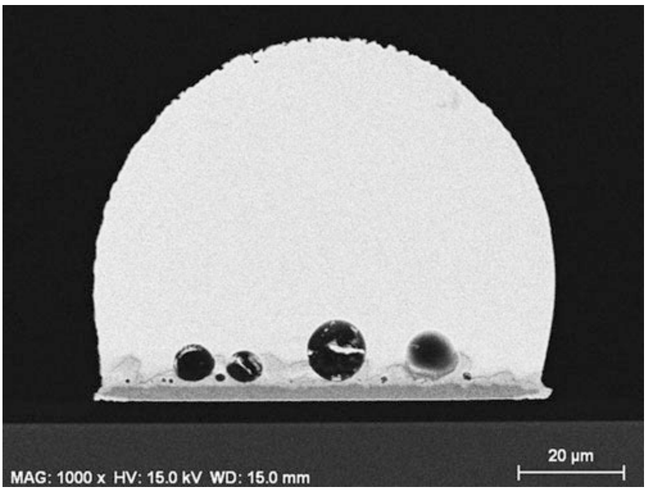


Fig. 2.8 Voids in a plated solder after reflow

2.2.6 Solder Void

Solder voids can be categorized in micro (0.1–1.0 μm) and macro (1–20 μm) scales. The macro voids are a result of entrapment of flux (for solder paste or ball drop) or plated volatiles (on electroplated solders). Figure 2.8 shows an example of a plated solder which was reflowed upside down, resulting in volatile entrapment at the UBM—solder interface.

Kirkendall voids are small voids which form during thermal aging. They start as microvoids within the Cu_3Sn IMC or at the Cu—Cu_3Sn interface when Cu is top layer of UBM, such as Cu OSP in the organic laminate. The Kirkendall voids formation results from a vacancy migration during inter-diffusion when the Cu in UBM and the Cu_3Sn reacts to form Cu_6Sn_5, with no Sn available from the bulk solder. As thermal aging continues, the small voids coalesce to form a crack [15]. Much has been written about Kirkendall voids, but reproduction of results does not always occur, indicating sensitivity to the Cu plated chemistry or Cu deposition parameters [16].

2.2.7 Soft Error and Alpha Emission

A soft error is a signal or datum which is wrong. After observing a soft error, there is no implication that the system is any less reliable than before. Soft error rate (SER) is the rate at which a device or system encounters or is predicted to encounter soft errors. It is typically expressed as either number of failures-in-time (FIT), or

mean-time-between-failures (MTBF). The unit adopted for quantifying failures in time is called FIT, equivalent to 1 error per billion hours of device operation. MTBF is usually given in years of device operation. To put it in perspective, 1 year MTBF is equal to approximately 114,077 FIT. Soft error can be caused by Alpha particles from package decay, Cosmic rays creating energetic neutrons and protons, and thermal neutrons. The mitigation of soft errors caused alpha particle emission of lead containing solder is of extreme importance to the electronic industry.

To give an overview, the problem with ^{210}Pb in solder is that it comes from the radioactive decay of ^{210}Po, which releases an energetic alpha particle (helium nucleus). Although it is easy to remove the ^{210}Po from the solder, the radioisotope ^{210}Pb will decay with a half life of 22.3 years to ^{210}Bi, which, in turn, decays with a half life of 5 days back to ^{210}Po. This point, where ^{210}Po is being replaced at about the same rate that it decays, is called "secular equilibrium" and is reached approximately in 2 years since refinement. Low alpha lead can be obtained from several sources, including "cold" lead ore, laser isotope separation process, and antiquity lead. Alpha emission rate is measured in terms of counts per centimeter square per hour. Lead-rich-solders with alpha emission rate of 0.02 per cm^2 per hour are commercially available for high-end-applications. Lead-free solders have alpha-emission due to trace impurities of Pb or Bi. Tin based Lead-free-solders with alpha emission rate of 0.002 per cm^2 per hour are commercially available for ultrahigh-end-applications. This level of alpha-emission presents challenges to reliable alpha-emission measurement technique, namely, stability of background, sample preparation, sample storage, measurement time, and sample size [17].

2.3 First Level Packaging Challenges

2.3.1 Weaker BEOL Structures

Semiconductor industry's continuous drive towards faster chips and smaller wiring has resulted in an increased capacitance and which drives need for insulators with lower dielectric constant. This is, generally, accompanied with reducing the insulator fracture toughness. Interconnect stress build-up during the cooling part of chip join cycle, due to the CTE mismatch between the device and the organic carrier coupled with the move to Pb-free solder, and increased solder hardness, can exceed device BEOL structure strength for specific design structures, resulting in interlayer delamination or cracking, see Fig. 2.9. The packaging trends, shown in Fig. 2.10, impose new restrictions on Pb-free systems. System optimization is needed to eliminate the stress cracks which occur during the cooling stage at the first chip joining [19]. Increase in polyimide thickness, acting as a cushion, and increase in thickness of interconnect metal Cu, are some of the variables used for increasing robustness of the BEOL Structure.

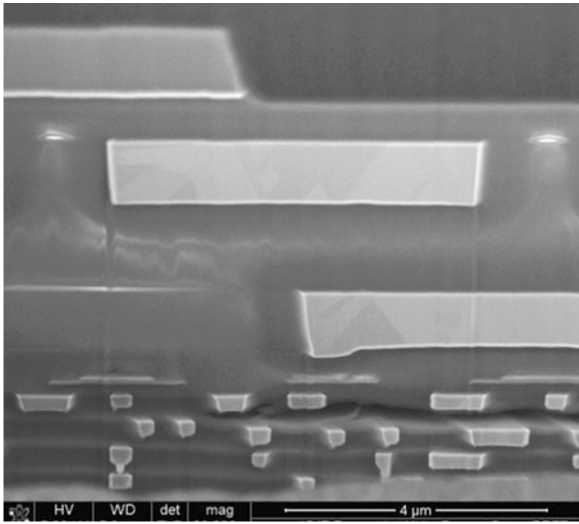


Fig. 2.9 X-section of ULK delamination

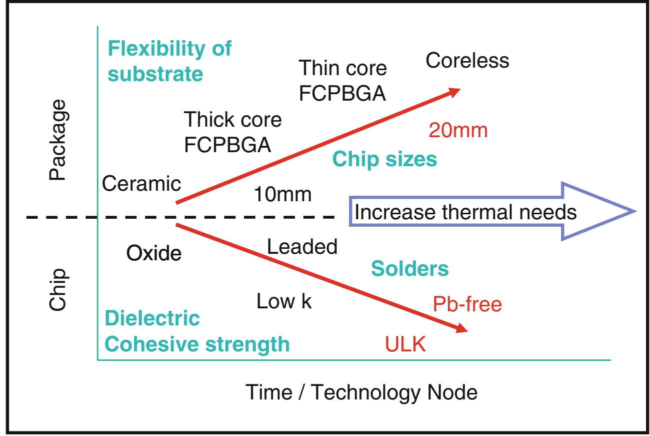


Fig. 2.10 Chip and package trends

2.3.2 *C4 Electromigration*

The EM performance of Pb-free interconnect has been the focus of recent studies due to the continuing demand for higher current density and circuit miniaturization. Different from the face center cubic structure of lead, tin (Sn) has a tetragonal crystal structure and tends to form large grains that exhibit highly anisotropic behavior in mechanical, thermal, electrical, and diffusion properties in the high Sn-based Pb-free

Fig. 2.11 Anisotropic elastic modulus and thermal coefficient of expansion on β-Sn at room temperature [18]

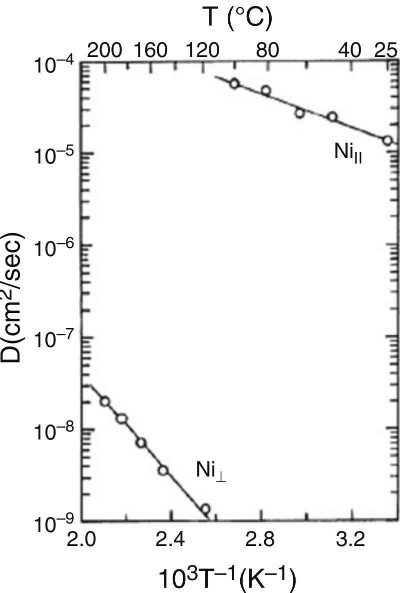


Fig. 2.12 Diffusion of Ni in Sn [20]

solder joint (see Fig. 2.11). Importantly, the noble and near noble metals are extremely fast diffusers in Sn and the diffusivity is highly anisotropic. For instance, the diffusivity of Ni along the tetragonal (*c*-) axis is ~7 × 10E4 times faster than that at right angles (*a*- or *b*-axis) at 120 °C (see Fig. 2.12). Experimentally it has been found that EM damage is strongly dependent on the Sn-grain orientation in Pb-free solders [21]. Figure 2.13 shows a SEM image of an EM stressed C4 with SnCu solder and the insertion of the Sn unit cell indicating the grain orientation. The grain on the right has the *c*-axis at nearly a right angle with respect to the current direction, where the rates of Cu and Ni diffusion in Sn are slow. Failure is characterized by Sn self-diffusion or lattice diffusion resulting in void formation between the IMC and solder,

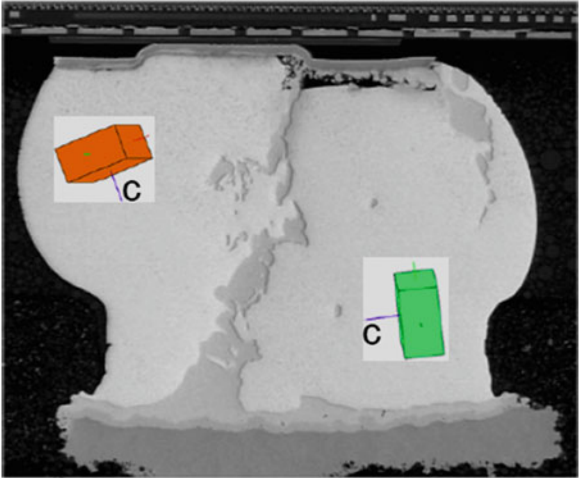


Fig. 2.13 SEM of an EM stressed C4 with SnCu solder failure. Courtesy of Minhua Lu of IBM Corp

referred to as mode-I failure [21]. The grain on the left, however, has its *c*-axis closely aligned with the current direction, which drives very fast Ni and Cu diffusion through the Sn grain. Mode-II failure is due to rapid depletion of intermetallic compounds and UBM. Since Mode-II failure usually occurs early, it is the mode which should be avoided.

In addition to the grain orientation, the solder and UBM interaction and solder alloy composition play an important role as well. Minhua Lu of IBM has reported the average time to failure for SnCu and SnAg(Cu) solders and three UBMs; Cu, Ni(P)/Au, Ni(P)/Cu surface finishes [22]. In general SnAg(Cu) solder performs better than SnCu solders and Ni UBMs are better than Cu UBMs. This study showed that Ag_3Sn intermetallic network is more stable than that of the Cu_6Sn_5 network under thermal and electrical stress, which is attributed to the much smaller diffusivity of Ag compared to Cu. Grain growth and grain reorientation is common in SnCu solder, while a much more stable cyclic twinning structure is more frequently found in SnAg solders, especially in high Ag solders. Although Cu_6Sn_5 is not stable against EM, this study shows that a certain amount of Cu on UBM to form a layer a Cu_6Sn_5 IMC after reflow is important to give an extra protection to Ni barrier layer. Effect of alloy composition such as Ag, Cu and other doping elements and UBM metallurgy on EM is complicated and convoluted.

2.3.3 Cu Pillar Technology

The Cu pillar technology adopted by Intel provides a uniform current distribution in the positive direction (current flowing from the chip to the package) where peak current has traditionally been a concern. The Cu pillar along with the lower volume

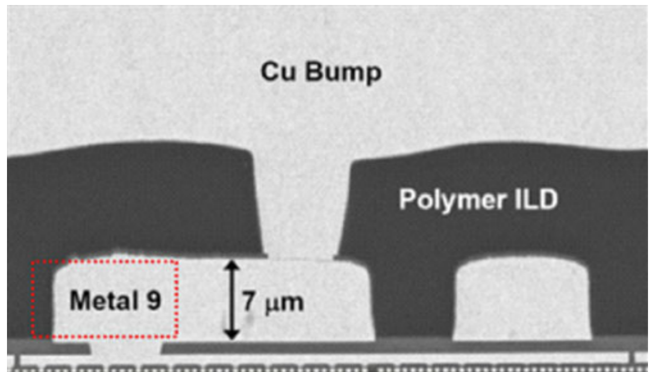


Fig. 2.14 Intel F-BEOL modifications for Cu Pillar technology

Pb-free solder results in a stiffer interconnect which requires a F-BEOL redesign involving three key elements: a 50+ µm of Cu pillar, a thick organic dielectric with a small via diameter, and a M9 metal (8 µm thick in the latest design) for better power distribution (See Fig. 2.14) [23]. Intel has selected not to put solder on the chip side, with all the solder being provided by the laminate side. This requires thicker and uniform solder on the laminate side. So far the implementation of the Cu pillar for leaded and lead-free solders has been limited to less than 15 mm chip size. The short distance between Cu pillar and laminate may improve EM as it approaches the Blech limit.

2.4 IC Technology Roadmaps: More Moore and More Than Moore

At the other end of the spectrum are portable devices, where main considerations are weight, size and cost. Portables mainly encompass handheld devices, such as smart phones, cameras, etc. It is these handheld devices that are driving the packaging innovation, such as WL-CSP (Fig. 2.15), wafer thinning, overmold, etc.

Gordon Moore postulated in 1965 that the semiconductor cost/performance ratio doubles every 18 months, now referred to as the Moore's Law [24]. For the next 40 years this trend was roughly maintained in great part due to the rapid scaling and miniaturization of the semiconductors pitch which has moved from micron-scale to nanoscale while increasing the wafer size from 25 mm to a current 300 mm. The effect of large wafer size and nanoscale lithography has placed a high cost burden in the industry's ability to keep up with the Moore's Law. Currently, a new FAB can cost around $5 billion, and research and development cost have maintained a 10 % increase over the last 10 years. How can this rate of technology progress, which has fueled an electronic revolution, keep up?

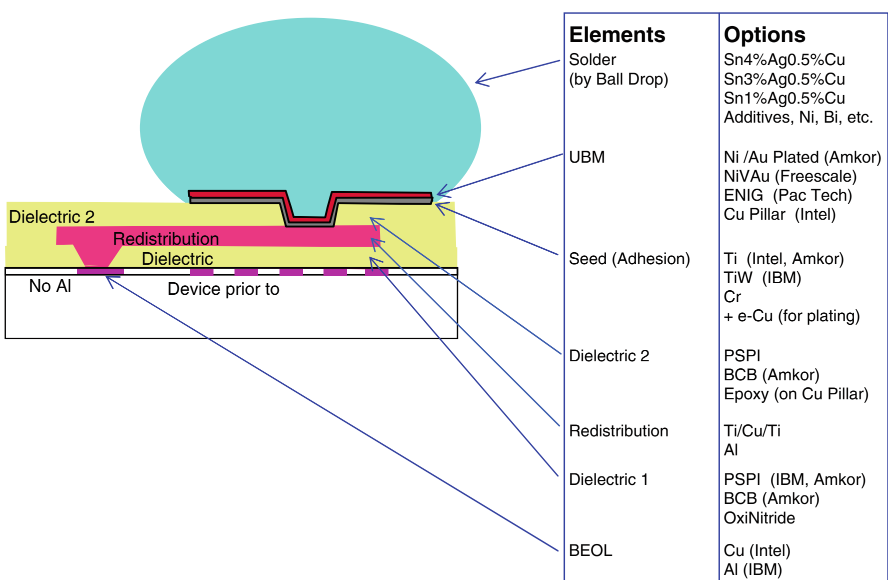


Fig. 2.15 Wafer Level Chip Scale Package elements

Research and Development consortiums are emerging as means to collaborate among companies while reducing the individual R&D costs. Both IBM and TSMC have formed joint development agreement with semiconductor FABs. Additionally, Sematech and SRC in the USA, IMEC in Europe, and ASAT in Japan, provide industry and university collaboration to train the workforce needed for future technology innovation. The European Union "Vision2020—Nanometrics at the Center for Change" is defining the path on how to become a global leader in nanoelectronics. Vision 2020 strategy is to integrate the supply chain with application side. ENIAC (the European Nanoelctronics Initiative Advisory Council), whose acronym was appropriately taken from the first multiple purpose general computer (Electronic Numerical Integrator And Computer), was created to define the means of achieving Vision 2020. ENIAC has proposed five key society areas for concentrated development focus: health, mobility, security, communications, and entertainment [25]. Technology drivers specific to these areas have been identified and grants are available for R&D.

"More Moore" refers to scaling specific technology advancements, in the device or associated structures, such as 3D. "More Than Moore (MtM)" is a new level of integration where non digital devices which were traditionally mounted to the printed-circuit wiring board (PWB) are migrated to package level. This reduction in distance also resulted in a reduction in passive power. Non digital devices such as RF, passive components, sensors and actuators, and power control have also reduced their dimensions, but not at the scale of semiconductor devices. Additionally, packaging improvements such as package on package (PoP), Multichip modules (MCM)

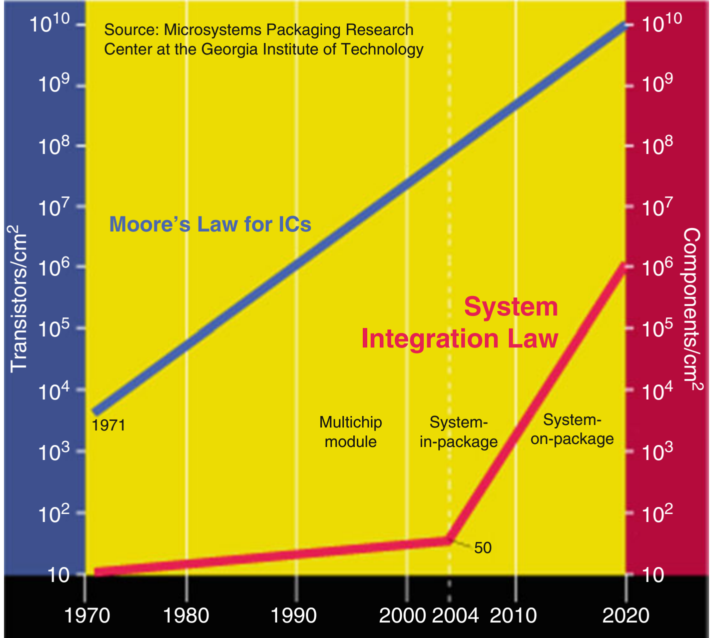


Fig. 2.16 Transistor evolution compared with the packaging integration

and system in package (SiP) have revolutionized system level interconnection. Figure 2.16 from Rao Tummala from GaTech University clearly depicts the effect of packaging integration in MtM, and the paradigm shift that it represents [26]. Packaging development focus is driven by the objective to reduce high module and assembly costs, when compared to the semiconductor devices.

The International Technology Roadmap for Semiconductors (ITRS) is a global organization that works diligently to provide semiconductor roadmaps, and more recently assembly and packaging roadmaps [27]. The ITRS has defined the "More Moore" applications to those where functional and performance scaling are enabled by both geometrical and equivalent scaling. Geometrical scaling refers to the vertical and horizontal shrinkage of features on the chip associated with speed and power advancements, while equivalent scaling refers to 3-dimensional device structures which complements geometrical scaling and enhances the electrical performance of the chip [27]. "More than Moore" integrates the More Moore innovations with the non-scaling technologies (such as analog/RF, passives, HV sensors, actuators and biochips) to produce high value system through functional diversification.

Heterogeneous integration, such as 3D or Multichip modules (vertical or horizontal connections), is providing the needed platform for new automobile, medical, aerospace and consumer applications which are emerging due to the lower costs and/or lower physical size. Software and application integrations are also contributing in improving cost/performance ratio.

Table 2.2 Ground rule evolution for the laminate technology

Item	Class	2006	2008	2010	2012	2014	2016
Min. line width (µm)	A	30	30	20	20	15	15
	B	20	15	15	7	7	7
	C	15	7	7	5	5	5
Min. space width (µm)	A	30	30	20	20	15	15
	B	20	15	15	7	7	7
	C	15	7	7	5	5	5
Land pad diameter for laser micro via (µm)	A	110	90	90	80	80	70
	B	90	70	60	60	50	50
	C	70	60	50	40	40	30

The non-digital devices integration in SiP and SoC applications started first with RF and passives components, followed closely by MEMS and Bio-devices, enabled by WL-CSP and embedded technologies. MEMS applications such as inductors, accelerometers and digital micro devices are being integrated within CMOS technology. By connecting the MEMS directly in the CMOS chip, thinner highly functional structures are achieved for applications such as Integrated RF, finger sensor, gyroscope, etc. [28].

2.4.1 *Improvements in Laminate Ground Rules*

As has been shown, the laminate technology is dominating the chip package industry, and with the assembly houses, it has enabled the WL-CSP revolution. The laminate technology improvements are not only in dimensional scaling as shown in Table 2.2, but it has been able to maintain the electrical parameters, such as impedance control, needed for high performance products [29]. The reduction in linewidth and in dielectric thickness are making the laminate substrate lighter, smaller and thinner, allowing for more function to be transfer in to the packages. One problem which the laminate substrate has is warpage control. Warpage occurs during the chip joining process affecting the laminate coplanarity. On large substrates, non-wetts or no C4 to laminate contact can result due to CTE mismatched coupled by the laminate warpage.

2.5 3D Flip Chip SiPs for Handhelds Require IC-Package-System Co-design

SIP refers to a semiconductor device that incorporates multiple components that make up a complete electronic system into a single package (Fig. 2.17). Electronic devices like mobile phones conventionally consist of several individually packaged IC's handling different functions, e.g., logic circuits for information processing, memory

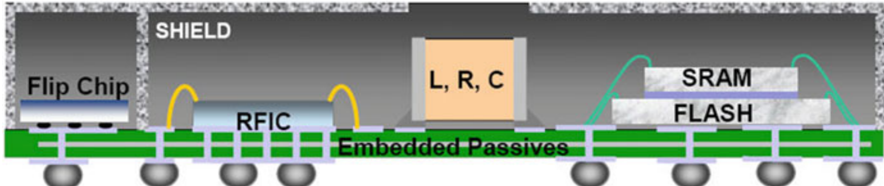


Fig. 2.17 SIP Concept; an Schematic SIP Package containing Chips, RFIC, passives (L, R, C), and memory [30]

for storing information, and I/O circuits for information exchange with the outside world. In a SIP all of these individual chips together with other devices like passives, filters, antennas are assembled into a single package using wirebond or/and, solder interconnect technologies, resulting into tremendous space savings and significant down-sizing of handheld wireless applications.

SIP is different from SoC or "System-on-a-Chip," which is a complete electronic system built on a single chip. SoC's suffer from long development cycle time, low yield, and high development costs; mainly because it is difficult to make an entire system of differently functioning circuit blocks work on a single chip. SoC's are generally, performance driven packages for specific applications such as medical or military. SiP technology, on the other hand, has proven to have lower cost, shorter development cycle, and unified solution to the system architect consisting of smaller system size, increased system density, and lower system power consumption.

SIP's are also designed using a 3D stacking of chips with following elements; thinned dies up to 40 μm thick, adhesive-with-silicon filled in between the chips and between the first chip and laminate, first chip may also be joined with underfill, wire bonding between the chips and flip chip between the first chip and carrier, low modulus mold compound encapsulating the wires, and organic carrier with ball grid array pitch >400 μm, (Fig. 2.18). H. Bottoms from ITRS has summarized classification for SIP structures, as shown in Fig. 2.19.

2.5.1 SIP Engineering Challenges and Co-design Tool

SIP's have engineering challenges like interaction among various stacked dies, variations of die stacking configurations, 3D nature of wirebonds, various configurations of bumps and balls, competing signaling architectures, heat dissipation, and test. Some of these engineering challenges can be addressed by a co-design tool encompassing design of various devices and the package. For example a co-design tool will create a design to address SIP heat dissipation and architecture testing in an economical and efficient way. The co-design tools enable package designers, package design service companies, and offshore assembly and test companies (OSAT) to participate in the multi-die SIP design chain using a

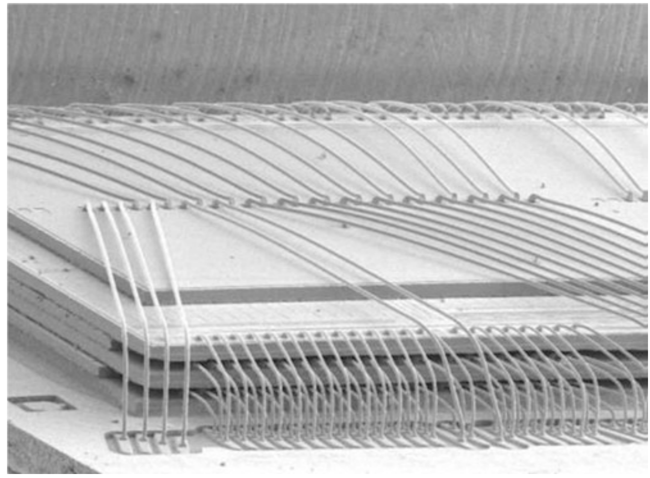


Fig. 2.18 Typical complexity of wire bond interconnections in a 3D chip stack [30]

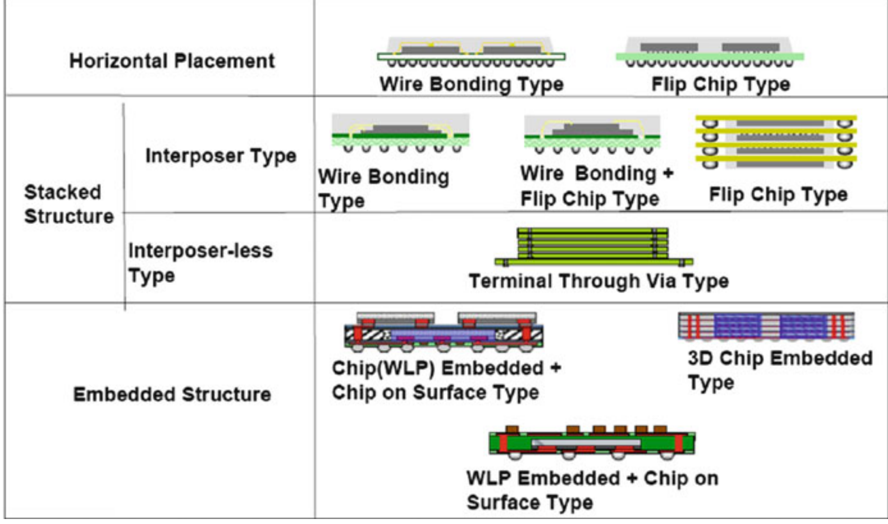


Fig. 2.19 Overview of SIP Categories [31]

co-design methodology. It enables data to be passed easily among design chain partners using co-design technology, such as Cadence [32].

For the test issues, open-architecture automated test equipment (OA-ATE) is considered a cost-effective solution that allows semiconductor manufacturers to specify their own test resource and instrumentation requirements. “Specialization” of test capability requires some standardized vital elements: for example an industry-standard bus structure, compatibility with industry-standard data formats, browser technology to access and control resources, a modular hardware and software

structure to enable re-configurability, and partitioned test supported by ATE and EDA tools.

SIP design allows manufacturers to bring together many IC, package assembly, and test technologies to create highly integrated products with optimized cost, size, and performance. EDA software suppliers provide advance technologies to help SIP design team members get control of the challenges, especially in the areas of co-design, advanced packaging, and RF module design. Successful implementation of SIP manufacturing offers many advantages that are important to the semiconductor industry of the future: shorter time-to-market, lower cost, flexibility, smaller size, etc.

2.6 PoP and Stacked Packages

PoP (Package-on-package) and PiP (Package in Package) technologies provide the least disruptive way to make 3D integrated micro-systems by vertically stacking fully tested prepackaged dies. Figure 2.20 depicts various PiP and PoP modules used in a cell phone application. Each individual package may have a different level of complexity. There are many versions of this 3D integration; wire-bonded-stack, perimeter BGA stack, or u-BGA through a laminate interposer (Fan-in PiP), flip chip C4's and a combination of these versions. One differentiation between PoP and PiP is that in a PoP the top and bottom packages are connected by BGA's, whereas in a PiP the top and bottom packages are connected by wirebond with or without using an interposer. Therefore laminate warpage can significantly affect the PoP connection yield [33]. Currently, 3D packages are widely used in many consumer portable

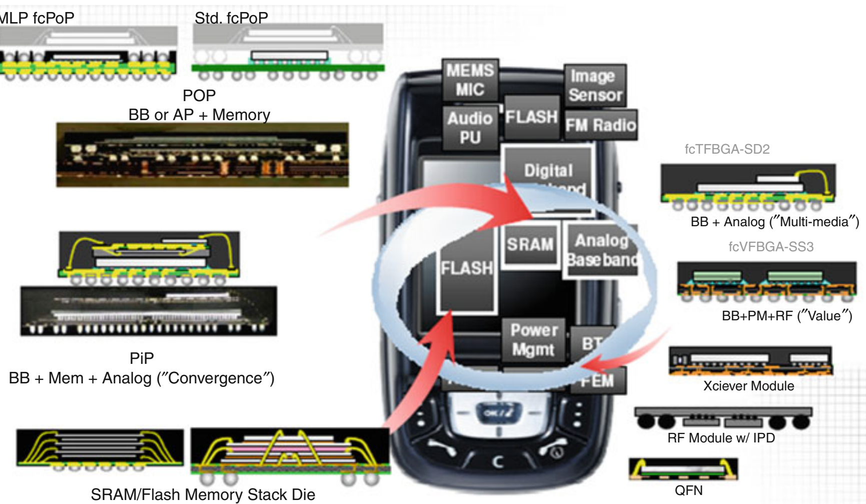


Fig. 2.20 Implementation of PoP and PiP in the cell phone. Courtesy of Raj Pense of StatsChipPac

applications; for example cell phones. Figure 2.20 shows many PoP's and PiP's, used in a cell phone assembly, varying in types, sizes of various devices, and interconnect methodologies used. A system level integration of various individual packages involving initial design validation and optimization in terms of mechanical, thermal and electrical performance is needed. Various software packages, such as, Ansoft SIP System Integration solution, are commercially available for such system level integration.

Stacked packages can contain all same die, as for memory PoP, or heterogeneous devices for graphics, ASICS, gamming, handheld devices where thin/light packages are needed. These dies can be without TSV (Through-Silicon-Via) or with TSV's. Stacked packages consist of either perimeter I/O (BGA) connections, or uBGA connection through a laminate interposer (Fan-in PiP). By using these package as an integration fabric to combine one or more ICs with discrete, embedded, and other packaged parts, a module can be created and used as a standard component for a POP assembly.

From a reliability perspective the PoP and PiP have shown that they can have a ball count of ~300 for 10 $\times$ 10 mm packages and ~600 for 15 $\times$ 15 mm packages and can have a junction temperature, T_j, as high as 110 °C. There is a need to improve/control package warpage to about 60 μm by materials and structural design. ITRS projection for 2014 stacked package include the following; Low cost/handheld 14 die/stack, High performance 5 die/stack, and low cost/handheld 15 die/SiP.

There is a large infrastructure of PoP with 10 OEMs in the portable Multi-media market, over 15 major (logic and memory) IDMs, and it is supported by the full service packaging houses (STATS ChipPAC, Amkor, & ASC)

Common design ground rules are allowing components from multiple vendors in a single PoP. All the flip chip structures discussed so far as characterized as fan-in, where the redistribution layer of the WL-CSP goes from perimeter array to u-BGA area array. A new kind of structure, called Fan-out, has been developed for cases where the number of required u-BGA exceeds chip area. Here the fan out redistribution occurs on molded areas around the chip [34].

2.6.1 Embedded Chip Packages

Embedded Chip technology was first developed by GE for military applications. The main idea of embedded chip packages is to build the laminate substrate around the chips. That is, the chip is placed first on a polymer film followed by a second polymer lamination cover the chip. Then, vias in the second polymer are opened through laser, exposing the metal pads on the chip. A metal wiring is defined which redistributes chip signal to the top of the second polymer film. The polymer lamination, via formation and wiring process is repeated several times, and the package is finished with u-BGA for further interconnection [34]. The alpha particles concerns are eliminated since the chips are surrounded by the laminated polymer. Reduced interconnection pitch is possible due to small via

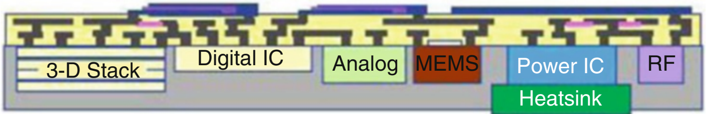


Fig. 2.21 Example of embedded chip package developed by GE

diameter. They are expected to have better EM since there is no solder to package connection.

The main limitation for embedded chips is high cost. Any yield losses in the fabrication of the package results in scrapping the electrically good chip. Additionally, unlike traditional laminate chip carriers which are fabricated on a ~350 × 500 mm panel size, the embedded chips are fabricated on wafer type tools on a 200 mm or 300 mm format. Most reported packages consist of a relative low number of I/O (<500 I/Os). Materials optimization for structural robustness is required to avoid chip, BEOL or organic dielectric cracks.

Commercialization has been slow but it is starting to take traction with cell phone applications where thin/light packages are needed, and high reliability is not required. Additionally, the embedded technology can be mapped with multiple chips per package, either by chips stacks or side by side. Intel published on embedded chips in 2002, but it has adopted the Cu-pillar joined to laminate technology instead. Fig. 2.21 shows a ACICs working test vehicle from GE containing four wiring layers and over 3,000 I/Os. Freescale. Shinko and the Institute of Microelectronics also have on-going development activities. IMEC is evaluating embedded chips on flex packages.

The embedded chip technology is extending to board manufactures with two new consortiums formed since 2007: HERMES, which is led by Fraunhofer and FlipChip, and Embedded Actives and Passives (EMAP) headed by PRC in GaTech.

2.6.2 Folded Stacked Packages

Folded Packages can have multiple flip chips joined to an organic tape which is folded onto itself with the net effect of stacking the chips, and significantly reducing the interconnection area. This represents a space reduction, similar to stacked packages, but the wiring lengths are similar to those in the 2D connection. Figure 2.22 shows a sample of this technology developed by Tessara. The prime features of this technology are described as following; a one-mil (25 μm) polyimide tape is used for low height, and better folding accuracy. The dies are thinned to 114 μm providing a smooth backside with laser dicing to eliminate die cracking. The die-to-die bond-line has a 12 μm thickness, and the back of die-to-tape attach adhesive is set to 25 μm. In a μBGA package with a compliant elastomer layer between active surfaces of die and substrate the 250 μm SAC solder ball results in a ball height of 140 μm after reflow.

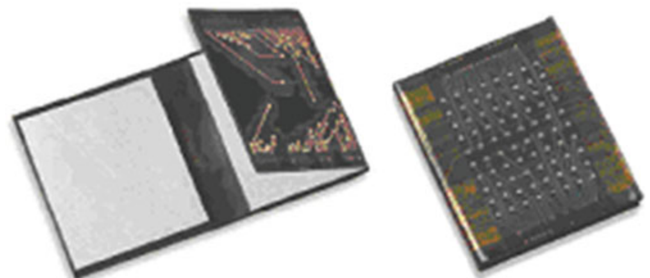


Fig. 2.22 Three dies Folded Stacked Package [36]

2.7 Emerging Flip Chip Technologies

Through-silicon-via (TSV) have reduced the chip-to-chip connection length to the minimum. The TSV length is between 10 and 100 μm, depending on the amount of Si removed during the grinding process. The shorter path translates in speed increase and lower power loss due to less resistance, and to less switching noise due to lower inductance. The typical TSV fill materials are copper or tungsten, with polysilicon being mainly used in memory applications due to its high resistance. Wire bonding used extensively in PoP and PiP, which have a length between 1 and 5 mm, suffer from high inductance in the power leads, is not the preferred interconnection for high performance applications.

The 3D Wafer to wafer Bonding can occur in three ways:

Dielectric Bonding	Silicon Dioxide (SiO_2) fusion or Polymer adhesive
Metal Bonding	Metal (Cu) fusion, Intermetallic bonding, Solder Connection,
or	or Au Thermosonic
Hybrid Bonding	Concurrent Dielectric and Metal bonding

Dielectric bonding assumes that the TSVs will be fabricated after the bonding is completed, or what is referred to as TSV last. However, due to the high planarity and smoothness needed to yield oxide to oxide connection, dielectric bonding is not the preferred bonding method. Polymer bonding is mainly practiced in hybrid bonding.

Cu to Cu bonding also requires fairly parallel surfaces, which can be created by using standard damascene process, with an etch-back of the temporary oxide dielectric. Cu–Cu thermocompression bonding occurs at 400 °C for 30 min at pressure per mm of >40 N/mm. A good Cu–Cu bonding occurs when the Cu grains extend through the connection interface (see Fig. 2.23). Lu, et al., from RPI, has shown excellent hybrid connection using Cu–Cu bonding and BCB as the adhesive [37]. Yu, et al., from IBM have also demonstrated Cu–Cu bonding with ×1 polyimide as the adhesive using the transfer-joining (TJ) method. Figure 2.24 shows the TJ connection on a 200 mm wafer. Alternatively, to add structural integrity to the Cu–Cu bonding without using hybrid joining, polymer underfill can be done

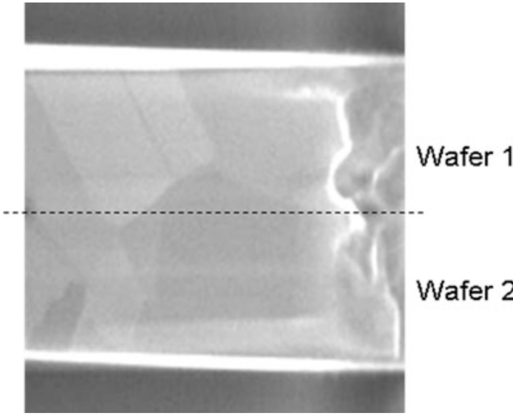


Fig. 2.23 Cu–Cu joining with shared grain growth

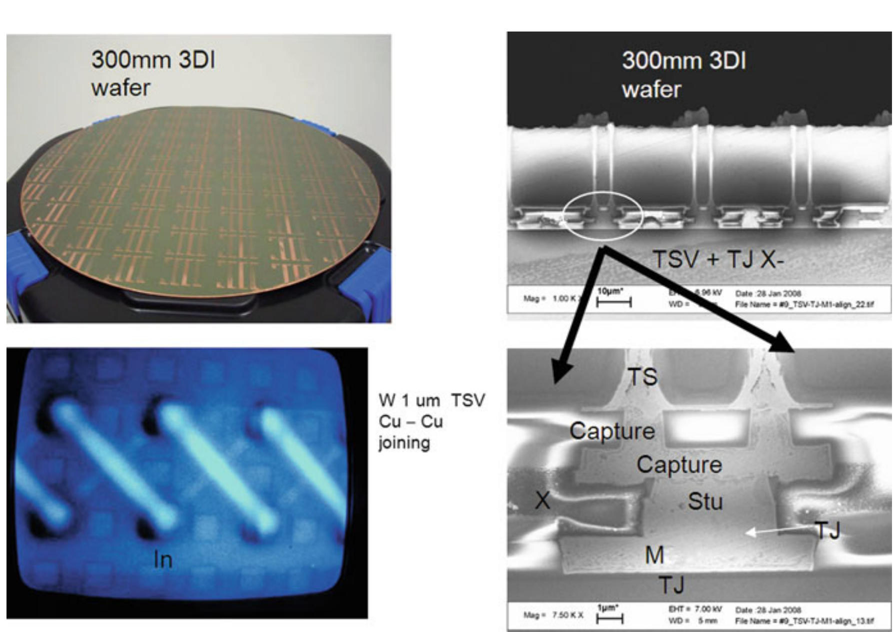


Fig. 2.24 The TJ connection on a 200 mm wafer. Courtesy of Roy Yu of IBM Corp

subsequently to the Cu–Cu joining. A gap of 10–15 μm is seeded for vacuum filled underfill, which is mainly practiced at the end of all wafer to wafer connections.

In some applications, the wafer is first bonded to a temporary carrier, such as Si or glass, with an adhesive. However, due to this limit on the maximum bonding temperature, a transient liquid phase (TLP) bonding is used, instead. During TLP or sometimes referred to as solid liquid interdiffusion, bonding results when solder (Sn or Sn3.5 %Ag) is used as the interconnection metal. The joining occurs at the solder melting temperature, with all the Sn converted to intermetallic, which has a

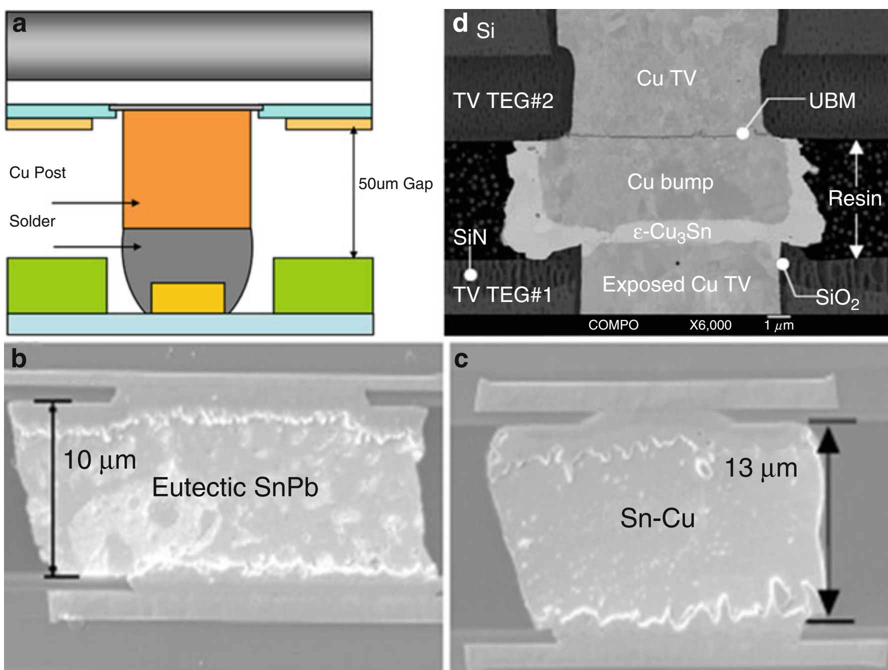


Fig. 2.25 Various fine pitch interconnects; (**a**) IBM's <70 μm pitch MPC-C2 process, (**b**) and (**c**) IBM 50 μm pitch report, and (d) ASAT 20 μm pitch TLP joining

much higher melting temperature, and can withstand the higher temperature needed for thermally decomposing the temporary adhesive. Intermetallic bonding by itself forms a brittle bond. Underfill is used for mechanical support and corrosion prevention purposes.

TLP and Cu-solder-Cu joints are most practiced on chip-to-Si interposer or chip-to-wafer. TLP joints should perform better in EM testing since all the Sn is consumed, eliminating a major EM failure mode. The Japanese Association of Advance Electronic Technologies (ASAT) has reported TLP bonding on 20 μm pitch [39]. They have reported importance of removing the Cu oxide to obtain good connection at joining. One main difference between TLP and solder joining is that TLP required thermal compression, while solder bonding does not. Solder joining is used on >40 μm pitch joining, mainly on chip-to-chip joining, and is a scaled down extension of the traditional chip joining, but on tight pitch requires a bonder to maintain placement accuracy during reflow. 50 μm pitch Cu-Solder-Cu evaluation was reported by IBM with excellent EM results [40]. The Metal Post Colder-chip connection (MPC-C2) process developed by IBM Japan, initially for chip to laminate can also be used on chip-to-chip connection for larger pitch and larger chip-to-chip gaps [41]. Figure 2.25 depicts various fine pitch technologies.

Emerging technologies, such as 3D, are further requiring the reduction of the solder pitch to 50 μm and lower, and in some cases elimination of the solder connection and replacing it with metal to metal connection.

2.8 Summary

Flip chip technology was invented and developed by IBM about 40 years ago, in response to the need for a relatively large number of highly reliable input/output connections in semiconductor devices. Since then, the continuously increasing need for even larger number of I/O's, for example 10,000 I/O's per chip, has led to the development of various solder deposition technologies enabling smaller solder bumps at finer pitch. Green movement requiring use of lead free solders has imposed additional challenges of developing solder materials in first level packaging for greater mechanical stability in view of weaker BEOL structure due to use of fragile ultralow-k dielectric materials, and the need for improvement in EM performance. Intel copper pillar structure is one of the possible solutions for lead-free interconnect challenges. Furthermore, from performance improvement point, the packaging development being in cost advantageous position when compared to semiconductor technology development has led to development of SIP, POP, PiP, versions of system level packaging. The co-design tools have enabled package designers to participate in the multi-die design chain using a co-design methodology. In response to meet increased number of interconnections and higher performance, the development of fine pitch flip chip interconnections, 3D with TSV's and without TSV's, temporary adhesive liquid phase connections, and Cu-solder-Cu connections is continuously striving for excellence.

References

1. Miller LF (1969) Controlled collapse reflow chip joining. IBM J Res Dev 13:239
2. Tsukada Y et al (1993) A novel solution for MCM-L utilizing surface laminar circuit and flip chip attach technology. Proceedings of the 2nd International Conference on Multichip Modules, Denver, CO, April 1993, pp 252–259
3. Perfecto ED et al (2008) C4NP technology: manufacturability, yields and reliability. 58th ECTC Proceeding, Orlando, May 2008
4. Schmidt R and Notohardjono BD (2002) High-end server lowtemperaturecooling. IBM J Res Dev Vol. 46 No. 6, November, 2002
5. Nowalk E (2002) Maintaining the benefits of CMOS scaling when scaling bogs down. IBM J Res Dev Vol. 46 No. 2/3, March/April, 2002
6. RoHS Enforcement Authority, NWML, Stanton Avenue, Teddington, Middlesex, TW11 0JZ. Information website and enquiry service: www.rohs.gov.uk
7. BERR (Department for Business Enterprise & ROHS Regulatory Reform) (2008) Government Guidance Notes SI 2008, No. 37, URN 08/582, Feb 2008
8. Cole M et al (2008) European Union RoHS exemption review case study. IBM Corporation
9. Technology Forecasters Inc. (2008) Report on Economic impact of the European Union RoHS Directive on The Electronic Industry, Jan 2008
10. Luijk P, et al (1991) Thermal degradation characteristics of high impact olystyrene/decabromodiphenylether/antimony oxide studied by derivative thermogravimetry and temperature resolved pyrolysis-mass spectrometry: formation of polybrominated dibenzofurans, antimony (oxy)bromides and brominated styrene oligomers, J Appl Pyrolysis 20, 303–319

11. Iji, M (1999) New environmentally conscious flame-retarding plastics for electronics products. Proceedings of EcoDesign ’99: First International Symposium On Environmentally Conscious Design and Inverse Manufacturing, 1–3 Feb 1999
12. de Sousa I, Henderson D, Patry L et al (2006) The influence of low level doping in thermal evolution of SAC alloys solder joins with Cu pad structures. Proceedings of 56th ECTC, San Diego, May 2006
13. Sylvestre J et al (2008) The impact of process parameters on the fracture of brittle structures during chip joining on organic laminates. Proceedings of 58th ECTC, Orlando, May 2008
14. Plumbridge WJ (2008) Recent observations on tin pest formation in solder alloys. J Electron Mater 37(2):218–223
15. Chiu T-C et al (2008) Effect of thermal aging on board level drop reliability for Pb-free BGA packages. Proceedings of 54th ECTC, Las Vegas, May 2008
16. Yin L et al (2009) Controlling Cu electroplating to prevent sporadic voiding in Cu3Sn. Proceedings of 59th ECTC, San Diego, May 2009
17. Communication with Michael Gordon, Watson Research Center, Dec 2008
18. Bieler TR et al (2006) Influence of Sn Grain Size and Orientation on the Thermomechanical Response and Reliability of Pb-free Solder Joints. Proceedings of 56th ECTC, Orlando, May 2006
19. UCHIBORI CJ et al (2006) Impact of Chip-Package Interaction on Reliability of Cu/Ultra low-k Interconnects for 65nm Technology and Beyond. 8th International Conference on Solid-State and Integrated Circuit Technology
20. Yeh DC, Huntington HB (1984) Extreme fast-diffusion system: nickel in single-crystal tin. Phys Rev Lett 53:1469
21. Lu M et al (2008) Effect of Sn grain orientation on electromigration degradation mechanism in high Sn based Pb-free solders. Appl Phys Lett 92(21):211909
22. Lu M et al (2008) Comparison of electromigration performance for Pb-free solders and surface finishes. Proceedings of 58th ECTC, Orlando, May 2008
23. Jan CH et al (2008) A 45 nm low power system-on-chip technology with dual gate (logic and I/O) high-κ/metal gate strained silicon transistors. IEDM Tech Digest, paper 27.4, Dec 2008
24. Moore G (1975) Progress in digital integrated electronics. IEDM Tech Digest, pp 11–13
25. Zhang GQ et al (2006) Strategic research agenda of more than moore. EuroSime 2006, 24–26 Apr 2006
26. Tummala RR (2006) Moore’s law meets its match. IEEE Spectrum magazine, Jun 2006.
27. Bottoms WR et al (2007) 2007 ITRS assembly and packaging report. 2007 ITRS Conference, Makuhari Messe, Japan, Dec 2007
28. Morimura H et al (2008) Integrated CMOS-MEMS technology and its applications. 9th International Conference on ICSICT, 20–23 Oct 2008
29. Utsunomiya HH (2008) Challenge on packaging substrate technologies towards next decade. 3rd IMPACT
30. Nunn WA, Soldo D (2006) Leading insight workshop, application workshops for high performance design. Ansoft Corp.
31. Bottoms WR, Chen W (2008) An overview of the innovations, emerging technologies and difficult challenges regarding the Assembly & Packaging chapter of the ITRS. Future Fab International, Issue 28, Jan 2008
32. Madisetti VK (2006) Electronic system, platform, and package codesign. IEEE Design Test 23(3):220–233
33. Kuo W-S et al (2007) POP package (Cavity BGA) warpage improvement and stress characteristic analyse. Proceedings of IMPACT 2007, 1–3 Oct 2007
34. Carson F et al (2008) The development of the Fan-in Package-on-Package. Proceedings of 58th ECTC, Orlando, May 2008
35. Fillion R et al (2007) Embedded chip build-up using fine line interconnect. Proceedings 57th ECTC, Reno, May 2007

36. Kim YG et al (2002) Folded stacked package development. Proceedings of 52nd ECTC, San Diego, May 2002
37. McMahon FJJ, Lu JQ, Gutmann RJ (2005) Proceedings of 50th ECTC, May 2005
38. Liu F et al (2008) A 300-mm wafer-level three-dimensional integration scheme using tungsten through-silicon via and hybrid Cu-adhesive bonding. IEDM, San Francisco
39. Umemoto M (2004) High performance vertical interconnection for high density 3D chip stacking packages. Proceedings 54th ECTC, May 2004
40. Gan H (2006) Pb-free micro joints (50 μm pitch) for next generation micro-systems. Proceedings 56th ECTC, May 2006
41. Orii Y (2009) Ultrafine-pitch C2 flip chip interconnections with solder-capped Cu pillar bumps. Proceedings of 59th ECTC, San Diego, May 2009

Chapter 3
Bumping Technologies

Michael Töpper and Daniel Lu

Abstract Electronic Packaging is more than housing of active and passive elements. Since the last decade packaging has evolved into the first step on system integration. In this sense Wafer Level Packaging (WLP) based on redistribution is the key technology for System in Package (SiP) and Heterogeneous Integration (HI) by three-dimensional (3D) packaging using Through Silicon Vias (TSV). Materials and process technologies are key for a reliable WLP. This chapter focuses on the materials and processes for WLP which are the basic for all new 3D integration technologies.

3.1 Introduction

Electronic packaging and assembly is the basic technology to link the small dimensions of the IC to an interconnecting substrate—usually the Printed Circuit Board (PCB) or a Multilayer Ceramic (MLC) [1–3]. These substrates combine a number of ICs and passive components to build the final microelectronic system for the user [4]. New applications with their expanding performance and functionality in conjunction with new device technologies are pushing the requirements and innovation for electronic packaging. Milestones for this progress have been Surface Mount Technology (SMT), Flip Chip in Package (FCIP), Flip Chip on Board (FCOB), and Wafer Level Packaging (WLP) which is evolving to System in Package (SiP) and Heterogeneous Integration (HI) by three-dimensional (3D) packaging using Through Silicon Vias (TSV) [5]. Therefore, the technology boundaries between front-end

M. Töpper (✉)
Fraunhofer IZM, Gustav-Meyer-Allee 25, 13355 Berlin, Germany
e-mail: toepper@izm.fhg.de

D. Lu
Henkel Corporation, Shanghai, China
e-mail: Daniel.lu@ieee.org

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_3, © Springer Science+Business Media New York 2013

semiconductor technology, packaging and system engineering are becoming seamless. Heterogeneous integration bridges the gap between nanoelectronics and its derived applications bringing together nanoelectronics, microsystem technologies, bio-electronic and photonic component technologies. The focus of this chapter is on the bumping technologies.

The evolution of Single Chip Packages (SCPs) has started from small metal boxes (TO packages) and developed over the Dual Inline Package (DIP) for through hole assembly and SMT packages such as the Plastic Quad Flat Package (PQFP) to the Ball Grid Array (BGA) [6]. BGA packages use rigid or flexible interposer for the redistribution from the peripheral pads to the area array. A further miniaturization to a maximum of 1.2× the chip size brought up the concept of the Chip Size Package (CSP) [7]. Typical area array pitch for CSP is currently 0.5 mm moving now to 0.4 mm which results in additional technology pressure on the PCB. DIP and PQFP represent packages with peripheral I/Os while BGAs and CSPs are area array packages which are assembled in Flip-Chip (FC) fashion. FC is a face-down assembly technique originally developed by IBM as C4 (Controlled Collapse Chip Connection) which provides excellent performance and represents a must for dice with high I/O counts like microprocessors.

A major requirement for the flip-chip interconnections are modified pads on the IC. The so called UBM (Under Bump Metallization) or BLM (Ball Limiting Metallurgy) is the basis for a low-ohmic electrical, mechanical, and thermal contact between chip and substrate. The self alignment function of the assembly is one of the major advantages of flip-chip assembly using solder. Chips can be misregistered as much as close to 50 % off the pad center and the surface tension of the molten solder will align the pads of the chip to the substrate metallization during the reflow process. The disadvantage of flip chip assembled IC is that the bumps are the only mechanical links between chip and substrate. As a consequence the stress caused by the CTE (coefficient of thermal expansion) mismatch of the semiconductor die and the substrate act only upon the bump interconnects. Therefore, underfillers (epoxy resins with filler particles) have to be filled in between the flip chip and the substrate which translates into extra costs in the assembly process.

3.2 Materials and Processes for Bumping Technology

3.2.1 Metals for Wafer Bumping

A bump formed in most cases on wafer level is defined as a usually conducting 3D interconnect element between die and substrate [8]. The interconnect process between chip and substrate is based on soldering, thermo-compression bonding and adhesive bonding [9]. Many different bump metallurgies are used ranging from pure Au, Cu, Sn, or In to alloys such as eutectic or high-melting PbSn, AuSn, AgSn, SnCu, and AgSnCu depending on the application. Legislations like the one by the

European Union and other countries demand the ban of lead from electronic products by 2006 with the exception of high-lead solders for microprocessor applications. This ban of lead for electronic products by RoHS (*Restriction of the use of certain hazardous substances in electrical and electronic equipment*) has changed the materials for choice dramatically.

3.2.1.1 Under Bump Metallization

The UBM has to provide a low contact resistivity to the chip pad and the solder, good adhesion to the chip metallization and the chip passivation and a hermetic seal between UBM and IC pad [10]. It has to be a reliable diffusion barrier between the IC pad and bump with low film stress and it needs to be sufficiently resistant to stress caused by thermal mismatch during die assembly. In the case of PbSn bumping, common UBM stacks are Cr–Cr:Cu–Cu–Au (original C4 from IBM); Ti–Cu; Ti:W–Cu; Ti–Ni:V; Cr–Cr:Cu–Cu; Al–Ni:V–Cu; Ti:W(N)–Au. Usually, these UBM stacks are sequentially deposited by sputtering or plating. The advantage of sputtering compared to evaporation is the kinetic energy of the deposited atoms which is in the range of 0.1–0.5 eV for evaporation and 1–100 eV for sputtering which guarantees a much higher adhesion. For 200 and 300 mm wafers the evaporation distance has increased to a nearly unacceptable level which decreases further the deposition efficiency being proportional to the square of distance.

The UBM etch process removes the UBM metallization between the bumps. For cost and technology reasons wet chemical etching is common. For a UBM stack consisting of different metal layers different etch chemistries are required for each layer. Among the requirements for the etching step are a uniform etching result, a minimum bump undercut and to monitor the remaining metallization thickness in order to stop the etching process or to switch the etch chemistry in case a layer of a UBM stack is fully removed. It is important to design the etching process in such a way that the bump surface is not oxidized or modified in any other way. In addition, the design of the UBM stack has to take the UBM etch process into account in order to achieve reliable and good process results.

A schematic draw of Ti:W–Cu for PbSn is given in Fig. 3.1.

In the case of Sn-based bumps deposited onto a copper based UBM intermetallics compounds (IMCs) between Sn and Cu are formed by the reflow process providing the required adhesion of the bump to the chip pad. IMCs are brittle in nature due to the ordered crystal structure which is in contrast to the solid solutions like the PbSn. The metals which are mostly used in packaging—Cu, Ni, Au, and Pd—are forming binary intermetallics with Sn-based solders of the Hume Rothery type [12]. These compounds are based on electron valence bonding. The crystal structure is controlled by the number of electrons in the bonding. The composition of each phase can be calculated based on the concentration of the valence electrons. For example the Cu_3Sn and the Cu_6Sn_5 phases are found for intermetallics of Cu and Sn and Ni_3Sn_4 and Ni_3Sn phases are formed between Ni and Sn. The growth rate depends on temperature, the different activation energies of

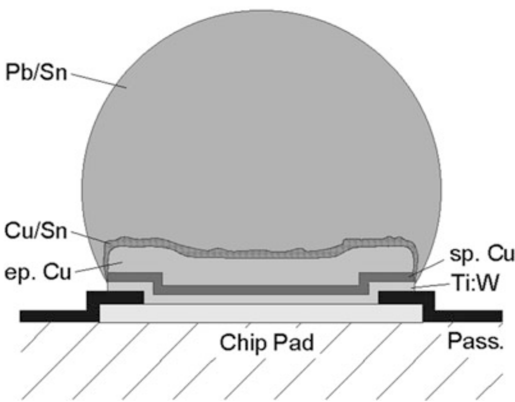


Fig. 3.1 Example of a UBM stack (Ti:W–Cu) for PbSn bumps [11]

compound forming and diffusion processes. In general, the intermetallic growth rate is much higher for Cu compared to Ni. This is becoming more important for the lead-free solder due to their higher Sn content.

3.2.1.2 Bumping Technologies

The main bumping processes are electroplating, stencil printing, evaporation, placing preformed solder spheres and C4NP. The selection of UBM and bump metal mainly depends on the melting point of the solder, the thermal and mechanical reliability of the interfaces between UBM and bump, the integrity of adjoining pad metallization, the bumping process capability, the operating conditions of the assembly, and the reliability demands on the whole package. The main process steps are summarized in Fig. 3.2.

A major requirement for all bumping deposition technologies is the control of the bump volume across the wafer as well as the solder composition in order to achieve a uniform bump height distribution and to avoid an incomplete reflow process. The maximum reflow temperature is typically a few 10 K above the melting point of the solder, and it is important that an optimized temperature profile is maintained during the reflow process. All interconnects of a chip are linked at the same time for the FC-assembly process. There is the risk that the chip will not work if one joint (in the case of a signal line) is failed.

Evaporation is the deposition method which was originally developed by IBM for FC bonding. It became popular under the acronym C4 (controlled collapse chip connection) summarizing the main steps of soldering (collapse) in combination with a necessary solder mask (controlled). The technology allows a wide range of solder materials with excellent quality. The transition to 200 mm wafer size was the end for C4 due to yield and cost.

Electrodeposition of metal depots through a lithographically defined plating mask introduced by Hitachi in 1981 [14] is the bumping method of choice to meet the requirements of defined bump shape and size. Mostly ECD

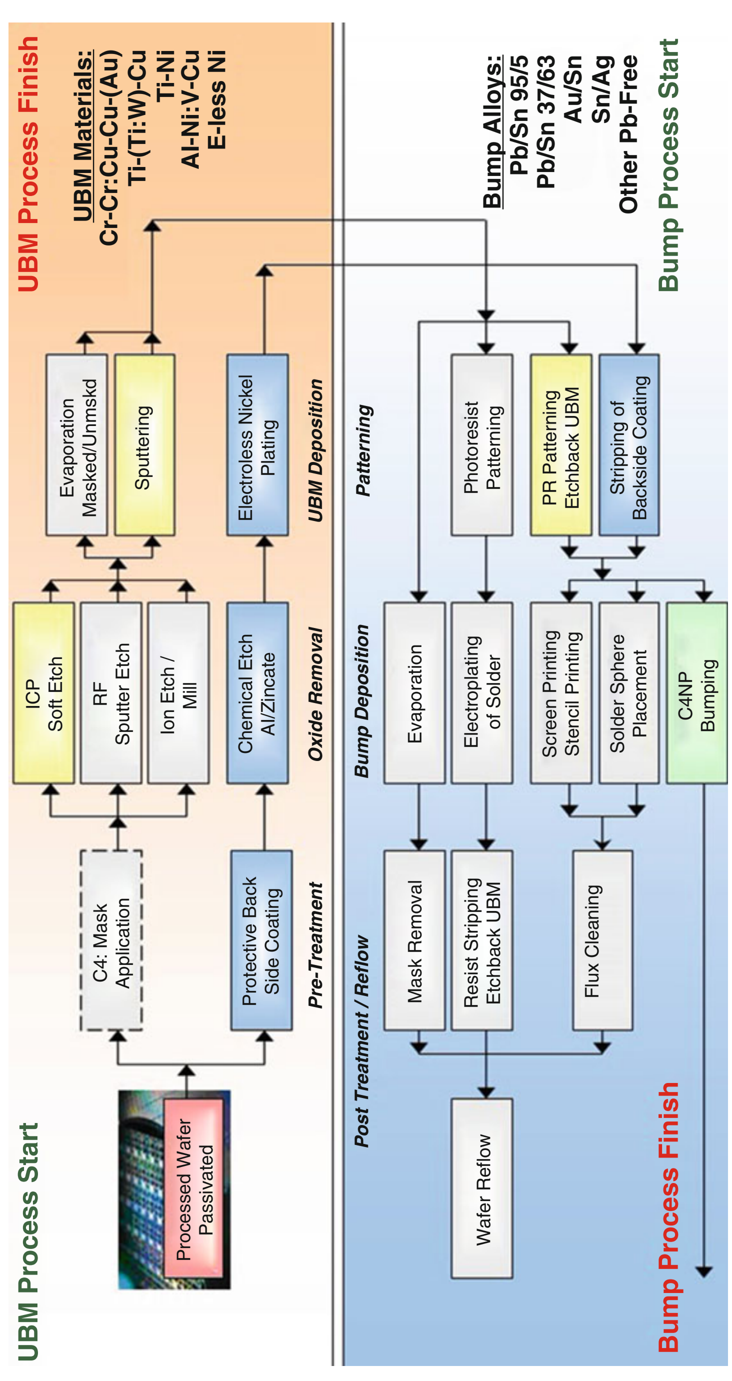


Fig. 3.2 Overview of bumping technologies [13]

(Electro-Chemical Deposition) is used as an acronym for this technology. In general electroplating is a relative slow deposition technique with typical plating speeds ranging from 0.2 μm to only a few μm per minute depending on the deposited material [15]. Plating techniques can use constant voltage (potentiostatic), constant current (galvanostatic), or pulse plating. Pulse plating is the method of choice for fine-pitch application providing a more uniform, smooth deposit with less porosity. Among the most important parameters that influence the uniformity of plating height and solder composition as well as bump morphology is the electrical field distribution across the wafer as it defines the plating current. Therefore, the voltage has to be applied over many points along the wafer parameter. In this case the photoresist is completely removed along the perimeter of the wafer (edge bead removal) and the electrode in form of a ring is attached to the wafer. A sealing ring is put on top of the resist surface to prevent the electrode from being contaminated by the electrolyte. The current distribution has approximately rotational symmetry but can show a radial variation. In this case the anode design in fountain platers offers another means to control the plating current uniformity by compensating for radial field variations. In addition, the ratio of open area (this is the total plated area) versus full wafer area influences the plating current uniformity. It is important to have a uniform distribution of bumps across the wafer surface which may require to deposit bumps in some areas on the wafer without any die underneath (dummy bumps). A proper tool design ensuring uniform current density and equalized bath agitation across the wafer allows a bump height homogeneity of less than ±5 % at 300 mm in diameter.

In the case of PbSn plating tin and lead salts are dissolved in the electrolyte and dissociate into their anions and cations. Sulfuric acid is added to increase the conductivity of the electrolyte. Additives are added to refine the deposited solder. The positively charged Sn^{2+} and Pb^{2+} cations migrate to the cathode (wafer) due to the applied voltage in the plating cell and are deposited by a discharge reaction on its surface. For this to occur the cations need to be reduced to the metallic state by accepting electrons from the cathode. The electroplating process is a complex process. The mechanism of the metal deposition consists of several steps: The hydrated metal ions have to diffuse to the wafer surface covered by a Helmholtz double layer. In addition the metal ions can be chemically attached to complexing molecules. Additives are controlling the metal growth to achieve fine-grained solders. Obviously, the plating time can be shortened by increasing the plating current density. However, the maximum current density is limited because high plating speeds translate into higher challenges in plating bath maintenance.

Though structure dimensions are not as critical as in advanced front-end processing like 65 nm node, they are tiny enough to be defined best by photoimaging. In contrast to subtractive etching or lift-off technique, ECD bump formation precisely replicates the photoresist pattern in lateral dimensions. The ECD bumping technique can accommodate a wide range of wafer types, previously applied materials and pattern configurations. All kinds of semiconductor such as silicon, SiGe, GaAs, and InP as well as ceramic and quartz substrates can be processed. Besides, there is no restriction in passivation types such as silicon oxide, oxy-nitride, silicon nitride, and

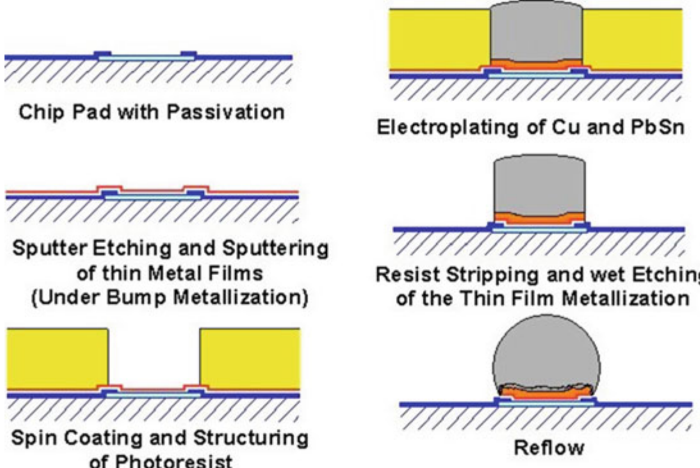


Fig. 3.3 Solder bumping process flow (electroplating) [15]

polymers like PI or BCB in the main. In case of standard I/C wafers the bumping sequence will be applied directly onto the I/O pads. If the original I/O layout has been redistributed, the bumping sequence will take place onto the routed metal layer covered by a dielectric solder resist mask. The placement of solder bumps on top of a polymer layer offers reduced self-capacitance, which is desirable for RF applications. As the ECD bump fabrication can be incorporated into the thin-film technology, the sequence can be divided into fundamental process steps which are sputtering of the UBM, lithographical printing, electroplating of the bump, removal of the photoresist, and differential etching of the plating base schematically shown in Fig. 3.3.

In this process the UBM that provides the adhesion layer in combination with a diffusion barrier and a wettable surface is sputtered and a thick liquid or dry film resist is coated onto the wafer. The resist is then exposed and developed. Electroplating allows to deposit the solder over the top of the resist layer to form mushroom like structures. Then the resist is stripped and the UBM is etched between the bumps. The reflow process transforms the solder into nearly ball shape and leads to the formation of intermetallic compounds at the UBM/solder interface which is important for a reliable adhesion of the bump to the UBM. The advantage of mushroom plating is that the photoresist layer can be significantly thinner than the final solder ball height and that solder deposition is fast because the solder surface is growing when being plated over the resist edge. A disadvantage is that with mushroom bumping plating control becomes more demanding. Photoresist thickness in mushroom plating typically is between 25 and 60 μm. On the other hand, mushroom plating becomes a problem for finer bump pitches. Thicker resist (approximately 100 μm) is used for fine pitch bumping where the solder is completely plated in the bump mold.

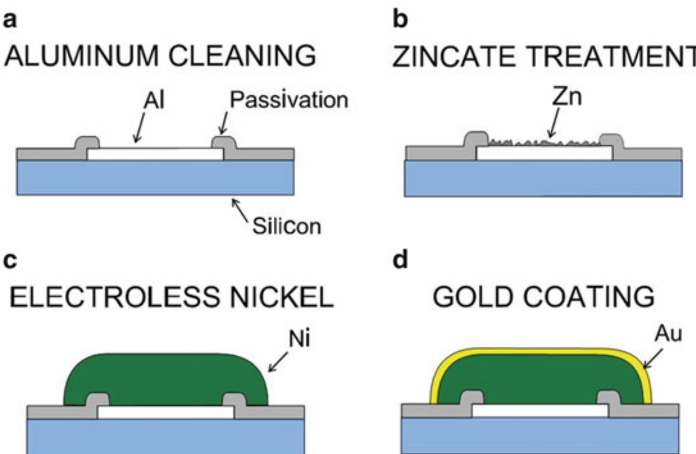


Fig. 3.4 Principle process of electroless Ni bumping: (**a**) bondpad in initial state, (**b**) after zinc deposition, (**c**) after growth of electroless Ni, (**d**) after plating of thin immersion Au [22]

If micromechanical elements are present on the wafer, one's attention must be directed to a possible damaging during thin film processing [21]. Cavity surface contamination can later impair MEMS performance, too. In some cases it will be necessary to protect the MEMS areas by covering with photoresist prior to the sputter step, or by local etching the plating base prior to the electroplating process. Some 3D structures with cavities like air bridges or acceleration transducers require more than one photoresist layer. In that case a primary coated and patterned resist layer has to fill the hollow space of the complex items and consequently serves as a smooth subsurface for the following UBM deposition.

For *solder printing* either a metal or resist stencils (for finer pitches) are used. The deposition of the UBM is not part of this process. It has to be deposited separately. A thin film process based on sputtering (sometimes in combination with plating) can be used. A low cost method is the electroless deposition of Ni (ENIG) on the Al pads. The ENIG process is based on the selective chemical deposition of metal on Al bond pads. Wafers are treated in a sequence of chemical solutions. After each treatment they have to be rinsed carefully in DI water. The principle of the process is shown in Fig. 3.4 [16].

First the surface of Al bondpads is cleaned by immersing the wafers into two baths. This passivation cleaner removes possible residues, while the second (Al cleaner) removes thick Al oxides and roughens the surface. In a zincate bath a thin Zn layer is deposited on Al by an exchange reaction in order to activate the surface for subsequent Ni plating. The electroless Ni bath contains mainly Ni ions and hypophosphite. A first Ni layer is deposited on the pads by an exchange reaction between Zn and the Ni ions. On this first layer additional Ni is plated by a continuous autocatalytic reaction which is necessary to plate more than monolayers without any current. The energy is supplied internally inside the plating bath by the

Table 3.1 Process steps of electroless Ni bumping and their function (Courtesy of Fraunhofer IZM)

Process step	Function
1. Protective resist coating	Protective resist coating on wafer backside
2. Passivation cleaning	Removes passivation residues from Al pads
3. Al cleaning	Removes thick Al oxides and prepares surface for metal deposition
4. Zincating	Activates Al for Ni deposition
5. Electroless Ni	Deposition of Ni layer (typ. 5 μm)
6. Immersion Au	Au finish on Ni (typ. 0.08 μm) to prevent Ni from oxidation
7. Backside cleaning	Removes protective coating from backside

oxidization of adsorbed hypophosphite. The released electrons are able to reduce the Ni^{2+}. The phosphorus converted from hypophosphite is built into the Ni-layer. This can change the mechanical and electrical property of the Ni. The plating rate is 25 μm/h. In the subsequent immersion Au solution, a thin Au film is deposited by an exchange reaction on the surface of the Ni layer. The Au has a thickness of 0.05–0.08 μm and is required to prevent Ni from oxidation. But the formation of brittle Au–Ni–Sn intermetallic phases at the interface between Ni and solder can impact the long-term operation reliability of the interconnections; therefore, the Au layer has to be kept as thin as necessary.

The complete process flow is shown in Table 3.1. In addition to the wet chemical treatments mentioned above, the backsides of the wafers have to be protected in order to prevent Ni deposition on Si. This is done by spin-coating of a protective resist on the backside of the wafer. After bump plating, the resist is stripped. All chemicals used in this process are commercially available. They are completely cyanide-free and no organic solvents are used.

For all wet-chemical treatments, 25 wafers are handled together in one carrier. The process requires tanks with seven different chemical baths and additional rinse tanks. The process times are relatively short. They range from 30 s (zincating) to 30 min (immersion Au). By handling the wafer cassettes manually from bath to bath a throughput of 25 wafers per hour can be achieved. In fully automatic systems 100 wafers per hour are possible. The Ni height uniformity is better than ±5 % over a 200 mm wafer. Within die variation is correspondingly lower. The specification of the Fraunhofer IZM ENIG process is given in Table 3.2

Control of the UBM quality is monitored by shear strength which has to be around 150 MPa (min 100 MPa). The Al etching process has to be restricted to less than 0.5 μm to avoid damages of the Silicon devices. The Ni UBM has been extensively tested. No failures were detected even after 10,000 h thermal storage at 300 °C, 10,000 cycles AATC (−55/+125 °C) and 10,000 h humidity storage (85 °C/85 % r.h.) [17].

The solder deposition is then done by printing process shown in Fig. 3.5 [18].

This process was transferred from PCB industry to wafer technology. The solder pastes that are screened into the apertures of the stencil consist of solder particles of 2–150 μm diameter in binders and flux. They are classified according to the particle size of the solder (Table 3.3).

Table 3.2 Specification of the e-less Ni/Au process (Courtesy of Fraunhofer IZM)

Property	Specification
Wafer material	Si
Bond pad material	AlSi1 %, AlSi1 %Cu0.5 %, AlCu2 %
Pad metal thickness	≥1 μm
Passivation	Defect-free nitride, oxide, oxi-nitride, polyimide, BCB
Residues on bond pads	<5 nm
Non organic	Not acceptable
Organic	
Wafer size	100–300 mm
Wafer thickness	>200 μm (>150 μm)
Bond pad geometry	Any (square, rectangular, round, octagonal)
Passivation opening	>40 μm
Bond pad spacing	>20 μm
Passivation overlap	5 μm
Wafer fabrication process	CMOS, BiCMOS, bipolar
Ink dots	Acceptable Stability depend on ink
Probe marks	Acceptable
Scribe lines	Must be passivated (thermal oxide) Test structures acceptable
Laser fuses	Not acceptable
AL fuses	Acceptable (with limitations)
Poly Si fuses	

An additional function of the paste is the prefixing of the placed components on the substrate before reflow. The viscosity of the pastes should be between 250 and 550 Pa s for screen printing and 400–800 Pa s for stencil printing.

The screens are typically made by laser drilling of metal sheets or by electroforming. The printer aligns the stencil to the substrate or wafer. The paste is pressed within seconds through the stencil and thousands of pads are bumped at a time. One of the key issues is the requirement that all the paste has to be transferred from the stencil to the wafer. Any solder paste residues in the screen will reduce the final bump height uniformity. The metal stencils have to be cleaned by solvents. Mostly water-soluble solder pastes are used in production. To calculate the size of the reflowed bump as a function of pad size and geometry the following equation is used for a reflowed bump as a truncated sphere: $V = (1/2)AH + (\pi/6)H^3$ with V as the solder volume, A as the pad area and H for the bump height.

Reflow is necessary to form the bump shape out of the printed solder depots. Due to the addition additives to the solder particles the formation of voids during the reflow processes is likely. There is no reliability reduction if they are kept under a certain volume depending on the design. Solder printing is a fairly simple and inexpensive process step when compared to electroplating or evaporation. An important advantage of stencil printing is the large variety of available solder pastes. This offers flexibility and is of particular importance for selecting lead free solders.

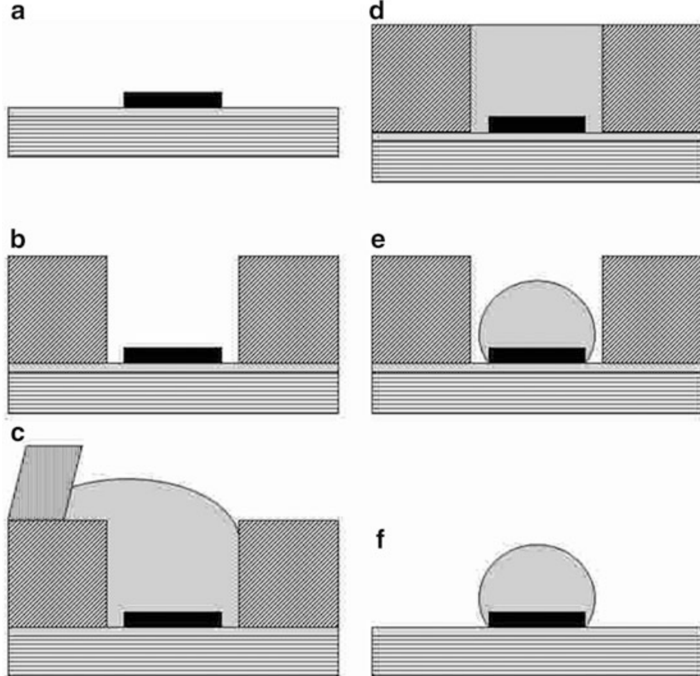


Fig. 3.5 Solder printing using photoresists (UBM formation (**a**), stencil or photo-mask (**b**), printing (**c**, **d**), reflow (**e**), and removal of the mask (**f**)) [24]

Table 3.3 Solder paste classes

Class	1	2	3	4	5	6	7
Solder sphere size [μm]	75–150	45–75	20–45	20–38	15–25	5–15	2–11

Even triple component solders such as SnAgCu (SAC) can be deposited, which is difficult to electroplate.

For fine pitch below 150 μm a photoresist mask of a thickness 70 μm and thicker gains a lot of advantages for the printing process but being a higher cost process compared to metal stencils. The process is based on work done by Flip Chip International (FCI, former FC Division of K&S, founded as FCT) and is widely used by companies that licensed their Flex-On-Cap-Process (FOC) [23]. As the UBM pad defines the final bump base the molds for the stencil process can have a larger footprint than the final bump in order to have more solder paste being screened into the mold to achieve a larger bump height. Before resist stripping the solder paste has to be heated to be transformed into solid solder. Examples of this process are given in Figs. 3.6 and 3.7 [18].

With this technology bumps of 25 μm diameter with a pitch of 60 μm have been achieved.

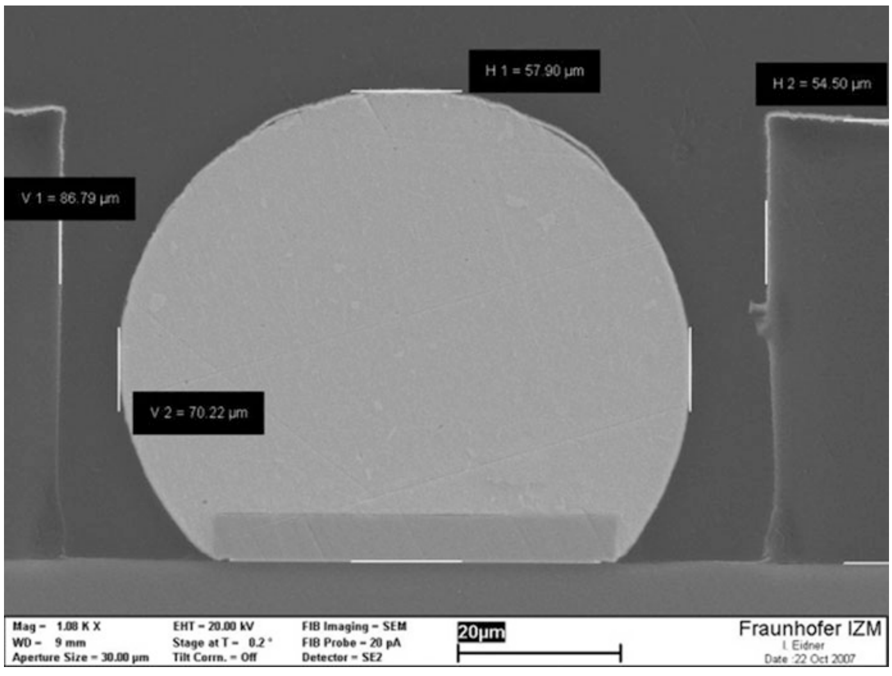


Fig. 3.6 Printed SnAgCu solder into Dry-Film resist [18]

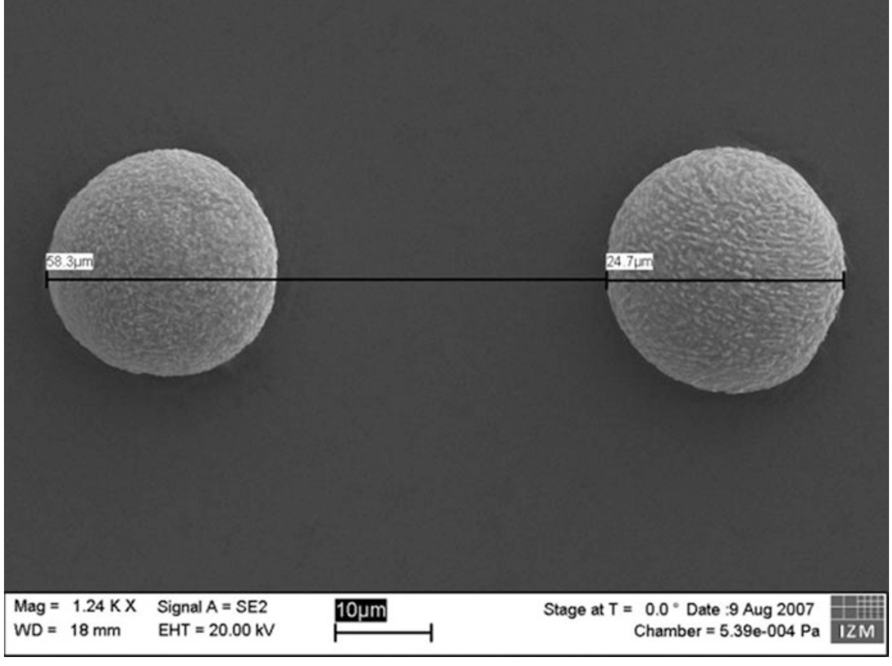


Fig. 3.7 Printed SnAgCu bumps (25 µm with 60 µm pitch) after resist stripping [18]

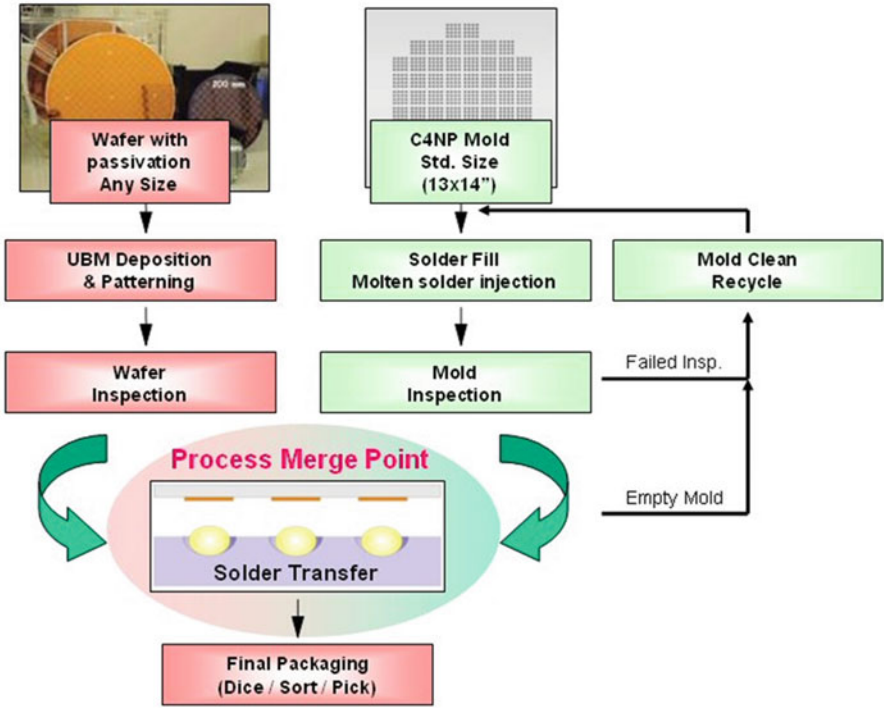


Fig. 3.8 C4NP bumping process sequence

C4NP is a novel solder bumping technology developed by IBM which addresses the limitations of existing bumping technologies by enabling low-cost, fine pitch bumping using a variety of lead-free solder alloys [19]. It is a solder transfer technology where molten solder is injected into prefabricated and reusable glass molds. The basic process flow is given in Fig. 3.8.

The glass mold contains etched cavities which mirror the bump pattern on the wafer. The filled mold is inspected prior to solder transfer to the wafer to ensure high final yields. Filled mold and wafer are brought into close proximity/soft contact at reflow temperature and solder bumps are transferred onto the entire 300 mm (or smaller) wafer in a single process step without the complexities associated with liquid flux.

The C4NP process was transferred to production by a cooperation between IBM and SussMicrotec. Different cost models are proposing low cost but the infrastructure of the glass mold is not established yet.

Other Bumping Technologies: Alternative bump technologies include stud bumping, solder jetting and ball attach.

Direct ball attach is only used for large solder balls of typical diameters of 150 μm or above. In some cases a vacuum head serves as a template for the ball layout on the wafer and picks up preformed solder spheres from a reservoir. The spheres are dipped into flux and are placed on the wafer and are reflowed. The advantage of ball attach is

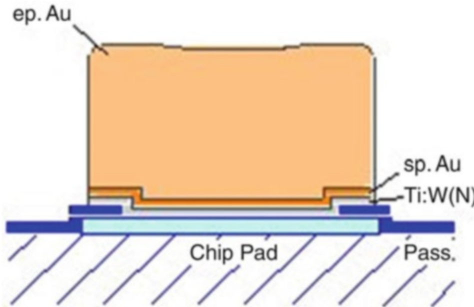


Fig. 3.9 Schematic of a Au bump [15]

that it is an inexpensive way to provide large solder volumes and that it is easily applicable to any solder type. Ball attach is often used to provide solder balls on redistributed dice were the pitch of the area array pad layout is larger or equal to 500 μm. Equipment makers are trying to push the limits to finer geometries.

Stud bumps are made with wire bonders by cutting of the wire right after bonding it to the IC pad. The bump can be either left with a spike or can be coined creating a flat surface or can be sheared off across the top directly after bonding. This technique is fairly flexible in regards to the desired bump metallurgy and can be used for example for gold bumping and even for solder bumping. Since stud bumping is a serial process it has little importance for mass production but is an important technique for flip-chip prototyping and low volume manufacturing. Stud bumping is the major technology applicable to singulated dice.

Solder jetting is a serial and maskless solder deposition technique where solder droplets are ejected from a print head onto the wafer. High ejection frequencies are possible. However, overall process control is difficult and solder jetting has not yet been adoption by the industry. Several attempts have been made in the past without any success.

3.2.1.3 Bump Metallurgy

Bumps Based on Gold and AuSn

Gold is a noble metal which does not oxidize or corrode. It has excellent electrical and thermal conductivity. Due to its high melting point gold bumps cannot be reflowed. The interconnection is mainly to flexible substrates or tapes (TCP: tape carrier packaging) by thermo-compression bonding (TAB: tape automatic bonding) or adhesively bonded to glass substrates (COG: chip on glass) using conductive films. Main application is for LCD drivers. The deposition is done by electroplating on a sputtered Au layer with an adhesion/diffusion layer of Ti or TiW or other transition metals. A schematic structure is given in Fig. 3.9.

Today, minimum gold bump pitch for LCD drivers in mass production is around 30 μm and less with a 10 μm gap between bumps (COG applications allow finer pitch than TAB). These are the tightest pitches of any bumping process in mass production.

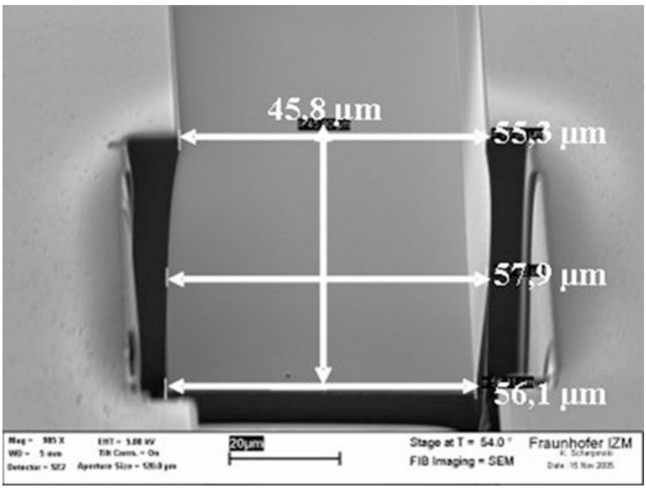


Fig. 3.10 FIB cut of a resist used for Au plating

For electroplating pure Au, both sulfitic as well as cyanidic electrolytes have been used. The advantages of the sulfitic bath type are its excellent compatibility to the applied photoresist system, its well-known non-toxicity, and the easy softening of the Au bumps during annealing. On the other hand, cyanidic Au electrolytes are easy to handle due to the more stable Au complex and offer deposits with strong adhesion even on Ni and Cu as base metal. Cyanidic solutions tend to underplating, which can become a critical fact for fine-pitch bumping. The metal content of most commercially available Au electrolytes is between 8 and 15 g/l, and the applicable current density of exemplarily 5 through 20 mA/cm^2 results in a deposition rate of 0.3 up to 1.2 µm/min. All kinds of Au bumping electrolytes work at higher temperatures, mostly between 50 and 70 °C. In many cases of thermo compression bonding, the ductility of as-plated Au bumps is not sufficient and has to be improved by subsequent annealing step. During thermal aging at 200 °C the initial microhardness of around 130 $HV_{0.025}$ is decreasing down to 70 through 50 $HV_{0.025}$ within a few minutes. The shape of the bump is totally controlled by the resist (thick range of 30 µm) due to the lack of the reflow. Therefore, the combination of resist and electrolyte has to be chosen carefully. Figure 3.10 shows a FIB (Focus Ion Beam) cut of a high structured resist. The resist has vertical side-walls with a variation in the opening of ±1 µm for 45.8 µm thickness

In case of solder bumps, a final reflow is regularly done to homogenize the tin alloy and to form the spherical bump shape. Each solder composition requires an adapted temperature–time profile, which depends on the melting point, the ambient medium, and the out-gassing behavior of the solder. This is not necessary for Au.

The electroplating process can create a lot of stress to the resist which can cause a deformation of the final structure. This is shown in Fig. 3.11 for a not optimized combination of plating electrolyte and photoresist (left-hand side) and for the right choice (right-hand side):

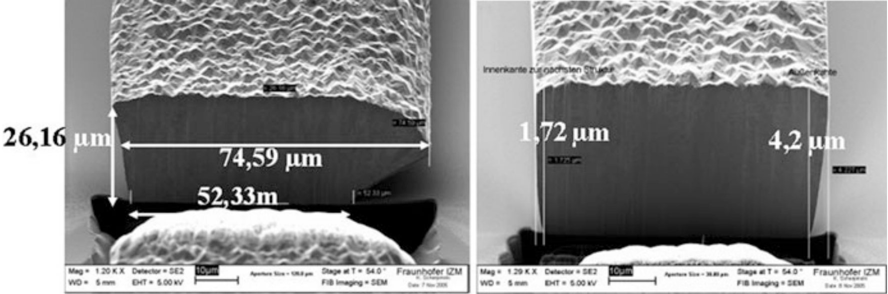


Fig. 3.11 Electroplated Au: Wrong combination of electrolyte/photoresist (*left*) and nearly perfect combination (*right*)

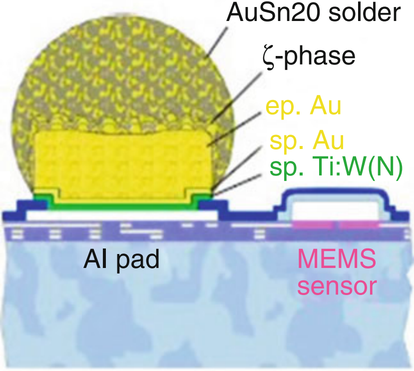


Fig. 3.12 Schematic of a AuSn bump [21]

The ECD can cause a deformation of the resist to over 20 µm if the wrong combination of materials is chosen.

The *eutectic Au/Sn* system provides a good corrosion resistance and enables a flux-free FC assembly. Hence, it is the most suitable interconnection material for optical and optoelectronic devices. The basic structure of the AuSn bump is given in Fig. 3.12.

Au and tin are sequentially electroplated. Evaporation in a sandwich-like fashion is also possible.

Bumps Based on Solder

An interconnection by soldering is the most common technology for all microelectronic systems [20]. The main advantage is the possibility to link uneven and/or rough surfaces and allows by the same time the possibility of repair. The basic requirement is a wettable surface on both interconnection sides. To avoid the flow of the solder over the whole surface a solder mask (mostly a non-wettable area like a solder mask based on epoxy) is needed. A high reliability for the lifetime of the

Table 3.4 Selection of solders for flip-chip interconnects

Solder	Melting point	Remark
63Pb37Sn	183 °C	Eutectic PbSn, low melting point, compatible with organic PCBs, commonly used for most SMDs. Not allowed anymore for ROOS
95Pb5Sn (or similar)	315 °C	High-lead, good electromigration behavior, highly reliable thermo-mechanical interconnect, flip chip on ceramic substrate; no reflow of high lead bumps during chip attach on PCB (eutectic PbSn on PCB side), flux free reflow in H_2 atmosphere
96.5Sn3.5Ag (or similar)	221 °C	Currently most common binary lead free solder for flip chip. Typically used in conjunction with electroplating
97Sn/3Cu	227 °C	Difficult to electroplate. Short bath lifetime
95.5Sn3.9Ag0.6Cu	218 °C	Common lead-free solder paste, Cu content reduces Cu consumption from UBM
80Au20Sn	280 °C	Common for flux free opto-electronic assembly on gold finishes, controlled standoff height
In	157 °C	Ideal for temperature sensitive electronic devices due to very low reflow temperatures
Sn	232 °C	Risk of tin whisker formation

product is given if the solder is highly elastic. If the substrate and the chip has a large CTE mismatch the solder is outside the elastic state. Therefore, an underfiller has to be used. Underfillers are highly filled epoxy polymers. The filler (small particles of SiO_2 for example) is necessary to reduce the CTE of the polymer which is much higher than inorganic materials like Si or Metals. An overview of the most common solder materials is given in Table 3.4.

Solder interconnects are formed during a reflow process with a reflow temperature which is 10+ °C higher than the theoretical melting temperature of the solder.

Fluxes are needed to remove oxides during the assembly and soldering process. In addition, the tackiness allows to hold the placed components in place before the reflow and the final joining process is completed. Fluxes can be inorganic acids, organic acids, rosin and no-clean resins. The J-STD classification describes both flux activity and flux residues activity as follows: L = Low or no flux/flux residue activity; M = Moderate flux/flux residue activity and H = High flux/flux residue activity. These classes are further labeled for activity or corrosiveness.

Copper Posts (Pillars)

Copper has approximately a ten times better electrical and thermal conductivity than PbSn solders. Copper posts are therefore attractive whenever high currents pass through the interconnect like the case with power devices. This is particularly important for fine pitch bumping which reduces the size of the bump base and

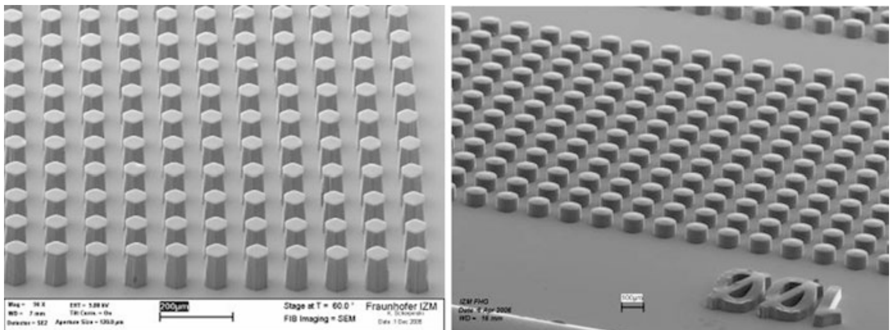


Fig. 3.13 Copper posts after resist stripping (*left*: 80 µm thick Cu plated inside dry film, *right* 60 µm thick Cu plated in liquid type resist)

increases the resistance of the bump. Cu pillars are more prone against electromigration, and they are therefore a promising candidate for lead-free alternatives especially for high-current applications with small bump size [21]. Electromigration is the transport of material which is caused by the gradual movement of the ions in a conductor due to the momentum transfer between conducting electrons and diffusing metal atoms. The effect is important in applications where high direct current densities are used, such as high density interconnects and high power applications. As the structure size in electronics such as integrated circuits (ICs) decreases, the practical significance of this effect increases. Cu pillar can be seen a thick UBM structure to alleviate current crowding at the interface between bump and chip pad.

The Cu bumps as well as the Cu routing layers are deposited from a sulfuric acidic bath containing an organic suppressor and an accelerator to achieve a bright and fine-grained Cu crystallization, especially developed for semiconductor application but primary for damascene processing.

Copper posts are not reflowed and therefore can be used to provide high aspect ratio structures for fine pitch bumping without reducing the bump stand-off height. For the interconnect to the substrate a solder tip is added on top of the post or on the board pad. Copper posts are plated and typical structures are shown in Fig. 3.13.

Unlike solder plating copper post require in-via plating with very thick resist layers of 70+ µm.

Copper deposition will attract more attention in the future by the upcoming trend to 3D integration. TSV (Through Silicon Vias) are a major step for these technologies. A schematic graph shows the importance of Cu in Fig. 3.14.

The Si-etching by Deep Reactive Ion Etching (Deep-RIE) is a well-known process common in the MEMS industry (Bosch process). Major issues are the passivation of the Si, the deposition of the Cu-seed layer inside the Si via and the Cu filling process.

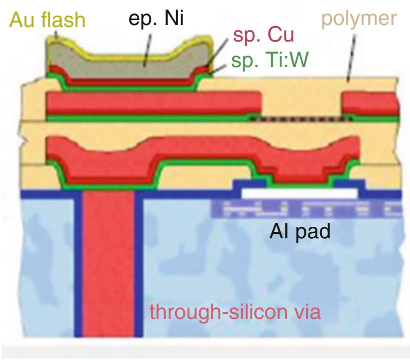


Fig. 3.14 Schematic graph of a TSV filled with Cu [15]

Plating of Ni

The used Ni electrolyte is chemically based on sulfamic acid and generally origins from electroforming industry, where very thick and low-stress layers are required. The Ni deposits look semi-bright, although a very low co-deposition of organic additives happens.

3.2.1.4 Plating of Alloys

For solder bumping, different kinds of stannous alloy electrolytes and a pure Sn bath are used, all of it basing on methane sulfonic acid. Here, a nonporous crystallization and a low co-deposition rate of organics are required to prevent bubble formation inside the bump during the final reflow process. The main issue of plating alloys at the same time is the difference of the electromotoric force (EMF). The difference in EMF is only 10 mV for PbSn. Pb is slightly more noble than Sn. The actual cell potential departs from the equilibrium value by the over-potential or cell polarization which includes all concentration, diffusion and other effects. These can be minimized by bath agitation, high ion content, lower current density etc. In addition the seed layer must be thick enough to avoid voltage drops below a few mV. A deposition rate of over 4 μm/min with a uniformity of less 3 % (1 sigma) over 300 mm can be achieved with optimized conditions using automatic plating systems.

The difference for the potentials for Sn, Ag, and Cu are given in Fig. 3.15.

Inhibitors (for example organic additives etc. or complexing agents) are acting different for each ion partner resulting in a different change in the deposition potential. For example a complexing agent can change the activity of the given ions. If cyanide ions are added to a Cu-ion solution the following reaction will take place:

$$Cu^{+} + 3\ CN^{-} \rightleftharpoons [Cu(CN)3]^{2-}$$

The concentration of Cu^{+}-Ion will be reduced. Free Cu(I) concentrations is in the range of ~0.5×10^{-7} mol/l but it will be reduced by ~0.25 mol/l NaCN or KCN

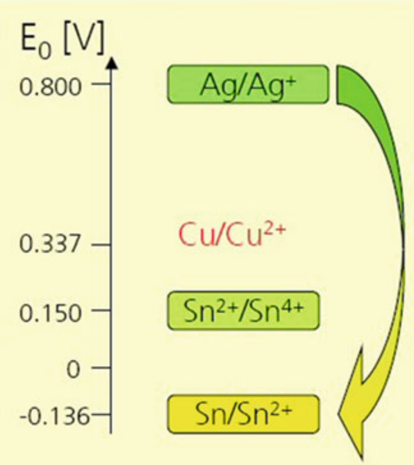


Fig. 3.15 EMF for Ag, Cu, and Sn

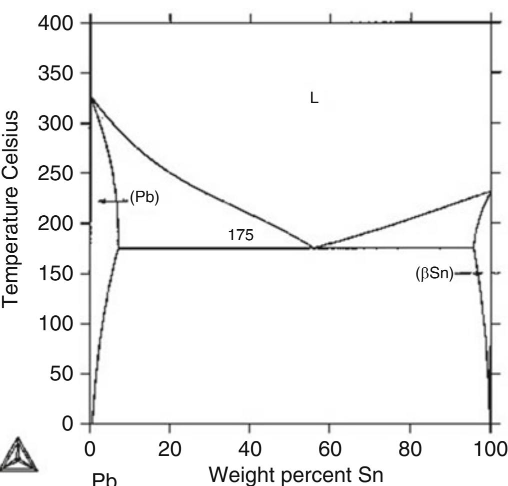


Fig. 3.16 Phase diagram of PbSn [23]

down to 10^{-26} mol/l. The EMF (Cu+/Cu): is +0.35 V but will change to −1.0 V by adding this cyanide salt. Such complexing agents are necessary to reduce the EMF of Ag/Ag^{+} to be co-deposited with Sn in case of SnAg ECD.

One advantage of processing PbSn compared to the lead-free alternatives is the melting temperature at the eutectic point shown in the phase diagram in Fig. 3.16 [23].

The eutectic melting temperature of 183 °C is changed only slightly by varying the solder composition. This is totally different for SnAg and SnCu as shown in Fig. 3.17 [22].

For SnAg3.5 % the eutectic melting temperature is 221 °C but will increase to 285 °C by adding another 2.3 % Ag. Similar for SnCu going from 227 °C melting

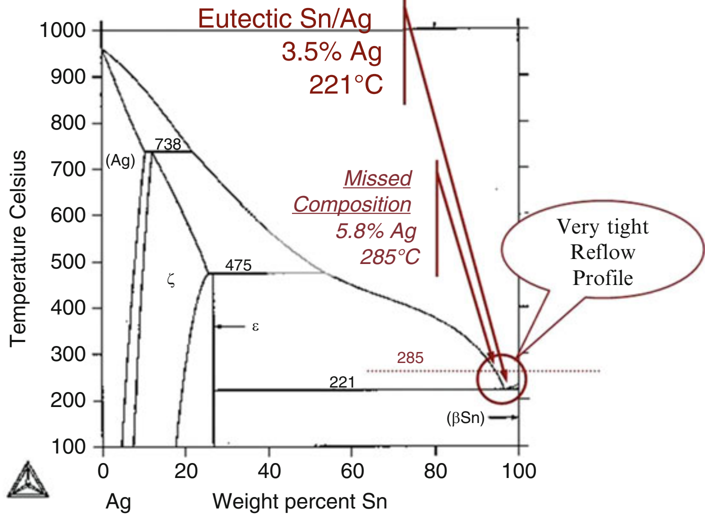


Fig. 3.17 Phase diagram of SnAg [23]

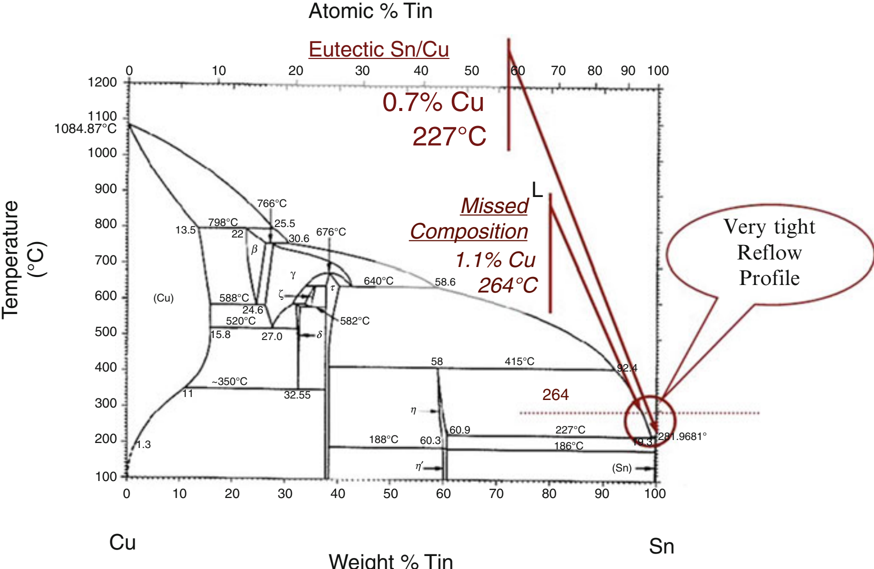


Fig. 3.18 Phase diagram of SnCu [22]

temperature for the 0.7 % Cu eutectic composition to 264 °C for 1.1 % Cu (Fig. 3.18). This means an exact control of the electroplating process to avoid cold joint in the assembly process. The schematic structure of a final SnAg bump is given in Fig. 3.19.

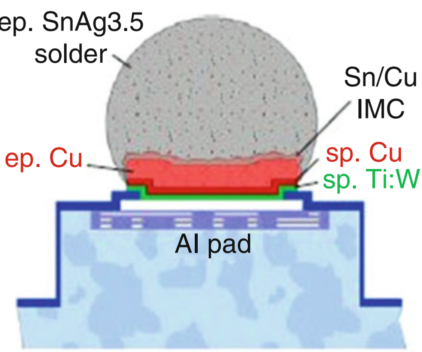


Fig. 3.19 Schematic of a SnAg bump

During ECD processing, several electrochemical reactions are taking place between the solution and both anode and cathode [21]. Permanent changes of the bath chemistry require an individual monitoring of each relevant organic as well as inorganic compound to ensure consistent plating results. The contents of metallic ions, acids, and other anionic compounds are commonly determined by titrimetric or spectrophotometric analysis. The organic additives are consumed during the electroplating process mainly by co-deposition, anodic decomposition, and drag-out loss. All these reactions are nearly in proportion to the current throughput and therefore predictable. Additionally, a time-dependent degradation of organics can often be observed. In the field of microelectroplating, cyclic voltametric stripping (CVS) is the most utilized method for the analysis of organic substances. The concentrations of all components mentioned above must be maintained by periodically adding the respective concentrates to the electrolyte. Today, fully automatic control units including the self-dosing function are commercially available for high-volume manufacturer. In the end, the changes of specific gravity, the bleed-out of photoresist and the accumulation of organic breakdown products limit the life-time of the plating bath.

The monitoring of the metal ions inside the lead-free alloy electrolytes is of utmost importance because of its small amount of Ag and Cu. Small variations would lead to significant changes of the solder composition and therefore to an impact of the reflow manner. Due to the pure tin anodes mainly used for lead-free solder bumping, the Ag and Cu is not continuously dissolving but must be frequently replenished. The composition of the deposited Sn/Ag and Sn/Cu alloys can be determined by differential scanning calorimetry (DSC) and energy-dispersive X-ray analysis (EDX).

A further issue for these high Sn lead-free alternatives is the growth of Whiskers which are fine, hair-like needles, mono-crystals. The diameter of these crystals is in the range of ~1 μm and the length can go up to the mm-range baring the risks of electrical shorts between interconnects. They result due to stress from the substrate or by intermetallic layers. Growth rate is in a range of 3 μm/month to 130 mm/month were tin atoms are diffusing to the bottom of the whisker being highest around 50 °C. In general alloys are reducing the risk of tin whiskers due to limiting the mobility of Sn atoms.

The compositions of selected electrolytes are summarized in Table 3.5.

Table 3.5 Overview of selected electrolytes [15]

	Cu electrolyte	Ni electrolyte	PbSn electrolyte	Au electrolyte	Sn electrolyte
Contents	$CuSO_4$, sulfonic acid, chloric acid, grain refiner and leveler, wetting agent	$Ni(NH_2SO_3)_2$, boric acid, grain refiner and wetting agent (if necessary)	$Sn(CH_3SO_3)_2$, $Pb(CH_3SO_3)_2$, methane sulfonic acid, grain refiner, wetting agent, oxidation inhibitor	$(NH_4)_3[Au(SO_3)]$, ammonium sulfite, ammonia, organic grain refiner and leveler, complexing agents and stabilizers	$Sn(CH_3SO_3)_2$, methane sulfonic acid, grain refiner, wetting agent, oxidation inhibitor
Metal concentration	20 g/l Cu	45 g/l Ni	Total of 28 g/l	12 g/l Au	20 g/l Sn
Temperature	25 °C	50 °C	25 °C	55 °C	25 °C
pH value	<1	4.0	<1	7.0	<1
Current density	10–30 mA/cm^2	10–30 mA/cm^2	20 mA/cm^2	5–10 mA/cm^2	7–15 mA/cm^2
Current efficiency	Nearly 100 %	>95 %	Nearly 100 %	>95 %	Nearly 100 %
Anode material	Phosphorus alloyed copper	S-activated nickel pellets	Appropriate Pb/Sn alloys	Platinum-covered titanium	Pure tin

3.3 Recent Advances on Bumping Technologies

3.3.1 Low Cost Solder Bumping Process: Solder Bump Maker

A low cost maskless bumping technology (called solder bump maker—SBM) was demonstrated by Bae et al. recently [24]. It does not need any special equipment and mask tools. The SBM is a paste material which is composed of the solder powder and polymeric resin which has fluxing capability to eliminate the oxide layers on the solder powder and electrode and to facilitate the formation of the solder bumps on the electrodes using the rheological behavior of the droplets without any aligning methods. Figure 3.20 shows the flow chart of a simple process to make solder bump using the SBM. As shown in Fig. 3.20, the SBM paste was applied on the metal pads through a stencil. When the printing was completed, the device was reflowed (to melt the solder powders). Then, the stencil was completely removed, and resin and excessive solder powders were cleaned off. Better bump height uniformity was achieved by coining process. The author also demonstrated the interconnect formation using SMB bumps in a five die TSV package.

3.3.2 Nanoporous Interconnect

Oppermann et al. successfully deposited nanoporous gold bumps on silicon wafers by electroplating a silver–gold alloy followed by etching away the silver [25]. An open-porous cellular structure of gold at meso-scale is left on top of the bumps (Fig. 3.21). Porous gold bump-to-bump bonding was demonstrated at low temperature (200 °C) and low force bonding conditions. The porous interconnects have very promising properties, like compressibility and reduced stiffness, which should result in higher bond yield and extended reliability.

3.3.3 Inclined Microbump

Park et al. demonstrated the fabrication of an inclined conductive bump (ICB) which provides the uniform electrical conductivity, elastic modulus, and controlled bump deformation [26]. The ICBs were fabricated on the test wafer with the inclined angles of the ICB were 70 and 80° and a pitch of 30 μm. The singulated chips assembled onto an organic substrates with Cu/Au pads using a thermo-compression bonding method, and acceptable contact resistance was achieved (Fig. 3.22).

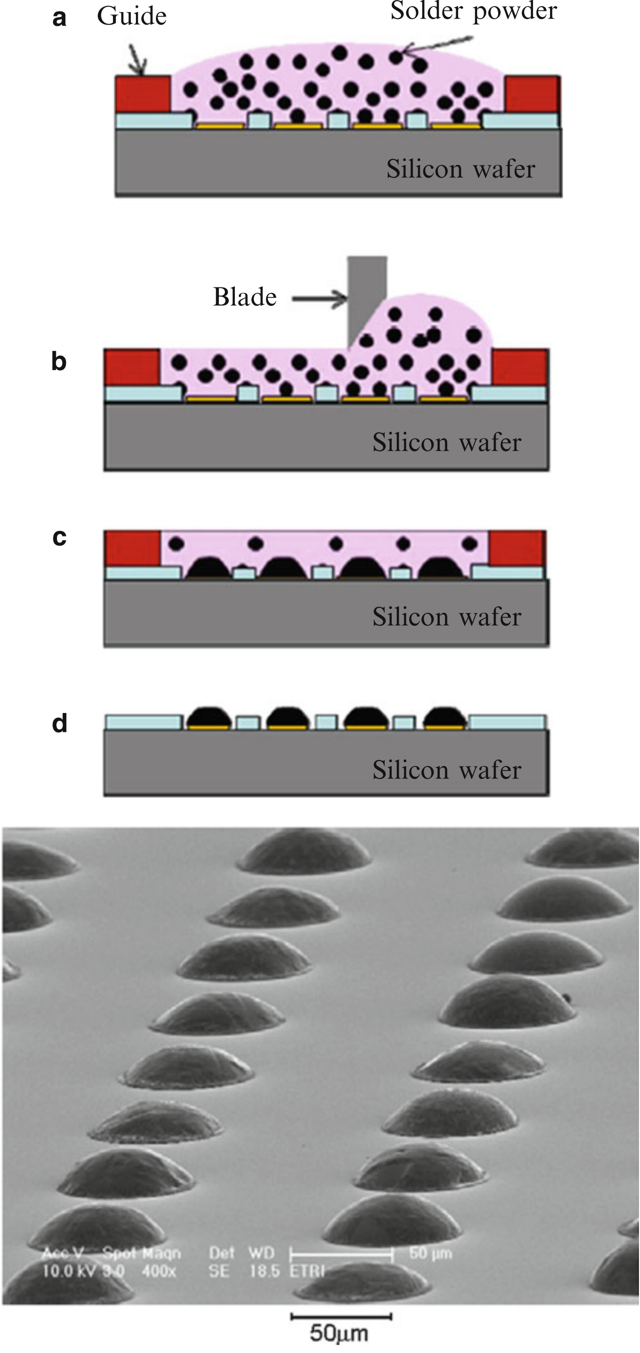


Fig. 3.20 *Top*: process flow of SBM: (**a**) SBM paste deposition, (**b**) flattening, (**c**) reflow, and (**d**) cleaning; *Bottom*: image of formed solder bumps [24]

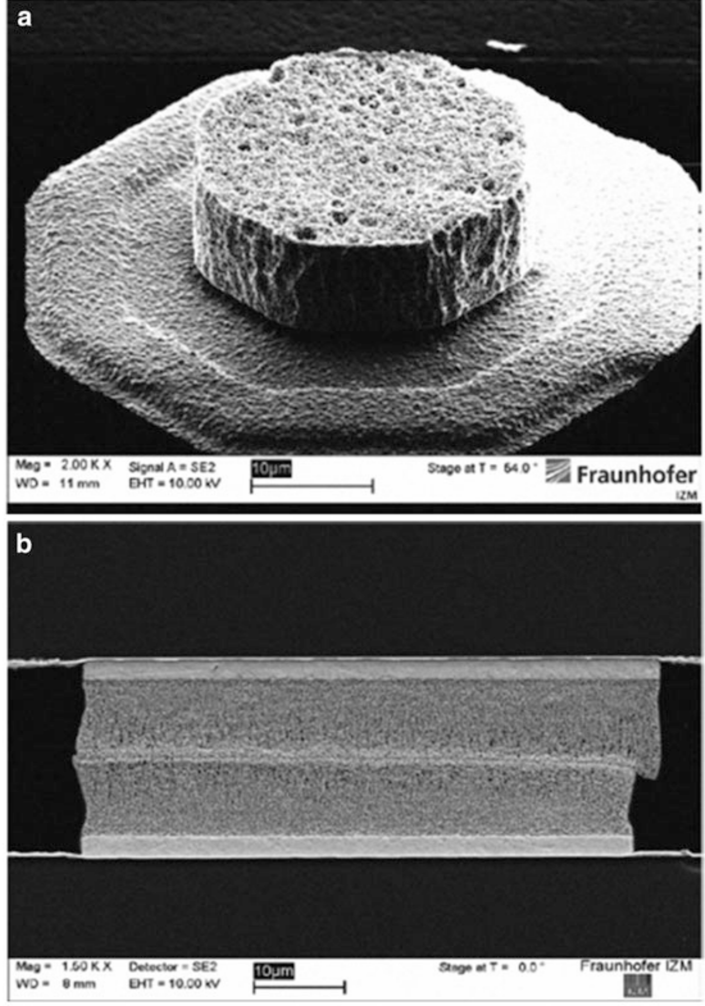


Fig. 3.21 (**a**) Nanoporous gold bump and (**b**) the interconnection formed between two nanoporous gold bumps [25]

3.3.4 Fine Pitch Imprinting Bumping

Corsat et al. utilize imprinting technique to make pyramid shaped In and Au–Sn bumps at pitch as low as 4 µm after the residue layer is etched away (Fig. 3.23) [27]. The pyramid solder could be reflowed and form bumps. The author also imprinted thermoplastic conductive polymer pyramid bumps, and then utilized the deformability of the polymer bump to make the electrical contact onto the metal pads on a substrate under a low bonding force and low temperature (i.e., 70 °C).

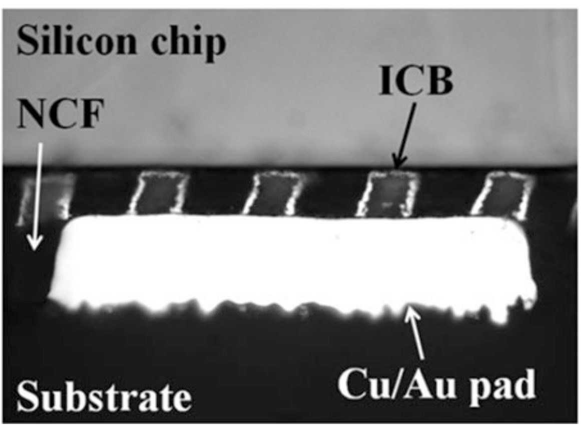


Fig. 3.22 Chip with ICBs bonded onto a Cu/Au pad of a substrate [26]

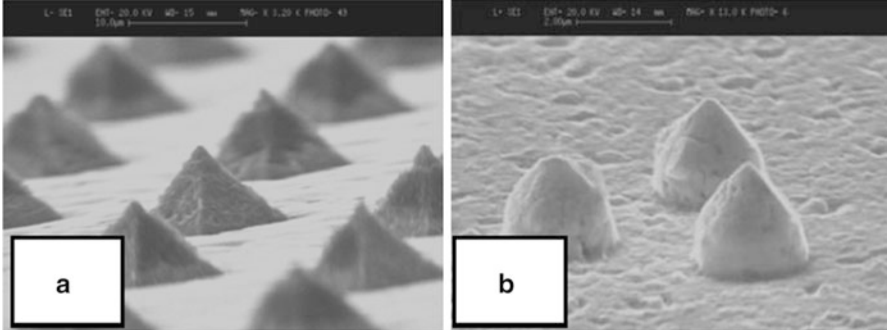


Fig. 3.23 Indium pyramid shaped bumps made by imprinting technique: (**a**) 30 μm pitch; (**b**) 4 μm pitch [27]

3.3.5 Solder Bumping by Liquid Droplet Microgripper

Figure 3.24(I), (II) shows schematic process steps for solder bump creation by using a droplet microgripper [28]. The droplet microgripper can be realized simply by patterning a hydrophobic layer (such as Teflon) on a hydrophilic substrate such as glass (Fig. 3.24a). A liquid droplet is held on each hydrophilic opening site by the attractive force between the solid and the liquid after the substrate is dipped into a liquid bath (Fig. 3.24b). Then, each droplet can catch and hold an encountered solder ball (Fig. 3.24c). After flux coating on a soldering substrate (Fig. 3.24d, e), the droplet microgripper, which is holding the solder balls, is aligned and approached to the soldering substrate (Fig. 3.24f). When the solder balls held by the droplet microgripper contact the flux of the soldering substrate, the solder balls

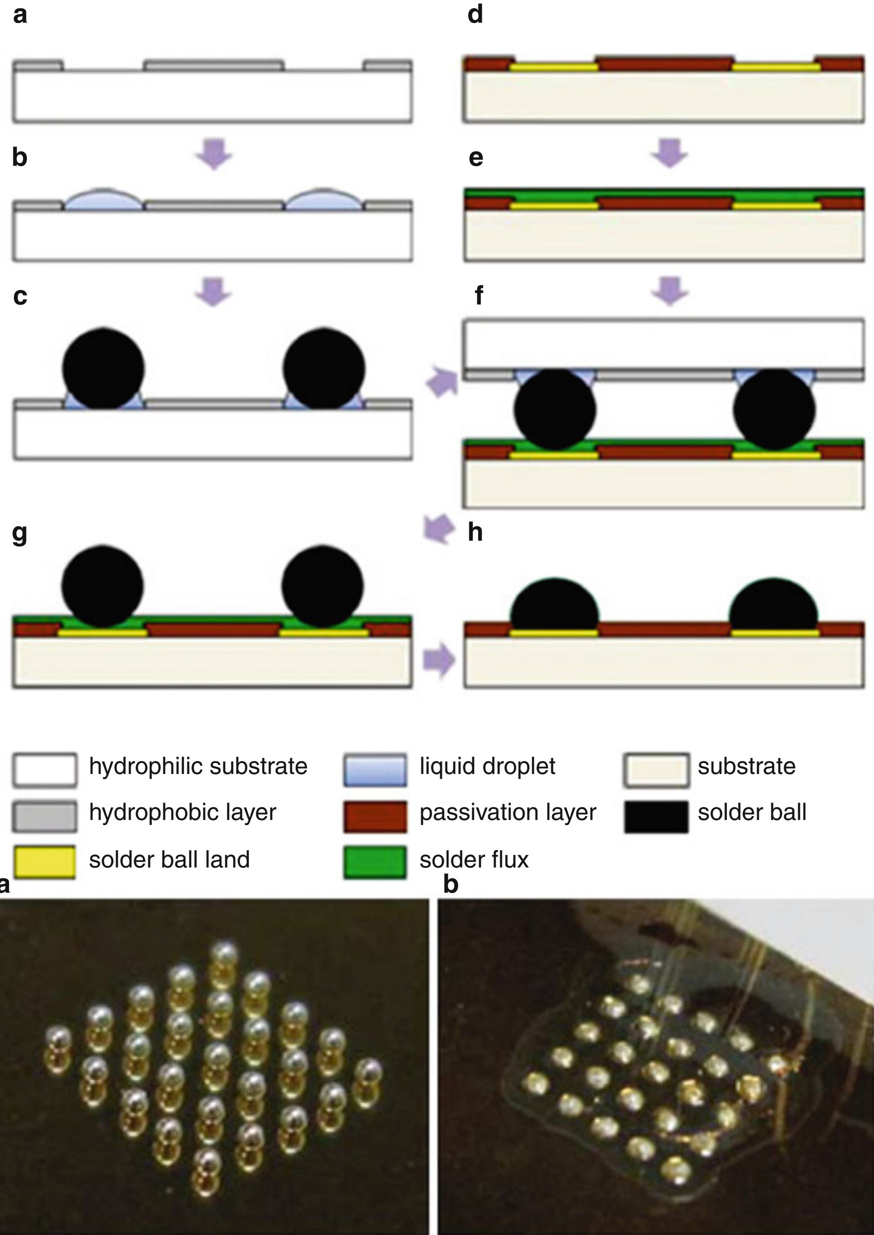


Fig. 3.24 (I): Schematic process step for solder bump creation by using liquid droplet microgripper: (**a**) patterning hydrophonic layer on the hydrophilic substrate; (**b**) Forming liquid droplets on hydrophilic openings; (**c**) Solder balls caught by liquid droplets; (**d**, **e**) Coating flux on the soldering substrate; (**f**) Align microgripper onto the solder substrate; (**g**) Solder balls were caught by the flux layer on the soldering substrate; (**h**) Reflowing solder ball to form bumps. (II): (**a**) transferred solder balls from droplet microgripper to soldering substrate; (**b**) Formed solder bumps after the reflowing [28]

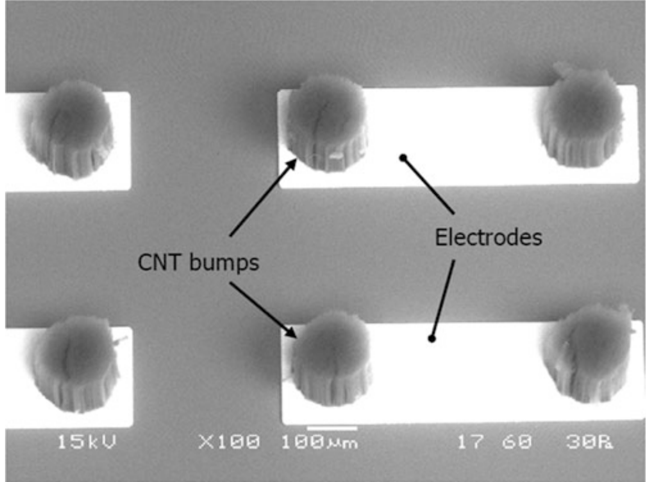


Fig. 3.25 SEM image of CNT bump formed by pattern-transfer process [29]

are transferred to the soldering substrate because the adhesive force of the flux is stronger than the liquid droplet (Fig. 3.24g). Finally, solder bumps are created on the solder ball lands after a reflow process (Fig. 3.24h).

3.3.6 CNT Bumps

Soga et al. reported on CNT bumps for the flip-chip interconnect instead of solder bumps [29]. The CNT bumps were formed on the electrodes on a chip by the CNT pattern-transfer-process (Fig. 3.25). It was found that coating the CNT bumps with gold greatly improved the contact resistance of the bumps with the chip and the substrate, resulting in a lower bump resistance of 2.3 Ω. The author also demonstrated the resilience and flexibility of CNT bumps which can realize thermal stress free flip-chip structure. The CNT bumps were able to absorb a displacement between the chip and substrate up to 10 ~ 20 % of the CNT bump height.

More recently, Wei et al. investigated a simple assembly process of carbon nanotube (CNT) bumps which could be used for flip-chip interconnects [30]. First, high density aligned CNT bumps were grown on both top and bottom substrates. By employing typical flip-chip bonding technique, the CNT bumps on the flipped top substrate were aligned with those on the bottom substrate. With applying a bonding force, the CNT bumps on the top substrate were pressed and penetrated into the CNT bumps on the bottom substrate. The penetration depth increased with increased bonding force (Fig. 3.26). After CNT insertion, the CNTs from both top and bottom substrates were held together with van der Waals force, and the CNT interconnection bumps were formed. It was found that the conductivity of CNT interconnection bump

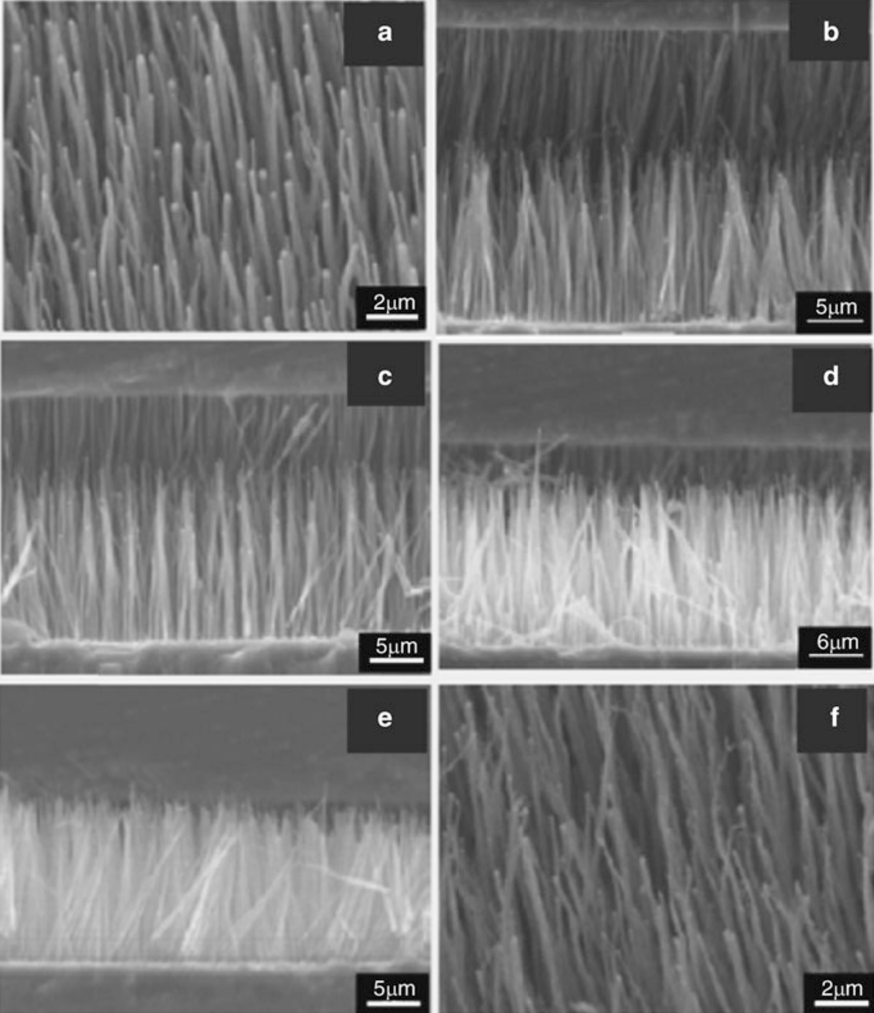


Fig. 3.26 SEM images of the as grown CNT bump (**a**), the cross-sectional views of the CNT structures after their assembly process: zero force loading (**b**), 1 kg force loading (**c**), 2 kg force loading (**d**), and 3 kg loading (**e**), and the CNT bump that has went through 50 cycles of interpenetration and detachment processes (**f**) [30]

was found to be of many magnitudes higher than that of the silver paste. It was demonstrated that the CNT bumps have a much superior electrical properties over the typical metallic bumps. Resistance of the inserted CNT structure is also found to have decreased with the increase in the penetration depth. The linear decrease in resistance with increase in penetration depth further proves that the CNT–CNT contacts are having metallic behavior and are ohmic. The lower resistance measured from the interpenetrated CNT structure has demonstrated the advantage of our approach with respect in creating low contact resistance, CNT-based interconnects.

References

1. Töpper M (2009) Wafer level chips scale package. In: Lu D, Wong CP (eds) Materials for advanced packaging. Springer, New York, pp 547–600
2. Töpper M, Tönnies D (2005) Microelectronic packaging. In: Geng MH (ed) Semiconductor fabrication handbook. McGraw-Hill, New York, pp 21.1–21.54
3. Tummala R, Rymaszewski E, Klopfenstein A (1997) Microelectronic packaging handbook, Part 1–3. Chapman & Hall, New York
4. Garrou P, Turlik I (1998) Multichip module technology handbook. McGraw-Hill, New York
5. van Roosmalen (2004) There is more than moore. Proceedings of 5th International Conference on Mech. Sim. and Exp. in Microelectronics and MST, EuroSim2004
6. Lau J (1995) Ball grid array technology. McGraw-Hill, New York
7. Kosuga K (1997) CSP technology for mobile apparatuses. Proceedings International Symposium on Microelectronics 1997, Philadelphia, Oct 1997
8. Seraphim DP, Lasky R, Li C-Y (1989) Principles of electronic packaging. McGraw-Hill, New York
9. Reichl H (1998) Direktmontage. Springer, Berlin
10. Puttlitz K, Totta P (2001) Area array interconnection handbook. Kluwer Academic Publishers, Boston
11. Dietrich L, Wolf J, Ehrmann O, Reichl H (1998) Wafer bumping technologies using electroplating for high-dense chip packaging. Proceedings Third International Symposium on Electronic Packaging Technology (ISPT'98), Beijing (China), 17–20 Aug 1998
12. Müller U (1996) Anorganische strukturchemie. Teubner, Wiesbaden
13. Ruhmer K, Laine E, O'Donnell K, Hauck K, Manessis D, Ostmann A, Töpper M (2007) UBM structures for lead free solder bumping using C4NP. EMPC, June 2007
14. Kawanobe T, Miyamoto K, Inaba Y (1981) Solder bump fabrication by electrochemical method for FC interconnection. Proceedings of IEEE Electronics Components Conference, S 149, May 1981
15. Dietrich L, Toepper M, Ehrmann O, Reichl H (2006) ECD wafer bumping to future demands on CSP, 3D integration, and MEMS. Proceedings of the ECTC 2006, San Diego
16. Ostmann A, Motulla G, Kloeser J, Zakel E, Reichl H (1996) Low cost techniques for flip chip soldering. Proceedings of Surface Mount International Conference, San José, Sept 1996
17. Anhöck S, Ostmann A, Oppermann H, Aschenbrenner R, Reichl H (1999) Reliability of electroless nickel for high temperature applications. International Symposium of Advanced Packaging Mat. Conference, Braselton, USA, Mar 1999
18. Baumgartner T, Manessis D, Töpper M, Hauck K, Ostmann A, Reichl H, Goncalo CT, Jorge P, Yamada H (2007) Printing solder paste in dry film—a low cost fine-pitch bumping technique. Proceedings of EPTC 2007, Singapore
19. Gruber PA, Belanger L, Brouillette GP, Danovitch DH, Landreville J-L, Naugle DT, Oberson VA, Shih D-Y, Tessler CL, Turgeon MR (2005) Low-cost wafer bumping. IBM J Res Dev 49 (4/5):621–639
20. Hwang J (1996) Modern solder technology for competitive electronic manufacturing. Mc Graw-Hill, New York
21. Tu K-N (2007) Solder joint technology materials, properties, and reliability. Springer, New York
22. Kim B (2004) Leadfree solder deposition for wafer level packaging applications. 5th Annual SECAP East Asia Seminar Series, Nov 2004
23. Chevalier P-Y (1988) A thermodynamic evaluation of the Ag-Sn system. Thermochimica Acta 136:p45–p54

24. Bae H-C, Choi K-S, Eom Y-S, Lim B-O, Sung K-J, Jung S, Kim B-G, Kang I-S, Moon J-T (2010) 3D SiP module using TSV and novel solder bump maker. Proceedings of 2010 Electronic Components and Technology Conference, pp 1637–1641
25. Oppermann H, Dietrich L, Klein M, Wunderle B (2010) Nanoporous interconnects. Proceedings of 3rd Electronic System-Integration Technology Conference (ESTC), pp 1–4
26. Park A-Y, Kim S-R, Yoo CD, Kim T-S (2011) Development of inclined conductive bump (ICB) for flip-chip interconnection. Proceedings of Electronic Components and Technology Conference (ECTC), pp 880–885
27. Corsat F, Davoine C, Gasse A, Fendler M, Feuillet G, Mathieu L, Marion F, Pron A (2006) Imprint technologies on conductive polymers and metals for interconnection and bumping purposes. Proceeding of 1st Electronics System integration Technology Conference, pp 1336–1341
28. Lee S-Y, Chang J-H, Kim D, Ju BK, Pak JJ (2010) Solder bump creation by using droplet microgripper for electronic packaging. Electron Lett 46(19):1336–1338
29. Soga I, Kondo D, Yamaguchi Y, Iwai T, Mizukoshi M, Awano Y, Yube K, Fujii T (2008) Carbon nanotube bumps for LSI interconnect. Proceedings of 2008 Electronic Components and Technology Conference, pp 1390–1394
30. Wei J, Yung K, Tay BK (2009) Formation of CNT bumps for interconnection applications. SIMTech Tech Rep 10(2):76–79

Chapter 4
Flip-Chip Interconnections: Past, Present, and Future

Sung-Kwon Kang, Da-Yuan Shih, and William E. Bernier

Abstract Flip-chip interconnection technologies have been extensively used in many microelectronic applications for high performance systems as well as consumer electronics in recent years. There are several types of flip-chip interconnects being used today in the industry, which include high-Pb solder bumps joined to a ceramic substrate, high-Pb bumps on chip joined to eutectic PbSn on a laminate substrate, all eutectic PbSn bumps, Pb-free bumps, Cu pillar bumps, and Au-stud bumps. Steady improvements have been made in high performance packages, such as having interconnects greater than 10,000 I/O's with a pitch less than 200 μm, migrating from ceramic to low-cost organic substrates, replacing high-Pb with Pb-free interconnects, and others. Fundamental reliability issues, especially with the Pb-free solder bumps, remain to be solved. As the expiration of European Union (EU) Reduction of Hazardous Substance (RoHS) exemption approaches soon, Pb-containing solder bumps for flip-chip interconnects are expected to be phased out in a few years. However, the recent introduction of fragile low-k, or very fragile ultra low-k interlayer dielectrics (ILD) into the back-end interconnect architectures in the advanced semiconductor devices has imposed serious technical challenges in integrating Pb-free technologies for high-performance systems. How to control the cracking of low-k ILD layers during chip joining has become an urgent issue to be resolved together by the semiconductor and packaging industries. Another reliability challenge associated with the implementation of Pb-free solder in flip-chip packages is the poor electromigration (EM) performance of Sn-rich solders. This is mainly due to the highly anisotropic crystal structure of Sn, causing the fast solute diffusion along its *c*-axis, combined with the aggressive interfacial reactions with under-bump metallurgy (UBM) and laminate pads.

S.-K. Kang (✉) • D.-Y. Shih
IBM Research Div, Yorktown Heights, NY, USA
e-mail: kang@us.ibm.com

W.E. Bernier
IBM Systems & Tech Group, Hopewell Junction, NY, USA

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_4, © Springer Science+Business Media New York 2013

In this chapter, the recent progress of flip-chip assembly processes is presented with ceramic substrates and organic laminates, including the new bumping technologies such as C4NP and Cu pillar bumping. In addition, various reliability challenges of Pb-free, flip-chip applications are extensively discussed along with possible solutions for specific applications.

4.1 Evolution of Flip-Chip Interconnection Technologies

Flip-chip interconnection technologies have been extensively used in many microelectronic applications for high performance systems as well as consumer electronics in recent years. There are several types of flip-chip interconnects being used today in the industry, which include high-Pb solder bumps joined to a ceramic substrate, high-Pb bumps on chip joined to eutectic PbSn on a laminate substrate, all eutectic PbSn bumps, Pb-free bumps, Cu pillar bumps, and Au-stud bumps. Steady improvements have been made in high performance packages, such as having interconnects greater than 10,000 I/O's with a pitch less than 200 μm, migrating from ceramic to low-cost organic substrates, replacing high-Pb with Pb-free interconnects, and others. Fundamental reliability issues, especially with the Pb-free solder bumps, remain to be solved. As the expiration of European Union (EU) Reduction of Hazardous Substance (RoHS) exemption approaches soon, Pb-containing solder bumps for flip-chip interconnects are expected to be phased out by 2016. However, the recent introduction of fragile low-k, or very fragile ultra low-k interlayer dielectrics (ILD) into the back-end interconnect architectures in the advanced semiconductor devices has imposed serious technical challenges in integrating Pb-free technologies for high-performance systems. The hard or less compliant Pb-free solder bumps coupled with the fragile ILD layers in a large chip can cause cracking in the ILD structure, which often leads to mechanical/electrical failures of the chip. How to control the cracking of low-k ILD layers during chip joining has become an urgent issue to be resolved together by the semiconductor and packaging industries. Another reliability challenge associated with the implementation of Pb-free solder in flip-chip packages is the poor electromigration (EM) performance of Sn-rich solders. This is mainly due to the highly anisotropic crystal structure of Sn, causing the fast solute diffusion along its *c*-axis, combined with the aggressive interfacial reactions with under-bump metallurgy (UBM) and laminate pads.

In this chapter, the recent progress of flip-chip assembly processes is presented with ceramic substrates and organic laminates, including the new bumping technologies such as C4NP and Cu pillar bumping. In the following, various reliability challenges of Pb-free, flip-chip applications are extensively discussed along with possible solutions for specific applications.

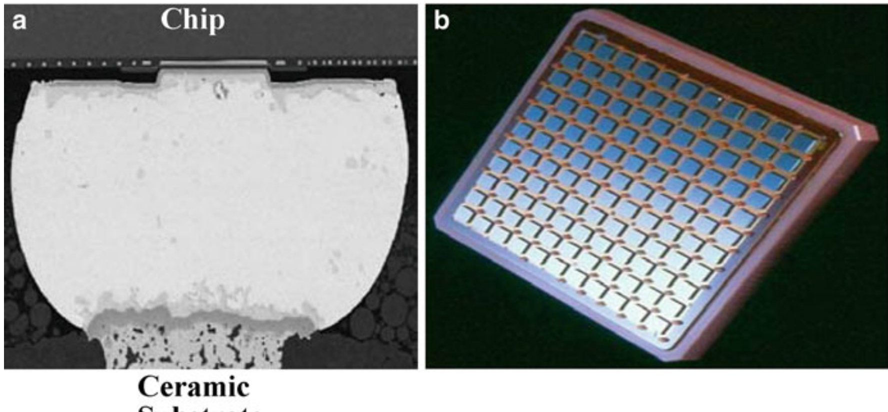


Fig. 4.1 (**a**) Flip-chip solder joint (97Pb/3Sn) formed between a chip and a ceramic carrier. (**b**) Multi-chip module containing 121 flip chips mounted on a glass ceramic substrate (CTE ~3 ppm/°C) without underfill [5]

4.1.1 High-Pb-Based Solder Joint

The first, and the original type of flip-chip interconnect structure, was introduced by IBM in 1964 [1], which is based on bumping high-melting temperature, high Pb–Sn solder (3–5 % Sn and 97–95 % Pb) on chip and subsequently joined to a ceramic carrier at a temperature greater than 320 °C. This is a fully melting Controlled-Collapse-Chip-Connection (C4) solder joint as shown in Fig. 4.1. The standoff gap between the chip and substrate is determined by the surface tension of the solder, solder volume and pad sizes on the chip and substrate. The C4 bump height, Sn content, and size of a chip are the major factors determining the fatigue life of the C4 joints during the power-on and -off machine cycles [2]. Due to the close CTE match between the chip (CTE, ~3 ppm/°C) and ceramic substrate (CTE, 3–6 ppm/°C), along with the highly compliant high-Pb soft solder, no underfill is needed to meet the reliability requirements. In this structure, a simple UBM consisting of a multilayer structure of Cr/CrCu/Cu/Au along with the high Pb solder, has been successfully used for several decades and demonstrated excellent reliability [3–5].

4.1.2 High-Pb Solder on Chip Joined to Eutectic Solder on Laminate Carrier

The move to organic substrates drives the need for a lower-melting-temperature solder to replace the high-Pb solder. The organic materials are typically epoxy-based and are not stable at temperatures in excess of 250 °C for a prolonged time. One way to circumvent this problem is to deposit low-melting-temperature eutectic

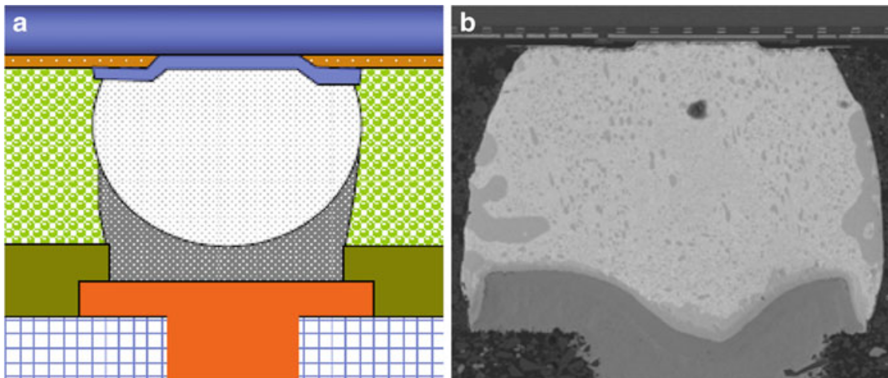


Fig. 4.2 The package assembly showing a high-Pb solder bump on chip joined to eutectic solder on the laminate and then underfilled; (**a**) schematic drawing and (**b**) a cross-sectional view of an actual solder joint

PbSn solder on laminate pads, or on the high-melting bumps on the chip. As shown in Fig. 4.2, a high-Pb solder bump on chip is joined to a low-melting-point solder on a laminate pad. This combination allows the assembly of chip and laminate at temperatures compatible with a low-cost organic laminate. This is a huge breakthrough that IBM (Japan) pioneered to allow direct chip attachment (DCA) to low cost organic laminate carrier [6, 7]. To ensure that the solder joint can survive the large strain generated from the global thermal expansion mismatch between the laminate (~17 ppm/°C) and the chip (~3 ppm/°C), a thermally compatible underfill material was developed. When the chip/underfill/laminate is bonded into one entity, it deforms simultaneously and the relative movement between the chip and laminate and, therefore, the strain on solder bumps is greatly reduced [8]. Figure 4.2 shows the package assembly with a high-Pb solder bump on a chip joined to eutectic solder on the laminate, followed by flux residue cleaning and then underfilled. The same UBM, Cr/CuCu/Cu, which has been successfully used for high-Pb C4 bumps for decades, cannot survive multiple reflows and high temperature storage tests. When the dual solder layers are reflowed for multiple times (up to seven times), the Sn content in the solder matrix near UBM continues to increase with each reflow, accelerating the reaction with UBM and eventually leading to complete separation of the Sn–Cu intermetallics from the base of UBM, as shown in Fig. 4.3. The solution to the problem is either the use of a more robust reaction barrier layer, such as Ni, or to increase the Cu thickness. NiV and NiFe barrier layers [9] are also reported to have good interface stability when joined to high-Sn solders, such as full-melt eutectic 37Pb/63Sn or Pb-free (Sn >95 %) solder C4 bumps. After joining a chip to laminate, the flux residues need to be thoroughly cleaned and the assembled module must be underfilled and cured to overcome the CTE mismatch between the chip and organic laminate. With underfill, even large chips with dimensions up to 15 mm, survive the thermal cycling tests (−45 to 100 °C, one cycle per hour). On the other hand, without underfill, even a small chip failed before 300 cycles [10, 11].

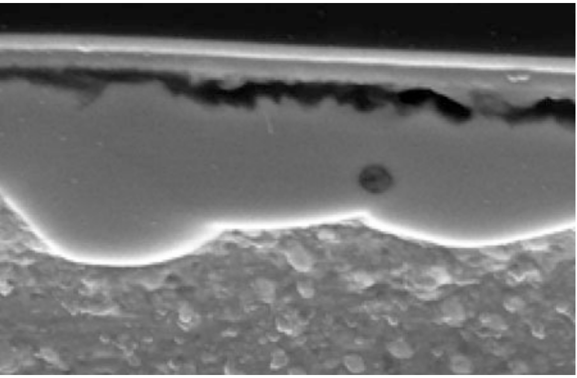


Fig. 4.3 Open failure caused by separation of solder/IMC from UBM

4.1.3 Pb-Free Solder Joint

Although the amount of solder volume used in flip-chip interconnects is much smaller than used in BGA or surface mount solder joints, the environmental legislation required to replace Pb-containing solders with Pb-free solders in flip-chip applications. Pb-free solders, such as popular SnCu, SnAg, and SnAgCu, all contain more than 95 wt% of Sn, which not only have higher melting temperatures than eutectic PbSn, but are highly reactive with UBM and substrate pads. This necessitates a thicker UBM and/or more robust reaction barrier layer, such as Ni or NiFe [9] to survive multiple reflows and various reliability tests without being totally consumed.

For Pb-free solder applications, Ni UBM is a popular choice as a reaction barrier layer because it reacts much slowly and wet well with Pb-free solders. As shown in Fig. 4.4, SnAg solder is joined between Ni UBM on the chip and Cu-OSP laminate pad, and subsequently subjected to high temperature storage at 170 °C for 1,000 h. The solder joint remains intact with controlled growth of intermetallic compounds at the interface between UBM and solder, as shown in Fig. 4.5. Another popular UBM choice for Pb-free solders is thick Cu UBM. On Cu, SnAg or SnAgCu solder forms two intermetallic phases; Cu_6Sn_5 adjacent to the solder and Cu_3Sn adjacent to the Cu [12–14]. The interfacial reactions between Cu, Ni, and Pb-free solders are further discussed later in other sections.

4.1.4 Cu Pillar Joint

Cu pillar bump is another flip-chip structure [15]. It takes an advantage of the mature Cu electroplating technology commonly used in BEOL structure. There are two versions; one with solder cap plated on top of Cu pillar, and another with solder totally applied on the laminate pad. Figure 4.6 is a SEM image of an area array of electroplated Cu pillars with Sn cap on top at 100 μm pitch [16]. The cross-sectional view is shown in Fig. 4.7, as used by Intel on the microprocessor chips [17].

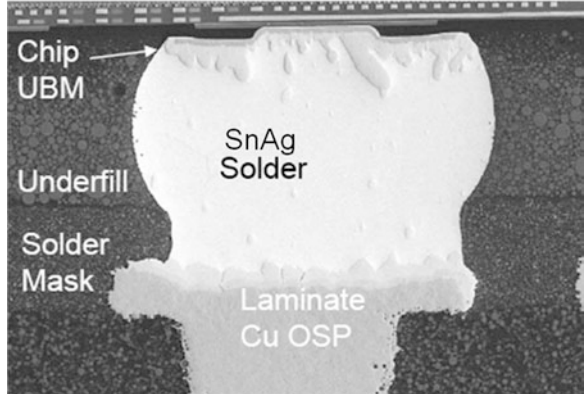


Fig. 4.4 SnAg solder bump joined between Ni UBM and Cu OSP, and tested for high temperature storage

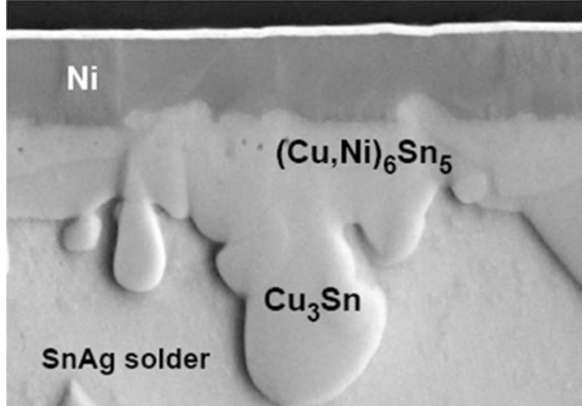


Fig. 4.5 Controlled IMC growth of Ni UBM with SnAg solder

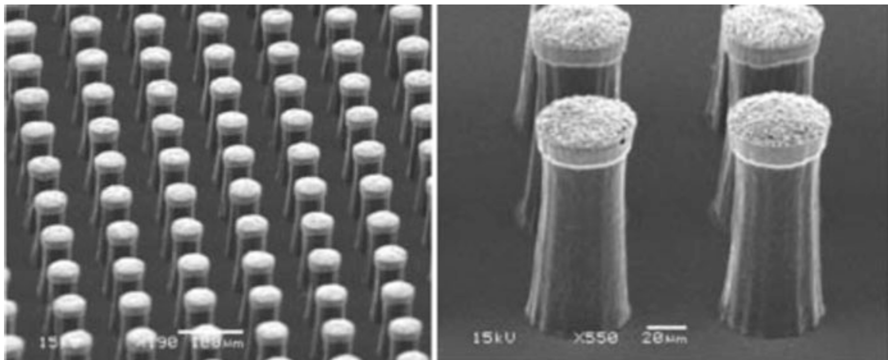


Fig. 4.6 An area array of electroplated Cu pillar bumps with Sn cap at 100 µm pitch and 80 µm height [16]

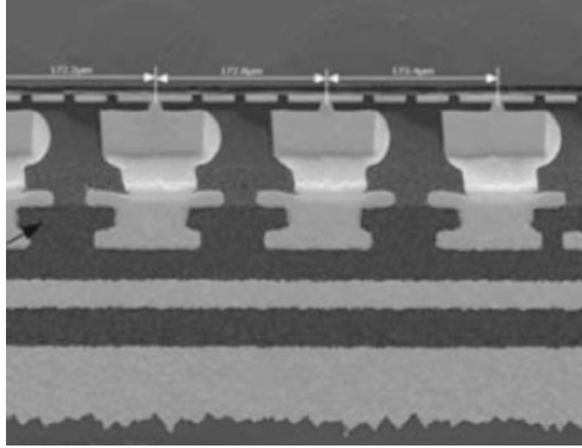


Fig. 4.7 Cross-sectional SEM image of Cu bumps joined to laminate substrate [17]

Its superior EM performance, along with the other advantages, makes it an attractive flip-chip interconnects technology. However, the stiff Cu pillar along with a thin solder joint is less compliant and, thus, makes it more difficult to integrate on the chip BEOL structure without cracking the fragile low-k ILD materials. The stress mitigation methods and its bumping process will be further discussed in the later sections.

4.2 Evolution of Enabling Assembly Technologies

4.2.1 Wafer Thinning and Wafer Dicing (Covering Cu/low k Devices)

Wafer thinning is practiced widely today for a variety of applications. Processing consists of placing a bumped wafer with bumps down onto a backer sheet. The wafer is immersed in deionized water and then ground with a rough grit polish or so-called kiss polish to complete the operation. Typically, 300 mm wafer have thickness of 780 μm, while 200 mm wafers or smaller wafers have 730 μm in thickness. Wafers may be ground with the kiss polish to remove as little as 50 μm of material. Combined with the rough grit polish, wafers as thin as 50 μm thick may be produced.

Wafer dicing today is practiced using two general processes—saw dicing and laser dicing. In dicing the wafer has already been bumped and is attached to a backer tape on the back side of the wafer. In saw dicing the diamond saw disk is selected in thickness depending on the width of the kerf and the thickness of the wafer. The wafer is typically immersed in deionized water and a stream of deionized water is directed on the active cut area. One or more passes may be used with the saw to separate individual devices which are held in place by the backer tape prior to

picking. Saw blades may be beveled or straight depending on the width of the cut and the cut finish desired. Laser dicing is similar in initial preparation. In addition, a protective coating material is applied to the bumped wafer surface to guard against ablated material from the laser action depositing on the bumps or passivation surface. One or more passes by the laser may be needed to dice the wafer. A combined process is also often used wherein initial laser passes are used followed by completion with saw dicing. The laser dicing protective coating is removed promptly after processing. Saw dicing has been used effective for years in manufacturing and can leave scallops and chipped silicon areas at the edge of the singulated devices. Laser dicing may leave a much rougher surface from the ablating action of the laser, but is much less likely to result in scalloping and edge chipping. The rough edge area on the singulated devices provides an excellent avenue for mechanical interlock of the underfill encapsulant in subsequent processing.

4.2.2 Wafer Bumping

There are a number of wafer bumping technologies available for flip-chip interconnects, which provide mechanical support for and electrical path to the device. The wafer bumping technologies currently used include; electroplating of solder bumps on the UBM on a wafer, screen printing of solder paste on a wafer, direct deposition of molten solder on the UBM using the C4NP (Controlled Collapse Chip Connection New Process), Cu pillar bumping, and conductive adhesives including both isotropic and anisotropic materials. Each has its own niche in the application space based on the consideration of cost, application suitability and flexibility.

In each bumping process, the UBM structure is used to establish the wetting area of the solder on the device and provide a base for chip joining through the formation of intermetallic compounds at the interface during solder reflow cycles. Common UBM types include TiWCrCuCu, TiWNiV, TiWNi, TiWNiCu, and others (more discussion in the later sections). Typical solders used with these UBM types are PbSn, SnAg, SnAgCu, and SnCu of various compositions (refer to Pb-free solders section).

Plating processes for solder bumping include a series of steps outlined in Fig. 4.8. Incoming wafers must have interconnection via's available to process. The UBM metallization as described above is applied by plating or an appropriate alternate technique. A photoresist is deposited and patterned on the wafer over the device pads for bumping. The solder metallurgy is then plated in one or more steps to deposit the appropriate solder composition on the chip pads. The photoresist is then stripped from the wafer. The UBM metallization is then removed by etching in areas where solder has not been deposited. The solder bumps are then reflowed to form characteristic truncated sphere shapes.

Plating is widely used for wafer bump processing because plating tools are widely available and plating solution chemistry is abundant and economical. There are some inherent challenges to the processing that can affect yield of the process. Undercut can occur during the UBM etch process; not only removing the excess UBM

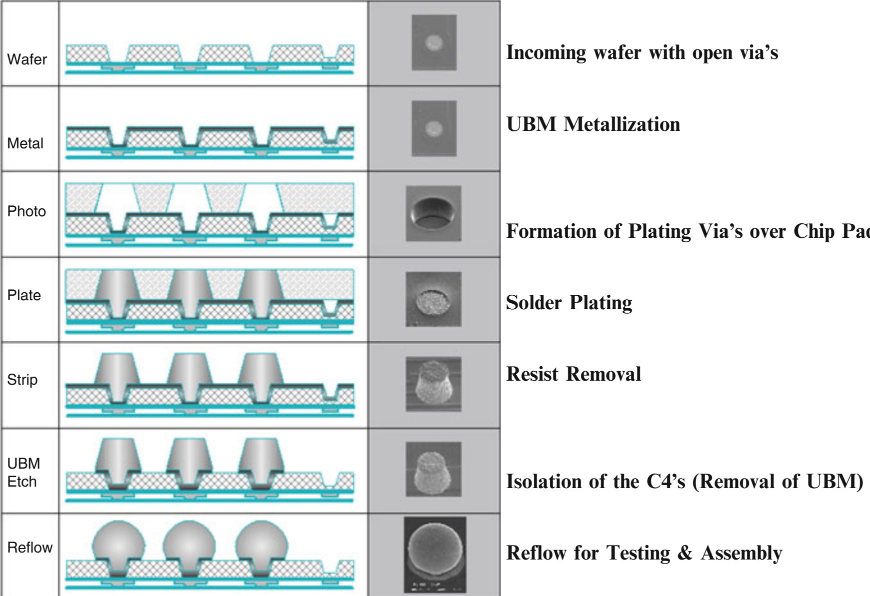


Fig. 4.8 Process outline for wafer bumping formed by electroplating. (Bernier WE, Pompeo F (2007) IC FCPBGA packaging: a tutorial. Presented at Georgia Institute of Technology Packaging Research Center, Atlanta, GA, 18 Sept 2007)

metallurgy but also undermining the UBM at the edges of the solder bump areas. Careful process control is required. Plating is also subjective to the current density in the plating process, causing significant variations in solder volume and height unless appropriate process control such as thieving is implemented. Photoresist stripping can leave residues if solvent selection, stripping solution agitation, and duration are not carefully controlled. Residues can result in metal contamination or electrical bridging to occur if the contamination extends between two or more solder bumps. Proper care in process set up and process control would facilitate excellent production results.

Solder paste screening has been widely used as a relatively inexpensive technique for applying bumps to wafers. The UBM metallurgy is initially applied to the wafer forming the appropriate metallization over the device pads. A screening mask can be formed from etched and/or laser processed metal sheets with holes formed in the pattern of the wafer bumps or, alternatively, photoresist can be applied to the wafer to provide the function of the mask and subsequently be removed after bump deposition. The solder paste contains very fine particles of solder balls, which are mixed with flux, solvent, and thixotropic agents. The solder paste is screened through the mask which is patterned with openings where the bumps are desired. After screening the solder paste is reflowed to form solder bumps. If solder resist were used for the mask it would subsequently be removed.

There are some key challenges with using solder paste screening which may limit its use and extendibility in manufacturing application. Solder pastes are well known

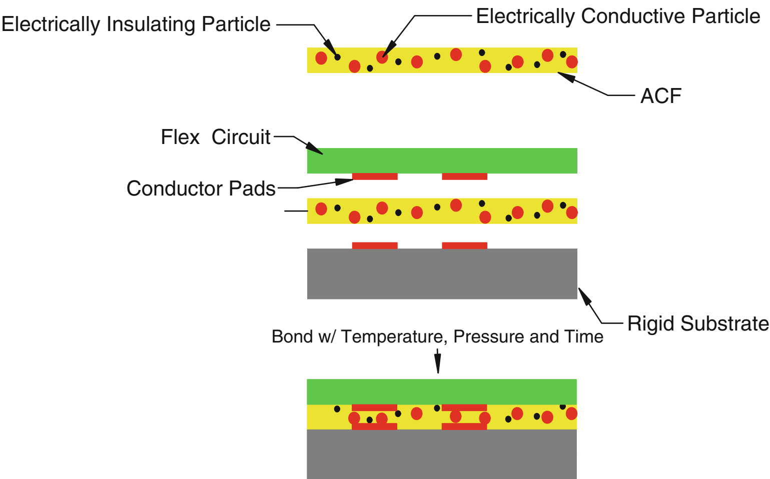


Fig. 4.9 A typical bonding process using anisotropic conductive film. (Bernier WE, Pompeo F (2007) IC FCPBGA packaging: a tutorial. Presented at Georgia Institute of Technology Packaging Research Center, Atlanta, GA, 18 Sept 2007)

to form voids interior to the solder bumps. Large voids can result in low solder volume which may have yield or reliability concerns. The openings of solder mask screens can be easily blocked, which may prevent adequate solder deposition, resulting in yield loss or rework.

With anisotropic conductive adhesives an electrical connection is made by a monolayer of electrically conductive particles that are deformed in a compressive manner during the bonding process. The particle volume fraction in the adhesive film is below the percolation in order to ensure that the insulation resistance in the *x–y* plane is maintained over distance greater than the conductor spacing. The conductive particles which are used are typically either solid metal or metal plated polymer spheres. The particles can be coated with an electrical insulating layer that ruptures during bonding. In the bonding process, either the device or the carrier used should be adequately flexible—as noted in Fig. 4.9, where a flexible circuit carrier is used in assembly. If both connections on either side of the anisotropic film are rigid, then any lack of planarity between them will drive elastic recovery after bonding, which can result in increased contact resistance or electrical opens. During the bonding process it is essential to maintain electrical contact, alignment, parallelism, and uniform pressure across the anisotropic film. Typically the formed bond line is between 3 and 10 μm in thickness. A minimum level of 15–20 conductive particles is generally required in the bond area to provide reliable contact resistance to be maintained through the life of the application. The anisotropic electrically conductive film thickness selection is routinely based on the circuit line width and the pitch of the electrical contacts. There needs to be sufficient adhesive between the contacts to fill the space. The circuit line height is typically less that 14 μm to accommodate the anisotropic electrically conductive film thickness used. The

adhesive material properties are important to control based on several factors. The coefficient of thermal expansion of the particles should be matched to the adhesive. The glass transition temperature, adhesion strength, modulus of the adhesive and its resistance to moisture are all important to maintain processing capability and reliability.

C4NP solder bumping process, a novel wafer bumping technology, is described separately in the following section. In addition, the fabrication process of Cu pillar bumps is also discussed in a separate section.

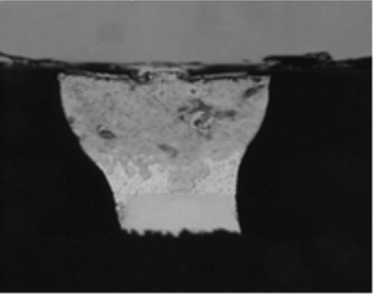
4.2.3 Flux and Flux Cleaning

Fluxes used in electronic packaging are formulated to react with stannous oxide and related species on the surface of solder bumps or pre-solder on the laminate bond pads. Species used in the flux system often include tertiary amines and carboxylic, dicarboxylic, or related organic acids. These species form complexes or chelates with the stannous oxide and related Sn(II) compounds. This reaction is important to provide wettable surfaces for joining at solder reflow temperatures. Typically, fluxes are not designed to react with stannic oxide or Sn (IV) compounds, which if formed, are customarily difficult to remove and thus may significantly inhibit solder wetting under reflow conditions. Commercially available fluxes have been classified into no-clean and solvent-clean categories. Solvent cleaning consists of two types: aqueous cleaning and nonaqueous cleaning fluxes. The no-clean flux provides the necessary chemical activity to sequester stannous oxide. The fact that the flux does not need cleaning does not mean there are no residues. Residues of the flux reaction products may indeed be present, but they satisfy insulation resistance requirements so the flux residues are not detrimental to the reliability of assembled products in temperature, humidity and bias testing. However, if they form a barrier to interfacial adhesion of underfill in subsequent processing, they may introduce additional concerns in other environmental stress testing for reliability. Therefore, the flux residues may need to be cleaned if possible. The alternative is to use a solvent-clean flux. For eutectic PbSn solder and many Pb-Free solders, water-soluble fluxes are preferred since their residues may be removed by deionized water rinsing at elevated temperatures or by aqueous solutions of surfactants or saponifiers washing and deionized water rinsing at elevated temperatures. The water-soluble flux is principally used in manufacturing and consists of an organic acid, surfactant, and a vehicle such as an alcohol based compound. It is preferable that the flux has a tacky nature to provide and maintain proper solder bump to pre-solder alignment. There are a variety of ways to apply flux for chip joining to a laminate chip carrier. The bumps of the chip can be dipped in flux, placed on the chip carrier and reflowed. Alternatively, flux can be dispensed or sprayed on the laminate chip carrier; the chip placed on the pre-solder pads of the laminate chip carrier and then reflowed. Finally, a combination of dip, dispense, or spray may also be employed. In the case of dipping, at least 50 % of the bump should be coated

with flux. In the case of the spray or dispensed flux on the laminate chip carrier, the complete chip footprint should be covered and the amount applied should be verified by weight. Time from dispense of the flux until reflow should be minimized to avoid excess contact time regardless of the flux used. Once flux application and reflow are completed using water-soluble flux, the flux residues must be removed within a limited process time window. The flux residues with water-soluble flux would be reactive and somewhat corrosive if left in contact indefinitely with the device joined to the laminate chip carrier. Various types of cleaning may be used including aqueous immersion and spraying. The cleaning action of spray is particularly amenable to inline processing through wash, rinse, and drying stages. Spray nozzle angle, pressure, temperature, and conveyor belt speed are all key parameters in setting up the cleaning profile. Other key mechanical parameters are the gap between chip device and laminate chip carrier, the size of chip and laminate chip carrier, and the pitch and density of solder interconnections.

4.2.4 Reflow Soldering, Thermo-compression Bonding

There are a variety of ways to connect conductive bumps electrically to laminate chip carriers. When solder bumping is employed on the device, solder reflow interconnection is often used. The purpose of solder reflow is to melt solder and to allow metallurgical reaction with tin and other constituents of the solder to form modest controlled amounts of intermetallic compounds to establish a solder joint with the laminate chip carrier. In order to carry out the solder reflow a thermal profile must be established; raising the temperature above the melting points of the solder bump and any pre-solder present on the laminate chip carrier. A time above liquidus is routinely established in the range of 40–220 s or longer. A range of peak temperature is established depending on solder type. For PbSn eutectic solders this peak temperature can range from 195 to 250 °C. For Pb-free solder the peak temperature can range from 230 °C to as high as 260 °C. Such thermal profiles are usually established by a thermocouple placed in the center of a chip footprint between the chip and the laminate chip carrier. Generally, a forced convection nitrogen oven is used although other furnace types such as infrared furnace and others may be used. Prior to chip joining a variety of perimeter visual inspections are performed when the device is placed on the laminate chip carrier. Rotational and *x*–*y* alignment are grouped as chip skew and are important for proper chip joining. After chip join reflow, additional perimeter visual inspections are performed. Chip skew is checked once again to verify that mechanical motion and vibration have not disturbed the initial bump alignment. Non contact of a solder bump with solder on the laminate chip pad can occur, particularly in the perimeter and corner locations of the device. This is highly dependent on product configuration such as chip size, chip construction, laminate carrier size as well as appropriate setup parameters for inspection, e.g., lighting or fixturing. Non-wetting joints are produced during the reflow when two solder surfaces come in contact, but due to a



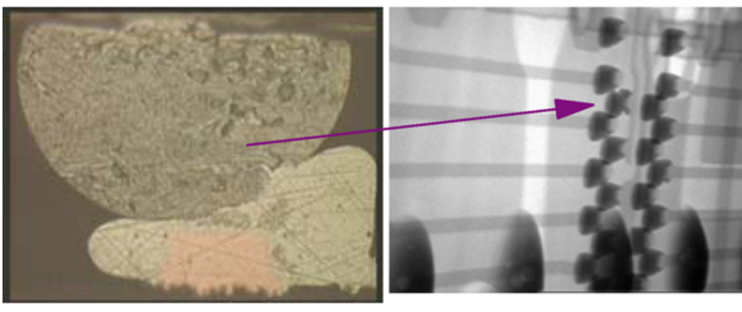


Fig. 4.10 Examples of flip-chip solder joints in FCPBGA packages. (Bernier WE, Pompeo F (2007) IC FCPBGA packaging: a tutorial. Presented at Georgia Institute of Technology Packaging Research Center, Atlanta, GA, 18 Sept 2007)

variety of factors may not be successfully joined together. Potential causes of non-wet joins include foreign materials or contamination on the joining surfaces, intractable oxides or inadequate flux on the chip bump or laminate pad, or laminate warpage in the chip site, or placement damage of a cracked device or tilted chip on the chip carrier. Non-wetting joints can be not only a yield detractor, but occasionally will form contact opens which may be missed in the final test and can fail in the field. A few examples of solder joints formed in FCPBGA packages are shown in Fig. 4.10.

As the visual inspection can only inspect the perimeter rows and maybe the second row of the solder joint array, nondestructive process control is commonly employed, such as X-ray and ultrasonic techniques. X-ray can sometimes be used, but resolving fine pitch solder joints can be challenging. Destructive analysis such as chip pull test is often performed to verify the integrity of solder joints. Taffy pull separation with

PbSn solder is acceptable. For Pb-free solders, various failure modes in chip pull tests are observed; taffy pull in solder, interfacial separation at the intermetallics, via separation in the chip device, UBM metal separation, or solder separation from the UBM. Chip pull test can be used to confirm noncontact failures in solder joints. However, due to the variety of separation modes with Pb-free solder joints, the tensile force data collected from chip pull test may not be consistent indicators.

Thermo-compression bonding is an alternate technique for joining bumped-chip devices to laminate chip carriers. It provides a combination of temperature and pressure over a specified duration of time to form solder joints. Techniques such as Tape Automated Bonding (TAB), anisotropic conductive adhesives reviewed earlier as well as isotropic conductive adhesive bonding, and wafer-level underfill device bonding all may use such a method. The combination of time and temperature provides the necessary energy to bond bumps to pads using appropriate joining metals (TAB), to promote chemical reaction and mechanical interlock (anisotropic conductive adhesive), to facilitate thermoplastic and thermosetting reactions (isotropic conductive adhesive), or to complete solder joining, flow and cure of underfill for wafer level underfill.

4.2.5 Underfill and Over-Mold

Careful selection of underfill encapsulant is important to maintain the integrity of the chip joint through mechanical handling and to minimize stresses associated with circuit package processing. Underfill encapsulation process consists of three stages: pretreatment, dispense, and cure. In pretreatment, the laminate chip carrier is subjected to a moisture bake out to minimize the residual content during processing and cure. Often the bake out is performed in nitrogen atmosphere to prevent oxidation of metal surfaces, particularly copper metal surfaces. Surface adhesion enhancements may be applied to key interfaces to promote underfill adhesion to the laminate chip carrier solder mask, chip passivation, and chip device sides and edges. The interfacial adhesion is particularly critical at the corners and perimeter of the device where mechanical stresses tend to be highest. The dispense pass requires control of the underfill encapsulant shot size customized to the chip dimensions. Bias heating of the laminate chip carrier and the underfill encapsulant is commonly used, and once again this is customized to both the chip device and the laminate chip carrier to promote the capillary flow of encapsulant. Various fill dispense patterns can be used such as a dot, line, of “L” pattern for optimum throughput while minimizing void occurrence. The dot fill consists of a single dot of material placed next to the chip and after a control time to allow for capillary flow another identical dot of material is placed in exactly the same place in the same manner. This sequence is repeated as necessary until the underfill encapsulant flows out from underneath all sides of the chip. This must occur before dispensing the fillet pass. In line dispense, the encapsulant is dispensed once along the long side of the chip device. The material must flow underneath the chip through to all four sides

prior to dispensing the fillet pass. In the “L-shaped” dispense, a continuous line dispense is performed along a long side and a short side of the chip device. Once again the underfill encapsulant must flow underneath the chip through all sides prior to the dispensing the fillet pass. With each dispense procedure the two key factors in consideration are the throughput and void population, which must be optimized for fastest throughput and highest yield. After under-chip dispense is completed, a final fillet pass is performed to provide adequate protection for edge and corner, high mechanical stress areas as determined by fillet height and shape requirements for the particular application. Normally the final fillet dispense avoids the areas involved in the initial fill dispense areas. The final stage is to cure the underfill encapsulant. The cure profile is established to achieve full underfill polymerization for proper time and temperature. The final cure temperature is also selected to minimize the induced warpage of the cured assembly at room temperature. By monitoring chip bending after cure and room temperature stabilization, the highest level of bending in the assembly is known to occur during the cure processing, and it is particularly high for Pb-free solders.

After the three process stages are complete, there is customarily post underfill encapsulation inspection. Visual criteria typically specify that a continuous fillet must be present around the chip device. The height of the fillet must meet minimum coverage of the chip height at the midpoints of the side. The fillet may be lower at the chip corners, but the lower portion of chip corners must not be exposed. The fillet extends from the edge of the chip to the minimum geometry onto the laminate surface. Underfill material on the back of the chip is allowed, provided that it does not interfere with lid bond line requirements. Missing fillets are not allowed. Bubbles in the fillet that exposed on the laminate or chip surface or sides are not allowed. The maximum size of any bubbles should be limited regardless of whether the laminate or chip is covered. No foreign material may be embedded in the cured underfill material. No vertical corner cracks are allowed in the underfill fillet. Horizontal edge cracks may be allowed depending on application requirements. A nondestructive inspection by CSAM (Computerized Scanning Acoustic Microscopy) is routinely performed to confirm the void criteria under the chip and others. Figure 4.11 highlights a large void in the underfill to chip device passivation interface which would fail typical inspection criteria. Figure 4.12 shows extensive delamination at the underfill to solder mask interface of the laminate chip carrier.

Over-mold processing consists of two general techniques: dispensed dam and fill material, and injection molding processing. Similar stages of processing are used for over-mold processing as for underfill encapsulation: pretreatment, dispense, and cure. Pretreatment requires laminate chip carrier bake out often in nitrogen environment to remove residual moisture. In the dam and fill dispense technique the under-chip dispense has already been completed. A high viscosity dam material is dispensed at the outline of the over-mold area. A separate lower viscosity fill material is then dispensed within the dam outline to fill the volume to adequate height in order to meet application requirements. Bias heating may be used for the laminate chip carrier as well as the dispensed dam and fill materials. Cure of the dispense materials is usually performed together. Injection mold processing requires special tooling to generate a

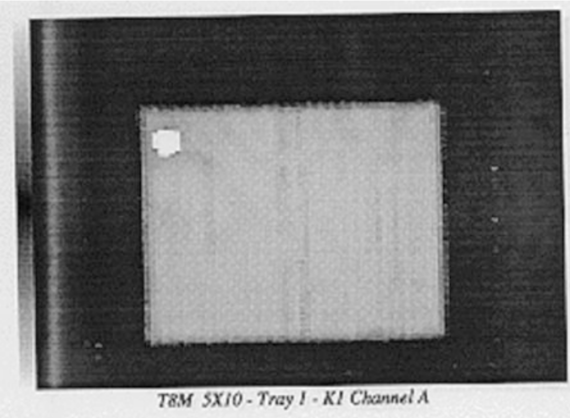


Fig. 4.11 Underfill void found with CSAM. Underfill encapsulant voids can be caused by dispense processing, dispense pattern, and surface contamination. (Bernier WE (2008) Flip-chip PBGA assembly—quality and reliability challenges. Presented at IMAPS Upstate NY and Garden State Chapter Fall 2008 Packaging Symposium, Endicott, NY, 2 Oct 2008)

Fig. 4.12 Delamination between underfill and solder mask found with CSAM. Delamination can be caused by underfill cure problems and surface contamination. (Bernier WE (2008) Flip-chip PBGA assembly—quality and reliability challenges. Presented at IMAPS Upstate NY and Garden State Chapter Fall 2008 Packaging Symposium, Endicott, NY, 2 Oct 2008)

mold representing the desired shape and size of the over-mold area. The over-mold material is injected into the mold to fill the mold cavity. The over-mold resin is then cured at controlled time, temperature and pressure to yield an encapsulated chip device with low warpage. As with underfill encapsulation, similar care must be exercised in set up and dispense so yield issues with voids do not occur.

4.2.6 Quality Assurance Methodologies

Several examples of typical quality assurance stress tests are listed in Table 4.1. These tests are industry standard procedures commonly used for qualification of FCPBGA applications. Normally, there are initial preconditioning tests including moisture soak

Table 4.1 Typical standardized stress testing performed for qualification

Test	Conditions	Specification	Duration	Note
Temp/humidity/bias	85 °C/85 % RH/3.6 V	JEDEC A101	1,000 h	1
Deep thermal cycle	−55 to 125 °C	JESD22-A104-B	700 cycles	1
Thermal cycle	−25 to 125 °C	JESD22-A104-B	1,000 cycles	1
Thermal cycle	0 to100 °C	JESD22-A104-B	3,000 cycles	1
Power cycle	25 to 125 °C	JEDEC Draft	1,000 cycles	1
Card level shock	100, 200, 340 g	JESD22-B-110	2/1.5/1.2 ms	1
Low temp storage	−65 °C	JESD22-A119	1,000 h	1
High temp storage	150 °C	JESD22-A103-C	1,000 h	1
Highly accelerated stress test (HAST)	130 °C/85 % RH/3.6 V	JEDEC A110-B	96 h	1
Vibration on card	1.04 G, 0–500 Hz, 3 axis	MIL STD 810F	3 h	1
Electromigration	150 °C/0.7 A	JEDEC Draft	2,000 h	1
Thermal cycle (Sn whisker)	−55 to 85 °C	JESD22A121.01	1,500 cycles	1
Temp and humidity (Sn whisker)	60 °C 87 % RH	JESD22A121.01	4,000 h	1

Note 1: JESD22-A113D Level 3 or 4 preconditioning is done prior to testing
(Bernier WE (2008) Flip-chip PBGA assembly—quality and reliability challenges. Presented at IMAPS Upstate NY and Garden State Chapter Fall 2008 Packaging Symposium, Endicott, NY, 2 Oct 2008)

and exposures to three reflow cycles. After this base test, other stress tests are performed such as thermal cycling, temperature/humidity/bias testing, extended temperature storage testing, shock and vibration testing, and tin whisker testing for new Pb-free solders. Various failure modes may result from these tests if the technology is not adequately robust. These failure modes include solder joint cracking, device cracking; solder bridging, electrical shorting due to electromigration, and others.

4.3 C4NP Technology

Controlled Collapse Chip Connection—New Process (C4NP) technology is a novel solder bumping technology developed by IBM [18–23] to address the limitations of existing wafer bumping technologies. Through continuous improvements in processes, materials and defect control, C4NP technology has been successfully implemented in the manufacturing of 300 mm Pb-free solder bumped wafers. Both 200 and 150 μm-pitch products have been qualified and are in volume production. Extendibility of C4NP to 50 μm-fine-pitch microbump application has been demonstrated with the existing C4NP manufacturing tools. Target applications for microbumps are three-dimensional (3D) chip integration and the conversion of memory wafers from wire bonding (WB) to C4 bumping.

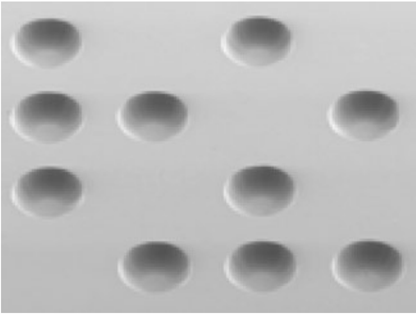


Fig. 4.13 C4NP glass mold with etched cavities [18]

4.3.1 C4NP Wafer Bumping Processes

C4NP process starts with the fabrication of a glass mold, as shown in the SEM image of Fig. 4.13, in which the I/O pads of UBM of an entire wafer is replicated as a mirror image of tiny cavities etched into the glass plate. These cavities are filled with solder as the mold is scanned beneath a fill head, as shown in Fig. 4.14. The fill head contains a reservoir of molten solder and a slot through which the solder is injected into the mold cavities. The cavity depth and diameter determine the volume of each solder bump that will be subsequently transferred to the wafer. The filled mold is inspected automatically and then aligned below a wafer with exposed UBM pads facing the mold. Mold and wafer are heated above the solder melting point in formic acid vapor to activate the UBM pads and solder surface, and then they are brought into contact. The solder forms spherical balls which transfer from the mold to the UBM pads on the wafer, where they wet and solidify. Subsequently, the wafer and mold are separated, and the mold is cleaned for reuse. Figure 4.15 describes this process flow.

4.3.2 Mold Fabrication and Solder Transfer to Wafer

C4NP mold is made of borofloat glass, which has a CTE closely matched to silicon. Photolithographically defined pattern is used to create the cavities by wet etching. The molds are scanned beneath a solder injection head (in Fig. 4.14) which fills the cavities with molten solder precisely to the top surface of the mold. Therefore, the solder volume transferred to the wafer at contact is a direct function of the glass cavity volume. The processes of the solder fill, auto inspection, solder transfer and cleaning are automated using Mold Fill, Mold Inspection, Solder Transfer and Mold Cleaning tools. The solders chosen for wafer bumping usually do not wet to the glass mold, so upon heating, the solder alloys form spherical balls in the cavities, as described in Fig. 4.16. The reflowed balls protrude above the surface of the mold by 10–20 μm depending on ball size and cavity. Note from Fig. 4.16 that the balls are not uniformly formed at the center of the mold cavity. The alignment of the mold

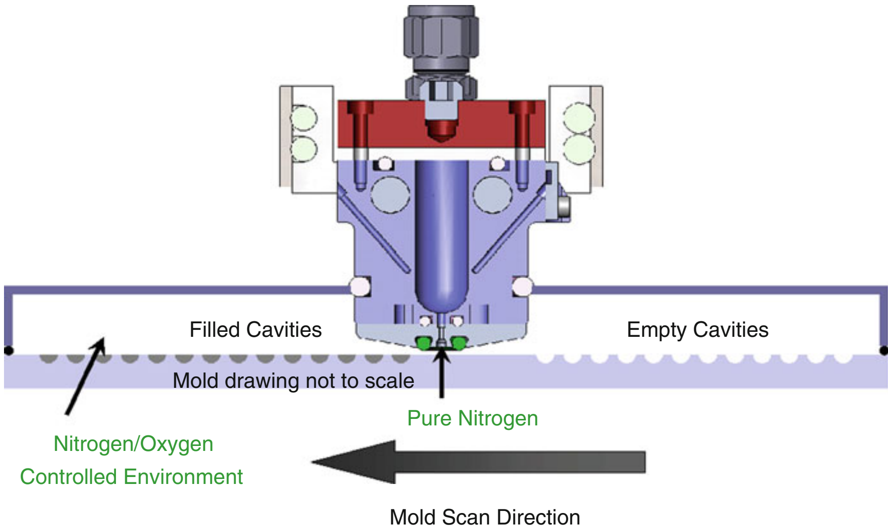


Fig. 4.14 Scan fill head to inject molten solder into cavities of glass mold [18]

cavities to the corresponding UBM pads is sufficient to assure that the solder wets to the correct UBM pads.

The filled molds are aligned with the wafer as shown in Fig. 4.15. After alignment, the mold and wafer are heated and are brought into close proximity/contact, allowing the molten solder balls to wet to the appropriate UBM pads where they preferentially remain when the wafer and mold are separated (Fig. 4.17).

For the 200, 150 and down to 50 μm pitch applications [21, 23], as-received glass molds have been successfully used for mold fill and wafer transfer. For further extension to very fine pitch applications, <50 μm, where solder volume becomes extremely small, to ensure a successful solder transfer, the flatness of the glass surface needs to be improved. And a technique to increase the depth and sidewall angle of mold cavities is essential to maximize the solder standoff height above the glass surface that would overcome a local non-flatness.

4.3.3 Wafer Bumping Yield Improvements

C4NP bumping yield has improved significantly to meet manufacturing yield requirements. Defect root cause analysis has resulted in process improvements in patterning the UBM pads, as well as in the mold and mold fill areas which contributed to robust yield improvements. Figure 4.18 illustrates the significant yield improvement over the years [20]. The yield data is derived from RVSI inspection of the 200 μm pitch product wafers. Yield learning model showed a 15 % defect reduction per month since the start of the C4NP program. Initially, the major yield detractor

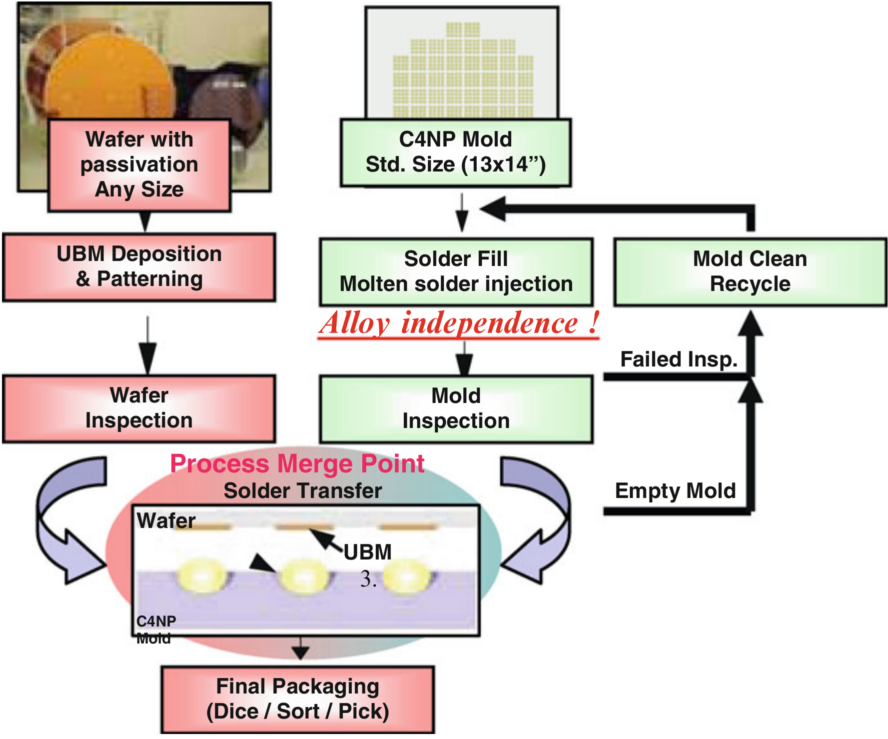


Fig. 4.15 C4NP process flow [18]

was contamination. With the installation HVM tools with FOUP to FOUP automatic handling and tightened cleanliness control, this type of defects was mostly eliminated. Also, as the quality of glass mold and mold fill processes improve, a significant reduction in missing C4s, along with improvements in volume uniformity and coplanarity, has contributed to improvements in transfer yield.

All these improvements have resulted in excellent bumping yield for both the 200 and 150 μm pitch applications. Best wafers have consistently achieved 100 % yield. The feasibility for 50 μm pitch wafer bumping was successfully demonstrated in the same manufacturing environment using the same set of C4NP tools for mold fill and wafer transfer [23]. The mold inspection tool (MIT) which works well for the 150–200 μm pitch was unable to handle the high density bumps (~11,000 bumps per chip at 50 μm pitch) due to lack of pixel density. Using an improved inspection tool developed by RVSI, very high bumping yield was demonstrated. The good results were largely attributed to improvements in the volume uniformity of mold cavities, eliminating contaminations and using pure nitrogen in the fill environment.

Figure 4.19 compares mold filling under N_2/O_2 ambient (Fig. 4.19a) to that under pure N_2 (Fig. 4.19b). More solder bridging defects between adjacent cavities are observed under N_2/O_2 mixture environment, while bridging is mostly eliminated in N_2 ambient. Solder bridging for standard C4s (pitch $\geq$150 μm) does

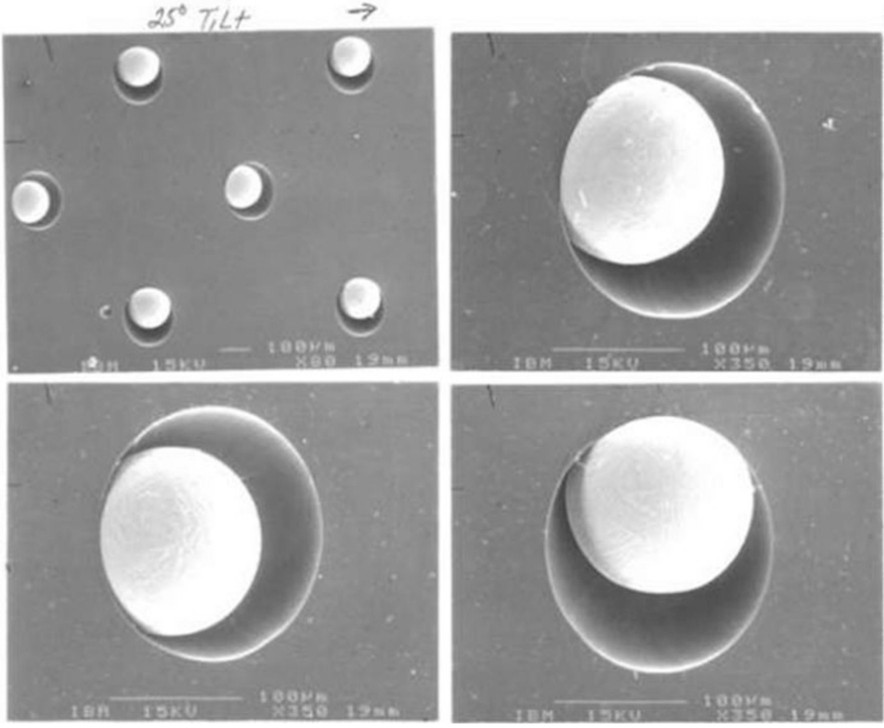


Fig. 4.16 Reflowed solder spheres in glass mold cavities prior to transfer to wafer [22]

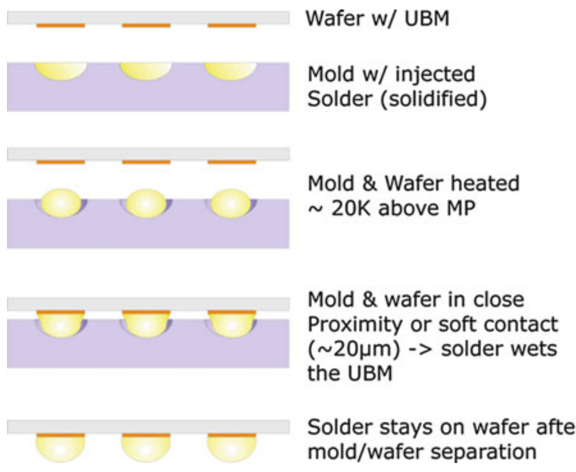


Fig. 4.17 Solder transfer process sequence [20]

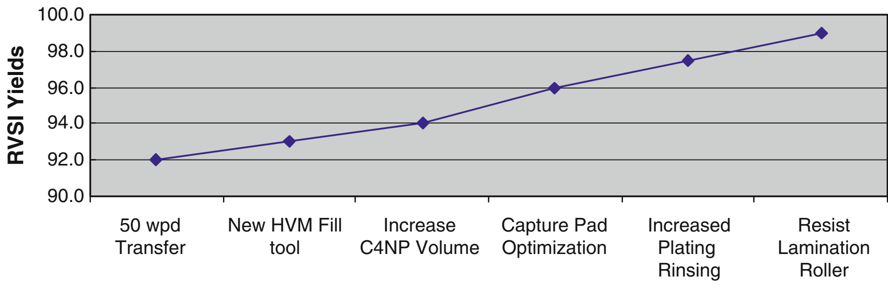


Fig. 4.18 Process improvement, start of manufacturing. (Best Wafers = 100 % yield) [20, 21]

not require special care because the spacing between adjacent cavities is longer. However, bridging is much more sensitive to micro-bumps because of shorter spacing between cavities.

With the high-yield molds, micro-bumps were successfully transferred from glass molds to both 200 and 300 mm wafers which were patterned with a three-layer UBM, as shown in Fig. 4.20. With the aid of formic acid vapor flux, excellent wetting was achieved for the SnAg solder micro-bumps. The UBM pads are ~28 μm in diameter. An additional reflow was performed to reshape the micro-bumps and uniform bump heights were obtained. Excellent height uniformity was achieved for the transferred microbumps (coplanarity <2 μm). The preliminary results suggest that C4NP technology can be well scaled down to 50 μm to meet ever increasing demand on the I/O density.

4.3.4 C4NP Advantages: Alloy Flexibility

For C4NP technology changing solder alloys for wafer bumping is a simple operation. It is accomplished by changing the fill head in the mold fill tool, which is usually done in less than an hour. The process temperatures of the solder reservoir and the mold can be adjusted to accommodate a particular solder alloy. This flexibility allows C4NP to be backward compatible with existing solder alloys and UBM stacks as well as to enable the use of any multicomponent new solder alloys and UBM stacks. As discussed later in the section of Pb-free solders, “dopants” can be added to Sn-rich solders to improve EM performance, suppress Kirdendall voiding, reduce copper pad consumption, suppress Sn pest, etc. The ability to maintain alloy flexibility with precisely controlled solder composition and alloy doping is critically important to deliver enhanced performance and reliability. The advantages of having alloy flexibility are described in the other sections.

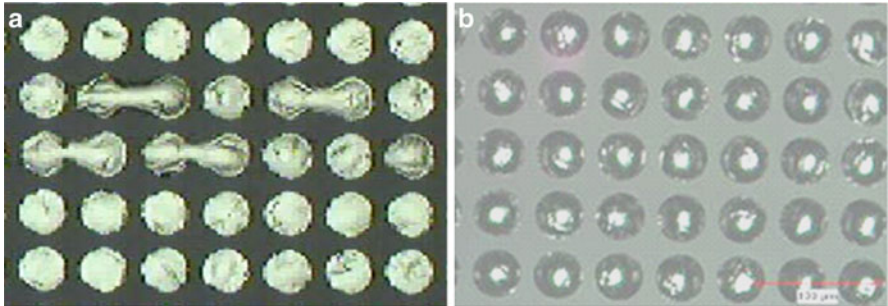


Fig. 4.19 (**a**) Mold fill in O_2/N_2 mixture gas [21, 23]. (**b**) Mold filled in Pure N_2 [21, 23]

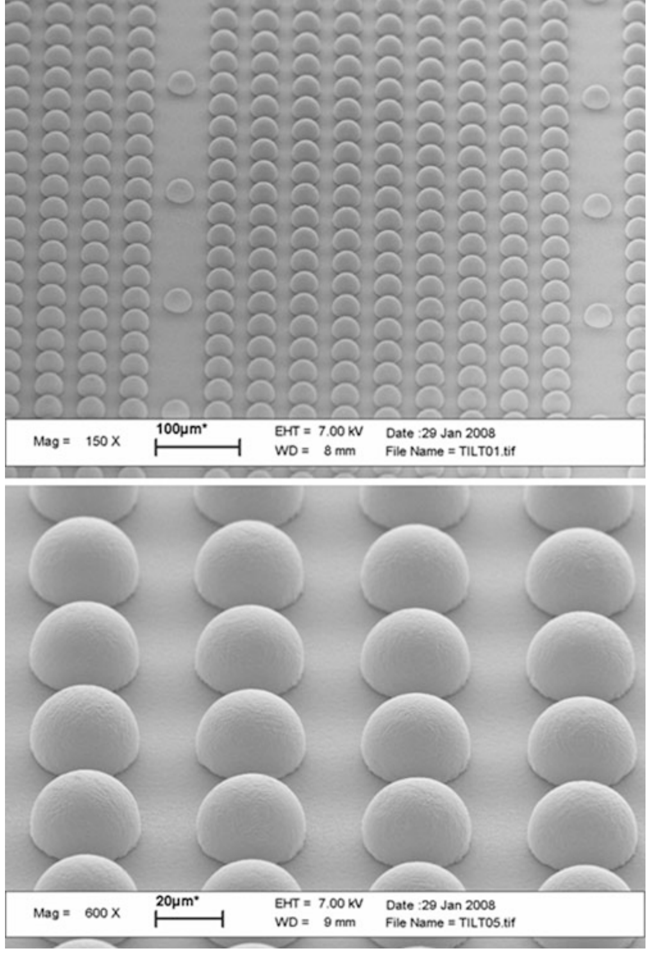


Fig. 4.20 SEM images of 50 µm microbumps [23]

4.4 Fabrication of Cu Pillar Bumps

Cu pillar bumps have been introduced as an alternative to the conventional collapsible solder bumps for first level flip-chip interconnections [15–17, 24]. The structure and production process of a pillar bump are analogous to an electroplated solder bump. The fabrication requires the integration of the photolithography and electroplating processes. A thin metal seed layer is first blanket-sputtered on the passivation surface and pads of the BEOL structure on a silicon wafer. It consists of an adhesion layer, normally Ti or TiW, and a Cu seed layer serving as a conducting layer for plating. A thick photoresist material is spin-coated over the Cu seed layer, normally with a thickness range of 40–100 μm. After exposure and development, the opening cavities with the Cu seed layer at the bottom are electroplated with Cu. A solder cap can be optionally plated on top of the pillar and subsequently reflowed to form a solder bump or cap. Resist stripping is performed in an environmentally friendly solvent system. For the electroplating process, the photoresist profile, plating durability, and strip ability after plating are important considerations. Cu pillar bumps can be fabricated over a variety of spacing on a Cu-seed layer. Figure 4.21 is schematic drawings of a conventional solder bump (a), and a copper pillar with solder cap (b) [25]. Figure 4.22 shows the full array of pillars arrayed at different spacing.

The copper pillar bump is an emerging technology as an option for high performance packaging since it offers fine pitch capabilities and has good electromigration performance. Some concerns of copper pillars on chip include a high risk of chip-to-package interaction (CPI) with ultra low-k dielectrics and a relatively high cost. The significant improvement in EM performance has been demonstrated [26–29]. Figure 4.23 shows a SEM image of a Cu pillar bump on Intel's Presler processor [30]. The stiff, non-deformable nature of Cu bumps requires a significantly engineered BEOL structure to mitigate the high stress applied to the ultra low-k layer.

4.5 Substrate Bumping Technologies

In the assembly of a chip to an organic laminate substrate, solder interconnects are formed by reflowing both solder bumps on the chip and pre-solder bumps on the substrate. The per-solder bumps on the substrate are needed to compensate the variations in the solder bump height on chip as well as the laminate substrate warpage. This is particularly critical for large chips with fine-pitch bumps. To improve the bump coplanarity on substrate, a coining process is usually applied to form flat tops on the solder bumps on a substrate [31].

There are several solder bump forming methods developed for substrate pre-soldering. The most popular manufacturing technology for forming solder bumps on organic substrates is the solder paste stencil printing method [32–34]. Solder paste, typically consisting of about 50 % of flux, is printed onto surface finishing pads on a laminate substrate through a patterned stencil mask. After stencil removal, the substrate is heated to melt the solder paste to form bumps, and then coined to make

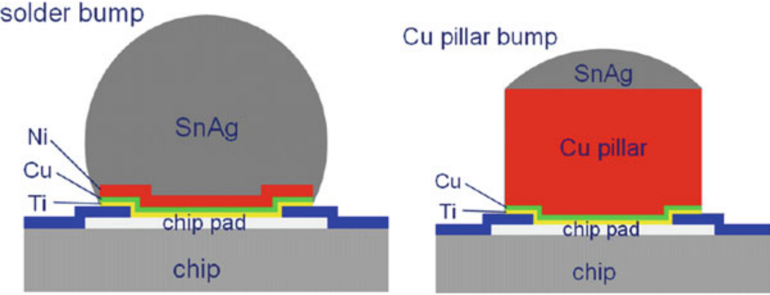


Fig. 4.21 Schematic drawing of a conventional solder bump (**a**), and a copper pillar with solder cap [25]

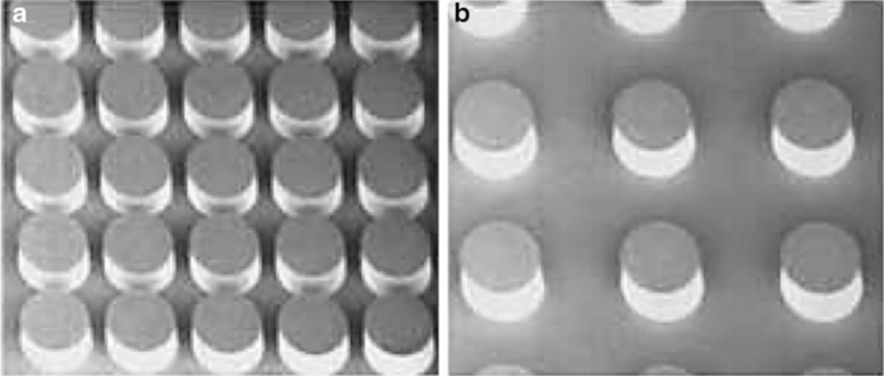


Fig. 4.22 Examples of Cu pillars at 50 μm pitch with (**a**) 12.5 μm spacing and (**b**) 50 μm spacing [16]

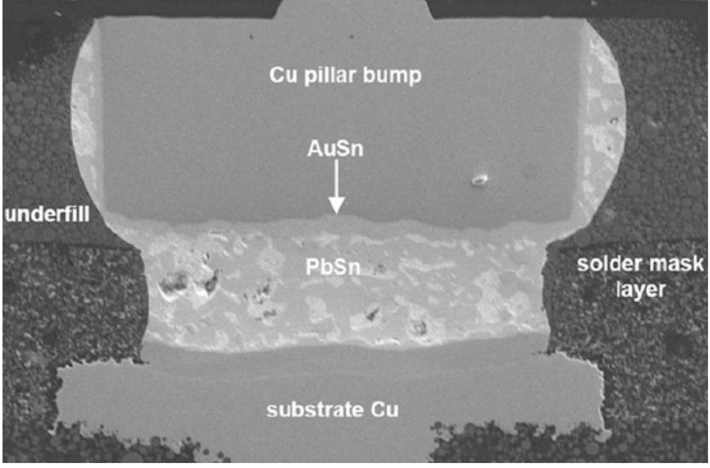


Fig. 4.23 SEM image of a Cu pillar bump on Intel's Presler processor [26]

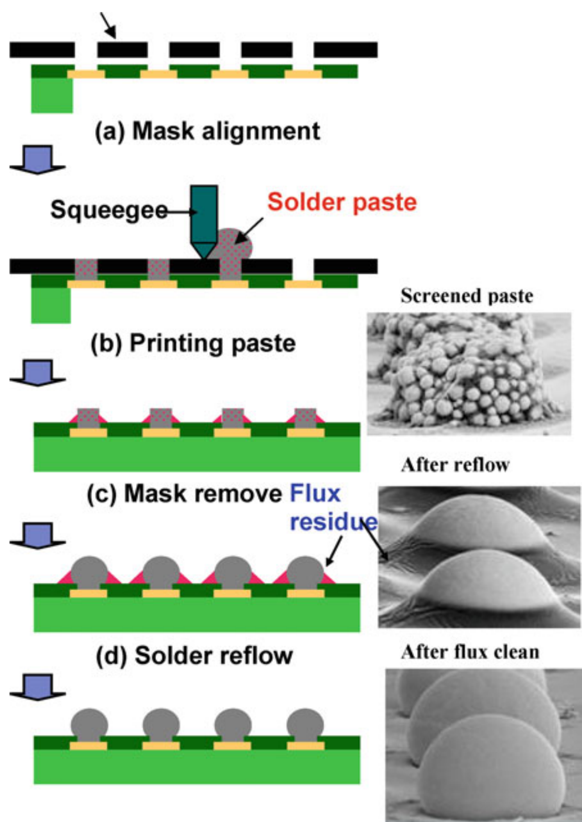


Fig. 4.24 Solder paste stencil printing processes [35]

flat tops. This method is simple and low cost, since it does not require using a photolithography process. However, it cannot be easily extended to high density interconnects having less than a 150 μm pitch. In the fine-pitch application, the flux tends to bridge after paste screening, which would cause a high defect rate of solder bridging after reflow. In addition, due to the significant volume reduction after reflow, the void population and bump height variation could be significant and, consequently affect die attachment yields negatively. The process steps describing the solder paste stencil printing method is schematically shown in Fig. 4.24 [35]. Stencil paste screening method has the advantage of one step solder bumping of different size solder resist openings on the same substrate by screen-printing different amount of solder paste on the stencil mask openings on the correspondingly mask holes. Due to the large height variation of the paste screened and reflowed bumps on substrate, a coining process is used to guarantee the coplanarity of all the solder bumps for higher assembly yield [31]. A substrate solder bump after the coining process is shown in Fig. 4.25. The process also offers alloy flexibility. Practically any solder alloys can be fabricated into the preformed solder spheres before mixing with the flux to form the paste.

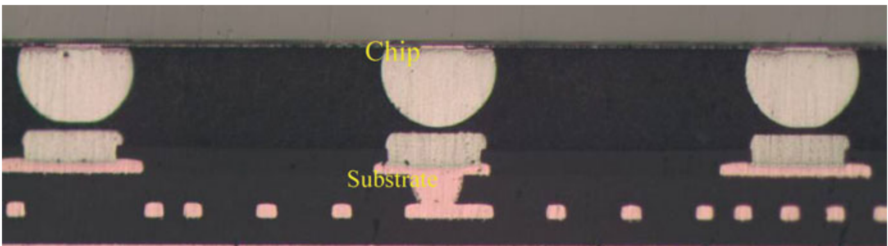


Fig. 4.25 A coining process was used to guarantee the coplanarity of height of all bumps on substrate within a tight range [36]

Recently, new solder bumping techniques have been developed to address the ever growing demands for fine pitch and the reduction of solder bump and solder resist opening (SRO) sizes, such as pitches down to 120 μm and SRO sizes of about 60–70 μm. One method involves the placement of prefabricated microspheres or micro-balls of solder onto substrate pads. Several solder ball placement methods have been developed [36–40]. According to one method, solder balls are sucked into a jig by vacuum suction, and then mounted onto flux-coated pads on substrate. Another micro-ball placement method involves the use of a stencil mask [39, 40]. In the later method, thick tacky flux is first printed through a stencil mask onto the substrate pads. Solder balls of a uniform size are dispensed onto a second stencil mask with holes aligned with the substrate pads. A squeeze brush is then used to disperse the balls and press them into the mask holes. The balls transfer and stick to the tacky flux which was previously screened on the pads. The stencil mask is removed after ball placement and then heated and melted to form bumps. Figure 4.26 describes the ball placement method and processes. This method, however, does not have the flexibility of the placement of different size balls on matching pads with different solder resist openings. It generally requires an additional mask process for placement of different size solder spheres or use paste screening method. The micro-ball placement method eliminates the volume reduction problem by using preformed solder balls and tacky flux in separate mask processes, as shown in Fig. 4.27 [41].

Micro-ball mounting method can form higher volume solder bumps and could be applied to finer pitch substrates than the stencil printing method. However, the micro-ball mounting processes are more complicated as compared to the stencil printing method. Three masking processes and alignments are needed, respectively, for tacky flux dispensing, micro-ball “brushing” placement on C4 pads, and paste dispensing for the larger capacitor pads [41]. Also, the cost of preformed solder balls significantly increases with size reduction for fine-pitch application. Micro-ball mounting method still has limitation in maximum volume of solder bumps at a given pitch even though it can provide higher solder volume than the stencil printing method. Extendibility to very fine pitch application is difficult because of the fluxing process and handling of very small balls. In addition, the micro-ball mounting method cannot handle multiple size pad openings because the same size balls must be used at each mask process. The coplanarity of solder bump height will degrade if one size solder balls are dispensed on different size pads.

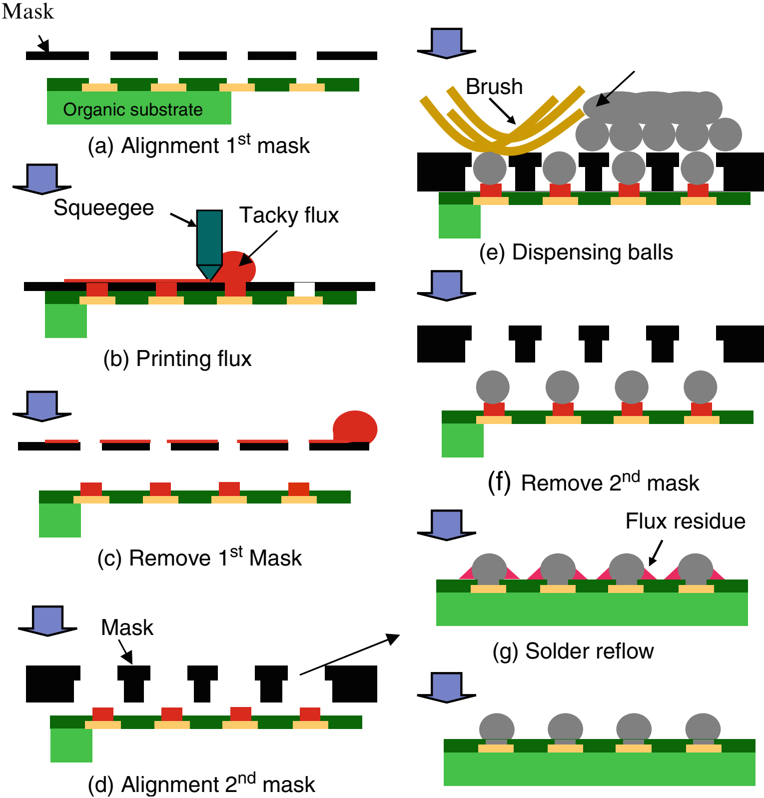


Fig. 4.26 The micro-balls placement processes [35]

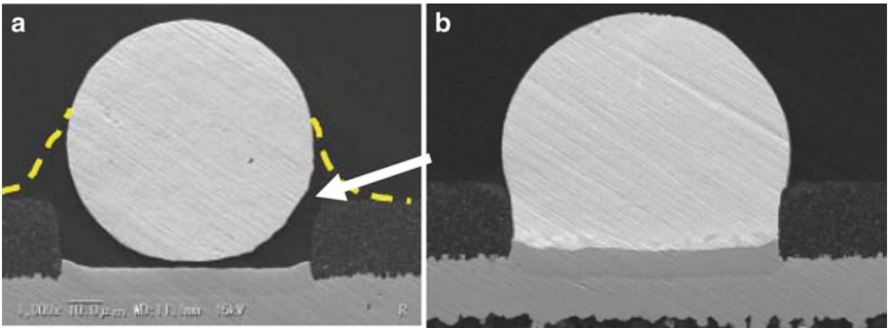


Fig. 4.27 The micro-ball placement method; (**a**) after ball mount, and (**b**) after reflow

Recently, Intel has integrated Cu pillar bumps on chip UBM using Cu electroplating process in high volume manufacturing and showed reliability benefits on electromigration and thermal conduction [29]. The Cu pillar bumps have flat tops which are different from round tops in the case of solder die bumps.

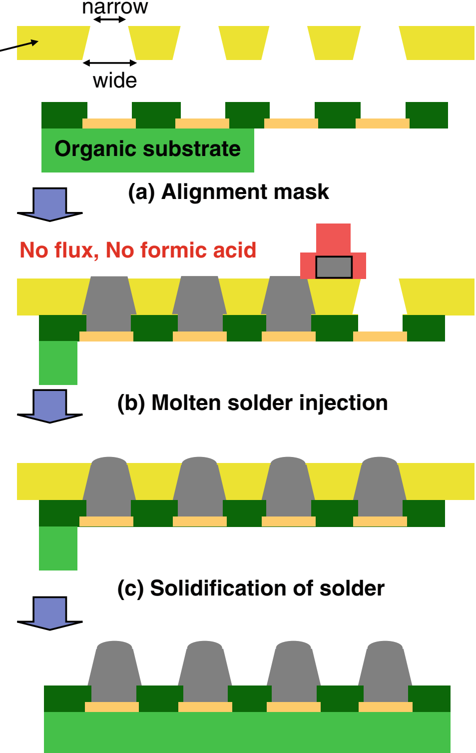


Fig. 4.28 Injection Molded Solder processes [41]

The Cu pillar bumps need round tops of substrate solder bumps to avoid the formation of voids at the Cu–solder interface during flip-chip assembly process. The round top solder bumps provide by the micro-ball placement method have worse bump coplanarity than the coined flat top solder bumps. Therefore, the round top solder bumps on substrate side need higher volume for flip-chip assembly with Cu pillar bumps for high assembly yield.

IBM has recently developed and qualified the C4NP wafer bumping technology for the manufacturing of all Pb-free solder 300 mm wafers at very fine pitch. Feasibility of wafer bumping of 50 μm pitch, on both 200 and 300 mm, has been demonstrated with excellent yield [18, 21]. The new wafer bumping method is extended to substrate bumping, where molten solders are directly injected on substrates pads by using a flexible mask [41].

The new substrate bumping technique, the Injection Molten Solder (IMS) technology, is shown in Fig. 4.28 using a patterned polyimide decal. First, the holes in the polyimide (PI) mask are aligned to the pads on the substrate. Using an optimized combination of pressure and temperature, molten solder is injected into the aligned holes and fills both the solder resist openings and holes in the mask. In this step, through careful control of the fill environment in low oxygen, neither flux nor formic acid is needed. After the molten solder has wetted and solidified on the pads, the PI mask is separated from the substrate.

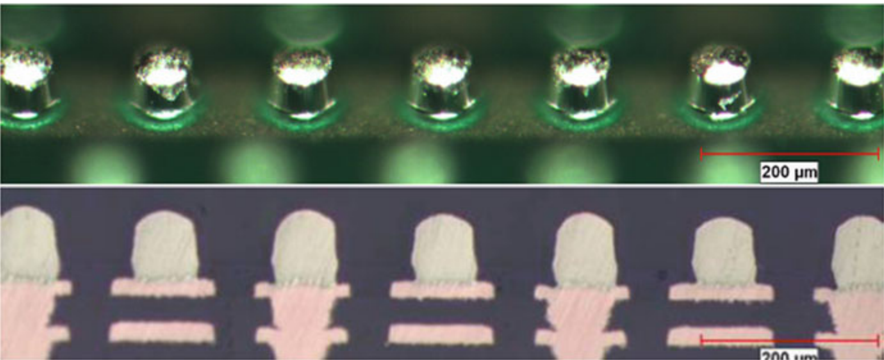


Fig. 4.29 Side and cross-sectional views of IMS bumps on ENIG Ni/Au pads [41]

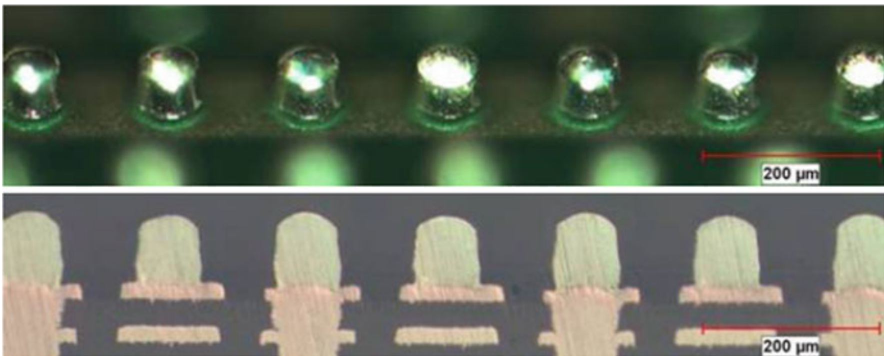


Fig. 4.30 Side and cross-sectional views of IMS bumps on Cu OSP surface finish [41]

The flexible PI mask follows the non-flat surface contour of the laminate substrate by enabling intimate contact between the mask and substrate to prevent solder bridging. The holes in the mask have tapered angle to facilitate mask separation after solder solidification that increases the reusability of the mask.

While the IMS process is very simple and similar to the stencil printing method, the advantage of IMS over the paste process is that it can make higher volume solder bumps for fine-pitch application due to the use of pure molten solder. The solidified solder bumps have column shape and round tops as shown in Fig. 4.29. The shape of the solder bump mimics the hole shape in the flexible mask. Solder volume can be easily changed by the changing of mask thickness or hole size. Any types of solder alloy can be used with no increase in materials cost when applied to very fine pitch applications. Figures 4.29 and 4.30 show the side views and cross-sectional images of IMS bumped Cu OSP and ENIG Au finished substrates, respectively. The cross-sectional images show that the IMS bumps have good interfacial microstructures and tight coplanarity on both surface finishes.

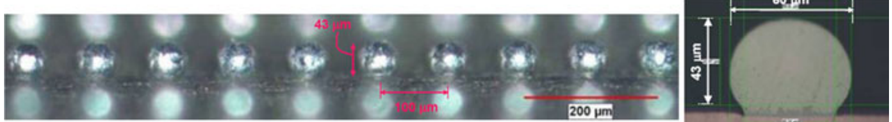


Fig. 4.31 Side view and cross sectional image of IMS bumps on 100 µm pitch substrates [41]

Another key advantage of IMS bumping is that IMS makes different solder volumes in a single pass and still maintains good coplanarity when the substrate has different sizes of resist openings, including the large capacitor pads. The technology allows the design flexibility of applying larger volume interconnects for power joints and smaller volume interconnects for signal to meet both performance and reliability requirements. Currently, the method has demonstrated the full area array bumping capability at 80 µm pitch with excellent bum height uniformity. Figure 4.31 shows a side view of an array of bumps at 100 µm pitch.

4.6 Pb-Free Solders for Flip-Chip Applications

Since July, 2006, following the EU's RoHS legislation [42–44], the consumer electronics industry has been offering "green" products by eliminating Pb-containing solders and other toxic materials. This transition has been relatively smooth, because the reliability requirements are less stringent. However, the Pb-free transition for high performance electronic systems (such as servers and telecommunication) is still ongoing due to their rigorous reliability requirements. Recognizing this situation, EU has extended the Pb-free exemption date for high performance electronic systems by July, 2016. The research and development efforts to implement Pb-free flip-chip interconnections for high end applications are still very active. A key technical element in implementing Pb-free, flip-chip interconnection is the wafer bumping technology. One new wafer bumping process, C4NP (C4 new process) for Pb-free solders is discussed in this chapter, while other wafer bumping technologies are covered in Chap. 3.

Pb-free solders in use are listed with their melting point, applications and concerns in Table 4.2. Two major Pb-containing solders used for flip-chip applications, eutectic 63Sn–37Pb and 97Pb–3Sn, are also included in Table 4.2. Pb-free solders used in consumer electronics are mainly applied by paste screen printing, ball mounting or electroplating method, while Pb-free solders for high-end flip-chip applications are applied by either electroplating or C4NP technology.

In the early stage of development, many Pb-free solders have been evaluated for flip-chip applications [9, 32, 45]. The near-ternary eutectic Sn–Ag–Cu (SAC) solder commonly used in printed circuit board (PCB) assembly, such as BGA or SMT solder joints is not readily accepted for flip-chip applications. This is largely due to several challenging issues, such as difficulty of electroplating a ternary composition,

Table 4.2 Examples of some Pb-free candidate solders and Pb-containing solders

Composition (wt%)	Melting point (°C)	Applications	Concerns
Sn–3.5Ag	221	SMT, flip chip	Cu dissolution, excessive IMCs, voids
Sn–3.8Ag–0.7Cu	217	SMT, PTH, BGA	Cu dissolution, excessive IMCs, voids
Sn–3.5Ag–3Bi	208–215	SMT	Cu dissolution, fillet lift, low mp phase
Sn–0.7Cu	227	PTH, flip chip	Cu dissolution, wetting, excessive IMCs
63Sn–37Pb	183	PTH, SMT, BGA, flip chip	Pb toxicity
97Pb–3Sn	317	Flip chip	Pb-toxicity

formation of large intermetallic plates of Ag_3Sn in near-ternary SAC, or yielding high modulus or stiff solder joints, and other concerns. To mitigate these concerns, the usage of Sn–Ag or Sn–Cu binary systems with a low Ag or Cu content, was proposed to obtain low modulus or more ductile flip-chip interconnections [46, 47]. However, reducing the alloying content in the binary and ternary Sn-based solders can significantly alter their microstructure, melting point, and consequently their mechanical or other properties.

4.6.1 Properties of Pb-Free Solders

Most of Pb-free solders used are Sn-rich solders that typically contain more than 90 % Sn. This suggests that the physical, chemical, and mechanical properties of Pb-free solders are heavily influenced by the properties of pure Sn, as opposed to eutectic Sn–Pb solder, a mixture of Sn-rich and Pb-rich phases. Pure Sn is polymorphic, capable of existing as three crystal structures (α, β, and γ) depending on temperature and pressure [48]. Since the white tin phase (β-Sn), stable at room temperature, has a body-centered tetragonal (BCT) crystal structure in contrast to the face-centered cubic (FCC) structure of Pb, the physical and mechanical properties of white tin are highly anisotropic compared to Pb. The white Sn crystal is optically birefringent, meaning an incident polarized light beam on a β-Sn crystal has a plane of polarization rotated upon reflection. Accordingly, polarized light microscopy is a convenient method to determine the crystal orientation of β-Sn dendrites or grain sizes for Pb-free solders in addition to EBSD (electron backscatter diffraction) technique [49, 50].

The melting point of most Pb-free commercial solders is within the range between 208 to 227 °C, which is about 30 °C higher than the melting point of eutectic Sn–Pb, 183 °C. The higher melting point or reflow temperature has serious implications on the performance of packaging materials and assembly processes, and can affect the integrity and/or reliability of Pb-free microelectronic packages. Another important issue related to the melting points is the difficulty of maintaining solder melting-point hierarchy, which is well established with Pb-containing solders. For example, since

Table 4.3 Anisotropic properties of β-Sn

Properties	*a*-axis	*c*-axis	Ratio of anisotropy (*c*/*a*-axis)	References
Lattice spacing (A)	5.83	3.18	0.54	[51]
Coefficient of thermal expansion (CTE) (ppm/°C)	15.45	30.50	1.97	[52]
Young's modulus (GPa)	22.9	68.9	3.01	[53]
Electrical resistivity (μΩ cm) at 300 °C	14.3	9.9	0.69	[54]
Self diffusivity of Sn at 150 °C (cm^2/s)	8.70×10^{-13}	4.71×10^{-13}	0.54	[55]
Diffusivity of Ag in Sn at 150 °C (cm^2/s)	5.60×10^{-11}	3.13×10^{-9}	56	[56]
Diffusivity of Cu in Sn at 150 °C (cm^2/s)	1.99×10^{-7}	8.57×10^{-6}	43	[57]
Diffusivity of Ni in Sn at 150 °C (cm^2/s)	3.85×10^{-9}	1.17×10^{-4}	30,390	[58]

the melting point of high Pb flip-chip solder bumps is well separated by more than 100 °C from the melting point of eutectic Sn–Pb used in the next level of assembly, the high Pb, flip-chip solder joints formed at a higher temperature (e.g., 350 °C) do not become molten, during the subsequent eutectic Sn–Pb card assembly reflow operation (e.g., 215 °C). However, the proposed Pb-free solders only allow a maximum possible separation of about 30 °C or less between the melting points of any two solders. Hence, the choice of Pb-free flip-chip solder compositions does not allow enough temperature separation for subsequent module or card assembly processes with another Pb-free solder composition.

Owing to the unique crystal structure of β-Sn, body-centered tetragonal (bct) crystal structure (a: 5.83, c: 3.18 Å), the physical, mechanical, thermal, and electrical properties of β-Sn are highly anisotropic. Some of selected properties are collected in Table 4.3. The ratio of anisotropy (defined as property along *c*-axis/ property along *a*-axis) is also calculated in Table 4.3. The coefficient of thermal expansion (CTE) is about two times larger along the *c*-axis than the *a*-axis, while the Young's modulus along *c*-axis is three times larger than along the *a*-axis. This is a quite unique situation for β-Sn, since the trends of CTE and modulus are usually in an opposite direction for most metals and alloys. The electrical resistivity of β-Sn is significantly different in each direction with the anisotropy ratio of 0.69, i.e., the resistivity is about 70 % smaller along the *c*-axis than the *a*-axis, consistent with the factor of a shorter lattice spacing along the *c*-axis.

More dramatic anisotropy is noticed with the atomic transport rates of Ag, Cu, or Ni in β-Sn, as shown in Table 4.3. The calculated ratios of anisotropy for Ag, Cu, and Ni are about 60, 40, and 30,000, respectively, at 150 °C, indicating enormous disparity in the solute diffusion rates along the *c*-axis vs. *a*-axis of β-Sn. The solute

diffusion is attributed to the interstitial diffusion of Ag, Cu, and Ni atoms, while the self-diffusion of Sn relies on a substitutional diffusion mechanism. The Sn self-diffusion rates are much slower than the solute diffusions, and are much less anisotropic compared to the solute diffusions in Sn.

When a solder joint contains a few grains or one single grain such as in flip-chip or BGA joints, the anisotropic properties of β-Sn would seriously impact on the integrity and reliability of Pb-free solder joints. In case of Pb-free, near eutectic SAC solder joints, a clear dependence of the thermo-mechanical response on Sn grain orientation was reported [59]. BGA solder balls with the *c*-axis orientation parallel to the substrate were observed to fail before neighboring balls with different orientations. This result was explained based on the disparity of CTE values; a maximal CTE mismatch would occur in shear at the joint interface when the *c*-axis orientation is parallel to the substrate. In Pb-free, flip-chip solder joints, the effect of Sn grain orientation on electromigration degradation mechanism was reported [60]. Premature electromigration damage was identified when the *c*-axis orientation is aligned with the electrical current flow direction. This was explained based on the fast solute diffusion along the *c*-axis of β-Sn.

4.6.2 Solidification, Microstructure, and Undercooling

Formation of solder joints involves with a metallurgical process of melting (reflow) and solidification. The resultant microstructure therefore reveals the unique characteristics of as-cast microstructure. The effects of cooling rates and alloying elements with various Pb-free solders were investigated [50, 61]. One of the distinct properties of Sn-rich solders is a propensity for large amount of undercooling of β-Sn during solidification. The undercooling is defined as the temperature difference between the melting temperature of a solder during heating and the solidification temperature during cooling. The undercooling required for solidification of near-ternary Sn–Ag–Cu solder spheres of a few hundred micrometers in diameter (such as BGA or CSP solder joints) was much larger than high Pb solders or Sn–Pb eutectic solders [61, 62]. This large undercooling is also responsible for the growth of large primary phases such as Ag_3Sn in near-ternary Sn–Ag–Cu solders [62, 63]. It was also reported that the amount of the β-Sn undercooling in Sn–Ag–Cu solders is inversely proportional to sample size, suggesting a larger undercooling in a smaller solder joint (such as flip-chip vs. BGA solder joints) [64].

A large undercooling in flip-chip solder bumps can have a serious impact on the reliability of solder joints, since random solidification among many solder bumps can cause a situation of some bumps already solidified while others not, which could lead to a stress concentration to some bumps and possibly to early mechanical failures within solder joints.

A systematic investigation was conducted to find critical factors affecting the undercooling of Pb-free, flip-chip solder bumps by DSC (differential scanning calorimetry) and the direct observation of individual solder bumps in a glass mold

during melting and solidification [65, 66]. The amount of the undercooling of Sn-rich solders is found to be strongly affected by solder volume, inversely proportional to its volume (or effective diameter of solder balls). It is also found that the solder composition and UBM (under bump metallurgy) significantly affect undercooling, but not so strongly as other factors such as the cooling rate and holding temperature of the mold plate used in C4NP. Sn–0.7Cu C4NP solder bumps in a mold plate were under-cooled by as much as 90 °C from its melting point, while a less amount of undercooling (40–60 °C) was observed on a Si chip with Cu/Ni UBM [65]. The direct observation of individual flip-chip-size solder bumps on a glass mold during their solidification process revealed the random nature of the molten solder nucleation process, and also confirmed the similar amount of the undercooling measured by DSC [66]. It was also found that some minor alloying elements such as Zn, Co, and Ni are quite effective in reducing the amount of undercooling in Sn-rich solders [66]. Minimizing the amount of the undercooling of Pb-free solder bumps is desirable to provide chip joining integrity as well as solder joint reliability.

4.7 Interfacial Reactions in Pb-Free, Flip-Chip Joints

4.7.1 Ball-Limiting Metallurgy or Under Bump Metallization

In flip-chip structures, the chip wiring is terminated by a plurality of metal films that form the ball-limiting metallurgy (BLM), which is also referred as under-bump metallurgy (UBM), as schematically shown in Fig. 4.32. The BLM defines the size of the solder bump after reflow, provides a surface that is wettable by solder and that reacts with the solder to provide good adhesion and acceptable reliability under mechanical, electrical, and thermal stress, and is a barrier between the integrated-circuit device and the metals in the interconnection.

When the chip is attached to a ceramic module, solder with a high Pb content is reflowed at a temperature around 350 °C. The Sn content of high Pb solders is typically less than 5 % by weight (Table 4.2). A typical BLM structure is a thin film stack consisting of Cr or TiW (at the chip surface), CrCu, and Cu. The thin film stack of BLM is typically deposited by sputtering on a passivated wafer. The sputtered Cr or TiW layer provides good adhesion of the interconnection structure to the silicon-wafer substrate. The "phased CrCu" is a second adhesion layer consisting of co-sputtered Cr and Cu that is high in Cr at the barrier layer interface and high in Cu at the solderable metal Cu interface. The BLM layer also serves as an electrical connection for the electroplating of solder bumps. After solder bump plating, the BLM is formed by selective electro and chemical etching of the blanket thin film layers. During reflow, the Sn in the solder reacts readily with Cu to form Cu–Sn intermetallics, which provide adhesion between the solder the BLM. For solders with less than about 5 % Sn by weight, a thin film of Cu (about 0.5 μm) is a suitable terminal layer for the BLM [67].

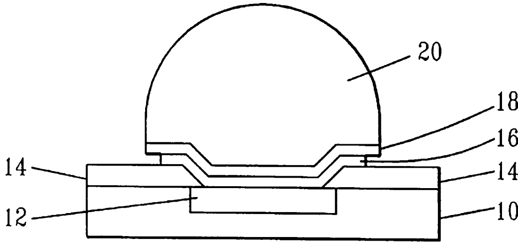


Fig. 4.32 Ball-limiting metallurgy (BLM) or under-bump metallization (UBM) in flip-chip structures [67]

Table 4.4 Solubility of Cu vs. Ni in Sn as a function of temperature (all in wt%)

Temperature	Cu solubility	Ni solubility
260 °C	1.41	0.216
250 °C	1.23	0.198
200 °C	0.0035	0.0009
150 °C	0.0011	0.0002
50 °C	3.79×10^{-5}	3.52×10^{-6}

Table 4.5 Solderable layer thickness after simulated reflows at 250 °C of Sn–3.5 %Ag solder (all values in μm), [9]

Solderable layer	0 min	2 min	6 min	20 min
Cu(4 μm)	4.0	1.0–2.3	0.7–1.7	0–1.7
Cu(2 mm)/Ni(2 μm)	4.0	3.3	3.1	3.1
Cu/50Ni–50Fe(2.7 μm)	6.7	6.1	6.1	6.0
Cu/Ni(P)(8 μm)/Au(50 nm)	8.3	5.8	5.5	4.9

Table 4.6 Thickness of intermetallic layer formed when Sn–3.5Ag solder reacts with solderable layer at 250 °C (all values in μm), [9]

Solderable layer	2 min	6 min	20 min
Cu(4 μm)	1.7–8.3	0.7–10.0	1.7–13.3
Cu(2 mm)/Ni(2 μm)	1.2–4.0	1.8–4.8	2.0–6.4
Cu/50Ni–50Fe(2.7 μm)	0.3–0.5	0.3–0.5	0.3–0.6
Cu/Ni(P)(8 μm)/Au(50 nm)	0.7–1.7	0.7–2.1	0.9–2.4

When the chip is directly attached to an organic laminate using SnPb eutectic solder, a thin layer of Cu (about 0.5 μm) would be essentially consumed to form Cu–Sn intermetallics, resulting in a loss of mechanical reliability. To control Cu consumption, a thick Cu (a few microns) was proposed as a solder wettable layer of BLM structures for SnPb eutectic [68] as well as Sn-rich solders. But this idea was abandoned after it was found that the thick Cu layer could facilitate the growth of voids during a high temperature aging such as at 150 °C.

The proper choice of BLM is a critical element for successful Pb-free, flip-chip development. Since most Pb-free solders have the Sn content more than 90 % by weight and hence they are required to reflow at a higher temperature (at least 30 °C) than eutectic Sn–Pb, the interfacial reactions are very aggressive in terms of UBM consumption and intermetallic formation, particularly with Cu metallization. A thin Ni barrier layer has been commonly used to control the aggressive interfacial reactions [13, 69, 70], because Ni has a much less solubility than Cu in the molten Sn at the same reflow temperature. Table 4.4 compares the solubility of Cu vs. Ni in

Sn as a function of temperature [71, 72]. The solubility of Cu in Sn is approximately seven times higher than Ni at a reflow temperature of 260 °C, and is about five times higher at the typical aging temperature of 150 °C.

Although Ni is an effective reaction barrier over Cu in Sn-rich solders, it is still desirable to control the consumption rate of Ni layer during reflow or high temperature storage or high current electromigration tests. Several other BLM structures are proposed for Pb-free, flip-chip structures [67, 73–75]. Among them, NiFe alloys were found to be an effective, solderable layer for Sn-rich solders [9, 73]. The NiFe was an electroplated thin film; compositions of 90 %Ni–10 %Fe, 80 %Ni–20 %Fe, and 50 %Ni–50 %Fe were investigated. The interfacial reactions of NiFe alloys with Sn–3.5 %Ag were investigated to compare with other solderable BLM layers, such as Cu, Ni, and Ni(P) for various reflow times up to 20 min at 250 °C. The dissolution of the solderable layers and the growth of intermetallic layers are listed for the case of 50 %Ni–50 %Fe in Tables 4.5 and 4.6. The 80 %Ni–20 %Fe behaved similarly to 50 %Ni–50 %Fe in terms of the dissolution of the NiFe films and the intermetallic growth. The 90 %Ni–10 %Fe, on the other hand, reacted about as fast as pure Ni with the molten Sn–3.5 %Ag; while a thick Cu layer of 4 μm was nearly consumed after 6 min, and a very thick intermetallic layer was formed. Thus, it was conclude that the composition range of NiFe up to 80 % Ni was effective as a barrier layer for the high Sn solders. The electroless Ni(P) layer of 8 μm was compared for its interfacial reaction as shown in Tables 4.4 and 4.5. In Sn–3.5Ag solder, Ni(P) dissolved more than the 50 %Ni–50 %Fe or Cu/Ni did, and its intermetallic growth was also more extensive than the NiFe. In a recent study, excellent solderability of electroplated Fe–Ni alloys was also reported with eutectic SnAgCu solder [76].

4.7.2 Substrate Metallization

The choice of substrate metallization either on laminates or ceramic modules is as important as UBM, since a solder joint is formed between two interfaces; UBM and substrate metallization. Most common substrate metallization used are Cu-OSP and electroless Ni(P) with immersion Au (ENIG) on laminates, and Ni(P)/Au on ceramic substrates.

Electroless nickel-phosphorous film, Ni(P) is widely used from under-bump metallization (UBM) of flip-chip and substrate metallization of ball-grid array packages. This is due to its superior characteristics, such as excellent solderability, corrosion resistance, uniform deposition thickness, and selective deposition process. The microstructure of a Ni(P) film changes significantly with its P content. A Ni(P) layer containing less than 5.5 wt% P is known to be nanocrystalline, while that of more than 9 wt% P is amorphous. The Ni(P) films with 5.5–8.5 wt% P are known to be a mixture of small crystallites and an amorphous phase [77].

When a Ni(P) film reacts with a Sn–Pb eutectic solder, some part of the film underneath the solder crystallizes into Ni_3P around the reflow temperature (200–240 °C). This low-temperature reaction is called “solder reaction-assisted

crystallization" in contrast to the self-recrystallization of Ni(P) at a higher temperature at 300–450 °C. The solder reaction-assisted crystallization is accompanied by the formation of the Ni–Sn IMC and Kirkendall voids [78]. These interfacial reactions result in brittle fracture whose paths are often found to be around the P-rich layer, causing a reliability issue of Ni(P) metallization. When a Ni(P) film reacts with pure Sn or Sn-rich solders, the extent of interfacial reactions significantly increase, often a severe IMC spalling from Ni(P) is observed [79, 80]. The IMC spalling is found to be strongly influenced by P content, solder volume, or deposition method [80], higher P content and larger solder volume causing more spalling. In addition, higher tendencies of IMC spalling are noted from screen printed solder paste over electrodeposited solder.

To prevent IMC spalling, a thin intermediate layer of Sn or Cu was deposited on top of Ni(P) by electro- or electroless plating. During the reflow reaction of Sn–3.5 %Ag solder paste, the intermediate layers effectively suppressed Ni–Sn IMC spalling during the reflow reaction at 250 °C, 30 min, while most IMC spalled off the Ni(P) film in a few minutes in the control samples without an intermediate layer [81]. The Sn layer provided protection of the Ni(P) surface and a good wettable surface during reflow. The thin Cu layer changed the chemical structure of the interfacial IMCs in addition to providing a good wettable surface.

In the study of the interfacial reactions of Pb-free solder joints in plastic ball grid arrays (PBGA), it was found that the choice of substrate metallization can also affect the IMC formation at the UBM interface [13]. By pairing three substrate metallization, Cu/OSP, Ni(P)/Au, and Ni(P)/Pd/Au, five groups of the PBGA laminate modules were produced with surface finishes of Cu–Cu, Cu–Au/Ni(P), Au/Ni(P)–Au/Ni(P), Cu–Au/Pd/Ni(P), and Au/Pd/Ni(P)–Au/Pd/Ni(P). PBGA laminates were assembled using 0.89 mm solder balls of Sn–3.8Ag–0.7Cu, reflowed up to 12 cycles at 260 °C. From the systematic analysis of their microstructure, composition and microhardness, it was found that the interfacial reaction of one side is significantly affected by the choice of surface finish at another interface, because the elements from one surface finish dissolve sufficiently into the molten solder and rapidly diffuse to the other. The microhardness in the BGA solder joints is strongly affected by the choice of surface finish, not much by reflow cycle. The surface finish composed of more Ni layers shows a higher hardness value than without a Ni layer, indicating dissolved Ni solute atoms are much more effective in hardening Sn-rich solder joints than Cu atoms [13]. Considering the big difference in solder volume between a BGA and flip-chip solder joint, the effect of substrate metallization on the interfacial reactions at the UBM side would be even more substantial in the case of flip-chip interconnection.

4.7.3 Interfacial Reactions in Pb-Free Solder Joints

The interfacial reactions in Sn-rich solder joints are very aggressive compared to those in eutectic SnPb joints, mainly due to the higher Sn content and the higher

Table 4.7 IMC phases formed in Cu(25 μm)/Ni/Sn(40 μm) samples as a function of Ni thickness and reflow time, [82]

Samples	Reflow time (min)	
Ni(ED) or Ni(EL)	2	10
0.3	$(Cu,Ni)_6Sn_5 > (Ni,Cu)_3Sn_4$	$(Cu,Ni)_6Sn_5 > (Ni,Cu)_3Sn_4$
0.6	$(Ni,Cu)_3Sn_4$	$(Cu,Ni)_6Sn_5 > (Ni,Cu)_3Sn_4$
1	$(Ni,Cu)_3Sn_4$	$(Ni,Cu)_3Sn_4 > (Cu,Ni)_6Sn_5$
3	Ni_3Sn_4	$(Ni,Cu)_3Sn_4$
10	Ni_3Sn_4	$(Ni,Cu)_3Sn_4$

Table 4.8 IMC formed phases in Cu(25 μm)/Ni/Cu/Sn (40 μm) samples as a function of Cu thickness and reflow time on electroplated Ni, [82]

Samples			Reflow time (min)	
Ni (μm)	Cu (μm)	Cu (wt%)	2	10
1	0.04	0.1	$(Ni,Cu)_3Sn_4 > (Cu,Ni)_6Sn_5$	$(Cu,Ni)_6Sn_5 \sim (Ni,Cu)_3Sn_4$
	0.3	0.8	$(Cu,Ni)_6Sn_5$	$(Cu,Ni)_6Sn_5 > (Ni,Cu)_3Sn_4$
	0.5	1.4	$(Cu,Ni)_6Sn_5$	$(Cu,Ni)_6Sn_5$
3	0.1	0.3	$(Ni,Cu)_3Sn_4 > (Cu,Ni)_6Sn_5$	$(Ni,Cu)_3Sn_4 > (Cu,Ni)_6Sn$
	0.3	0.8	$(Cu,Ni)_6Sn_5$	$(Ni,Cu)_3Sn_4 > (Cu,Ni)_6Sn$
	0.5	1.4	$(Cu,Ni)_6Sn_5$	$(Cu,Ni)_6Sn_5 > (Ni,Cu)_3Sn_4$

reflow temperature used in Pb-free soldering [69, 70]. During the reflow, there are two basic reactions occur at the soldering interfaces: dissolution of BLM and substrate metallization into a molten solder, and concomitant intermetallic formation at the interfaces. Both reactions have serious impacts on the integrity of solder joints, and should be controlled for reliable solder joints. The intermetallic phases continue to grow in the solid state during a high temperature aging, often accompanied with interfacial void formation when a disparity of diffusing species across the interface would exist. Numerous studies have been conducted to understand the fundamental characteristics of the interfacial reactions in Pb-free solders (such as intermetallic identification, growth kinetics, etc.), as well as to control the interfacial reactions to improve the reliability of solder joints. A few examples of the intermetallics formed at Cu and Ni metallization are summarized in Tables 4.7 and 4.8 [82]. With a model joint of Cu(25 μm)/Ni/Sn(40 μm), the interfacial reactions were studied in terms of Ni plating type (electroplated vs. electroless), Ni thickness, and reflow condition. The IMC phases formed in Cu/Ni/Sn joints are determined by Ni thickness and reflow time due to the involvement of Cu under-layer. Ni_3Sn_4 is detected in all samples for a short reflow time, indicating that. Ni_3Sn_4 is the first-forming IMC phase on Ni layer. For the long reflow time of 10 min, when Ni is less than 1 μm thick, Cu under-layer facilitates to form $(Cu,Ni)_6Sn_5$. The IMC morphology is strongly affected by the type of Ni plating. The angular or faceted Ni_3Sn_4 is observed on electroplated Ni, while needle-like IMC is on electroless Ni(P). The presence of Cu under-layer changes the interfacial microstructure significantly by forming the $(Cu,Ni)_6Sn_5$ phase, which

grows faster than the binary Cu_6Sn_5 or Ni_3Sn_4. The effect of Cu over-layer was investigated with the model joint of Cu(25 μm)/Ni/Cu/Sn(40 μm) samples as shown in Table 4.8. For thin Cu, $(Ni,Cu)_3Sn_4$ forms dominantly, while for thick Cu, $(Cu,Ni)_6Sn_5$ is the major phase. In Table 4.8, the Cu concentration in Sn–Cu alloys is estimated for each Cu thickness by assuming the entire Cu over-layer to dissolve into 40 mm Sn. This result is quite consistent with the previous work on the effect of Cu concentration on the interfacial reactions between Ni and Sn–Cu solders [83]. The transition of the IMC phases from $(Ni,Cu)_3Sn_4$ type to $(Cu,Ni)_6Sn_5$ occurred around 0.5 wt% Cu.

Recently, many research works have been reported to control their interfacial reactions by adding minor alloying elements to Sn-rich solders, such as Co, Fe, Ge, Ni, Mn, Ti, Zn, rare-earth metals (Ce, La), and others [12, 84–94]. Among them, Zn is found to be the most effective with Cu UBM by providing multiple benefits, such as reducing Cu consumption and IMC formation [12, 91], suppressing void formation during aging [92], reducing the undercooling of Sn-rich solder alloys [90], improving impact strength [93], enhancing electromigration resistance [94], and others. The beneficial effects of Zn addition on the interfacial reactions with Cu metallization was explained by detecting the accumulation of Zn atoms at the interface between Cu and Cu_3Sn IMC layer [12, 92]. This Zn accumulation layer was claimed to be a Cu–Zn solid solution alloy rather than a Cu–Zn IMC layer, based on a chemical etching experiment as well as a thermodynamic calculation of the driving force of Cu_3Sn phase [92]. In addition, a couple of possible mechanisms for suppressing void formation on electroplated Cu during high temperature aging was also proposed, such as reduction of Cu_3Sn growth rate or direct diffusion of Zn to the interface to fill up any vacancy sites before they grow into a large size of voids [92].

Since Zn addition has some concerns in soldering such as propensity for oxidation or degradation of wettability, it should be very careful in controlling the amount of Zn addition to Sn-rich solders. It is reported that around 0.4 wt% Zn was effective without degrading other properties. Since the direct addition of Zn into Sn-rich solders is a challenging task in solder manufacturing process, it is recently proposed to use Zn-containing UBM to control the interfacial reactions [95].

In a recent study, the Ni consumption during reflow with Sn-rich solders were evaluated with various Ni-based (electroplated, electroless and sputtered) UBMs [96]. A minor addition of Ni (0.2 wt%) to Sn-rich solders was found to be beneficial in suppressing Ni consumption for pure Sn and Sn–2Ag, but not for Sn–0.7Cu. In fact, the Ni addition has increased the Ni UBM consumption in Sn–0.7Cu. This enhanced Ni consumption in Sn–0.7Cu was attributed to the formation of $(Cu,Ni)_6Sn_5$ intermetallic phase in the solder matrix [96]. A trace amount of Ni and Ge has been added to eutectic Sn–Cu solder to improve fluidity and wettability, to prevent shrinkage cracks (hot tears) or to enhance impact fracture strength [97]. In Table 4.9, some beneficial effects of other minor alloying additions are also summarized.

Table 4.9 Effects of minor alloying elements on properties of Sn-rich solders

Properties	Ag	Cu	Bi	Co	Fe	Ge	In	Mn	Ni	RE	Sb	Ti	Zn
1) Undercooling decrease				X	X			X	X			X	X
2) Microstructure to refine	X	X				X		X		X		X	
3) Shear strength increase	X	X	X						X	X	X		
4) Ductility increase							X			X			
5) Cu dissolution decrease		X	X						X				X
6) Intermetallics control			X								X		X
7) Void formation decrease				X	X								X
8) Impact strength increase									X			X	X
9) Fatigue life increase		X									X		
10) Electromigration improve	X												X
11) Wettability improve						X				X			
References	[50, 94]	[50, 91]	[89]	[88]	[88]	[97]	[85]	[87, 98]	[87, 97]	[86]	[99, 100]	[87, 98]	[12, 90–94]

4.8 Reliability of Flip-Chip Interconnect Structure

4.8.1 Thermal Fatigue

Thermal fatigue performance of flip-chip solder joints is a key issue in developing reliable flip-chip packages both on organic and ceramic substrates. Development of flip-chip packages on organic laminates has been only possible by successfully applying underfill encapsulation technology for solder bumps. The underfill encapsulation provides compressive stress to each solder bump as well as reduces the effect of the global thermal expansion mismatch between the silicon chip (2.5×10^{-6}/°C) and the organic FR-4 PCB (18.5×10^{-6}/°C) by deforming together as a unit to reduce the relative deformation between the chip and the PCB [101]. Without the underfill encapsulation, flip-chip packages on organic substrates would not survive more than a few hundred cycles in a typical thermal fatigue test condition. Flip-chip packages on ceramic substrates would generally not need the underfill encapsulation owing to their small or closely matched thermal expansion mismatch with ceramic substrates. However, as the chip size is continuously increasing for high performance flip-chip interconnection, the underfill encapsulation could be an option to extend the thermal fatigue life or other reliability performance of ceramic flip-chip packages at the expense of reworkability of flip-chip solder joints.

In the early Pb-free, flip-chip development, several research works were reported on the thermal fatigue performance of Pb-free solders compared to Pb-containing flip-chip solder joints [32, 45, 102]. Three Pb-free solders, Sn–0.7Cu, Sn–3.8Ag–0.7Cu, and Sn–3.5Ag, formulated as solder paste, were evaluated for flip-chip applications [102]. Test dies of 12.6×7.5 mm^2 with TiW–Cu and Ni (P)–Au UBM were directly attached to organic boards with Cu-OSP or Ni(P)–Au pad finish. To accelerate solder bump fatigue, no underfill encapsulation was used on the assembled parts. The flip-chip packages were then tested in air under 0 to 100 °C and −40 to 125 °C thermal cycling conditions. Among the solders/UBM evaluated, Sn–0.7Cu bump on both Ni(P) and TiW/Cu UBMs had the longest fatigue life, while Sn–3.5Ag on Ni(P) UBM the shortest life [102]. The Sn–3.8Ag–0.7Cu on TiW/Cu UBM had a better fatigue life than Sn–3.5Ag, worse than Sn–0.7Cu, and a similar life to Sn–37Pb on Ni(P). The better fatigue performance of Sn–0.7Cu joints were explained by the fatigue crack initiation and propagation mechanism through the grain boundaries. The cracks were observed to propagate at the grain boundaries, significantly removed from the UBM–bump interface near the center of the joint. This solder was claimed to be most compliant in thermal fatigue and to undergo massive deformation before failing by crack propagation [45].

The effect of silver content on thermal fatigue life of Sn–Ag–Cu flip-chip interconnects was systematically investigated for flip-chip packages assembled to FR-4 substrate with SAC solder balls of 300 μm in diameter [103]. From the thermal cycling test in air between −45 and 125 °C, it was found that solder joints with a high Ag content, 3 and 4 wt%, had a longer fatigue life compared to solder joints with a

lower Ag content, 1 and 2 wt%. The better fatigue performance of solder joints with a higher Ag content was attributed to the stable microstructure owing to the dispersion of fine Ag_3Sn intermetallic particles. In the low Ag solder joints, significant microstructure coarsening was observed during the thermal cycling test [103]. In another study, the effect of Ag content on thermal fatigue life of ceramic BGA modules mounted on an organic substrate was evaluated in terms of Ag content, cooling rate and thermal cycling conditions [104]. It was found that the fatigue life was influenced by Ag content as well as thermal cycling test condition. The low-Ag joints (2.1 %) had the best thermal fatigue life for the thermal cycling condition of 0–100 °C with a long cycle time of 120 min, while the high-Ag joints (3.8 %) had the best life over other joints for the condition of 0–100 °C with a short cycle time of 30 min. The slow cooling rate (0.5 °C/s) used during assembly was beneficial for the thermal fatigue life of SAC joints regardless of Ag content or thermal cycling conditions in comparison to the fast cooling rate (1.7 °C/s). Extensive failure analysis was conducted with thermal-cycled solder joints to propose the failure mechanisms operating during the accelerated thermal cycling (ATC) tests [105].

The fatigue performance of Pb-free flip-chip solder joints would be strongly affected by Sn crystal orientation and its directionality in physical/mechanical properties, since flip-chip solder bumps are expected to be either a single crystal or composed of a few grains, similarly as reported in much larger BGA joints of Pb-free solders [59, 106]. When Sn crystal orientation plays in thermal failure process, the conventional fatigue failure mechanism, largely affected by the DNP-related factors (such as chip size, solder bump height, CTE mismatch or thermal cycling temperature difference, etc.), would be more complex to understand. In the conventional thermal cycling test, the solder bumps at the corner of a Si chip with the largest DNP would expect to fail first, but this situation would not be warranted when Sn crystal orientation plays into the failure process. A similar case has been already reported in the thermal cycling test of Pb-free, BGA solder joints [106], when the *c*-axis of Sn-crystal is oriented parallel to the substrate direction, premature fails were observed regardless of the position of solder joints [59].

4.8.2 Drop Impact Reliability

With the recent advent of ubiquitous portable or mobile electronics, drop impact resistance of solder joints has been recognized as a critical reliability issue, especially for chip scale packages (CSP), wafer level packages (WLP), micro-BGA or BGA, where underfill encapsulation is not normally applied [107–110]. For flip-chip-on-board (FCOB) applications, underfill encapsulation is necessary to meet thermal fatigue requirements, and is also beneficial for its drop impact reliablity [109].

The root cause of brittle fracture of solder joints during drop test has been identified as the weak interface between IMC layers and Cu pad [111]. It was reported that upon thermal aging solder joint strength can drastically degrade due to the

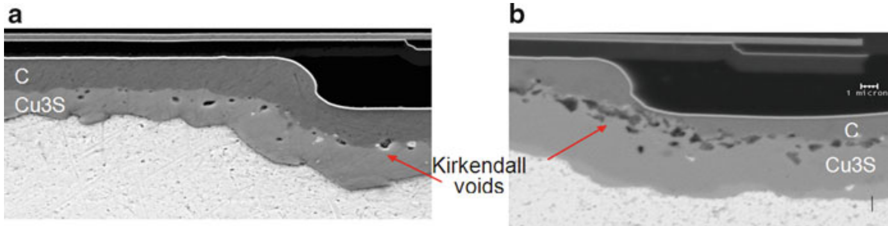


Fig. 4.33 (**a**) Thick Cu BLM—incoming, (**b**) Thick Cu BLM—HTS 1,000 h

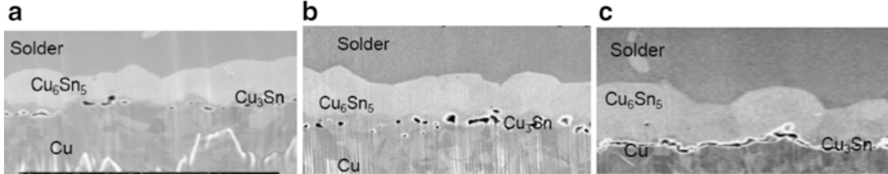


Fig. 4.34 SnAgCu solder joints on Cu and annealed at 125 °C for (**a**) 3 days, (**b**) 10 days and (**c**) 40 days [111]

formation of Kirkendall voids at the interface between IMC and Cu pad in Pb-free solder joints [111, 112]. However, this weak interface is not easily detected by the mechanical testing with a slow strain rate, such as ball shear test, but by drop or impact test of a high strain rate [111, 112]. It is also reported that the formation of interfacial voids is only sporadic with certain electroplated Cu pads, rarely observed in high purity or wrought Cu foils even after extensive thermal aging [92, 113].

In case of Pb-free flip-chip interconnects having thick Cu UBM, similar interfacial voids were observed during a high temperature storage test at 150 °C, 1,000 h, as shown in Fig. 4.33. The concentration of voids between the Cu and Cu_3Sn IMC layer suggests the void formation mechanism being related to the growth of Cu_3Sn layer. This was one of the reasons why the thick Cu UBM structure was not adopted for Pb-free flip-chip applications.

Kirkendall void is a reliability issue when Sn containing solders, both Pb-containing and Pb-free, are joined to thick Cu pad and annealed for a prolonged time. Figure 4.34a shows a SnAgCu alloy joined to Cu and annealed at 125 °C for 3, 10, and 40 days, respectively. The voids are formed in the Cu_3Sn intermetallic layer near the interface with Cu. It can form an almost continuous layer after annealing for 40 days and seriously impact the drop reliability of the package assembly, as shown in Fig. 4.35 [111, 114]. The formation of voids is highly variable. Depending on the property of the plated Cu the void density may increase dramatically. The voiding is faster at higher temperatures but even products that do not get very hot in service may still be endangered.

In order to improve the drop reliability of Pb-free solder joints, two approaches have been taken: either reducing Ag or Cu content in Sn–Ag–Cu solders, adopting a low Ag version, such as SAC105 (Sn–1.0 %Ag–0.5 %Cu) [109, 115, 116], or

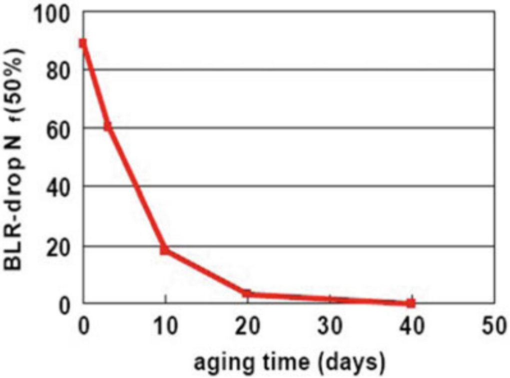


Fig. 4.35 Drastic degradation of the drop reliability of a SAC solder joint after aging at 125 °C for up to 40 days [111]

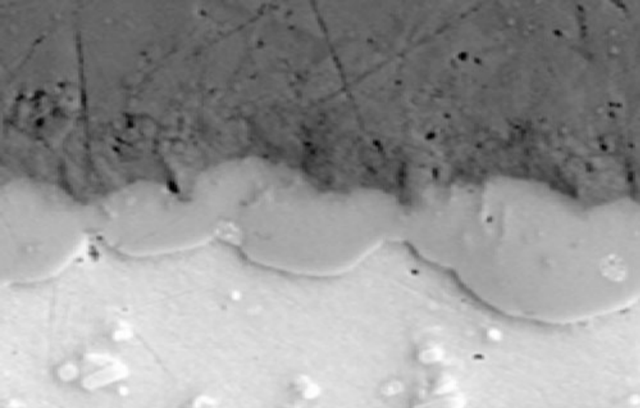


Fig. 4.36 SnAgCu with minor Zn addition joined to Cu and annealed at 150 °C, 1,000 h, showing no void [12]

adding minor alloying elements to control their interfacial reactions and thereby to suppress the interfacial void formation [88, 93, 97, 98, 109, 115]. The additions of Ti, Mn, Ni, or In were reported to improve the impact resistance of Sn–Ag–Cu joints [98, 115], while Ni and Ge were added to Sn–Cu [97], and Zn added to Sn–Ag [56] for the same purpose. One solution to the problem is by doping the solder with a small amount of Zn and the voids are mostly eliminated, as shown in Fig. 4.36 [12].

4.8.3 Chip-Package-Interaction: Interlayer Dielectric Cracking During Module Assembly

Critical reliability challenges were encountered during the qualification of Pb-free solders in microelectronics packages. The challenges are associated with the inability of accommodating the large stress/strain induced during chip joining by

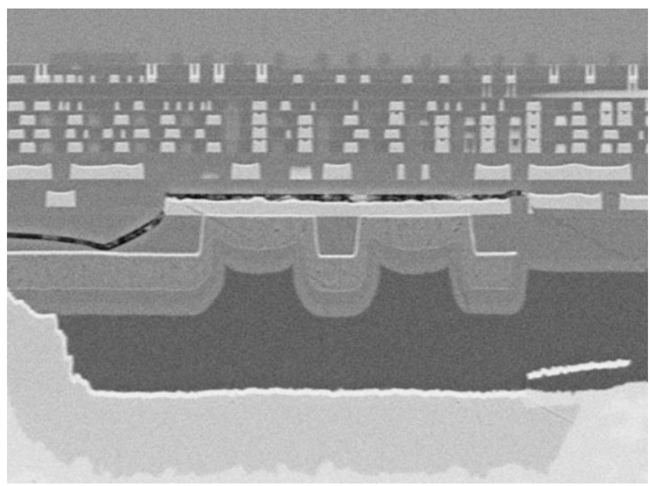


Fig. 4.37 SEM image of interfacial delamination between the Cu pad and the FTEOS layer [21]

high-strength Pb-free solders. This situation often leads to cause either delamination between the BEOL layers or cracking within the low-k layer in the back-end-of-line (BEOL) structure [21, 47, 117]. The first type is commonly referred to as ILD (Interlayer Dielectric) delamination [21, 47] caused primarily by the greater thermo-mechanical stresses imposed on the BEOL structure due to the use of high strength Pb-free solder. The problem is further aggravated by the use of large chips which generate higher stresses due to the global differential thermal expansion mismatch between the chip (CTE ~3 ppm) and laminate carrier (CTE ~17 ppm). Delamination failure between BEOL layers is shown in Fig. 4.37. The open circuit can be detected by using acoustic scan imaging and is identified as a "white bump," as shown in Fig. 4.38. To mitigate the problem, a solder with a higher creep rate can be chosen to deform more easily during chip joining, and therefore it reduces the stresses transmitted to the BEOL layers. In addition, improving the adhesion strength between the layers was shown to effectively mitigate the problem. The other type of commonly observed white bump failures involves the coherent cracking within the ultra low dielectric (ULK) layer, as shown in Fig. 4.39 [24]. Corner C4's due to higher DNP (distance from neutral point) stresses are particularly susceptible to cracking. As a consequence of the continued reduction in dielectric constant, typically accompanied with increased porosity, the ILD materials become increasingly fragile. Elastic modules and fracture toughness of the low-k layers decrease rapidly as porosity increases. Cracking in the ILD layers under the flip-chip bump is frequently observed. The move to higher CTE organic laminate (~16 ppm) from the ceramics (3–6 ppm) further aggravates the problem due to increased CTE mismatch and warpage of both the laminate and chip.

To mitigate the white bump problems the fundamental solutions are to reduce the thermo-mechanical stresses, in particular, the tensile stress transmitted to the ILD layers. Reported stress mitigation methods include optimization of the

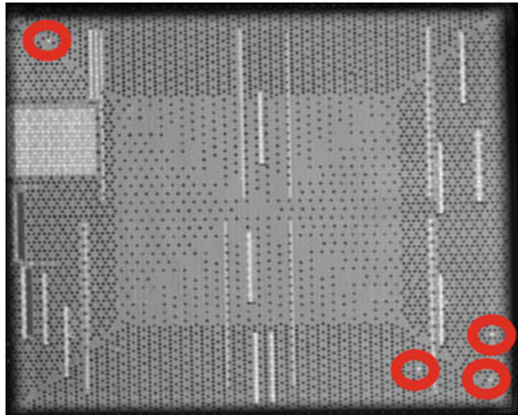


Fig. 4.38 Acoustic scan image showing the "white bumps" [21]

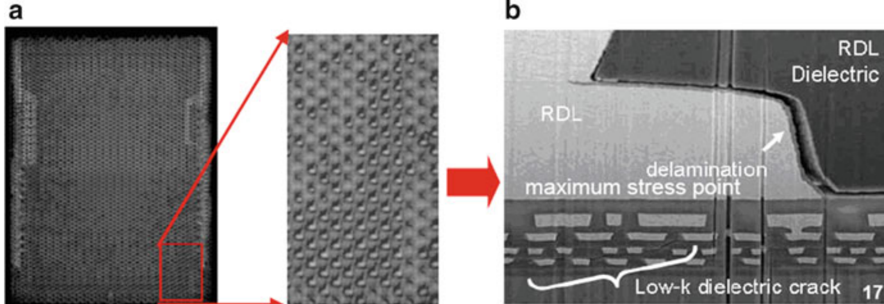


Fig. 4.39 Open failures detected by (**a**) the acoustic scan image at chip corner, was caused by (**b**) coherent cracking in the uLK layer [24]

mechanical properties and microstructure of Pb-free solders [47, 117], temperature profiles during chip joining to the laminate substrate [21, 117], improving the BEOL structure, reducing laminate and chip warpages, and use low CTE laminate material. Stress reduction to the ILD layer can also be achieved by designing a geometrically compliant interconnect which can readily absorb the stresses before transferred to the chip.

Reducing the cooling rate during reflow is shown to reduce the stresses and the incidents of WB. It, however, increases the process time and reduces manufacturing throughput. Solder with a higher creep rate is desired because it deforms more easily and, therefore, reduces the stresses transmitted to the BEOL structure. A faster creep rate often translates into a longer thermo-mechanical fatigue life. Sn–0.7Cu bumps creep more easily than Sn–3.5Ag bumps and were found to cause less warpage in chip and laminate and, therefore, less stresses to the BEOL layers. To understand the solder stress issues, C4NP bumping technology was used to facilitate quick debug, optimization and, finally, identification of the solder solution among a comprehensive list of solder candidates. Table 4.10 correlates the solder compositions with their

Table 4.10 Module level solder joint microhardness and indentation measurements

wt% Sn	wt% Ag	wt% Cu	Hardness, HV mean (standard deviation)
97.6	2.2	0.2	**16.0** (0.6)
98.5	1.3	0.2	**14.5** (0.8)
98.5	0.9	0.6	**14.0** (0.0)
98.6	1.2	0.2	**14.0** (0.6)
99.5	0.3	0.2	**12.0** (0.9)
99.3	0	0.7	**11.5** (0.5)

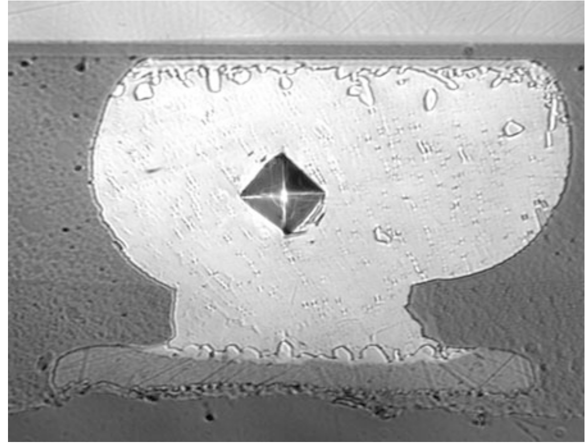


Fig. 4.40 A typical module level microhardness indentation measurement on solder joint

hardness data, which agree with the laminate warpage data. Low Ag and Cu content solders have lower hardness and more desirable mechanical properties to absorb the stresses before transmitted to the ILD layer to cause cracking. As a result the use of low Ag and Cu was shown to significantly reduce white bumps than the conventional near-eutectic SAC solder alloys. Figure 4.40 shows a microhardness indentation on the C4 solder joint.

Improving the far-BEOL structure is another effective method to reduce the stresses transmitted to the low dielectric material layer. As shown in Fig. 4.41, several critical elements in Intel's BEOL structures processors have been changed to manage the high stresses of the Cu pillar bumps, including a thick polyimide layer (~16 μm) under the Cu pillar bump, small via and a thick M9 Cu pad [24]. The structure optimization is designed to reduce the stresses to the low-k, uLK layer. The cross-sectional view is shown in Fig. 4.41.

With the continued movement to ultra low-k materials in the 32 and 22 nm nodes of BEOL silicon technology which is more porous with low modulus, the chip package interaction (CPI) problem will become even more challenging.

To provide a low-stress interconnect for flip-chip BGA package with Pb-free solders, the creep properties of Sn–0.7Cu vs. Sn–3.5Ag bumps were measured by nano-indentation technique at various temperatures [47]. It was found that Sn–0.7Cu bumps crept more easily than Sn–3.5Ag bumps, and the difference in the creep rate between them increased at a higher temperature. The different creep rates were

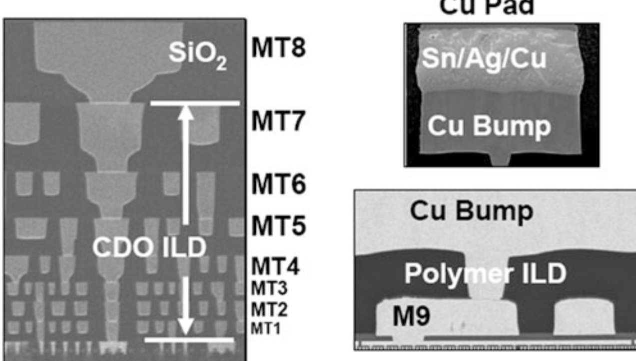


Fig. 4.41 The BEOL of Intel's processor chips have been drastically changed to prevent cracking of the low k materials [24]

explained by the different characteristics of IMC particles formed in each solder system. The chip warpage and the shear stress at the low-k layer were also calculated and measured; finding that Sn–0.7Cu bumps had the smaller chip warpage and shear stress than Sn–3.5Ag bumps. The maximum shear stress with Sn–0.7Cu bumps was about 11 % less than the delaminating stress of the low-k layer. In addition, further reduction in the maximum shear stress (36 %) was demonstrated by applying post-annealing at 200 °C for 10 min with Sn–0.7Cu bumps, but not much with Sn–3.5Ag. This positive result with Sn–0.7Cu bumps was again attributed to the higher creep rate of Sn–0.7Cu compared to Sn–3.5Ag.

As already discussed in the section of Drop Impact Reliability, low-Ag version of Sn–Ag solders were proposed to reduce the propensity for CPI damages [117]. Combining various characterization techniques such as dynamic chip warpage measurement (DCWM), confocal scanning acoustic microscope (CSAM), and EBSD (electron backscatter diffraction), it was able to demonstrate that solder creep deformation is the limiting process for CPI damages. Lowering the creep resistance (by reducing Ag content) and reducing the strain rate by slow cooling were confirmed to be beneficial in reducing the propensity of CPI dames on actual parts. The reduction in Ag content (from 2 to 0 %) in Sn–Ag bumps was very effective to reduce the peak stress by 40 %, down to levels comparable to SnPb solders. It was also noted that the significant local variations of CPI damages, without a strong dependence on the distance to the chip center, were explained by the "local" parameters of Sn grain size and orientation by analyzing EBSD data of damaged and undamaged sites [24].

4.8.4 *Electromigration Reliability*

Electromigration (EM) of metallic atoms in device conductors, such as Al or Cu has been well recognized as critical reliability and design issues in advanced integrated circuits (IC), while electromigration in flip-chip solder joints is not considered as a

serious concern simply because of their much larger dimensions and corresponding low current density. However, in the recent years, due to the continuing trend of miniaturization in IC and their interconnects, the electromigration of flip-chip solder joints became an important reliability issue, especially owing to the effects of current crowding or joule heating [118].

In the most recent development of Pb-free flip-chip technology, the electromigration in Sn-rich solder joints has become a critical reliability challenge, largely because of lower melting temperatures of Sn-rich solders in comparison to high Pb solders as well as the remarkable anisotropy in diffusion rates of common solute atoms such as Cu, Ni, or Ag in the Sn matrix as discussed earlier. Since the general subjects of electromigration in flip-chip solder joints are covered in a separate chapter, only a few selected topics regarding the microstructure effects of Pb-free solders on electromigration are briefly discussed in this section.

Early EM test results obtained from actual flip-chip solder joints are often complicated to compare each other because their solder temperature and current density significantly vary among samples due to current crowding and local joule heating. In order to avoid these complications, a model wire test structure, providing uniform current density and minimal gradients, was adopted to compare pure electromigration effects of solder composition, UBM, and surface finish UBM [60, 119]. From this study, it was found that Sn–Ag joints have a superior EM performance over Sn–Cu under the same conditions of other variables [119]. In addition, two failure mechanisms are identified [60]; Mode-I, probably dominated by Sn self-diffusion resulting in separation between IMC and solder. Mode-II is responsible for premature fails in EM tests, which is clearly dominated by a fast diffusion process of Ni and/or Cu in Sn, when the *c*-axis of Sn grain orientation is parallel to the direction of electron flow. EM failure mechanism in Sn–Ag joints with higher Ag content is dominated by Mode-I fail, while more Mode-II was observed in Sn–Cu joints. In the subsequent study, the alloying effects of Ag, Cu, and Zn on the EM performance of Sn-rich solders have been systematically investigated using the Cu wire structure [94]. For Sn–Ag joints, the frequency of early EM fails (associated with Mode-II) significantly decreases as Ag content increases, while Mode-II fails are more commonly observed in Sn–Cu joints for all Cu concentrations tested. The EM lifetime of Sn–Cu joints is generally shorter than Sn–Ag or Sn–Ag–Cu joints. The better EM performance of Sn–Ag over Sn–Cu is explained by the stable network of Ag_3Sn particles during EM or high temperature aging experiment [120], and also supported by EBSD work on Sn grain orientations.

The beneficial effect of Zn doping on EM performance is also reported for Sn–Ag joints [94]. It is observed that Zn combines closely with Cu and Ag that stabilizes the IMC network and effectively slows down Cu diffusion. Hence, EM reliability is significantly improved in Zn-doped Sn–Ag joints. The effects of other alloying elements such as Ni, Sb, or Bi were also investigated, but finding no appreciable improvement on EM performance.

EM degradation mechanisms in Pb-free solder joint are highly dependent on crystal orientation. Sn, with a compressed body center tetragonal structure, as shown in Figs. 4.42 and 4.43 [106], has a highly anisotropic electrical, mechanical

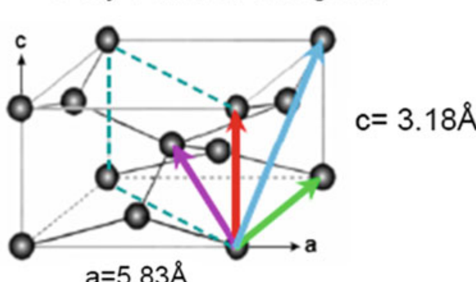


Fig. 4.42 The compressed Sn structure shows *c*-axis is much shorter than the *a*- and *b*-axis [106]

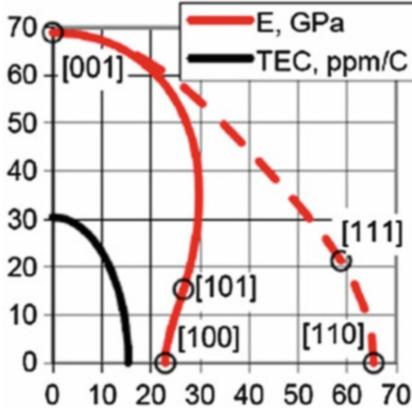


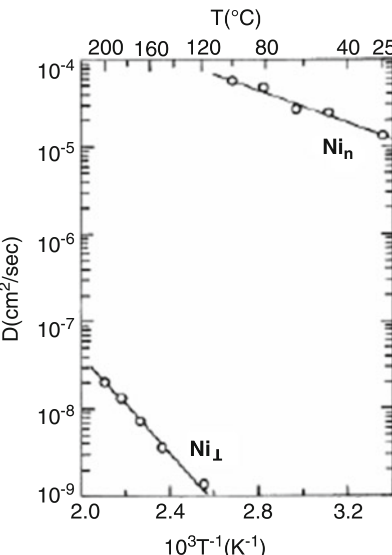
Fig. 4.43 The CTE and modulus of Sn are highly anisotropic [106]

of diffusion properties. For fast diffusers, such as Ni and Cu in Sn, the difference in diffusivities between *c*-axis and *a*- or *b*-axis can be very large as shown in Fig. 4.44 [55–58].

EM degradations were found to depend on grain orientations [60]. As shown in Fig. 4.45, when the *c*-axis is not closely aligned with the current direction (the right grain), the failure is mostly driven by Sn self-diffusion that causes cavitation at the solder–IMC interface. On the other hand, a much faster EM induced degradation mechanism occurs when the *c*-axis of the Sn grain is closely aligned with the current direction (on the left grain). With this alignment, the solute atoms such as Cu or Ni from the UBM and interfacial IMC are swept away by the fast interstitial diffusion along the *c*-axis of the Sn crystal, resulting in rapid consumption of UBM metallurgy and catastrophic failure at an early stage of EM test [60].

Statistically, the rapid failure driven by interstitial diffusion of Cu and Ni is more commonly observed in SnCu than in SnAg solder joints, as shown in Fig. 4.46. Kinetic study showed that the activation energy and current density exponent for SnAg solders (Mode-I dominated) are about 0.95 eV and 2, respectively; while those for SnCu solders are about 0.54 eV and 1, respectively [60, 119].

Fig. 4.44 The rate of diffusion of Ni in Sn is much faster along the *c*-axis [56, 57]



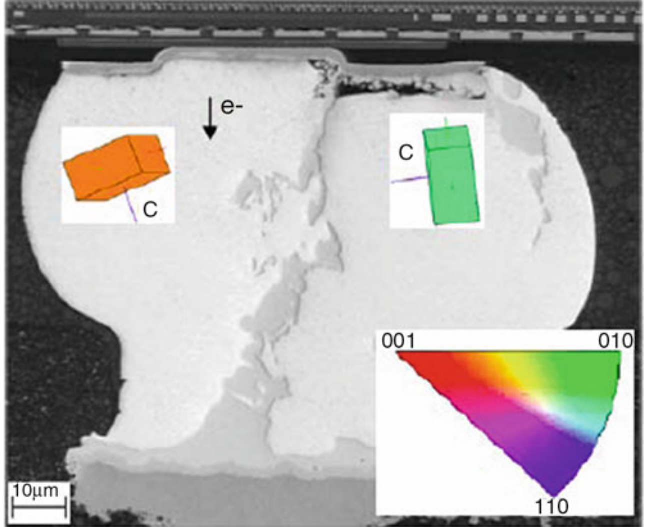


Fig. 4.45 SEM image of a bi-crystal C4 bump with SnCu solder [60]

The Blech length effect was observed in SnAg solders, but not in SnCu solders [121, 122]. The saturation in resistance change due to the Blech effect can effectively suppress the diffusion processes and extend the EM lifetime of flip-chip solder joints. The effect of special alloy doping is important on the EM performance as shown in Fig. 4.47, for SnAgCu alloy [123]. The increase in resistance is plotted

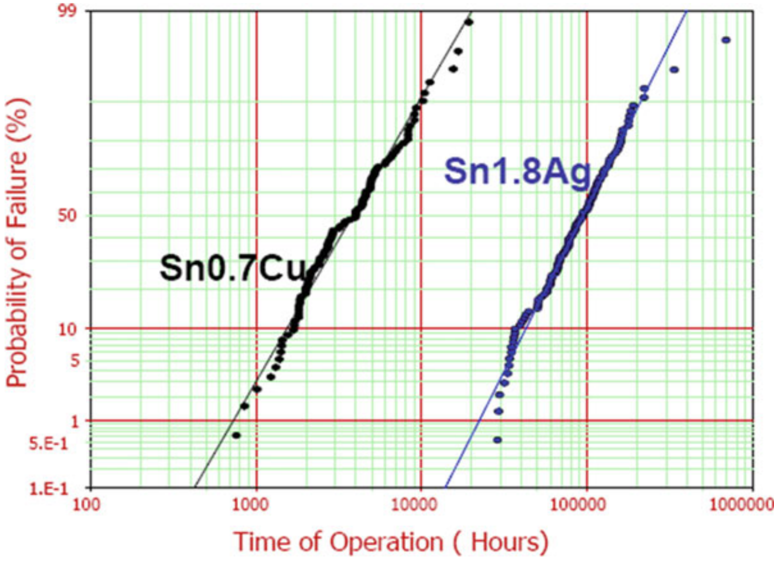
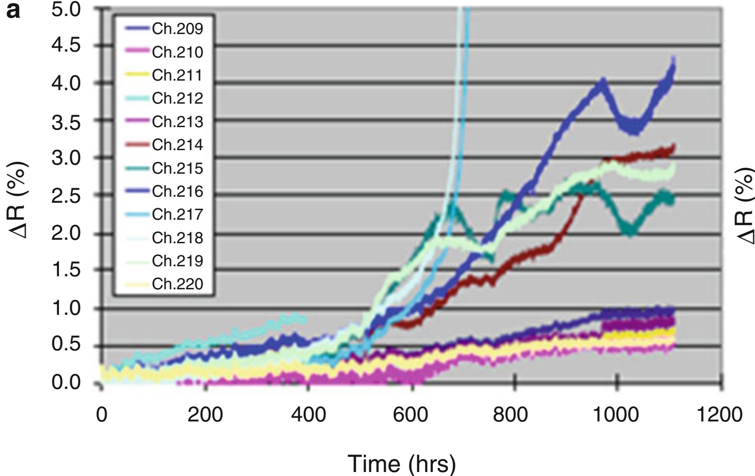


Fig. 4.46 Cumulative failure probability plot comparing SnCu vs. SnAg. Both solders were tested at 90 °C at 200 mA [119]

as a function of test time when the solder joints were stressed at 5.2×10^3 A/cm^2 and 150 °C for 1,100 h. For SAC alloy, shown in Fig. 4.47a, some samples showed early failures due to resistance increases that exceeded failure criteria. In comparison, SAC solder doped with minor Zn alloying element significantly enhanced EM performance by eliminating the early failures, as shown in Fig. 4.47b. While it is difficult to control the Sn grain orientation, minor alloy doping can influence the Sn microstructure and, thus, the EM performance. For example, Zn reacts strongly with the alloying elements, such as Cu, Ag, and Ni, in the Sn-rich solder matrix, as well as at the solder–UBM interfaces. The strong binding of Zn with Cu effectively slows down Cu migration. It is shown that Zn stabilizes both Ag_3Sn and Cu_6Sn_5 IMC networks and suppresses the formation of Cu_3Sn IMC. Although Zn doping does not seem to control the grain orientation in bulk solder, the random grain orientations at the interfaces and strong binding with Cu effectively eliminate the mode-II type early failures and suppress the mode-I type fails. Consequently, the solder EM lifetime with Zn doping is greatly improved [123].

In addition to the effects of grain orientation and solder alloying, EM performance can be significantly improved by reducing current crowding [124, 125]. Figure 4.48a is a two-dimensional simulation of current distribution in a solder joint. Figure 4.48b is a display of current density distribution in the joint, where the cross section of the joint is plotted on the *x*–*y* plane and the current density is plotted along the *z*-axis. It is the current crowding or the high current density shown at the upper right corner in Fig. 4.48a, b that leads to electromigration damage in the solder joint. Consequently, electromigration damage in a flip-chip solder joint can occur near the cathode contact on the chip side.



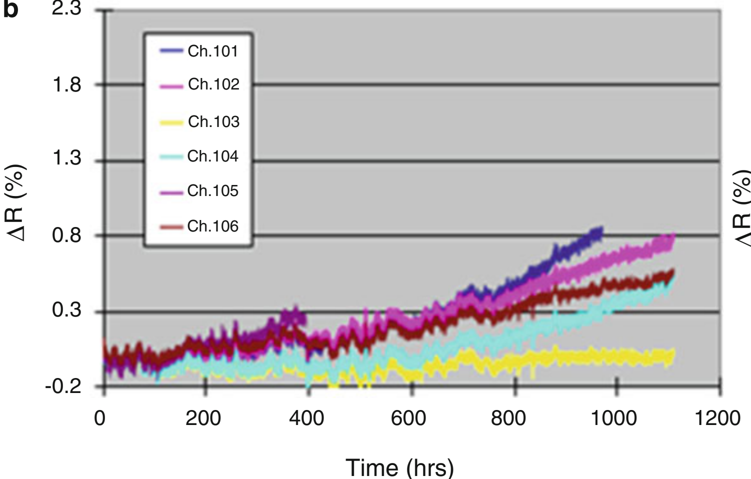


Fig. 4.47 (**a**) SnAgCu solder joint stressed at 5.2×10^3 A/cm^2, at 150 °C for 1,100 h. Some samples had early failures. (**b**) A minor Zn doped SnAgCu solder joint stressed under the same condition. Early failures have been eliminated [123]

Current distribution in a flip-chip solder joint can be studied using finite element analysis as a function of geometry and resistance of all the conducting elements associated with a solder joint, including the Al or Cu interconnect, the UBM, and the solder bump itself. The factors that affect current distribution the most have been found to be the thickness and resistance of under-bump metallization. If Cu is part of the UBM, the thickness of the Cu can affect current crowding greatly. For a thick Cu UBM of 5 μm, the highest current density due to current crowding will

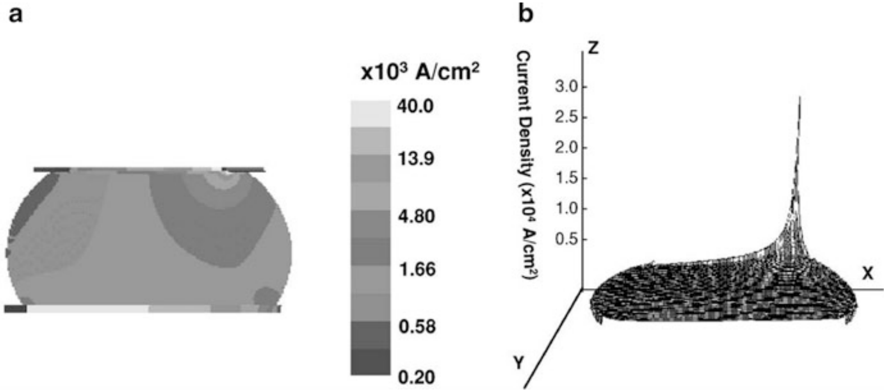


Fig. 4.48 (**a**) Two-dimensional simulation of current distribution in a solder joint. (**b**) Current density distribution in the joint, where the cross section of the joint is plotted on the x–y plane and the current density is plotted along the z-axis [124]

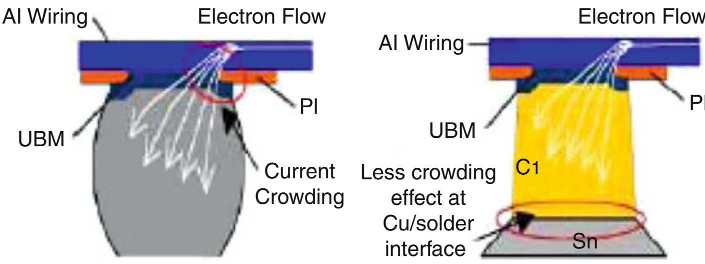


Fig. 4.49 Schematic representation of current spreading mechanism comparing solder bump with cu pillar bump [126]

occur within the Cu. Simulation has shown that the maximum current density can be reduced by at least 20 times with a 20 μm Cu UBM as compared to that of a thin UBM. More importantly, the thick Cu will enable a redistribution of current laterally in the entire Cu UBM, so the current density in the solder bump will be much closer to the average value, i.e., only a slight current crowding occurs in the solder near the Cu–solder interface. By using 3D simulation, the current redistribution in the thick Cu and the solder is found to be very uniform. As a consequence, compared with the conventional solder joint with thin UBM, the benefit of thick Cu pillar bump to enhance EM performance has been widely reported [124, 125].

Figure 4.49 shows a schematic representation of the EM mechanism in conventional solder vs. Cu pillar [126]. The thick pillar, normally more than 40 μm in thickness, significantly improves the current distribution uniformity by eliminating current crowding and hot spot, and, therefore, enhances the EM performance, as indicated in the plot of Fig. 4.50. The enhanced current distribution can be seen in the SEM images of a cross section through PbSn (Fig. 4.51a) and Cu (Fig. 4.51b) bumps

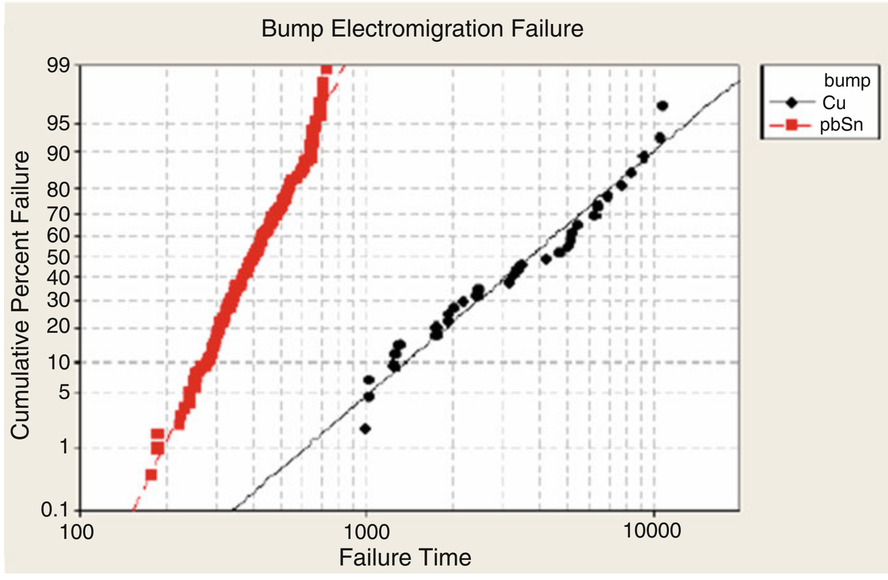


Fig. 4.50 Cumulative failure plot comparing a conventional PbSn bump with a Cu pillar bump [29]

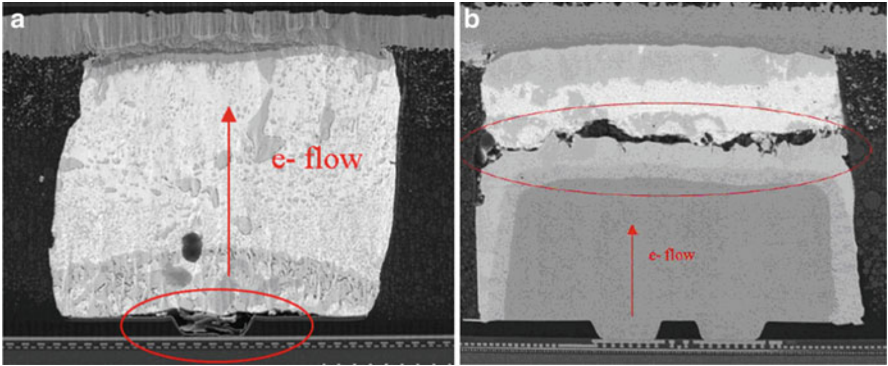
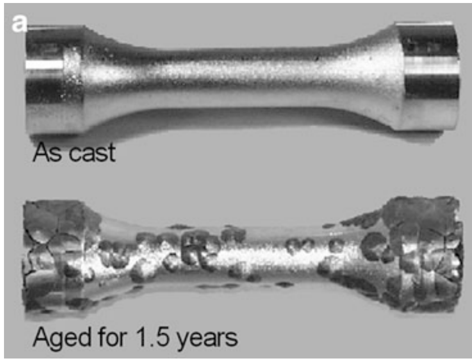


Fig. 4.51 SEM images of a cross section through (**a**) a conventional PbSn bump and (**b**) Cu pillar bump [29]

which have exhibited failure during EM stressing. For the conventional solder bump, the failure occurs in the pad via opening indicating high current crowding in that small region. For the Cu pillar bump, failure occurs away from the via region and shifted to a much greater area at the Cu–solder interface. As reported, despite the significant improvement in EM performance, the extremely stiff structure and, therefore, the high stresses generated by the tall Cu pillar bumps can make the implementation of low-k dielectric materials more challenging. It can often lead to partial or complete



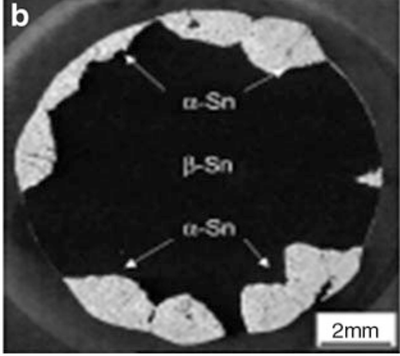


Fig. 4.52 (**a**) Top view Sn0.5Cu samples aged for 1.5 years at 8 °C, compared to an as machined sample [127]. (**b**) Cross-sectional view shows Sn pest transformation starts from the surface in the highly stressed grip area [127]

die cracking. A special stress mitigation architecture in the silicon back-end structure was necessary, as shown in Fig. 4.41 [27, 42], to accommodate the Cu pillar bump on low-k materials. Controlling the growth of CuSn IMCs and Kirkendall void formation, especially on the side wall of the Cu column, is another reliability issue that needs to be resolved.

4.8.5 Sn Pest

Sn pest, the allotropic transformation of β-Sn (body centered tetragonal) into α-Sn at temperatures below 13 °C, is another reliability issue. It has been observed in Pb-free solders at low temperatures, shown in Fig. 4.52a, b, [127]. The transformation normally takes long time to happen. It is accompanied by an increase in volume by 26 % and solder joint could practically disintegrate very rapidly. The presence of residual stress in solder joint accelerates the transformation process. Doping the solder with a small amount of Bi and Sb was reported to suppress Sn pest formation [62, 128].

4.9 Future Trends in Flip-Chip Technology

As feature size continues to scale down, the number of transistors and interconnects on a chip has increased continuously. As a result, the number of chip to package input/output (I/O) interconnects have also increased significantly in the past decades. On the one hand, flip-chip I/O pitch is being reduced continuously to meet the requirement of I/O counts in high-performance and high-bandwidth applications. On the other hand, chip stacking with fine-pitch wire bond

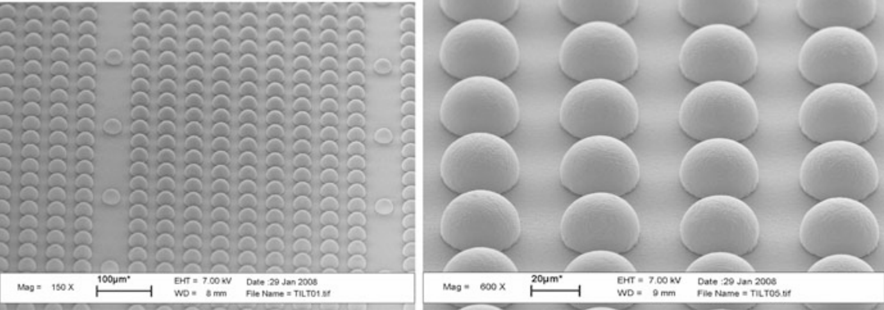


Fig. 4.53 A full array of micro-bumps at 50 µm pitch

interconnection in low-cost memory and handheld applications are currently being replaced by fine-pitch area interconnection due to the high cost of Au and performance limitation of wire-bond technology in high-frequency regime. Furthermore, fine-pitch interconnection is highly demanded for 3D integration of semiconductor chips. Therefore, manufacturing of fine-pitch flip-chip (50 µm pitch or below) interconnects needs to be explored. Integrated circuit (IC) system performance is significantly enhanced by 3D integration of chip and packages because of the benefits of high bandwidth, low latency, low power, and small form factor for a variety of applications. As demand for higher performance and higher bandwidth continue to increase, chip stacking with high-density thru-silicon-vias (TSV) interconnection is being developed and receiving more attention [129–131].

4.9.1 Conventional Micro-Solder Joint

Flip-chip interconnects can be produced by a number of methods. However, not all of the flip-chip bumping technologies are extendable to fine-pitch applications for volume production at low cost. Figure 4.53 shows SEM images of full area arrays of micro-bumps made of SnAg solder transferred to a 200 mm wafer at 50 µm pitch using C4NP technology. With the full area arrays, each test chip (6.5 × 5.4 mm) contains approximately 11,000 micro-bumps and a wafer contains ~9 million micro-bumps in total. To be able to inspect all the bumps, a high resolution tool is required with large memory for data processing. Figure 4.54 shows SEM image of a micro-bump in cross section after chip joined to an organic laminate with flux [23].

Figure 4.55a schematically shows stacked chips joined on a substrate with a sequential reflow process. In sequential reflow, the bottom chip is joined first onto the substrate, followed by subsequent chip joining. A sequential process can avoid relative displacement between chips as each subsequent chip is joined into the stack. Figure 4.55b shows examples of 2-, 3-, and 4-layer stacks of thinned TSV chips, utilizing sequential reflow of C4 interconnections. A major drawback of a sequential reflow process is that multiple reflows are necessary to complete the stack assembly.

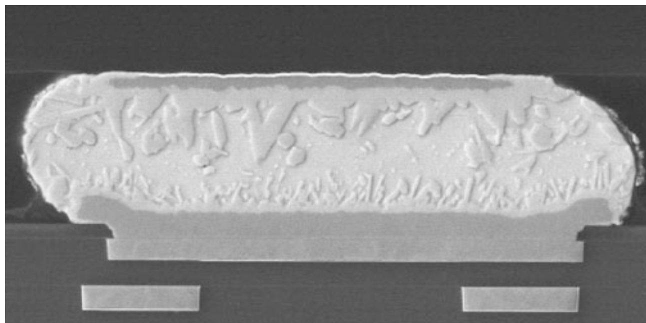


Fig. 4.54 A SEM image of a cross-sectioned micro-bump at 50 μm pitch with 28 μm diameter UBM pad

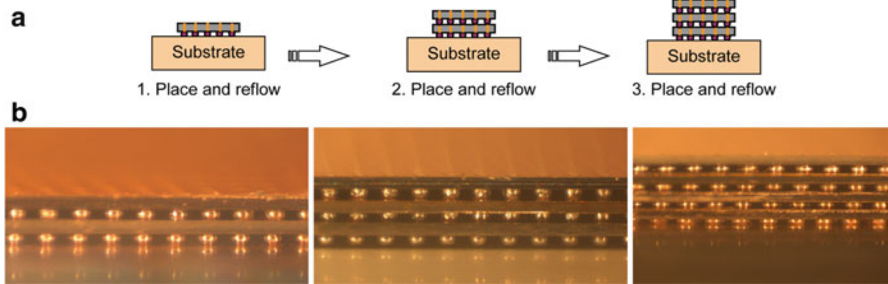
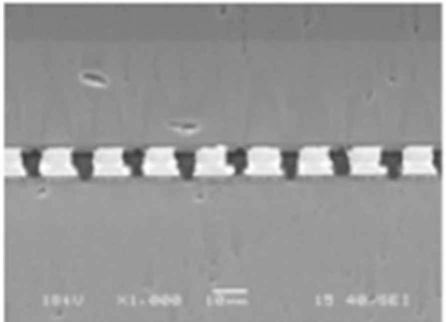


Fig. 4.55 (**a**) Schematic drawing of a sequential reflow process and 16 (**b**) photos of 2-, 3- and 4-layer stacks of thinned TSV Chip [132]

Multiple reflows require more processing time and lead to more dissolution of UBM, especially for the C4s contained in the lowest level of the stack, a concern for high reliability and high-performance applications. An alternative, parallel reflow process has also been demonstrated. A tacky flux is used to hold the stacked chips in place before the reflow process. Up to four layers of thinned TSV chips have been successfully formed with a single reflow step. With the self-centering effect of C4 bumps, a small amount of displacement between the chips is well compensated during reflow. To meet the demand for high I/O counts in high-performance and high-bandwidth applications, flip-chip I/O pitch needs to be reduced continuously. According to the International Technology Roadmap for Semiconductors (ITRS), the area-array flip-chip I/O bump (C4) pitch will be less than 70 μm for high-performance applications by 2018 [132].

To offer high degree of miniaturization, in particular, the 3D packaging technology, the assembly of thin IC's with thickness down to 10–50 μm and pitch down to less than 20 μm has been realized. The conventional 50–100 μm high solder bumps cannot be applied. Smaller solder volume, typically in the range of a few micrometers, by electroplating or immersion soldering, can be deposited on fine-pitch Cu, Au, or Ni UBM pad. Due to the thin solder layer between chip pads the mechanical stability of



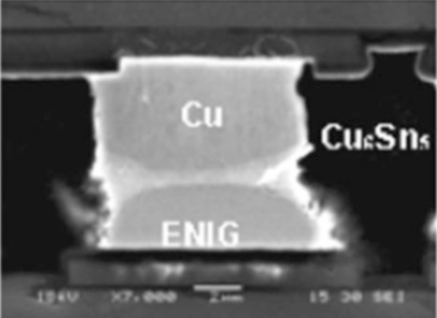


Fig. 4.56 A cross section of Cu/Sn micro-bump joined to ENIG at 15 μm pitch [133]

solder joint is mostly dominated by the intermetallic compounds. These thin interconnects are one of the key element in realizing thin modules and 3D assemblies to meet the requirements of a variety of applications. In one extreme 3D integration may require stacking of multiple dies without the disturbing of the previous bonded dies [129, 131, 132]. This can be achieved by transient liquid phase (TLP) die-to-die or die-to-wafer bonding [133]. The most commonly used eutectic system is Au–Sn, Cu–Sn, Cu–In and Au–In systems. Sn or In can be applied on one or both sides of the bonding pads. When pressure and temperature are applied, the Sn or In will melt at low temperature which react with Cu or Au at their respective eutectic temperatures. As the reaction continues the thin solder joint eventually all converts to intermetallics which melt at much higher temperatures. The TLP bonding process has advantages over the solid-state diffusion bonding of Cu–Cu or Au–Au because of its lower process temperature and less sensitivity to surface topography and roughness. Figure 4.56 shows cross section of micro-joints at 15 μm pitch with Cu/Sn bump joined to ENIG. Cu–Sn–Cu bonding has becoming more popular with prototyping IC-stacks assembled using this technology [134, 135]. A 3-die stack using this technology is shown in Fig. 4.57 [134]. Figure 4.58 shows a 16 Gb memory module made with 8-die stack from Samsung [136].

4.9.2 Metal-to-Metal Solid-State Diffusion Bonding

Metal-to-metal solid-state diffusion bonding of Cu to Cu [135, 137, 138] and Au to Au [139] has been widely practiced for IC stacking. The advantages are the formation of a non-melting micro-joint during any of the sequential stacking processes, good heat transfer, and strong mechanical support. Cu–Cu bonding is achieved by thermo-compression, which requires temperature and force. Surface cleanliness prior to bonding is very critical. The effects of surface oxide, contamination, surface roughness and hardness all play a critical role in achieving successful bonding. An annealing step in N_2 or N_2–H_2 gas is needed to achieve higher bond

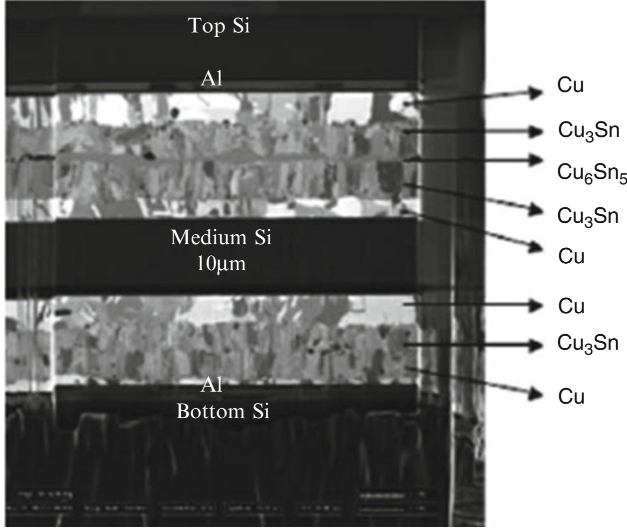


Fig. 4.57 Field Ion Beam cross section of a 3-die stack using Cu–Sn–cu technology [134]

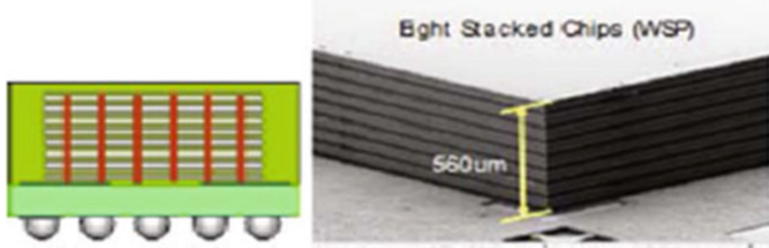


Fig. 4.58 Samsung's 16 Gb memory module made with an 8-die stack using Cu–Sn–Cu technology [136]

strength by allowing Cu interdiffusion and grain growth. Figure 4.59 shows the cross section of (a) a schematic drawing and (b) TEM images of Cu–Cu bond after bonding and (c) after annealing (no visible interface)

Au–Au can be bonded at lower temperature with surface planarization and plasma cleaning. Figure 4.60 shows cross-sectional images of Au–Au micro-bumps bonded at 20 μm pitch with good shear strength and yield.

The main driving force for future flip-chip technologies will continue to be dominated by miniaturization, 3D packaging, and 3D die-stacking. Flip-chip interconnect technology has expanded very rapidly. The conventional solder bumps will continue to dominate and moving toward finer pitch of 50 μm or less. While ultra-fine pitch bonding technologies, namely, direct bonding of Cu–Cu and Au–Au, along with thru-Si via (TSV) and indirect bonding through a solder intermediate layer will be

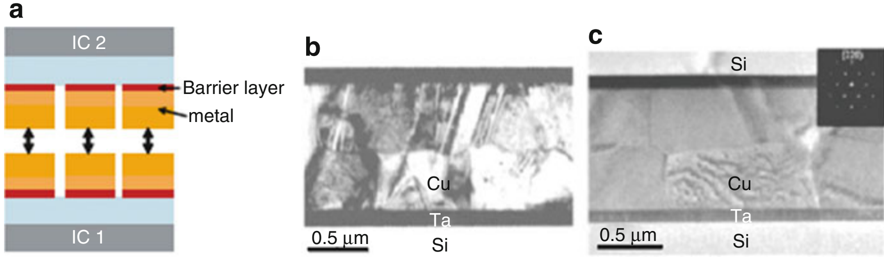


Fig. 4.59 (**a**) IC stacking using Cu–Cu bonding, (**b**) cross-sectional TEM image of Cu–Cu bond (**c**) after annealing (no visible interface) [140]

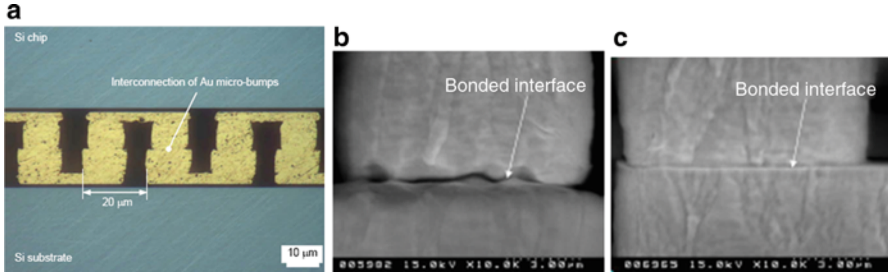


Fig. 4.60 Cross-sectional images of (**a**) Au–Au bumps at 20 μm pitch (**b**) before planarization, and (**c**) after planarization [139]

pursued at pitch of 20 μm or less to enable 3D packaging and integration. Significant progress has already been made in these areas. Major road blocks still exist. However, global alliance through the joint efforts between the equipment suppliers, device manufacturers, packaging houses, materials suppliers, and research institutions is in play to achieve the common goals of small form factor, high performance, and low cost.

4.10 Concluding Remark

The recent proliferation of flip-chip interconnects from high-performance microelectronics to low-end consumer electronics is largely owing to the development of several breakthrough technologies, such as high-density interconnect organic laminate (e.g., surface laminar circuitry), underfill encapsulation, direct-chip attachment using low-melt solder, low cost wafer bumping, and others. In this chapter, the evolution of enabling assembly technologies were discussed with emphasis on their key processes, such as wafer bumping, underfill encapsulation, substrate bumping, copper pillar fabrication, and others.

As the physical feature sizes of the on-chip logic and memory functions are continually shrinking according to the Moore's law, the flip-chip technology has steadily advanced to accommodate the interconnection requirements by decreasing bump size and pitch as well as increasing the number of interconnecting bumps.

Recently, two key technological transitions have imposed significant challenges on the further progress of flip-chip technology. The first transition is coming from the new Pb-free solder technology required by the EU environmental legislations. The second one is associated with the implementation of low-k or ultralow k dielectric materials in the BEOL structure in the advanced semiconductor devices. The application of Pb-free solders in flip-chip interconnects has raised several critical reliability issues, such as interfacial reactions, electromigration, drop impact resistance, or thermal fatigue. In this chapter, the fundamental issues of Pb-free solders for flip-chip applications were discussed in terms of the microstructure, solidification, physical/mechanical properties, and interfacial reactions. Subsequently, an in-depth discussion was presented on several key reliability issues including thermal fatigue, impact drop reliability, electromigration, and others.

The most challenging issue encountered in flip-chip joining is chip-to-package interaction (CPI), leading to interlayer dielectric cracking in the advanced semiconductor devices. This failure mode is becoming more serious when Pb-free solders are implemented together with the ultralow-k dielectric materials. Various solutions proposed or under development have been discussed in this chapter.

References

1. Miller LF (1969) Controlled collapse reflow chip joining. IBM J Res Dev 13:239
2. Goldmann LS (1969) Optimizing cycle fatigue life of controlled collapse chip joints. Proceedings of the 19th Electronics Components Conference, pp 404–423
3. Totta PA (1980) Flip chip solder terminals. Proceedings of the 21st Electronics Components Conference, p 89
4. Totta PA, Khadpe S, Koopman N, Reiley T, Sheaffer M (1997) Chip-to-package interactions. In: Tummala R, Rymaszewski E, Klopfenstein A (eds) Microelectronics packaging handbook, Part II, 2nd edn. Chapman & Hall, London, pp 129–283
5. Shih D-Y (1996) Effect of polyimide processing on multi-chip glass ceramic module fabrications. IEEE Trans Compon Packag Manuf Tech C 19(4):315
6. Tsukada Y, Mashimoto Y, Watanuki N (1993) A novel chip replacement method for encapsulated flip chip bonding. Proceedings of the 43rd Electronics Components and Technology Conference, pp 199–204
7. Tsukada Y, Tsuchida S, Mashimoto Y (1992) Surface laminar circuit packaging. Proceedings of the 42nd Electronics Components and Technology Conference, pp 22–27
8. Tsukada Y, Maeda Y, Yamanaka K (1993) A novel solution for MCM-L utilizing surface laminar circuit on flip chip attach technology. Proceedings of the 2nd International Conference and Exhibition on Multichip Modules, pp 252–259
9. Kang SK, Horkans J, Andricacos P, Crruthers R, Cotte J, Datta M, Gruber P, Harper J, Kwietniak K, Sambucetti C, Shi L, Brouillette G, Danovitch D (1999) Pb-free solder alloys for flip chip applications. Proceedings of 49th ECTC, IEEE, Piscataway, NJ, p 283

10. Rai A, Dotta Y, Tsukamoto H, Fujiwara T, Ishii H, Nukii T, Matsui H (1990) COB (Chip on Board) technology. ISHM Proceedings. pp 474–481
11. Rai A, Dotta Y, Nukii T, Ohnishi T (1992) Flip chip COB Technology on PWB. Proceedings of IMC, pp 144–149
12. Kang SK, Leonard D, Shih DY, Gignac L, Henderson DW, Cho S, Yu J (2006) Interfacial reactions of Sn-Ag-Cu solders modified by minor Zn alloying addition. J Electron Mater 35(3):479–485
13. Kang SK, Choi WK, Shih DY, Lauro P, Henderson DW, Gosselin T, Leonard DN (2002) Interfacial reactions, microstructure and mechanical properties of Pb-free solder joints in PBGA laminates. Proceedings of 52nd Electronic Components and Technology Conference, San Diego, CA, pp 147–153
14. Laurila T, Vuorinen V, Kivilahti JK (2005) Interfacial reactions between lead-free solders and common base materials. Mater Sci Eng 49:1–60
15. Wang T, Tung F, Foo L, Dutta V (2001) Studies on a novel flip-chip interconnect structure-pillar bump. Proceedings 51st Electronic Components and Technology Conference, pp 945
16. Kloeser J (2006) High-performance flip chip packages with Cu pillar bumping. Global SMT & Packaging, pp 28–31
17. Mallik D, Mahajan R, Wakharkar V (2008) Flip chip packaging for nanoscale silicon logic devices: challenges and opportunities. In: Morris JE (ed) Nanopackaging: nanotechnology and electronics packaging. Springer, New York, pp 491–516
18. Gruber P et al (2005) Low cost wafer bumping. IBM J Res Dev 49(4/5):621
19. Laine E, et al (2007) C4NP technology for lead free solder bumping. Proceedings 57th Electronic Components and Technology Conference, p 1320
20. Perfecto E et al (2008) C4NP technology: manufacturability, yields and reliability. Proceeding 58th Electronic Components and Technology Conference, p 1642
21. Shih D-Y, Dang B, Gruber P, Lu M, Kang S, Buchwalter S, Knickerbocker J, Perfecto E, Garant J, Knickerbocker S, Semkow K, Sundlof B, Busby J, Weisman R (2008) C4NP for Pb-free solder wafer bumping and 3D fine-pitch applications. Proceedings of International Conference on Electronic Packaging Technology, China
22. Busby J et al (2008) C4NP lead free solder bumping and 3D micro bumping. ASMC conference, pp 333–339
23. Dang B, shih D-Y, Buchwalter S, Tsang C, Patel C, Knickerbocker J, Gruber P, Knickerbocker S, Garant J, Semkow K, Ruhmer K, Hughlett E (2008) 50 μm Pitch Pb-free microbumps by C4NP technology. 58th Electronic Component and Technology Conference, pp 1505–1510, May 2008
24. Kasim R, Connor C, Hicks J, Jopling J, Litteken C (2000) Reliability for manufacturing on 45 nm logic technology with high-k + metal gate transistors and Pb-free packaging. International Reliability Physics Symposium, pp 350–354
25. Ebersberger B, Lee C (2008) Cu pillar bumps as a lead-free drop-in replacement for solder-bumped, flip-chip interconnects. Proceedings 59th Electronic Components and Technology Conference, pp 59–66
26. Yeoh A, Chang M, Pelto C, Huang T, Balakrishnan S, Leatherman G (2006) Copper die bumps (First Level Interconnect) and low-K dielectrics in 65 nm high volume manufacturing. 56th ECTC, pp 1611–1615
27. Nah JW, Suh JO, Tu KN, Yoon SW, Rao VS, Kripesh V, Hua F (2006) Electromigration in flip chip solder joints having a thick Cu column bump and a shallow solder interconnect. J Appl Phys 100:123513-1
28. Xu L, Han JK, Liang JJ, Tu KN, Lai YS (2008) Electromigration induced fraction of compound formation in SnAgCu flip-chip solder joints with copper column. Appl Phys Lett 92:262104-1
29. Lai YS, Chiu YT, Chen J (2008) Electromigration reliability and morphologies of Cu pillar flip-chip solder joints with Cu substrate pad metallization. J Elecron Mater 37(10):624

30. Longford A, James D (2006) Cu pillar bumps in Intel microprocessors. Semicon Europa-Advanced Packaging Conference, Germany, April 2006
31. Nah J-W, Paik KW, Hwang TK, Kim WH (2003) A study on coining processes of solder bumps on organic substrates. IEEE Trans Electron Packag Manuf 26(2):166–171
32. Elenius P, Leal J, Ney J, Stepniak D, Yeh S (1999) Recent advances in flip chip wafer bumping using solder paste technology. Proceedings of the 49th ECTC, pp 260–265
33. Li L, Wiegels S, Thompson P, Lee R (1998) Stencil printing process development for low cost flip chip interconnect. Proceedings of the 48th ECTC, pp 421–426
34. Kloeser J et al (1998) Low cost bumping by stencil printing: process qualification for 200 μm pitch. Proceedings of International Symposium on Microelectronics, pp 288–297
35. Photo courtesy of Dr. Jae-Woong Nah
36. Hashino E, Shimokawa K, Yamamoto Y, Tatsumi K (1998) Micro-ball wafer bumping for flip chip interconnection. Proceedings of the 51th ECTC, pp 421–426
37. Rinne GA (1997) Solder bumping methods for flip chip packaging. Proceedings of 47th Electronic Components & Technology Conference, pp 240–247
38. Tatsumi K, Yamamoto Y, Iwata K, Hashino E, Ishikawa S, Kohnox T (2005) An application of micro-ball wafer bumping to double ball bump for flip chip interconnection. Proceedings of 55th Electronic Components & Technology Conference, pp 855–860
39. Hazeyama I (2003) Micro-bump formation technology for flip chip LSIs using micro-solder-ball. NEC Res Dev 44(3):219–224
40. Inoue K et al (2003) Method of forming bumps. US Patent 6,213,386 (granted to Hitachi)
41. Nah J-W, Gruber PA, Lauro PA, Shih D-Y (2009) Injection molded solder—a new fine pitch substrate bumping method. Proceedings of the 59th Electronic Components & Technology Conference, pp 61–66
42. Directive 2002/95/EC of the European Parliament and of the Council of Jan 27 (2003) Official Journal of the European Union, 13.2.2003, L 37/19-23
43. Directive 2002/96/EC of the European Parliament and of the Council of Jan 27 (2003) Official Journal of the European Union, 13.2.2003, L 37/24-38
44. ROHS (Restriction of Hazardous Substances Directive). http://en.wikipedia.org/wiki/Restriction_of_Hazardous_Substances_Directive
45. Frear DR, Jang JW, Lin JK, Zhang C (2001) Pb-free solders for flip-chip interconnects. JOM 53(6):28–33
46. Kim D, Suh D, Millard T, Kim H, Kumar C, Zhu M, Xu Y (2007) Evaluation of high compliant low Ag solder alloys on OSP as a drop solution for the 2nd level Pb-free interconnection. Proceedings of 57th ECTC, pp 1614–1619
47. Uchida M, Ito H, Yabui K, Nishiuchi H, Togasaki T, Higuchi K, Ezawa H (2007) Low-stress interconnection for flip chip BGA employing lead-free solder bump. Proceedings of 57th ECTC, pp 885–891
48. Barnett JD, Bennion RB, Hall HT (1963) X-ray diffraction studies on tin at high pressure and high temperature. Science 141(3585):1041
49. Henderson DW, Woods JJ, Gosselin TA, Bartelo J, King DE, Korhonen TM, Korhonen MA, Lehman LP, Cotts EJ, Kang SK, Lauro PA, Shih DY, Goldsmith C, Puttlitz KJ (2004) The microstructure of Sn in near eutectic Sn-Ag-Cu alloy solder joints and its role in thermomechanical fatigue. J Mater Res 19(6):1608
50. Seo SK, Kang SK, Shih D-Y, Mo Lee H (2009) An investigation of microstructure and microhardness of Sn-Cu and Sn-Ag solders as a function of alloy composition and cooling rate. J Electron Mater 38(2):257–265
51. Sullivan MJ, Kilpatrick SJ (2004) Degradation phenomena. In: Puttlitz KJ, Stalter KA (eds) Handbook of lead-free solder technology for microelectronic assemblies. Marcel Dekker, New York, p 920
52. Mason WP, Bommel HE (1956) Ultrasonic attenuation at low temperatures for metals in the normal and superconducting states. J Acoust Soc Amer 28:930

53. Everhart JL (1967) Properties of tin and tin alloys. In: Lyman T (ed) Metals handbook. American Society of Metals, Metals Park, p 1142
54. Rayne JA, Chandrasekhar BS (1960) Elastic constants of β tin from 4.2 K to 300 K. Phys Rev 120:1658
55. Dyson BF (1966) Diffusion of gold and silver in tin single crystals. J Appl Phys 37:2375
56. Dyson BF, Anthony TR, Turnbull D (1967) Interstitial diffusion of copper in tin. J Appl Phys 37:3408
57. Yeh DC, Huntington HB (1984) Extreme fast-diffusion system: Nickel in single-crystal tin. Phys Rev Lett 53(15):1469
58. Huang FH, Huntington HB (1974) Diffusion of Sb^{124}, Cd^{109}, Sn^{113}, and Zn^{65} in tin. Phys Rev B 9(4):1479
59. Bieler TR, Jiang H, Lehman LP, Kirkpatrick, Cotts EJ (2006) Influence of Sn grain size and orientation on the thermomechanical response and reliability of Pb-free solder joints. 46th ECTC Proceedings, 1462,
60. Lu M, Shih D-Y, Lauro P, Goldsmith C, Henderson DW (2008) Effect of Sn grain orientation on electromigration degradation mechanism in high Sn-based Pb-free solders. Appl Phys Lett 92:211909
61. Kang SK, Lauro P, Shih DY, Henderson DW, Puttlitz KJ (2005) The microstructure, solidification, mechanical properties, and thermal fatigue behavior of lead(Pb)-free solders and solder joints used in microelectronic applications. IBM J Res Dev 49(4/5):606–620
62. Kang SK, Choi WK, Shih DY, Henderson DW, Gosselin T, Sarkhel A, Goldsmith C, Puttlitz K (2003) Formation of Ag_3Sn plates in Sn-Ag-Cu alloys and optimization of their alloy composition. Proceedings of 53rd ECTC, New Orleans, LA, May 2003, p 64–70
63. Kang SK, Choi WK, Shih DY, Henderson DW, Gosselin T, Sarkhel A, Goldsmith C, Puttlitz KJ (2003) Study of Ag_3Sn plate formation in the solidification of near ternary eutectic Sn-Ag-Cu alloys. J Mater 55(6):61–65
64. Kinyanjui R, Lehman LP, Zavalij L, Cotts E (2006) Effect of sample size on the solidification of temperature and microstructure of SnAgCu near eutectic alloys. J Mater Res 20(11):2914–2918
65. Kang SK, Cho MG, Lauro P, Shih D-Y (2007) Study of the undercooling of Pb-free, flip-chip solder bumps and in-situ observation of solidification process. J Mater Res 22(3):557
66. Kang SK, Cho MG, Lauro P, Shih D-Y (2007) Critical factors affecting the undercooling of Pb-free, flip-chip solder bumps and in-situ observation of solidification process. Proceedings 2007 ECTC, Reno, NV, May 2007, p 1597
67. Andricacos PC, Datta M, Horkans WJ, Kang SK, Kwietniak KT (1999) Barrier layers for electroplated SnPb eutectic solder joints. US Patent 5,937,320, 10 Aug 1999
68. Jang SY, Wolf J, Ehrmann O, Gloor H, Schreiber T, Reichl H, Park KW (2001) CrCu based UBM study with electroplated Pb/63Sn solder bumps-interfacial reaction and bump shear strength. IEEE Trans Compon Packag Technol 26(1):245
69. Kang SK, Rai RS, Purushothaman S (1996) Interfacial reactions during soldering with lead-tin eutectic and Pb-free, Tin-rich solders. J Electron Mater 25(7):1113–1120
70. Kang SK, Shih DY, Fogel K, Lauro P, Yim MJ, Advocate G, Griffin M, Goldsmith C, Henderson DW, Gosselin T, King D, Konrad J, Sarkhel A, Puttlitz KJ (2002) Interfacial reaction studies on lead (Pb)-free solder alloys. IEEE Trans Electron Packag Manuf 25 (3):155
71. Moon K-W, Boettinger WJ, Kattner UR, Biancaniello FS, Handwerker CA (2000) Experimental and thermodynamic assessment of Sn-Ag-Cu solder alloys. J Electron Mater 29(10):1122–1136
72. Ghosh G (1999) Thermodynamic modeling of the nickel-lead-tin system. Metall Mater Trans A 30A:1481–1494
73. Andricacos PC, Datta M, Horkans WJ, Kang SK, Kwietniak KT, Mathad GS, Purushothaman S, Shi L, Tong H-M (2001) Flip-chip interconnections using lead-free solders. US Patent 6,224,690 B1, 1 May 2001

74. Cheng Y-T, Chiras SR, Henderson DW, Kang SK, Kilpatrick SJ, Nye HA, Sambucetti CJ, Shih D-Y (2007) Ball limiting metallurgy, interconnection structure including the same, and method of forming an interconnection structure. US Patent 7,273,803, 25 Sept 2007
75. Fogel KE, Ghosal B, Kang SK, Kilpatrick S, Lauro PA, Nye HA, Shih D-Y, Zupanski-Nielsen DS (2008) Interconnections for flip-chip using lead-free solders and having reaction barrier layers. US Patent 7,410,833 B2, 12 Aug 2008
76. Guo J, Zhang L, Xian A, Shang JK (2007) Solderability of electrodeposited Fe-Ni alloys with eutectic SnAgCu solders. J Mater Sci Technol 23(6):811
77. Dietz G, Schneider HD (1990) Decomposition of crystalline ferromagnetic particles precipitated in amorphous paramagnetic Ni-P. J Phys Condens Matter 2:2169
78. Jang JW, Kim PG, Tu KN, Frear DR, Thompson P (1999) Solder reaction-assisted crystallization of electroless Ni–P under bump metallization in low cost flip chip technology. J Appl Phys 85:8456
79. Sohn YC, Yu J, Kang SK, Choi WK, Shih DY (2003) Study of the reaction mechanism between electroless Ni-P and Sn and its effect on the crystallization of Ni-P. J Mater Res 18(1):4–7
80. Sohn YC, Yu J, Kang SK, Shih DY, Lee TY (2004) Spalling of intermetallic compounds during the reaction between lead-free solders and electroless Ni-P metallization. J Mater Res 19(8):2428–2436
81. Sohn YC, Yu J, Kang SK, Shih DY, Lee TY (2005) Effect of intermetallics spalling on the mechanical behavior of electroless Ni(P)/Pb-free solder interconnection. Proceedings of 55th ECTC, pp 83–88
82. Choi WK, Kang SK, Sohn YC, Shih DY (2003) Study of IMC morphologies and phase characteristics affected by the reactions of Ni and Cu metallurgies with Pb-free solder joints. Proceedings of 53rd Elec. Comp. & Tech Conference, New Orleans, LA, CA, pp 1190–1196
83. Chen WT, Ho CE, Kao CR (2002) Effects of Cu concentration on the interfacial reactions between Ni and Sn-Cu solders. J Mater Res 17(2):263
84. Tsai JY, Kao CR (2002) The effect of Ni on the interfacial reaction between Sn-Ag solder and Cu metallization. Proceedings of 4th International Symposium on Electronic Materials and Packaging, Kaohsiung, Taiwan, Dec 2002, p 271
85. Choi WK, Kim JH, Jeong SW, Lee HM (2002) Interfacial microstructure and joint strength of Sn-3.5Ag-X (X = Cu, In, Ni) solder joint. J Mater Res 17(1):43
86. Yu DQ, Zhao J, Wang L (2004) Improvement on the microstructure stability, mechanical and wetting properties of Sn-Ag-Cu lead-free solder with the addition of rare earth elements. J Alloys Compd 376(1/2):170–175
87. Kim KS, Huh SH, Suganuma K (2003) Effects of fourth alloying additive on microstructure and tensile properties of Sn-Ag-Cu and joints with Cu. Microelectron Reliab 43:259–267
88. Anderson IE, Harringa JL (2006) Suppression of void coalescence in thermal aging of tin-silver-copper-X solder joints. J Electron Mater 35(1):94–106
89. Rizvi MJ, Chan YC, Bailey C, Lu H, Islam MN (2006) Effects of adding 1 wt% Bi into the Sn-2.8Ag-0.5Cu solder alloy on the intermetallic formation with Cu substrate during soldering and isothermal aging. J Alloys Compd 407:208–214
90. Kang SK, Shih DY, Leonard D, Henderson DW, Gosselin T, Cho SI, Yu J, Choi WK (2004) Controlling Ag3Sn plate formation in near-ternary-eutectic Sn-Ag-Cu solder by minor Zn alloying. JOM 56:34
91. de Sousa I, Henderson DW, Patry L, Kang SK, Shih DY (2006) The influence of low level doping on the thermal evolution of SAC alloy solder joints with Cu pad structures. Proceedings of 56th Electronic Components and Technology Conference, San Diego, CA, p 1454
92. Cho MG, Kang SK, Shih D-Y, Lee HM (2007) Effects of minor addition of Zn on interfacial reactions of Sn-Ag-Cu and Sn-Cu solders with various Cu substrates during thermal aging. J Electron Mater 36(11):1501–1509

93. Jee YK, Yu J, Ko YH (2007) Effects of Zn addition on the drop reliability of Sn-3.5Ag-xZn/Ni(P) solder joints. J Mater Res 22(10):2776–2784
94. Lu M, Lauro P, Shih DY, Kang SK, Chae SH, Seo SK (2009) The effects of Cu, Ag compositions and Zn doping on the electromigration performance of Pb-free solders. Proceedings 2009 ECTC, San Diego, 26–29 May 2009, p 922
95. Cho MG, Seo S, Lee HM (2009) Wettability and interfacial reactions of Sn-based Pb-free solders with Cu-xZn alloy under bump metallurgies, J Alloys Compd 474:510
96. Kang SK, Cho MG, Shih D-Y, Seo SK, Lee HM (2008) Controlling the interfacial reactions in Pb-free interconnections by adding minor alloying elements to Sn-rich solders. Proceedings of 58th ECTC, p 478
97. Sawamura T, Komiya H, Inazawa T, Nakagawa F (2004) Pb-free soldering alloy. US Patent 6,692,691B2, 17 Feb 2004
98. Liu W, Bachorik P, Lee NC (2008) The superior drop test performance of SAC-Ti solders and its mechanism. Proceedings of 58th ECTC, p 452
99. Li GY, Chen BL, Tey JN (2004) Reaction of Sn-3.5Ag-0.7Cu-xSb solder with Cu metallization during reflow soldering. IEEE Trans Elecron Packag Manuf 27(1):1521
100. Lee HT, Lin HS, Lee CS, Chen PW (2005) Reliability of Sn-Ag-Sb lead-free solder joints. Mater Sci Eng A 407:36–44
101. Lau JH (2000) Low cost flip chip technologies. McGraw-Hill, New York, 183
102. Zhang C, Lin JK, Li L (2001) Thermal fatigue properties of Pb-free solders on Cu and NiP under bump metallurgies. Proceedings of 51st ECTC, p 463
103. Terashima S, Kariya Y, Hosoi T, Tanaka M (2003) Effect of silver content on thermal fatigue life of Sn-xAg-0.5Cu flip-chip interconnects. J Elecron Mater 32(12):1527
104. Kang SK, Lauro P, Shih D-Y, Henderson DW, Bartelo J, Gosselin T, Cain SR, Goldsmith C, Puttlitz K, Hwang TK, Choi WK (2004) The microstructure, thermal fatigue, and failure analysis of near-ternary eutectic Sn-Ag-Cu solder joints. Mater Trans (The Japan Inst of Metals) 45(3):695–702
105. Kang SK, Lauro P, Shih D-Y, Henderson DW, Gosselin T, Bartelo J, Cain SR, Goldsmith C, Puttlitz K, Hwang TK (2004) Evaluation of thermal fatigue life and failure mechanism of Sn-Ag-Cu solder joints with reduced Ag contents. Proceedings of 54th ECTC, p 661
106. Arfaei B, Xing Y, Woods J, Wolcott N, Tumne P, Borgesen P, Cotts E (2008) The effect of Sn grain number and orientation on the shear fatigue life of SnAgCu solder joints. Proceedings of 58th ECTC, p 459
107. Quinones H, Babiarz A (2001) Flip chip, CSP, and WLP technologies: a reliability perspective. IMAPS Nordic, p 1, Sep 2001
108. Jang SY, Hong SM, Park MY, Kwok DO, Jeong JW, Roh SH, Moon YJ (2004) FCOB reliability study for mobile applications. Proceedings of 2004 ECTC, p 62
109. Syed A, Kim TS, Cha SW, Scanlon J, Ryu CG (2007) Effect of Pb-free alloy composition on drop/impact reliability of 0.4, 0.5, 0.8 mm pitch chip scale packages with NiAu pad finish. Proceedings of 2007 ECTC, p 951
110. Farris A, Pan J, Liddicoat A, Toleno BJ, Maslyk D, Shanggun D, Bath J, Willie D, Geiger DA (2008) Drop test reliability of lead-free chip scale packages. Proceedings of 2008 ECTC, p 1173
111. Chiu TC, Zeng K, Stierman R, Edwards D, Ano K (2005) Effect of thermal aging on board level drop reliability for Pb-free BGA packages. Proceedings of 2004 ECTC, p 1256
112. Date M, Shoji T, Fujiyoshi M, Sato K, Tu KN (2004) Impact reliability of solder joints. Proceedings of 2004 ECTC, p 668
113. Borgesen P, Yin L, Kondos R, Henderson DW, Servis G, Therriault J, Wang J, Srihari K (2007) Sporadic degradation in board level drop reliability—those aren't all Kirkendall voids!. Proceedings of 2007 ECTC, p 136
114. Mei Z, Ahmad M, Hu M, Ramakrishna G (2005) Kirkendall voids at Cu/solder interface and their effects on solder joint reliability. Proceedings of 55th Electronic Component and Technology Conference, pp 1256–1262

115. Amagai M, Toyoda Y, Ohnishi T, Akita S (2004) High drop test reliability: lead-free solders. Proceedings of 2004 ECTC, p 1304
116. Kim H, Zhang M, Kumar CM, Suh D, Liu P, Kim D, Xie M, Wang Z (2007) Improved drop reliability performance with lead free solders of low Ag content and their failure modes. Proceedings of 2007 ECTC, p 962
117. Sylvestre J, Blander A, Oberson V, Perfecto E, Srivastava K (2008) The impact of process parameters on the fracture of device structures. Proceedings of the 58th Electronic Components and Technology Conference, pp 82–88
118. Zeng K, Tu KN (2002) Six cases of reliability study of Pb-free solder joints in electronic packaging technology. Mater Sci Eng R Rep 38:55–105
119. Lu M, Shih D-Y, Polastre R, Goldsmith C, Henderson DW, Zhang H, Cho MG (2008) Comparison of electromigration performance for Pb-free solders and surface finishes with Ni UBM. Proceedings of 58th ECTC, p 360
120. Seo S-K, Kang SK, Shih D-Y, Lee HM (2009) The evolution of microstructure microhardness of Sn-Ag and Sn-Cu solders during high temperature aging. Microelectron Reliab 49:288
121. Blech IA (1976) Electromigration in thin aluminum films on titanium nitride. J Appl Phys 47:1203
122. Lu M, Shih D-Y, Lauro P, Goldsmith C (2009) Blech effect in Pb-free flip-chip solder joint. J Appl Phys Lett 94:011912
123. Lu M, Shih D-Y, Kang SK, Goldsmith C, Flaitz P (2009) Effect of Zn doping on SnAg solder microstructure and electromigration stability. J Appl Phys 106:053509
124. Tu KN (2007) Electromigration in flip chip solder joint, Chapter 9. In: Tu KN (ed) Solder joint technology. Springer, New York, pp 245–287
125. Chen C, Liang SW (2007) Electromigration issues in lead-free solder joints. In: Subramanian KN (ed) Lead-free electronic solders. Springer, New York, pp 259–268
126. Lee S, Guo X, Ong CK (2005) Electromigration effect on Cu-pillar (Sn) bumps. Electronics Packaging Technology Conference, pp 135–139
127. Kariya Y et al (2000) Tin pest in lead-free solders. Soldering Surf Mount Technol 13:39
128. Henderson DW, Gosselin TA, Goldsmith CC, Puttlitz KJ, Kang SK, Shih DY, Choi WK (2004) Lead-free tin-silver-copper alloy solder composition. US Patent 6805974, 19 Oct 2004
129. Knickerbocker JU, Andry PS, Dang B, Horton RR, Interrante MJ, Patel CS, Polastre RJ, Sakuma K, Sirdeshmukh R, Sprogis EJ, Sri-Jayantha SM, Stephens AM, Topol AW, Tsang CK, Webb BC, Wright SL (2008) Three-dimensional silicon integration. IBM J Res Dev 52 (6):553–569
130. Meindl JD, Davis JA, Zarkesh-Ha P, Patel CS, Martin KP, Kohl PA (2002) Interconnect opportunities for gigascale integration. IBM J Res Dev 46:245–263
131. Dang B, Wright SL, Andry PS, Sprogis EJ, Tsang CK, Interrante MJ, Webb BC, Polastre RJ, Horton RR, Patel CS, Sharma A, Zheng J, Sakuma K, Knickerbocker JU (2008) 3D chip stacking with C4 technology. IBM J Res Dev 52(6):599–609
132. Dang B, Wright S, Andry P, Sprogis E, Ketkar S, Tsang C, Polastre R, Knickerbocker J (2009) 3D chip stack with integrated decoupling capacitors. 59th Electronic Component and Technology Conference, pp 1–5, May 2009
133. Yu A, Lau J, Ho S, Kumar A, Hnin W, Yu D, Jong M, Kripesh V, Pinjala D, Kwong D (2009) Study of 15 μm pitch solder microbumps for 3D IC integration. 59th Electronic Component and Technology Conference, pp 6–10, May 2009
134. Krupp A et al (2005) 3D integration with ICV-solid technology. Conference on 3D Architecture for Semiconductor and Packaging, Tempe, AZ, June 2005
135. Ramm P (2004) Vertical integration technologies. Workshop on 3D Integration of Semiconductor Devices, San Diego, Oct 2004
136. Lee K (2006) The next generation package technology for higher performance and smaller systems. Conference on 3D Architecture for Semiconductor and Packaging, Burlingame, CA, Oct 2006

137. Schaper L (2003) 53rd Electronic Component and Technology Conference, pp 631–633, May 2003
138. Schaper L, Burkett S, Spiesshoefer S, Vangara G, Rahman Z, Polamreddy S (2005) Architectural implications and process development of 3-D VLSI Z -axis interconnects using through Silicon Vias. Trans IEEE Adv Packag 28(3):356, Aug 2005
139. Wang Y, Suga T (2008) 20-μm pitch Au micro-bump interconnection at room temperature in ambient air. 58th Electronic Component and Technology Conference, pp 944–949, May 2008
140. Chanchani R (2009) 3D Integration technology—an overview. In: Lu D, Wong CP (eds) Materials for advanced packaging. Springer, New York, pp 1–50

Chapter 5
Flip Chip Underfill: Materials, Process, and Reliability

Zhuqing Zhang, Shijian Luo, and C.P. Wong

Abstract In order to enhance the reliability of a flip chip on organic board package, underfill is usually used to redistribute the thermomechanical stress created by the Coefficient of Thermal Expansion (CTE) mismatch between the silicon chip and organic substrate. However, the conventional underfill relies on the capillary flow of the underfill resin and has many disadvantages. In order to overcome these disadvantages, many variations have been invented to improve the flip chip underfill process. This paper reviews the recent advances in the material design, process development, and reliability issues of flip chip underfill, especially in no-flow underfill, molded underfill, and wafer-level underfill. The relationship between the materials, process, and reliability in these packages is discussed.

5.1 Introduction

The brain of the modern electronics is the integrated circuit (IC) on semiconductor chip. In order for the brain to control the system, interconnects need to be established between the IC chip and other electronic parts, power and ground, and inputs and

Z. Zhang (✉)
Hewlett-Packard Company, 1000 NE Circle Blvd, Corvallis, OR 97330, USA
e-mail: zhuqing.zhang@hp.com

S. Luo
Micron Technology, Inc., 8000 S. Federal Way, P O Box 6, Boise, ID 83707, USA
e-mail: sluo@micron.com

C.P. Wong
Dean of Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China

on leave from:
Georgia Institute of Technology, School of Materials Science & Engineering,
771 Ferst Drive, Atlanta, GA 30332-0245, USA
e-mail: cp.wong@mse.gatech.edu

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_5, © Springer Science+Business Media New York 2013

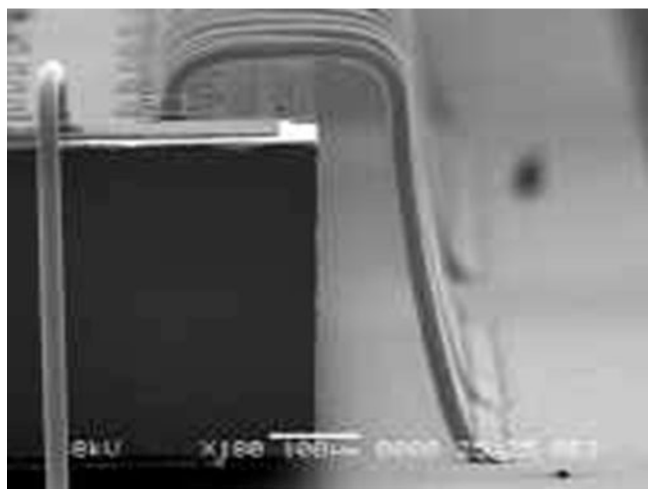


Fig. 5.1 First level interconnect using wire-bonding

outputs. The first-level interconnect usually connects the chip to a package made of either plastics and or ceramics, which in turn is assembled onto a printed circuit board (PCB). Three main interconnect techniques are used: wire-bonding, tape automated bonding (TAB), and flip chip. In a wire-bonded package, the chip is adhered to a carrier substrate using die-attach adhesive with the active IC facing up. A gold or aluminum wire is bonded between each pad on the chip and the corresponding bonding surface on the carrier as shown in Fig. 5.1. The chip and the wire interconnections are usually protected by encapsulation. TAB, on the other hand, uses a prefabricated lead frame carrier with copper or alloy 42 leads adapted to the IC pads. The copper is usually gold-plated and alloy 42 is usually tin–lead plated to provide a finish for bonding to the IC chip pads. The chip is attached onto the carrier and either thermosonic/thermocompression bonding or Au/Sn bonding is used to establish the interconnect. Both wire-bonding and TAB interconnects are limited to peripheral arrangement and therefore low input/output (I/O) counts. Flip chip, however, can utilize the entire semiconductor area for interconnects. In a flip chip package, the active side of an IC chip is faced down towards and mounted onto a substrate [1]. Interconnects, in the form of solder bumps, stud bumps, or adhesive bumps, are built on the active surface of the chip, and are joined to the substrate pads, in either a melting operation or adhesive joining or thermosonic or thermocompression process. Figure 5.2 shows an example of solder bumped chip surface for flip chip interconnect. Since flip chip was first developed over 40 years ago, many variations of the flip chip design have been developed, among which, the controlled collapsed chip connection (also known as C4) invented by IBM in 1960s is the most important form of flip chip [2]. Compared with conventional packaging using wire-bonding technology, flip chip offers many advantages such as high I/O density, short interconnects, self-alignment, better heat dissipation through the back of the die, smaller footprint, lower profile, and high throughput, etc. The outstanding merits of flip chip have made it one of the

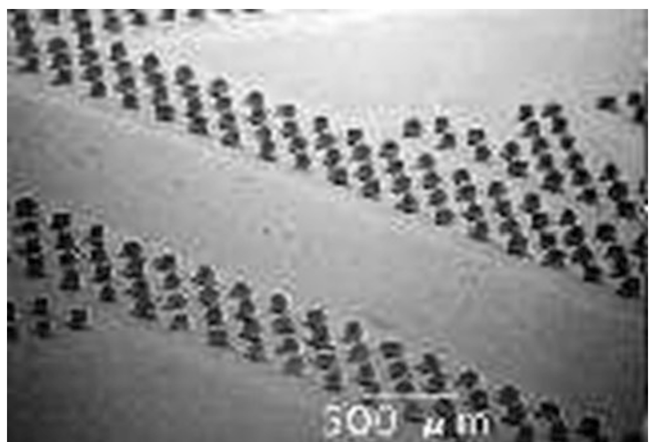


Fig. 5.2 Area array solder bumps for flip chip interconnect

most attracting techniques in modern electronic packaging, including MCM (multichip module), high frequency communications, high performance computers, portable electronics, and fiber optical assemblies.

Until the late 1980s, flip chips were mounted onto silicon or ceramic substrates. Low cost organic substrates could not be used due to the concern of the thermal–mechanical fatigue life of the C4 solder joints. This thermal-mechanical issue mainly arises from the CTE mismatch between the semiconductor chip (typically Si, 2.5 ppm/°C) and the substrate (4–10 ppm/°C for ceramics and 16–24 ppm/°C for organic FR4 board). As the distance from the neutral point (DNP) increases, the shear stress at the solder joints increases accordingly. So with the increase in the chip size, the thermal–mechanical reliability becomes a critical issue. Organic substrates have advantages over ceramic substrates because of their low cost and low dielectric constant. But the high CTE differences between the organic substrates and the silicon chip exert great thermal stress on the solder joint during temperature cycling.

In 1987, Hitachi first demonstrated the improvement of solder fatigue life with the use of filled resin to match solder CTE [3]. This filled resin, later called "underfill," was one of the most innovative developments to enable the use of low-cost organic substrate in flip chip packages. Underfill is a liquid encapsulate, usually epoxy resins heavily filled with fused silica (SiO_2) particles, that is applied between the chip and the substrate after flip chip interconnection. Upon curing, the hardened underfill exhibits high modulus, low CTE matching that of the solder joint, low moisture absorption and good adhesion towards the chip and the substrate. Thermal stresses on the solder joints are redistributed among the chip, underfill, substrate and all the solder joints, instead of concentrating on the peripheral solder joints. It has been demonstrated that the application of underfill can reduce the all-important solder strain level to 0.10–0.25 of the strain in joints, which are not encapsulated [4, 5]. Therefore, underfill can increase the solder joint fatigue life by 10–100 times. In addition, it provides an environmental protection to the IC chip and solder joints.

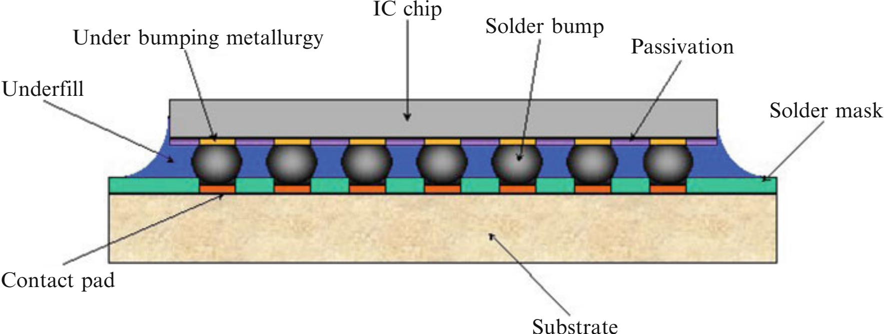


Fig. 5.3 Generic configuration of C4 with underfill

Underfill becomes the practical solution to extending the application of flip chip technology from ceramics to organic substrates, and from high-end to cost sensitive products. Today, flip chip is being extensively studied and used by almost all major electronic companies around world including Intel, AMD, Hitachi, IBM, Delphi, Motorola, Casio, SAE, Micron, FreeScale, etc.

5.2 Conventional Underfill Materials and Process

The generic schematic of a flip chip package is shown in Fig. 5.3. Conventional underfill is applied after the flip chip interconnects are formed. The resin flows into the gap between the chip and the substrate by a capillary force. Therefore, it is also called "capillary underfill." A typical capillary underfill is a mixture of liquid organic resin binder and inorganic fillers. The organic binders are often epoxy resin mix, although cyanate ester or other resin has been used for underfill application as well. Figure 5.4 shows the chemical structure of some commonly used epoxy resins. In additional to epoxy resin, a hardener is often used to form cross-linking structure upon curing. Sometimes a latent catalyst is incorporated to achieve long pot life and fast curing. Inorganic fillers typically used in underfill formulation are micron-sized silica. The silica fillers are incorporated into the resin binder to enhance the material properties of cured underfill such as low CTE, high modulus, and low moisture uptake, etc. Other agents that can be found in an underfill formulation include adhesion promoters, toughening agents, surfactants, and dispersing agents, etc. These chemicals are incorporated to help the resin mixing and enhance the cured underfill properties.

Figure 5.5 shows the process steps of flip chip with conventional underfill. Separate flux dispensing and cleaning steps are required before and after the assembling of the chip, respectively. After the chip is assembled onto the substrate, the underfill is usually needle-dispensed and is dragged into the gap between the

Bis A epoxy resin

Bis F epoxy resin

N,N-diglycidyl-4-glycidyloxylaniline

Naphthalene type epoxy resin

Fig. 5.4 Typical epoxy resin structures used in underfill

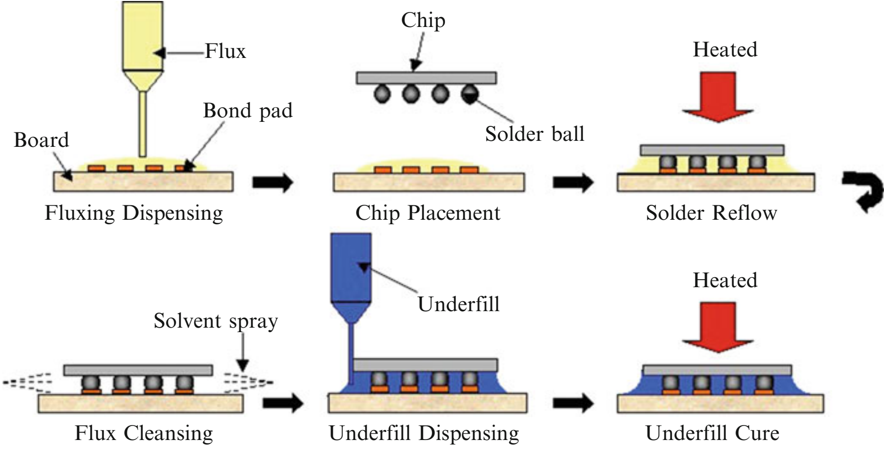


Fig. 5.5 Flip chip process using conventional underfill

chip and the substrate by a capillary force. Then a heating step is needed to cure the underfill resin to form a permanent composite.

The flow of the capillary underfill has been extensively studied since it is considered to be one of the bottlenecks for the flip chip process. The capillary flow is usually slow and can be incomplete, resulting in voids in the packages and also non-homogeneity in the resin/filler system. The filling problem becomes even more serious as the chip size increases. The flow modeling of flip chip underfill is often approximated as viscous flow of the underfill adhesive between two parallel plates. One can use the Hele-Shaw model to simulate the underfill flow with the above approximation. The time required to fill a chip of length L can be calculated as [6]:

$$t_{\text{fill}} = \frac{3\eta L^2}{\sigma h \cos\theta} \tag{5.1}$$

where η is the underfill viscosity; σ is the coefficient of the surface tension; θ is the contact angle; and h is the gap distance. It is easily seen that a larger chip with a smaller gap distance would require longer time to fill.

The above approximation does not take the existence of solder bumps into account. It is shown that the approximation breaks down when the spacing between bumps is comparable to the gap height [7]. Therefore, this model cannot apply to high density area array flip chip applications. Using transparent quartz dies assembled onto different substrates, Nguyen et al. observed the flow of commercial underfills and used a 3D PLICE-CAD to model the underfill flow front [8]. A comparison between the peripheral and area array chips showed that the bumps enhanced the flatness of the flow front by providing periodic wetting sites. A racing effect along the edges was observed. Voids can be formed at the merging of flow fronts. The merging of the fronts also produced streaks, which are zones of no- or slow-moving fluids, leading to higher potential for filler settling.

Recent development in underfill flow models has also considered the effect of the contact angle on solder and bump geometry. A study by Young and Yang used a modified Hele-Shaw model considering the flow resistance in both the thickness direction between the chip and substrate, and the plane direction between solder bumps [9]. It was found that the capillary force parameter would approach a constant value at very large pitch for the same gap height. As the bump pitch reduces, the capillary force will increase to a maximum as a result of underfill wetting on the solder given the contact angle on the solder is small, and then quickly drops to zero as the pitch approaches the bump diameter. Their study also showed that a hexagonal bump arrangement is more efficient to enhance the capillary force at critical bump pitch.

5.3 Characterizations of Underfill Materials

5.3.1 Differential Scanning Calorimeter Measured Curing Kinetics

In a typical differential scanning calorimeter (DSC) curing experiment, a small amount of underfill sample (~10 mg) is placed into a hermetic DSC sample pan and heated in the DSC cell from room temperature to an elevated temperature such as 300 °C at a certain heating rate. N_2 purging can be applied. When the sample goes through a chemical reaction that is exothermic in nature, the heat produced can be detected and recorded by DSC. The heat flux versus temperature diagram constructs the curing profile of the sample.

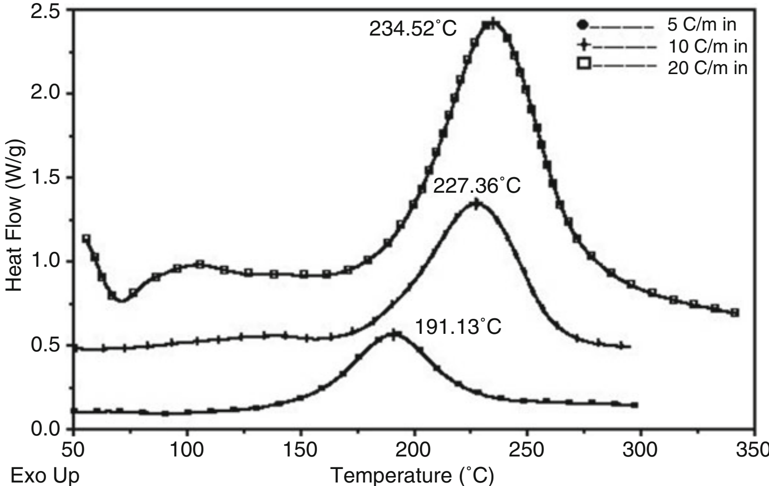


Fig. 5.6 Curing profile dependent on heating rate

The curing profile is sensitively dependent on the heating rate. Figure 5.6 shows the curing profiles of a particular underfill material at three different heating rates. Higher heating rate would push the curing reaction to a higher temperature range. A heating rate of 5 °C/min is commonly used in curing study. This heating rate can provide an accurate measurement of the chemical reactions that take place. The latency in curing of different underfill formulations can be compared according to the onset temperature and the peak temperature of the curing profile at the same heating rate. However, the curing behavior of an underfill formulation is much dependent on the curing kinetics. So even two formulations might have the same peak curing temperature, they could behave differently during the process. As a result, the curing profile obtained by a constant heating rate DSC experiment can only be a reference for judging the latency of curing.

To better understand the reaction kinetics of underfill curing process, isothermal DSC experiment can be used. In an isothermal DSC experiment, the sample is kept at a certain temperature for a sufficient amount of time and the reaction heat is recorded as a function of time. A typical DSC isothermal curing diagram and its analysis using autocatalytic model is as shown in Fig. 5.7.

The choice of DSC isothermal temperature is very important. Since the equipment takes certain amount of time to reach equilibrium, typically around 30–60 s, there is a period of initial instability in the isothermal diagram as illustrated in Fig. 5.7. If the reaction happens too quickly at this temperature, the initial heat flow cannot be analyzed. On the other hand, if the isothermal temperature is too low, the reaction takes a long time to finish and the heat flow is so low that DSC is not sensitive enough to recognize the signal. Usually a dynamic DSC curing at 5 °C/min is run before the isothermal experiment. The isothermal temperatures are chosen within ±20 °C of the onset curing temperature.

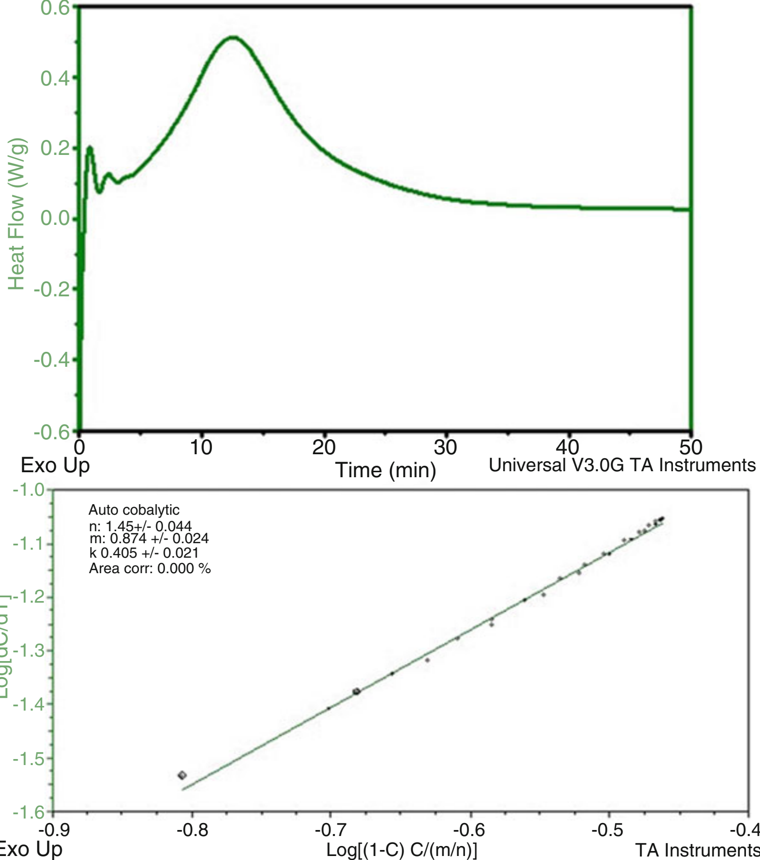


Fig. 5.7 DSC isothermal curing diagram (*top*) and kinetic analysis (*bottom*)

Two common kinetic models are used, nth order reaction model and autocatalytic model as shown in (5.2) (nth order) and (5.3) (autocatalytic).

$$\frac{dC}{dt} = k(1 - C)^n \tag{5.2}$$

$$\frac{dC}{dt} = k(1 - C)^n C^m \tag{5.3}$$

where C is the concentration of the product; k is the rate constant; n and m are reaction orders.

The rate constant of a reaction is related to temperature according to Arrhenius Equation:

$$k = K \exp\left(-\frac{E_a}{RT}\right) \tag{5.4}$$

where K is the frequency factor; E_a is the activation energy; R is the gas constant; and T is the absolute temperature.

5.3.2 *DSC Measured* T_g

Since most cross-linked polymers are amorphous, the glass transition temperature (T_g) is an important transition temperature of the material. The glass transition temperature is the temperature below which free rotations of the polymer chain cease because of the intramolecular energy barrier [10]. The physical properties of the polymer changed dramatically at the glass transition temperature. Below T_g, the polymer is in a glassy state, when the CTE of the polymer is lower and the modulus is higher than that of the polymer in a rubbery state beyond T_g. The T_g of an amorphous material can be measured by a number of characterization methods including differential scanning calorimeter (DSC), thermal-mechanical analyzer (TMA), and dynamic mechanical analyzer (DMA), etc. However, the glass transition temperatures measured using different methods are usually different for the same material. The same method shall be used for comparison of all the samples.

Traditionally, the temperature of the DSC furnace is raised or lowered in a linear fashion. If a heat-related change takes place in the sample, it either evolves or absorbs heat and causes the difference of sample temperature and reference temperature. The temperature between sample and reference from such a heat change is directly related to the differential heat flow by (5.5):

$$\frac{dQ}{dt} = \frac{\Delta T}{R_D} \tag{5.5}$$

where dQ/dt is the heat flow; ΔT is the temperature difference between the sample and the reference; and R_D is the thermal resistance of constantan disc.

In a modulated DSC, a sinusoidal temperature oscillation (modulation) is overlaid on the conventional linear temperature ramp. The DSC heat flow is then decomposed into two parts as shown in (5.6):

$$\frac{dQ}{dt} = C_p \frac{dT}{dt} + f(t, T) \tag{5.6}$$

where dQ/dt is the heat flow; dT/dt is the heating rate; C_p is the sample heat capacity; $f(t,T)$ is a function of time and temperature which govern the kinetic response of any physical or chemical transition observed in DSC.

Equation 5.6 shows that the total DSC heat flow in a modulated DSC experiment consists of two components: one which is heating rate dependent ($C_p(dT/dt)$), called reversible heat, and the other which is dependent only on absolute temperature ($f(t,T)$), called irreversible heat flow. In other words, the reversible heat directly follows the modulated heating rate and the irreversible heat does not follow the heating rate. When the polymer goes through a glass transition, the heat capacity of the material increases due to the onset of chain mobility. Using modulated DSC experiment, the heat capacity change can be separated from other kinetic heat flow involved during heating, so that the glass transition of the polymer can be easily identified.

In a modulated DSC experiment, a cured underfill sample is pressed into a hermetic DSC pan. Then the sample can be heated from room temperature to an elevated temperature significantly exceeding the expected T_g (such as 200 °C) at a heating rate of 5 °C/min, with a temperature modulation of ±1 °C/min. N_2 purging can be applied to the DSC cell. The initial temperature of the heat flow step in the reversible heat flow is defined as the glass transition temperature.

5.3.3 TMA Measured Coefficient of Thermal Expansion

Coefficient of Thermal Expansion (CTE) of the underfill material is an important material property related to thermal mechanical stress of a flip chip underfill assembly. It can be measured using thermal mechanical analyzer (TMA), which measures the displacement of a probe placed on a solid sample when the sample is heated in the TMA furnace. From the expansion profile, the CTE can be calculated. Usually during a TMA experiment, a static force can be placed on the probe, which lies on top of a cured underfill sample of a defined dimension. A typical forced applied can be 0.05 N. The sample is then heated in the TMA furnace from room temperature to an elevated temperature exceeding the expected T_g (such as 200 °C) at a constant heating rate (such as 5 °C/min). N_2 purge can be applied to the TMA furnace during the experiment.

Figure 5.8 shows a typical expansion profile of a cured underfill sample. The inflection point of the thermal expansion is defined as the heat distortion temperature (also called TMA T_g). As observed in Fig. 5.10, after the TMA T_g, the CTE increases by two orders of magnitude. However, at an even higher temperature, the CTE seems to decrease. If the same experiment is performed again on the same material after the first scan, the CTE before TMA T_g would be similar to that in the first scan, but the CTE after TMA T_g is not as high as shown in Fig. 5.8. The CTE after TMA T_g in the second scan is close to the CTE of the first scan at high temperature. It is believed that the dramatic jump in CTE after TMA T_g in the first scan is related to the stress relaxation of polymer after curing. In common literatures, the CTE below TMA T_g is defined as α_1 and that above TMA T_g in the second scan is defined as α_2.

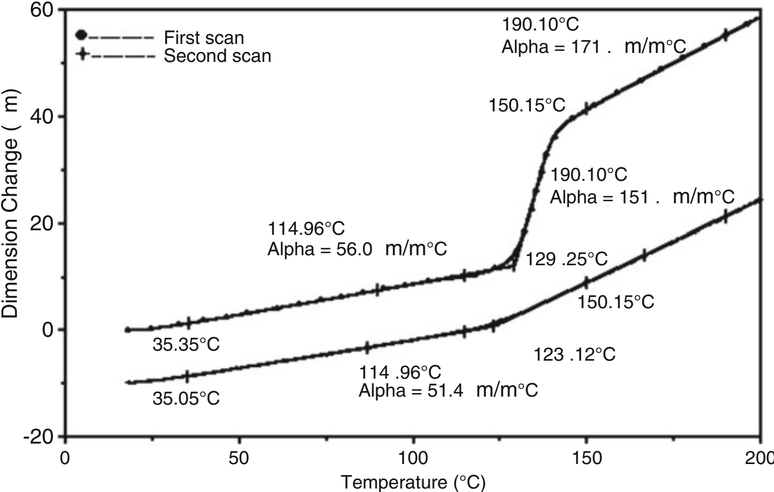


Fig. 5.8 A typical TMA thermal expansion profile of a cured underfill sample

5.3.4 DMA Measured Dynamic Moduli

One of the distinct characters of polymer is the viscoelasticity, which combines both solid-like and liquid-like features depending on the temperature and the experimentally chosen time scale [11]. The dynamic mechanical analyzer is a useful instrument to characterize the viscoelastic behavior of a polymer. In a typical DMA experiment, a sinusoidal stress or strain is applied on the sample as shown in (5.7):

$$\varepsilon = \varepsilon_0 \sin \varpi t,$$
$$\sigma = \sigma_0 \sin(\varpi t + \delta) \tag{5.7}$$

where ε is the strain; σ is the stress; ω is the angular frequency; t is the time; δ is the phase lag.

The stress–strain relationship can be defined by a quantity G' in phase with the strain and a quantity G'' which in 90° out of phase with the strain, as described by (5.8):

$$\sigma = \varepsilon_0 G' \sin \varpi t + \varepsilon_0 G'' \cos \varpi t \tag{5.8}$$

where $G' = (\sigma_0/\varepsilon_0) \cos \delta$;

$$G'' = (\sigma_0/\varepsilon_0) \sin \delta$$

A complex modulus G^* can be defined in (5.9):

$$G^* = \frac{\sigma}{\varepsilon} = \frac{\sigma_0}{\varepsilon_0}(\cos\,\delta + i\sin\,\delta) = G' + iG'' \tag{5.9}$$

G', which is in phase with the strain, is called the storage modulus because it defines the energy stored in the specimen due to the applied strain. G'', which is 90 ° out of phase with the strain, is called the loss modulus, for it is related to the energy dissipated per cycle as shown in (5.10):

$$\Delta E = \oint \sigma\,\mathrm{d}\varepsilon = \pi G'' \varepsilon_0^2 \tag{5.10}$$

The ratio of G'' to G' is tan delta:

$$\tan\,\delta = \frac{G''}{G'} \tag{5.11}$$

Figure 5.9 shows a typical DMA diagram for a cross-linked polymer. As can be seen, the storage modulus is usually several orders of magnitude higher than the loss modulus. So the magnitude of the complex modulus ($|G^*|$) is approximately the same as the storage modulus (G') at room temperature. The modulus of underfill at room temperature is important in determining the performance of the flip chip assembly. Ideally, a modulus around 8–10 GPa is desirable for underfill. The room temperature modulus of a typical epoxy resin without silica filler is around 2–3 GPa. As temperature arises, the storage modulus experiences a dramatic drop, which corresponds to the glass transition. The initial temperature of the step change in the storage modulus is defined as DMA T_g in this work. Around the glass transition, both the loss modulus and the tan δ show a maximum, indicating that there is a maximum in energy dissipated when the material experiences the glass transition.

After the glass transition, the storage modulus shows a plateau called rubbery plateau. According to the kinetic theory of rubber elasticity, the cross-linking density ρ can be determined from the storage modulus in rubbery state by (5.12) [12]:

$$\rho = \frac{G'}{3\Phi RT} \tag{5.12}$$

where G' is the rubbery storage modulus; ϕ is a front factor which is assumed to be 1; R is the gas constant; T is the absolute temperature.

5.3.5 TGA Measured Thermal Stability

Thermal gravimetric analyzer (TGA) measures the weight changes with respect to the temperature or time in a controlled atmosphere. It is commonly used to

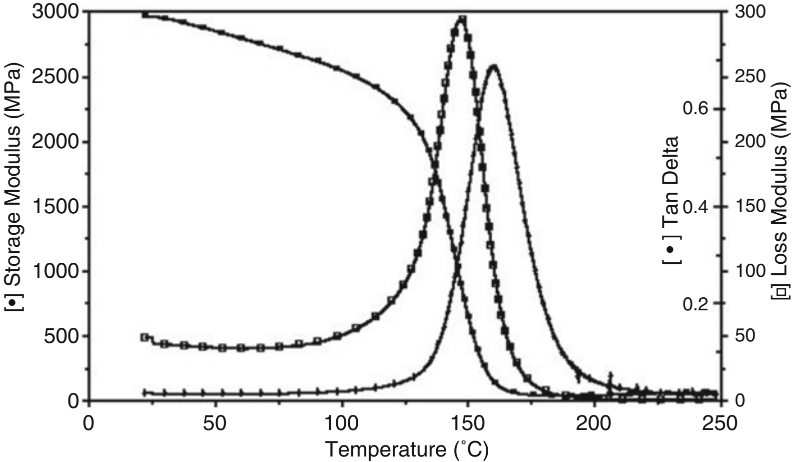


Fig. 5.9 A typical DMA diagram for a cross-linked polymer

investigate the outgassing of a formulation or a component of a formulation, and the thermal stability of cured resin. In a TGA experiment, a small amount of underfill sample (~20 mg) is placed in a platinum pan and heated to a specified temperature (usually 200 °C for an uncured sample and 600 °C for a cured sample) at a heating rate of 10 °C/min under N_2 purge. A curve of weight loss versus temperature is obtained and the thermal stability of the sample can be determined. For a cured sample, the onset temperature of the weight loss step was defined as the initial temperature of the decomposition of the material.

5.3.6 Flexure Test

A flexure test is often used to measure the Young's modulus of a solid sample. A universal test machine (UTM) can be used flexure test. The test is usually conducted in a 3-point bending mode at room temperature. From the test data, the stress–strain curve of the sample is graphed. The maximum stress and the strain at break of each specimen are recorded and the flexure modulus is calculated. Two typical stress–strain curves are shown in Fig. 5.10. Sample A shows a linear stress–strain relationship until near the break point, while Sample B shows a yield behavior. After yielding, Sample B continues to deform until break. The stress–strain curves indicate that Sample A experienced a brittle failure while Sample B experienced a ductile failure.

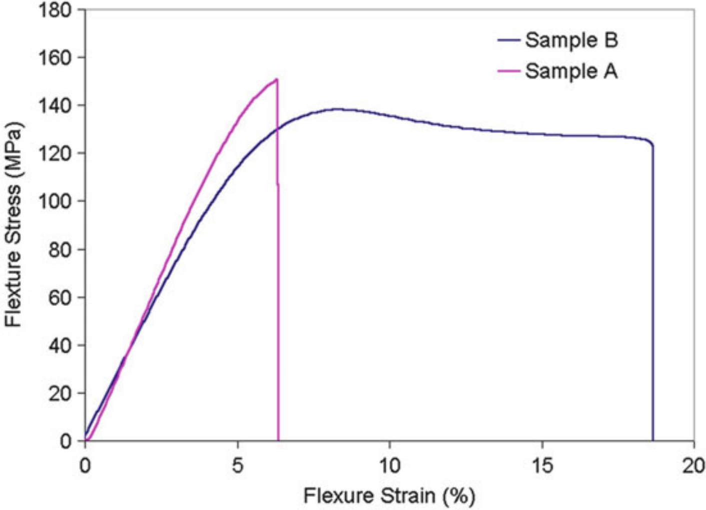


Fig. 5.10 Stress–strain curves for two polymers in a flexure test

5.3.7 Viscosity Measurement

Viscosity of a underfill is an important parameter in underfill application. A cone-and-plate geometry is typically used in unfilled material system since the material experiences a continuous shear rate or shear stress. In a filled material system where the filler size is significantly large, a parallel plate geometry can be used. Most liquid underfills without silica fillers display a Newtonian behavior while filled underfills can be shear-thinning or shear-thickening. If the liquid is Newtonian, constant shear rate experiment can be used to determine the viscosity at specific temperature. The viscosity of a non-Newtonian fluid is dependent on the shear rate and therefore viscosity has to be measured at different shear rates.

5.3.8 Adhesion of the Underfill to Die Passivation

Adhesion of the underfill to various interfaces is of critical importance to the performance of the flip chip assembly. In general, the interface between the underfill and the silicon chip is weaker than the interface between the underfill and the substrate. In most cases, delamination happens between the underfill and the die passivation during the reliability test.

One way to evaluate the adhesion strength of underfill is die shear test. In a die shear test, silicon dies (for instance, $2 \times 2\ \mathrm{mm}^2$) and substrates passivated with the same material are used. A thin layer of underfill material is applied in between the dies and the substrate. To ensure the consistent gap size, special glass beads of

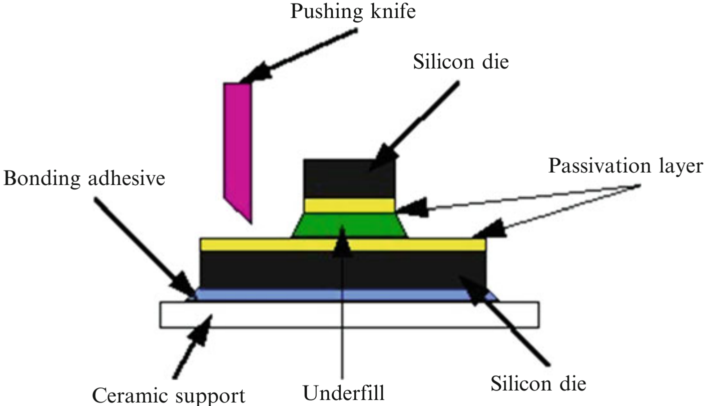


Fig. 5.11 A schematic of die shear test

consistent diameter can be added into the underfill. After the underfill is cured, the assembly can be tested using the bond tester. The structure of the test vehicle is shown in Fig. 5.11. The adhesion results are calculated from the shear forces used to shear the die off divided by the area of the die, and represented as the apparent strength of adhesion (ASA).

5.3.9 Moisture Absorption

Moisture absorption study can be conducted in an 85 °C/85 % RH temperature humidity chamber. The cured underfill samples are prepared to have similar initial weight and consistent geometry. Then the samples are put into the chamber and at different time intervals, the samples are taken out of the chamber and weighed. The percentage weight gain is defined as the moisture absorption at a certain time.

5.4 Reliability of Flip Chip Underfill Packages

The reliability of a flip chip package can be evaluated in a number of different methods, including thermal cycling, thermal shock, pressure-cook test, etc. The life time of a solder joint interconnect during temperature cycling can often be described by statistical models such as Weibull distribution. The probability density function (PDF) of the Weibull distribution is given by:

$$f(x) = \left(\frac{\beta}{x}\right)\left(\frac{x}{\theta}\right)^{\beta} \exp\left(-\left(\frac{x}{\theta}\right)^{\beta}\right) \tag{5.13}$$

where x is the thermal cycling life as a random variable; θ is the characteristic life; β is the shape parameter. The mean time to failure (MTTF), which is the expectation of the time to failure, for the Weibull distribution is:

$$\mathrm{MTTF} = \theta \cdot \Gamma\left(1 + \frac{1}{\beta}\right) \tag{5.14}$$

where Γ is the gamma function. It is generally believed that fatigue of the solder joints is a major reason for structure and electrical failures (Tummala 2001). The solder fatigue life can be described as a function of inelastic shear strain in Coffin–Manson equation (Manson & Coffin 1965, 1954):

$$N_f = \frac{1}{2}\left(\frac{\Delta\gamma}{2\varepsilon'_f}\right)^{1/c} \tag{5.15}$$

where N_f is the number of cycles to fatigue failure, $\Delta\gamma$ is the inelastic shear strain, ε'_f is the fatigue ductility coefficient, and c is the fatigue ductility exponent. Other strain based fatigue equations have been proposed, among which Solomon's model is often used (Soloman 1986):

$$N_f = \left(\frac{\theta}{\Delta\gamma_{\mathrm{p}}}\right)^{1/\alpha} \tag{5.16}$$

where $\Delta\gamma_{\mathrm{p}}$ is the percentage inelastic shear strain, θ and α are constants.

It has been shown that the use of the underfill can increase the lifetime of the solder joints by at least an order of magnitude during thermal cycling [13]. It was found that in an underfilled flip chip package, the fatigue life is highly dependent on the material properties of the underfill. The analytic model by Nysaether et al. [14] showed that while an underfill without filler increased the lifetime by a factor of 5–10, a filled underfill with a lower CTE gave a 20–24-fold increase in lifetime. For both filled and non-filled sample, the lifetime is nearly constant regardless of distance to the neutral point (DNP), indicating that the underfill effectively couples the stress among all the solder joints.

Many numerical models have been developed to study the solder fatigue life of a flip chip package with or without underfill. The polymeric nature of the underfill material requires careful characterization for correct material property input to the numerical models. The modulus of a polymeric material is not only a function of temperature, but also a function of time, i.e., it is a viscoelastic material. As mentioned in the previous section, the Thermal mechanical analyzer (TMA) and dynamical mechanical analyzer (DMA) are typically used to characterize the viscoelastic properties of the underfill material. Dudek et al. characterized four commercial electronic polymers and used finite element (FE) analyses to study the effect of die size and underfill material properties on the thermomechanical reliability of the flip

chip on board (FCOB) package [15]. They found that although the use of underfill can effective reduce the shear strain, it can also cause bump creep strain in the transverse board direction due to stretching and compressing of the bump during thermal cycling. This load is due to the CTE mismatch between the solder and underfill/ solder-mask layer. Underfill with CTE that matched the solder material (22–26 ppm/° C) gives the best thermal cycling life based on the creep strain criterion.

The function of the underfill in a flip chip package is stress redistribution, not stress reduction. A rigid underfill material mechanically couples the device and the substrate, changing partially the shear stress experienced by the solder joints into bending stress on the whole structure. Shrinkage of the underfill during cure and the CTE mismatch during cooling after cure can generate large stress on the Si chip, resulting in die crack in some cases. Palaniappan et al. performed in-situ stress measurements in the flip chip assemblies using a test chip with piezoresistive stress sensing devices [16]. The study concluded that the underfill cure process generates large compressive stress on the active die surface, indicating a complex convex bending state in the flip chip. The level of stress measured can lead to Si fracture. The residual die stress was found to be strongly dependent on underfill CTE, modulus and T_g. A finite element analysis by Mercado et al. on the die edge cracking in flip chip PBGA packages also concluded that the energy release rate for horizontal Si facture increases with underfill modulus and CTE [17].

In addition to temperature related thermomechanical failure, moisture induced failures such as delamination and corrosion are common for a flip chip underfill package. The adhesion of the underfill to the passivation layer on the IC chip is crucial to the integrity and reliability of the flip chip assembly. Delamination (total loss of adhesion) at the interface between underfill and die can lead to cracking of the interconnection. Moisture can then diffuse through the delaminated area, leading to corrosion of metal pad, line, and interconnection [18]. It has been shown that high adhesion strength at the interface between the underfill and the die is strongly correlated to improved solder joint fatigue life and alleviates underfill fillet crack problem.

Interfacial adhesion can degrade because of the degradation of underfill itself or due to the degradation of interfacial interaction.

When a flip chip package is subjected to aging in an environment with a high temperature and a high humidity, such as in an environmental chamber at 85 °C/ 85 % relative humidity (85/85) or a pressure cooker at 121 °C, 2 atm., and 100 % relative humidity, to test its reliability, the underfill material absorbs moisture in the humid environments, which increases its weight, increases the dielectric constant of the underfill material, and reduces the glass transition temperature of the underfill (as water is a plasticizer for the polymer matrix) [19]. As a result, the maximum use temperature of the underfilled flip chip package may be reduced. In addition, absorption of moisture leads to swelling of the underfill and this swelling introduces swelling stress at the interface between the underfill and the passivation layer as well as at the interface between the underfill and the solder mask [20, 21]. Moisture diffusion through epoxy underfill is the major cause of corrosion of metals in electronic packaging. More importantly, the adhesion of the underfill to the passivation layer decreases due to moisture adsorption during this kind of aging test, particularly to the hydrophilic passivation layer.

There are three major mechanisms for adhesion degradation due to moisture. Water can replace hydrogen bonding between the underfill and the passivation layer, reduce dipolar and dispersive interactions between the underfill and the passivation layer, and break chemical bonds at the interface through possible hydrolysis reactions. The adhesion stability in a humid environment can be improved by use of a coupling agent, which introduces a stronger chemical bond at the interface.

In a review paper, Luo et al. systematically discuss the influence of hot-wet aging on the adhesion of an underfill material (epoxy resin cured with acid anhydride) to the passivation layer in flip chip packaging [22]. Adhesion degradation in a hot-wet environment is correlated to passivation layer hydrophilicity. The rate of adhesion degradation is discussed in terms of mobility of both polymer chains and absorbed water in the polymer matrix. An approach to improving the adhesion stability during hot-wet aging is demonstrated.

Two epoxy formulations were used in this study. Formulation-A consisted of one equivalent of cycloaliphatic epoxy resin ERL4221 (7-oxabicyclo[4,1,0]heptane-3-carboxylic acid 7-oxabiscyclo[4,1,0]hept-3-ylmethyl ester) from Union Carbide, 0.8 equivalent of hardener 4-methylhexahydrophthalic anhydride (MHHPA) from Aldrich, and catalyst cobalt (II) acetylacetonate (0.4 % of total weight of epoxy resin and hardener). Formulation-A was cured in an oven at 250 °C for 30 min, and its glass transition temperature was about 190 °C (determined with a differential scanning calorimeter at a heating rate of 5 °C/min). Formulation-B consisted of one equivalent of epoxy resin 1,4-butanediol diglycidyl ether from Aldrich, 0.8 equivalent of hardener MHHPA, and catalyst 2E4MZ-CN (1-cyanoethyl-2-ethyl-4-methylimidazole, 1 % of total weight of epoxy resin, and hardener). Formulation-B was cured at 175 °C for 30 min, and its glass transition temperature was about 65 °C (determined with a differential scanning calorimeter at a heating rate of 5 °C/min). When a coupling agent (CA) was used, either epoxysilane or γ-glycidoxypropyltrimethoxysilane (CA-1) or aminosilane, γ-aminopropyltriethoxysilane (CA-2) was added into formulation-A. The concentration of the silane coupling agent in the underfill was 1.5 wt.%.

5.4.1 Effect of Passivation Layer

Table 5.1 shows the adhesion strength of underfill formulation-A to the four different passivation layers used before and after hot-wet aging. Adhesion strength of formulation-A to SiO_2 or Si_3N_4 decreased greatly after aging in a humid environment at a high temperature. The adhesion strength of formulation-A underfill to SiO_2 was 50 MPa before aging. It decreased to 13 MPa after 85/85 aging for 500 h, and to almost zero after pressure cook test (PCT) for 24 h. The adhesion strength of formulation-A to Si_3N_4 decreased from more than 60 MPa before aging to less than 10 MPa after 85/85 aging for 500 h or PCT for 24 h. However, BCB and PI showed different behavior. The adhesion strength of formulation-A to BCB and PI

Table 5.1 Adhesion strength of underfill formulation-A to different passivation layers

Passivation layer	Adhesion strength before aging (MPa)	Adhesion strength after 85/85 aging for 500 h (MPa)	Adhesion strength after PCT aging for 24 h (MPa)
SiO_2	50.0 ± 8.3	12.9 ± 8.9	5.4 ± 2.7
Si_3N_4	65.8 ± 12.0	8.7 ± 4.9	2.8 ± 2.8
BCB	37.0 ± 8.3	37.3 ± 8.5	11.5 ± 4.4
PI	58.8 ± 10.5	63.7 ± 9.4	11.2 ± 6.1

did not decrease significantly after 85/85 aging for 500 h, indicating that adhesion of formulation A to BCB and PI is more stable than that of formulation-A to SiO_2 and Si_3N_4 passivation layers. PCT aging is more severe than 85/85 aging. After PCT aging for 24 h, the adhesion strength of formulation-A to BCB dropped from an initial value of 38 to 12 MPa, while the adhesion strength of underfill to PI dropped from 58 to 12 MPa. Overall, BCB and PI showed better adhesion retention after PCT aging than SiO_2 and Si_3N_4.

Adhesion retention after hot-wet aging is related to the hydrophilicity of passivation layer. Hydrophobic organic passivation layers BCB and PI have better adhesion retention after hot-wet aging than hydrophilic inorganic passivation layers SiO_2 and Si_3N_4. Water tends to stay at the interface between the underfill and SiO_2 or Si_3N_4 rather than at the interface between the underfill and BCB or PI passivation layer. The presence of water at the interface reduces the interfacial adhesion by three mechanisms: reducing hydrogen bonding interactions between the underfill and passivation layer, reducing dipolar and dispersive interactions between the underfill and passivation layer, and breaking chemical bonds at the interface through possible hydrolysis reactions. Thus, it is necessary to improve the hydrolytic stability for adhesion between epoxy underfill and hydrophilic passivation layer.

5.4.2 Adhesion Degradation Versus 85/85 Aging Time

Moisture diffusion into the underfill material during 85/85 aging was studied, and the moisture diffusion coefficient in formulation-A during 85/85 aging (0.011 mm^2/h) was determined. During 85/85 aging of die shear samples, moisture primarily diffuses through the four edges into the underfill, while the much slower moisture diffusion through silicon can be ignored. The rate of moisture diffusion into the die shear sample would be the same as the rate of moisture diffusion into an infinitely long orthogonal bar with dimensions of 2 × 2 mm in cross section. To study moisture diffusion into the die shear samples during 85/85 aging, a long orthogonal bar (2 × 2 × 50 mm) of cured underfill was prepared, and the moisture absorption in this sample under 85/85 condition was measured. The results are shown in Fig. 5.12. The orthogonal underfill bar became saturated with moisture in less than 24 h during 85/85 aging. This implies that the moisture absorption into the underfill of the die shear sample (the area of the small die is 2 × 2 mm) reaches an equilibrium value in

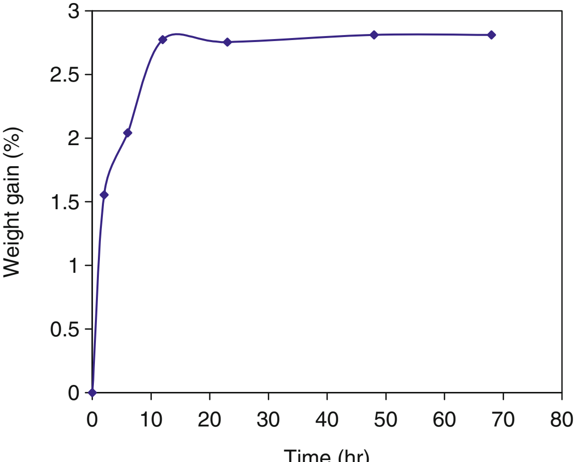


Fig. 5.12 Moisture absorption in a 2 × 2 × 50 mm long orthogonal bar of underfill formulation-A during 85/85 aging

less than 24 h. The moisture diffusion into underfill formulation-B is faster than into underfill formulation-A, and the die shear sample prepared with formulation-B was also saturated with moisture in less than 24 h during 85/85 aging.

The retention of adhesion strength between underfill formulations and Si_3N_4 passivation layer under 85/85 aging was measured after various periods of aging (Fig. 5.13). Adhesion strength decreased after 24 h of 85/85 aging for both formulation-A and formulation-B. However, a dramatic difference in the degradation rate of adhesion strength can be seen between formulation-A and formulation-B. Half of the adhesion strength was lost after 24 h of 85/85 aging for formulation-A. Further aging in 85/85 condition further reduced the adhesion strength. The adhesion strength reached an equilibrium value after 96 h of 85/85 aging. This showed that the adhesion degradation of formulation-A was a slower process than moisture absorption.

The degradation of the adhesion of underfill formulation-B with a relatively lower glass transition temperature (T_g) was very fast. About 85 % of the original adhesion strength was lost after 24 h of 85/85 aging. Further aging for 48, 96, and 200 h seemed to have no additional effect on the adhesion strength. It seems that the adhesion strength reached an equilibrium value after 24 h of 85/85 aging.

The difference in the rate of adhesion degradation is related to the mobility of polymer chains of the underfill material and the activity of the absorbed water under 85/85 aging condition. During 85/85 aging, formulation-A is in its glassy state (T_g of formulation-A after saturated with water under 85/85 condition is 130 °C). However, the underfill formulation-B is in its rubbery state during 85/85 aging (its T_g after becoming saturated with water under 85/85 condition is 45 °C). Solid state ^{1}H NMR spectra (Fig. 5.14) were recorded for deuterated water soaked formulations to determine the mobility of polymer chains at 85 °C [23]. The peak width indicates the mobility of the polymer chain. The polymer chain mobility of formulation-A during 85/85 aging is much lower than that of formulation-B. The peak widths at half height of the solid state ^{1}H NMR spectra were 56.1 and 3.4 kHz for D_2O soaked formulation-A and formulation-B, respectively.

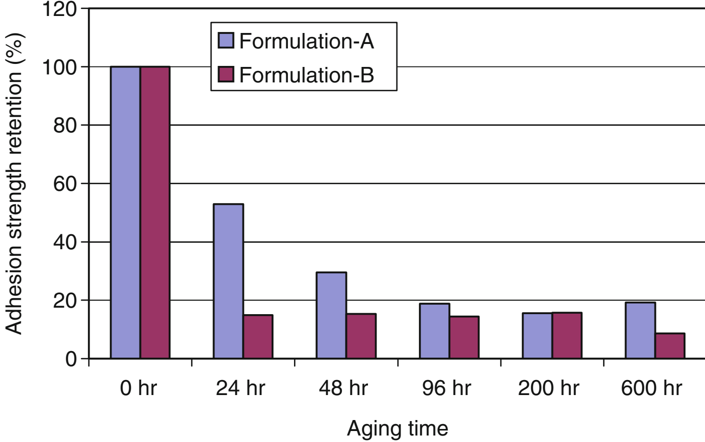


Fig. 5.13 Adhesion strength retention of underfills during 85/85 aging (passivation layer: Si_3N_4)

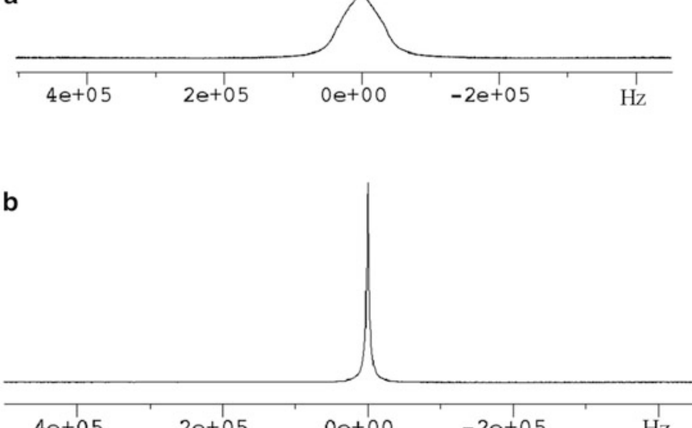


Fig. 5.14 Solid state ^{1}H NMR spectra of D_2O soaked underfill formulation-A (**a**) and underfill formulation-B (**b**) (the axis represents frequency)

The absorbed water in different polymers has different mobilities during 85/85 aging too. Solid state ^{2}H NMR experiments on deuterated water soaked cured underfill formulations were performed at 85 °C to determine the mobility of water in the polymer matrix (Fig. 5.15). At 85 °C, the absorbed water in formulation-B which in its rubbery state has a much higher mobility than in formulation-A which is in its glassy state. The higher mobilities of both absorbed water and of polymer chains during 85/85 aging contribute to the faster adhesion degradation of rubbery polymer compared to that of glassy polymer.

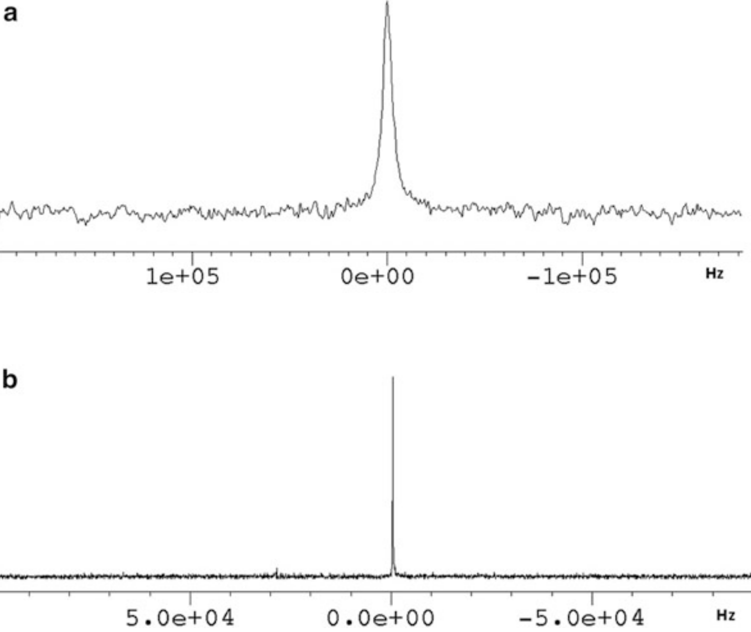


Fig. 5.15 Solid state ^{2}H NMR spectra of (**a**) D_2O soaked formulation-A and (**b**) D_2O soaked formulation-B at 85 °C (the axis represents frequency)

5.4.3 Improvement of Adhesion Hydrolytic Stability Through Coupling Agents

In order to improve the hydrolytic stability of adhesion between the underfill material and the hydrophilic passivation layers, an organofunctional silane coupling agent can be used as an additive in the underfill. The generic structure of a silane coupling agent is shown in Fig. 5.16, where X is a hydrolyzable group such as—OCH_3 or OCH_2CH_3, R is a short hydrocarbon segment which is normally two to three carbon atoms in length, and Y is an organofunctional group such as amino, epoxy, etc.

The Si–X bonds in a silane coupling agent hydrolyze in the presence of water to give –Si–OH group, which then react with the silanol group –Si–OH on the surface of SiO_2 or Si_3N_4. The organofunctional group Y reacts with the functional groups in the underfill during curing to form chemical bonds with the polymer matrix. Thus, Si–O–Si chemical bond bridges are formed between the SiO_2 or Si_3N_4 surface and the underfill through coupling agents [24–26].

The adhesion strengths of the underfill formulation-A with and without coupling agents to SiO_2 and Si_3N_4 are shown in Table 5.2. When silane coupling agents were added, the adhesion strength before aging was not improved. The hydrolysis of Si–X bond is not likely to take place without aging in a humid environment, so chemical bond bridges through the silane coupling agent between epoxy underfill and SiO_2

Fig. 5.16 Generic structure of a silane coupling agent

X
|
X —Si— R —Y
|
X

Table 5.2 Adhesion strength (MPa) of underfill formulation A with different coupling agents before and after aging tests

	Before aging	After 85/85 aging for 500 h	After PCT aging for 24 h
For Si_3N_4			
No CA	65.8 ± 12.0	8.7 ± 4.9	2.8 ± 2.8
With CA-1	63.3 ± 12.3	43.5 ± 7.1	32.1 ± 10.1
With CA-2	62.0 ± 11.4	50.5 ± 8.9	7.7 ± 6.2
For SiO_2			
No CA	50.0 ± 8.3	12.9 ± 8.9	5.4 ± 2.7
With CA-1	51.2 ± 11.4	44.1 ± 4.3	25.4 ± 8.2
With CA-2	48.7 ± 9.0	39.3 ± 7.7	9.7 ± 7.0

CA-1: epoxysilane; CA-2: aminosilane

or Si_3N_4 passivation layer could not be established. Adhesion strength retention after 85/85 aging was improved dramatically for underfill with the addition of silane coupling agents. A slight decrease in adhesion strength was observed after 85/85 aging for systems with coupling agents, which resulted from the reduced dispersive interaction, dipolar interaction, and hydrogen bond interaction between the epoxy underfill and passivation layers due to the adsorption of water at the interface. There was no significant difference in adhesion retention performance between CA-1 and CA-2 after 85/85 aging. However, the epoxysilane (CA-1) had much better adhesion retention ability than the aminosilane (CA-2) after PCT aging.

The difference in adhesion retention performance between CA-1 and CA-2 can be explained by the conformation of the silane coupling agent at the interface. In order for the silane coupling agent to form a chemical bond between the underfill and the SiO_2 or Si_3N_4 surface, the Si–X group in the coupling agent must be close to the passivation layer surface. The adsorption of silane coupling agent onto glass was measured by a flow micro-calorimeter (Table 5.3), showing that aminosilane had a stronger adsorption onto glass than epoxysilane [27]. This difference is caused not by the alkoxy group, but by the organofunctional groups: amino and epoxide. This indicated that the amino group is strongly adsorbed onto the glass surface due to hydrogen bonding as well as acid/base interaction, as glass surface is acidic in nature. Similarly, amino group should have stronger adsorption onto SiO_2 or Si_3N_4 passivation surface than epoxy groups in silane coupling agents. Because of this, the amino groups in aminosilane coupling agent have less chance to react with the polymer matrix during curing of underfill, and the Si–OR groups in aminosilane coupling agent have less chance to react with silanol group on passivation surface to form strong chemical bonds in the presence of water. In contrast, the epoxy group has less affinity to SiO_2 or Si_3N_4 surface. Thus, the epoxy groups in

Table 5.3 Measured heats of adduct formation of borosilicate glass with two silanes used

Silane	Heat of adduct formation (kJ/mol)
CA-1 (epoxysilane)	18.01 ± 1.73
CA-2 (aminosilane)	145.12 ± 11.53

epoxysilane have more chance to react with the underfill matrix, and the Si–OR groups in the epoxysilane have more chance to react with the silanol group on the SiO_2 or Si_3N_4 surface to form strong chemical bonds. As a result, there were more stable chemical bond bridges between epoxy underfill containing epoxysilane and the SiO_2 or Si_3N_4 surface after PCT aging.

In conclusion, the hydrolytic stability of adhesion depends on the hydrophilicity of the passivation layer. Hydrophilic passivation layers such as SiO_2 and Si_3N_4 show much more severe adhesion degradation than the hydrophobic passivation layers such as benzocyclobutene and polyimide. Higher mobilities of both underfill polymer chains and absorbed water contribute to a faster adhesion degradation of rubbery polymer than that of glassy polymer during 85/85 aging. The adhesion hydrolytic stability for hydrophilic passivation layer can be improved by use of a silane coupling agent that introduces stable chemical bonds at the interface. Between epoxysilane and aminosilane, the former is more effective in improving adhesion stability during the pressure cooker test. The adsorption of silane on the passivation layer surface plays an important role in the improvement of hydrolytic stability of adhesion.

In summary, many studies have concluded that underfill material property is one of the key factors determining the reliability of the package. The general guideline on the material properties of underfill for flip chip package can be summarized in the following table. However, one has to keep in mind that different failure modes coexist in a reliability test, which sometimes present conflicting requirements on underfill. For instance, to effectively couple the stress on the solder joints, high modulus underfill is desired. On the other hand, high underfill modulus can lead to high residual stress and therefore die crack. Another example is the filler loading. Low CTE requirement on underfill indicates high filler loading. However, an underfill with higher filler loading typically has a higher viscosity, causing difficulty in underfill dispensing. The result might be underfill voids and nonuniformity, which would cause reliability issues. Therefore, the choice of underfill highly depends on the application, e.g., die size, passivation material, substrate material, type of solder, and environment conditions the package will be subjected to during application, etc. (Table 5.4).

5.5 New Challenges to Underfill

As silicon technology moves to sub 0.1 μm feature size, the demands for packaging also involve as the bump pitch gets tighter, bump size smaller, die size larger for future flip chip. As a result, the capillary underfill process faces tremendous

Table 5.4 Desirable underfill properties for flip chip packages

Property	Value
Curing temperature	<150 °C
Curing time	<30 min
T_g	>125 °C
Working life (viscosity double @ 25 °C)	>16 h
CTE (α_1)	22–27 ppm/°C
Modulus	8–10 GPa
Fracture toughness	>1.3 MPa × $m^{1/2}$
Moisture absorption (8 h boiling water)	<0.25 %
Filler contents	<70 wt.%

challenges. As it was discussed previously, underfill flow problem aggravates as the size of the chip becomes larger and the gap between the chip and substrate gets smaller. Among the emerging development of flip chip, lead-free solder and low-K (dielectric constant) ILD (interlayer dielectric)/Cu present new challenges to underfill [28].

High-lead and lead–tin eutectic solders have been widely used for chip-package interconnections. Recent environmental legislations towards toxic materials and consumers' demand for green electronics have pushed the drive towards lead-free solders. Alternatives have been proposed using multiple combinations of elements like tin, silver, copper, bismuth, indium, and zinc, most of which require increased reflow temperature profiles during the soldering process relative to the well-known tin–lead alloys. The following table shows some of the common lead-free solders (Table 5.5).

Among the several Pb-free candidate solders, the near ternary eutectic Sn–Ag–Cu (SAC) alloy compositions, with melting temperatures around 217 °C, are becoming consensus candidates. The optimal composition 95.4Sn/3.1Ag/1.5Cu has provided a combination of good strength, fatigue resistance and plasticity [29]. In addition, the alloy has sufficient supply and adequate wetting characteristics.

The use of Sn/Ag/Cu solder presents two major challenges on the flip chip assembly process. First, since the melting point of the alloy is more than 30 °C higher than that of the eutectic Sn/Pb alloy, the process temperature is raised by 30–40 °C. The high process temperature has a great impact on the substrate since the conventional FR-4 material has a T_g at around 125 °C and also subjects the attached components to a higher thermal stress. Higher warpage is introduced when the board is subjected to higher temperature reflow. There have been considerable researches in high T_g substrate for lead-free process. The second challenge comes from the flux chemistry. Since the current fluxes in use are usually designed for eutectic Sn/Pb solder, they either do not have high enough activity or do not possess sufficient thermal stability at high temperature. So generally, the wetting behavior of the lead-free solders is not as good as that of the eutectic Sn/Pb solder [30, 31].

With the trends of lead-free solder interconnect, the underfill for flip chip in package application faces new challenges of compatibility with higher reflow temperature. High temperature reflow causes component damage due to enhanced level of materials degradation, moisture ingress and mechanical expansion. Therefore, the thermal stability, adhesion to various interfaces, strength and fracture

Table 5.5 Possible lead-free alloys

Alloy	Melting point
Sn96.5/Ag3.5	221 °C
Sn99.3/Cu0.7	227 °C
Sn/Ag/Cu	217 °C (Ternary eutectic)
Sn/Ag/Cu/X(Sb, In)	Ranging according to compositions, usually above 210 °C
Sn/Ag/Bi	Ranging according to compositions, usually above 200 °C
Sn95/Sb5	232–240 °C
Sn91/Zn9	199 °C
Bi58/Sn42	138 °C

toughness of the underfill need to be improved. The SAC alloy does not plastically deform as much as the eutectic PbSn solder. The creep deformation is less at lower stress level and more at higher stress level compared to the PbSn solder, which indicates that the choice of the underfill would depend on the application needs. A temperature cycle with a large temperature difference and lower dwelling times could induce more creep and therefore requires more protection from the underfill [32]. An evaluation of underfill materials for lead-free application conducted by Intel Corporation [33] showed that majority of the failure occurred during the moisture sensitive level (MSL) three followed by 260 °C reflow. Delamination seemed to be the common failure mechanism after the high temperature reflow. This failure was also correlated to materials with low filler content and low coupling agent content. In general, materials with high filler content (and therefore low CTE, high modulus, and low moisture uptake) and good adhesion were compatible with the lead-free process.

As the IC fabrication moves towards small feature and high density, the interconnect delay becomes dominant. This calls out for new interconnect and interlayer dielectric (ILD) materials. The Cu metallurgy and low-K ILD has been successfully implemented to increase device speed and reduce power consumption. These low-K materials tend to be porous and brittle, having high CTE and low mechanical strength, compared to the traditional ILD materials such as SiO_2. The CTE mismatch between the low-K ILD and the silicon die creates a high thermomechanical stress at the interface. Therefore, the choice of underfill becomes critical since it not only protects the solder joints by stress redistribution, but also need to protect the low-K IDL and its interface with the silicon. Critical material properties of underfill to achieve reliability requirement for the low-K ILD package include the T_g, CTE, and the modulus. However, the optimal combination of these properties is still controversial.

Five underfills were evaluated for the low-K flip chip package by Tsao et al. [34]. Both modeling and experimental evaluation indicated that the low T_g and low stress-coupling-index underfills yielded better reliability in the low-K flip chip package. Two moderately low T_g underfills (T_g between 70 and 120 °C) showed good potential in protecting both the solder joints and low-K interface. An underfill with a very low T_g (lower than 70 °C), on the other hand, failed to protect the solder joints during the thermal cycling test. A study conducted by LSI Logic Corp and

Henkel Loctite (now called Henkel) Corp [35], on the other hand, indicated that underfills with high T_g and low modulus is advantageous for the low-K flip chip. The low modulus of the underfill exerts lower stress on the package and therefore reducing the stress on the low-K layer, preventing underfill delamination and die cracking. The high T_g prevents solder bump fatigue by maintaining a low CTE over the temperature cycling. The high T_g, low modulus underfill developed by Henkel exhibited good manufacturability and reliability in the package qualification testing including JEDEC preconditioning, thermal cycling, biased humidity testing and high temperature storage.

With all the new challenges to the flip chip and underfill, capillary underfill is still the main packaging technology for flip chip devices. However, the continuing shrinking of pitch distance and gap height will eventually post limitation on the capillary flow. The industry has started to look for alternatives to capillary underfill. The following sections describe the new development in underfill material and process.

5.6 No-Flow Underfill

The idea of integrated flux and underfill was patented by Pennisi et al. in Motorola back in 1992 [36]. It triggered the research and development of no-flow underfill process. The first no-flow underfill process was published by Wong et al. in 1996 [37]. The schematic process steps are illustrated in Fig. 5.17. Instead of underfill dispensing after the chip assembly in the conventional process, in a no-flow underfill process, the underfill is dispensed onto the substrate before the placement of the chip. Then the chip is aligned and placed onto the substrate and the whole assembly goes through solder reflow where the interconnection through solder balls is established while the solder melts. This novel no-flow process eliminates the separate flux dispensing and flux cleaning steps, avoids the capillary flow of underfill, and finally combines the solder bump reflow and underfill cured into a single step, hence, improves the production efficiency of underfill process. It is a step forward for the flip chip to be compatible with surface mount technology (SMT).

The key to the success of a no-flow underfill process lies in the underfill material. The first patent on the no-flow underfill material was by Wong and Shi in Georgia Institute of Technology [38]. The two critical properties of the no-flow underfill to enable this new process are the latent curing ability and the build-in fluxing capability. The nature of the no-flow underfill process requires that the underfill have enough reaction latency to maintain its low viscosity until the solder joints are formed. Otherwise, gelled underfill would prevent the melting solders from collapsing onto the contact pads, resulting in low yield of solder joint. On the other hand, elimination of the post cure is desired since post cure takes additional off-line process time, adding to the cost of this process. Many latent catalysts for epoxy resins have been explored for the application of no-flow underfill. In the material system that Wong and Shi designed, Co(II) acetylacetonate was used as the latent

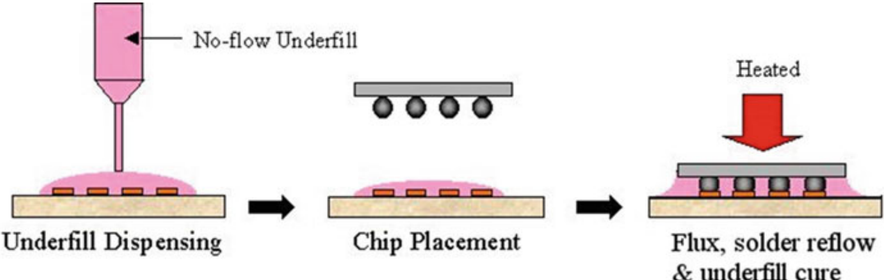


Fig. 5.17 Flip chip process using no-flow underfill

catalyst [39, 40], which gave enough curing latency for no-flow underfill. The advantage of metal chelates lies not only in its latent acceleration, but also in the wide curing range they offer. By exploring different metal ions and chelates, the curing behavior of different epoxy resins could be tailored to the application of no-flow underfill for lead-free solder bumped flip chip [41]. Since lead-free solders usually have a higher melting point than eutectic SnPb solder, no-flow underfill for lead-free bumped flip chip requires higher curing latency to ensure the wetting of the lead-free solder on the contact pad. Zhang et al. explored 43 different metal chelates and developed no-flow underfill compatible with lead-free solder reflow [41]. Successful lead-free bumped flip chip on board package using no-flow underfill process has been demonstrated [42].

Despite the importance of the curing process of no-flow underfill, there is little study on the curing kinetics and its relation to the reflow profile. In an attempt to develop systematic methodology to characterize the curing process of no-flow underfill, Zhang et al. used an autocatalytic curing kinetic model with temperature dependent parameters to predict the evolution of degree of cure (DOC) during the solder reflow process [43]. Figure 5.18 shows the result of DOC calculation of a no-flow underfill in eutectic SnPb and lead-free solder reflow process. If the DOC of the underfill at the solder melting temperature is lower than the gel point, the molten solder would be allowed to wet the substrate and make the interconnection. Another approach is the in-situ measurement of viscosity of no-flow underfill using microdielectrometry by Morganelli et al. [44]. Since the viscosity is related to the ionic conductivity, the dielectric properties of the underfill can be used for the in-situ analysis of the no-flow underfill in the reflow process, which can be used to predict the solder wetting behavior.

The other key property for no-flow underfill is the fluxing capability. In a conventional flip chip process, flux is used to reduce and eliminate the metal oxide on the solder and to prevent it from being reoxidized under high temperature. Instead of applying flux, no-flow underfill is dispensed before the chip placement. Hence, the self-fluxing capability is required to facilitate solder wetting. To achieve this goal, research has been done to develop reflow-curable polymer fluxes [45]. A comprehensive study on the fluxing agent of no-flow underfill material was carried out by Shi et al. [46–48], which included the relationship between the surface

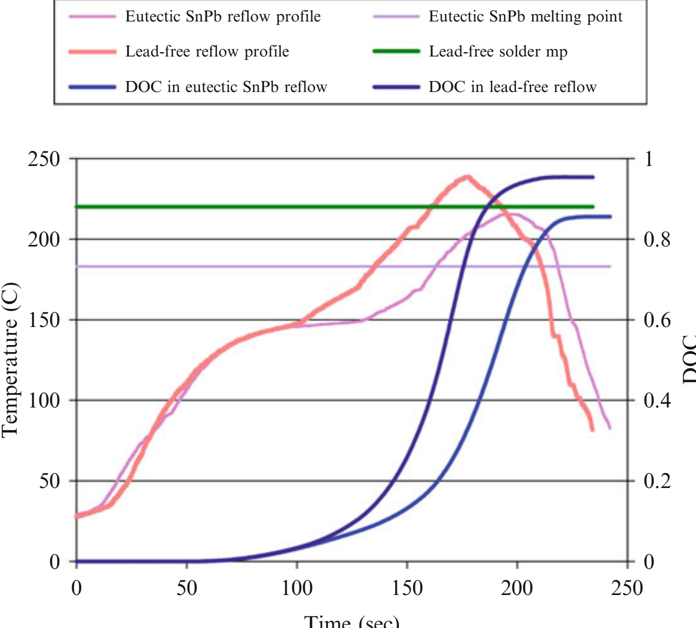


Fig. 5.18 Degree of cure (DOC) evolution of a no-flow underfill in eutectic SnPb and lead-free reflow process

composite on Cu pad and the fluxing capability of no-flow underfill, and also the effect of the addition of the fluxing agent on the curing and material properties of no-flow underfill.

The process of no-flow underfill has always attracted much attention in the assembly industry. Voids formation is often observed in many flip chip no-flow underfill packages. The origin of the voids could be the out-gassing of the underfill, moisture in the board, and trapped voids during assembly, etc. They are usually tacked to a solder bump or in between two bumps [49, 50]. Figure 5.19 shows an example of underfill voids in a no-flow underfill package observed with scanning acoustic microscope. Voids in the underfill, especially voids near the solder bumps lead to early failure through a number of ways including stress concentration, underfill delaminate, and solder extrusion. Study has showed that solder bridging might result from the solder bump extrusion through the micro voids trapped between adjacent bumps [51]. The material and process factors influencing the voiding behavior are complicated and interacting. It has been shown that the outgassing of anhydride could cause severe voiding, if the curing latency is high and also the reflow temperature is high; hence, the voiding becomes more prominent in a lead-free reflow process [52]. The important process parameters that affects underfill voiding in a no-flow process include the underfill dispensing pattern, the solder mask design, the placement force and speed, and the reflow profile, etc. [53, 54]. Before assembly, the PWB substrate needs to be baked to dry out any moisture to prevent voiding from

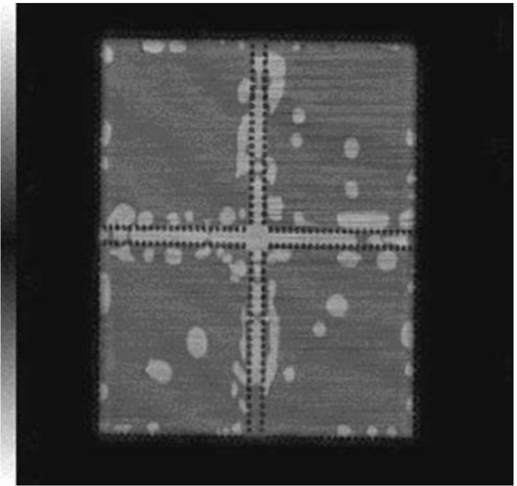


Fig. 5.19 An example of voids in a no-flow underfill package

the board [50]. It has been shown that in some cases a fast gelation of underfill is desired to minimize the voiding while in other cases, extending duration at high temperature can "push" out the voids [50, 55]. In short, with the right material and process parameters, voiding in no-flow underfill can be minimized. However, the process window is usually very narrow. An important point was raised by Zhao et al. [31] that for a small circuit board where the temperature distribution is more homogenous, it is relatively easy to develop a "good" reflow profile while for complex SMT assemblies involving multiple components and significant thermal mass difference across the board, the optimization of the reflow process presents great challenge.

The reliability of flip chip no-flow underfill package has been evaluated in many occasions. Discrepancies exist among these reports because the process and reliability of the no-flow underfill package depend largely on the package designs including the size of the chip, the pitch, the surface finish of the pad, etc. Among the earliest reporters on no-flow underfill, Gamota and Melton compared the reliability and typical failure mode of conventional underfill package and no-flow underfill packages [56]. They found that in a conventional underfill assembly, the failure of the assembly mainly resulted from the interfacial delamination between the underfill and the chip passivation. However, with unfilled materials in no-flow underfill, good interfacial integrity was observed, and the assembly failed mainly due to the fracture through the solder interconnects near PCB. Since the no flow underfill was unfilled, the CTE was high. They argued that the relative localized CTE mismatch between the chip, the underfill, and the PCB resulted in a high local stress field which initiated fracture in the solder interconnects. No-flow underfill without silica fillers or very low filler loadings is not only high in CTE, but also low in fracture toughness [57]. Combined with high CTE mismatch, the low fracture toughness leads to early underfill cracking both inside the bulk and in underfill fillet. Fillet cracking causes delamination between the underfill and the die passivation and/or between the underfill and the board, while bulk cracking can initiate solder

joint cracking and solder bridging [58]. These all become the common failure modes for flip chip no-flow underfill package. Efforts have been made to enhance the toughness of the no-flow underfill materials through the incorporation of the toughening agents [59]. The effect of the glass transition temperature (T_g) of the no-flow underfill on the reliability of the package has been controversial. It is usually believed that the T_g of the underfill should exceed the upper limit of the temperature cycling (125 °C, or 150 °C) to ensure consistent material behavior during the reliability test. However, some tests have shown that low T_g (~70 °C) underfill material performed better in liquid-to-liquid thermal shock (LLTS) [60]. The research by Zhang et al. on the development of non-anhydride based no-flow underfill [61] also showed that high T_g is not critical to reliability. Although the CTE of the underfill above T_g is much higher than that below T_g, the modulus of the underfill decreases dramatically; so the overall stress in the underfill does not necessarily increase when the environment temperature exceeds its T_g. But high T_g might result in a higher residue stress inside the underfill after the material cools down after curing, which leads to early crack in the underfill.

5.6.1 Approaches of Incorporating Silica Fillers into No-Flow Underfill

The previous research has shown that the correlation between the material properties and package reliability in the case of flip chip underfill is very complicated. It is difficult to separate the effect of each factor since the material properties are often correlated with each other. However, it is generally agreed that low CTE and high modulus are favorable for high interconnect reliability [62]. Hence, the inclusion of silica fillers into the underfill is critical to enhance the reliability. However, since the underfill is pre-deposited on the board before the chip assembly in a no-flow process, the fillers are easily trapped in between the solder bump and contact pad and hinder the interconnection [63]. Thermocompression reflow (TCR) has been used to exclude the silica filler from solder joint [64]. The process step is illustrated in Fig. 5.20. In a TCR process, the underfill is dispensed on to a preheated substrate. The chip is then picked and bonded to the substrate and held at an elevated temperature under force for a certain period of time for solder joint formation. The assembly is post-cured afterwards. It was found that the bonding force and temperature were important factors influencing yield. A detailed study was carried out by Kawamoto et al. at NAMICS Corporation to determine the effect of filler on the solder joint connection in a TCR-like no-flow underfill process [65]. The study used two different sizes of silica fillers at different loading levels. It was found that good solder connection can be made with underfill with up to 60 wt.% filler loading without filler surface treatment. Smaller filler tended to increase the viscosity of the underfill and more fillers remained at the solder interface due to larger number of fillers at the same

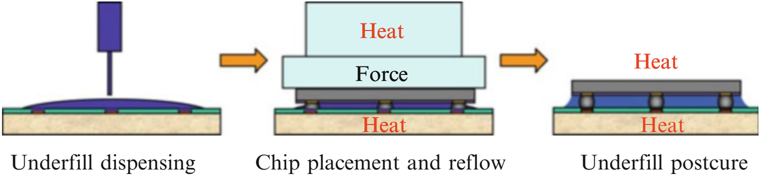


Fig. 5.20 Thermocompression reflow for flip chip

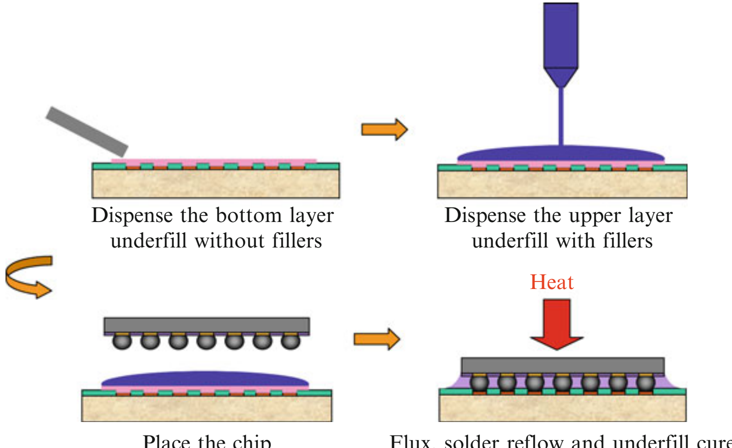


Fig. 5.21 Double-layer no-flow underfill process

weight percentage loading. The study also found that proper surface treatment of the fillers can lower the underfill viscosity and increase yield at high filler loading.

Other approaches have been explored to incorporate silica fillers into no-flow underfill. In a novel patented process, Zhang et al. used a double-layer no-flow underfill [66], in which two layers of no-flow underfill are applied. The bottom layer underfill is relatively high in viscosity and is not filled with silica fillers. It is applied onto the substrate first; then the upper layer underfill which is filled with silica fillers is dispensed. The chip is then placed onto the substrate and reflowed, during which the solder joints are formed and the underfill is cured or partially cured. The process flow chart is illustrated in Fig. 5.21. It was demonstrated that high yield was achieved using upper layer underfill of 65 wt.% silica filled [67]. Further investigation on the process indicated that factors affecting the interconnection yield of the double-layer no-flow underfill are complicated and interacting with each other [68]. The process window is narrow and the thickness and the viscosity of the bottom layer underfill are essential to the wetting of the solder bumps. And of course, it adds on another step in the flip chip process and has a higher process cost.

The recent advances in nano-science and nano-technology have enabled innovative research in materials for electronic packaging. It was found that nano-sized silica fillers with surface modification can be mixed with thermosetting resins to

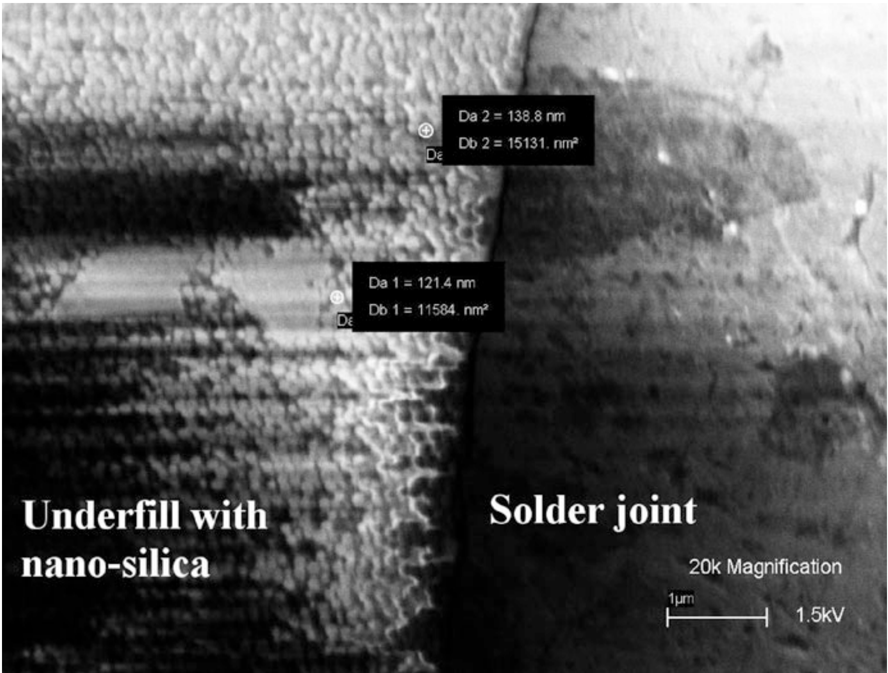


Fig. 5.22 SEM picture of a solder joint with nano-silica incorporated no-flow underfill

provide a uniform dispersion of non-agglomerated particles. Used as no-flow underfill, the nano-composite materials allowed 50 wt.% filler loading with good interconnect yield [69]. This high performance no-flow underfill developed by 3 M used 123 nm silica filler. With filler loading of 50 wt.%, the CTE of the material was 42 ppm/°C and the good interconnect yield was achieved using PB10 die (5×5 mm, 64 peripheral bumps). A joint research study was conducted by 3 M and Georgia Tech in the process and reliability evaluation of the nano-silica incorporated no-flow underfill [70]. Figure 5.22 shows a SEM picture of the solder joint in presence of no-flow underfill with nano-silica fillers. A 1.5× increase of characteristics life was observed in the air-to-air thermal cycling (AATC) reliability test with the nano-silica fillers. Although the nano-composite no-flow underfill material shows good potential for a highly reliable flip chip package using a SMT-friendly no-flow underfill process, the fundamental mechanism of the nano-silica interaction with the solder joints and the underfill is still not well understood. Since nano-size particles have a large surface area and tend to form irregular agglomerations which increase the difficult to be incorporated into a binder, surface treatment of nano-silica is of great importance in formulating an underfill. A fundamental study on the surface modification of nano-size silica for underfill application was carried out by Sun et al. [71]. They found that the type of the

surface treatment was the primary factor affecting the property of the formulation. Using an epoxy silane, the authors showed that the viscosity of the composite underfill was greatly reduced.

In summary, the invention of no-flow underfill greatly simplifies the flip chip underfill process and draws flip chip towards SMT. A successful no-flow underfill process requires careful investigation on the materials and process parameters. A lot of research efforts have been devoted to the materials, process, and reliability of flip chip no-flow underfill assembly. Since the underfill does not contain silica filler and hence behaves differently from the conventional underfill, the failure modes and reliability concerns are sometimes also different from the conventional flip chip underfill assembly. There are several ways to enhance the reliability of a flip chip no-flow underfill package. One way is to enhance the fracture toughness of the underfill without degrading other material properties to prevent underfill cracking in the thermal cycling. Or, low T_g and low modulus materials have been used to decrease the stress in the underfill. However, this approach diminishes the role of the underfill as a stress redistribution layer, and although it does decrease the stress in underfill, it cannot prevent solder joint fatigue failure from the thermomechanical stress, especially in the case of the large chip, high I/O counts and small pitch size applications. The other way is to add silica fillers into the underfill and match the properties of a conventional underfill. In order to overcome the difficulty of filler entrapment, different approaches have been explored. However, these approaches are less SMT transparent and diminish the low cost purpose of a no-flow underfill process. Nano-silica incorporated no-flow underfill showed potential of a highly reliable flip chip package with a SMT-friendly no-flow underfill process. However, fundamental understanding of the nano-silica and its interaction with underfill and solder is still lacking and further development is needed to optimize the materials and processes.

5.7 Molded Underfill

Epoxy molding compounds (EMCs) have been practiced in component packaging for a long time. The novel idea of combining over-molding and the underfill together results in a molded underfill [72, 73]. Molded underfill is applied to a flip chip in package via a transfer molding process, during which, the molding compound not only fills the gap between the chip and the substrate but also encapsulates the whole chip [74]. It offers the advantages of combining the underfilling and transfer molding into one step for reduced process time and improved mechanical stability [75]. It also utilizes EMCs which have long been proven to provide superior package reliability. Compared with the conventional underfill which is usually filled with silica at around 50–70 wt.%, molded underfill can afford a much higher filler content up to 80 wt.%, which offers a low CTE closely match with the solder joint and the board. Also, compared with the conventional molding compound, molded underfill requires fillers

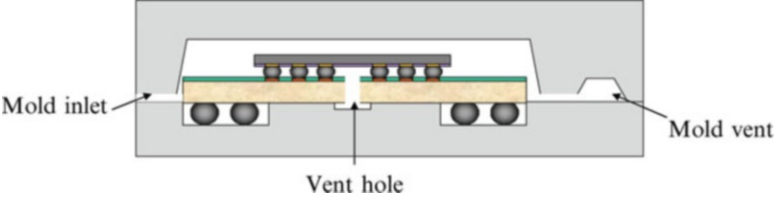


Fig. 5.23 Design of flip chip BGA with molded underfill

in smaller size, which also can contribution to lower the CTE of the material [76]. Molded underfill is especially suitable for flip chip in package to improve the production efficiency. It was reported that a fourfold production rate increase can be expected using molded underfill versus a conventional underfill process [77].

Molded underfill resembles the pressurized underfill encapsulation [78] in the mold design and process except that the materials in use are not liquid encapsulants that only fill up the gap between the chip and the substrate, but rather molding compounds that over-mold the entire components. Figure 5.23 shows a design of the mold for flip chip ball grid array (FCBGA) components using molded underfill.

The design of the mold faces the challenge that the flip chip geometry has a higher resistance to the mold flow so that air might be trapped under the chip. In fact, voids have been observed in the molded underfill packages using acoustic microscope [79]. Several molding processes can be used to minimize this geometry effect [80]. One way is to use mold vents as shown in Fig. 5.23 and to use also geometrical optimization to create similar flow resistance over and under the chip. One can also use vacuum assisted molding to prevent air entrapment. Another approach is to design a cavity in the substrate as shown in Fig. 5.23. Though it requires a special design on the substrate, this method has proved to be a robust process and is commonly adopted.

Important process parameters in a molded underfill process include the molding temperature, clamp force, and injection pressure [53]. High temperature molding is favored for lower viscosity of the molding compound and hence better flow properties and less stress on the solder joint. However, the upper limit of the molding temperature is the melting point (T_m) of the solder material. Temperature near T_m combined with high injection pressure might cause the solder to melt and even the die to be “swept” away from the site. Also low T_g substrate is likely to be damaged at high molding temperature and high clamp force. Flash is affected by both the clamp force and the injection pressure. The overflow of the molding compound might contaminate other contact pads or testing pads on the substrate. Bump cracking and die cracking are likely to occur as a result of high injection pressure. In short, a successful molded underfill process requires a combined effort in material selection, mold design, and process optimization. But the potential cost reduction and reliability enhancement of molded underfill is attracting great efforts in the industry.

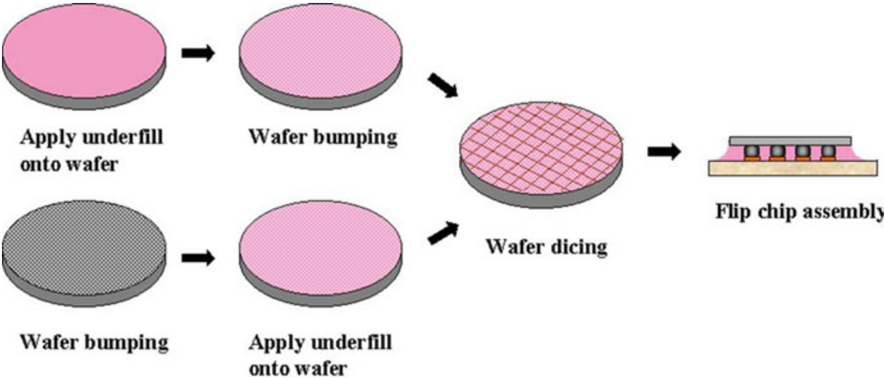


Fig. 5.24 Process steps of wafer level underfill

5.8 Wafer Level Underfill

The invention of no-flow underfill eliminates the capillary flow and combines fluxing, solder reflow and underfill curing into one step, which greatly simplifies the underfill process. However, as pointed out in the previous text, no-flow underfill has some inherent disadvantages including the unavailability of a heavily filled material, which is a big concern for high reliability packages. Also, no-flow process still needs individual underfill dispensing step and therefore is not totally transparent to standard SMT facilities. An improved concept, wafer level underfill, was proposed as a SMT-compatible flip chip process to achieve low cost and high reliability [81–84]. The schematic process steps are illustrated in Fig. 5.24. In this process, the underfill is applied either onto a bumped wafer or a wafer without solder bumps, using a proper method, such as printing or coating. Then the underfill is B-staged and wafer is diced into single chips. In the case of unbumped wafer, the wafer is bumped before dicing when the underfill can be used as a mask. The individual chips are then placed onto the substrate by standard SMT assembly equipment.

It is noted that in some types of WLCSP, a polymeric layer is also used on the wafer scale to redistribute the I/O and/or to enhance the reliability. However, this polymeric layer usually does not glue with the substrate and cannot be considered as underfill. The wafer level underfill discussed here is an adhesive to glue chip and substrate together and functions as a stress-redistribution layer rather than a stress-buffering layer. The attraction of the wafer level underfill lies in the potential low cost potential (since it does not require a significant change in the wafer back-end of the line process) and high reliability of the assembly enhanced with the underfill. However, the wafer level underfill faces critical material and process challenges including uniform underfill film deposition on the wafer, B-stage process for the underfill, dicing and storage of B-staged underfill, fluxing capability, shelf-life, solder wetting in presence of underfill, desire for no post-cure and reworkability, etc. Since the wafer level underfill process suggests a convergence of front-end and

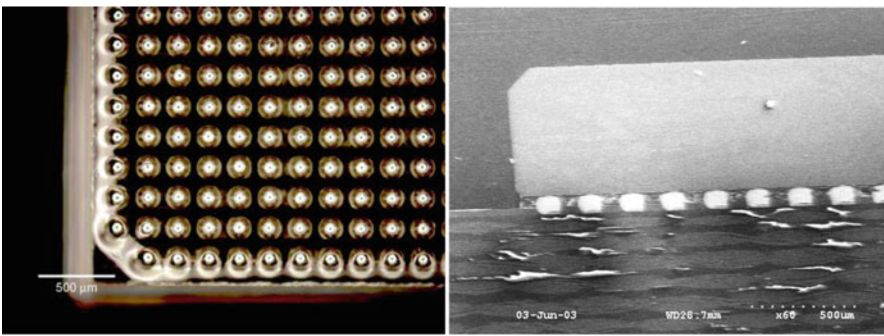


Fig. 5.25 200 μm pitch wafer level underfill process: underfill coated wafer (*left*) and assembly on the substrate (*right*)

back-end of the line in package manufacturing, close cooperation between chip manufacturers, package companies, and material suppliers are required. Several programs been carried out the cooperated research in this area [85–87]. Innovative ways of addressing the above issues and examples of wafer level processes are presented in this chapter.

In most wafer level underfill process, the applied underfill must be B-staged before the singulation of the wafer. The B-stage process usually involves partial curing, solvent evaporation, or both, of the underfill. In order to facilitate dicing, storage and handling, the B-staged underfill must appear solid-like and possess enough mechanical integrity and stability after B-stage. However, in the final assembly, the underfill is required to possess "reflowability," i.e., the ability to melt and flow to allow the solder bumps to wet the contacting pads and form solder joints. Therefore, the control of the curing process and the B-stage properties of the underfill is essential for a successful wafer level underfill process. A study conducted in Georgia Tech utilized the curing kinetics model to calculate the degree of cure (DOC) evolution of different underfills during solder reflow process [88]. Combined with the gelation behavior of the underfills, the solder wetting capability during reflow was predicted and confirmed experimentally. Based on the B-stage process window and the material properties of the B-staged underfill, a successful wafer level underfill material and process was developed. Full area array at 200 μm pitch flip chip assembly with the developed wafer level underfill was also demonstrated [89] as shown in Fig. 5.25.

The above study shows that the control of B-stage process of the wafer level underfill is critical to achieve good dicing and storage properties and the solder interconnect on board-level assembly. One way to avoid dicing in presence of non-fully cured underfill is presented in Fig. 5.26, a wafer scale applied reworkable fluxing underfill process developed by Motorola, Loctite and Auburn University [85]. Since uncured underfill materials are likely to absorb moisture that leads to potential voiding in the assembly, in this process, wafer is diced prior to underfill coating. Two dissimilar materials are applied; the flux layer coating by screen or stencil printing and the bulk underfill coating by a modified screen printing to keep

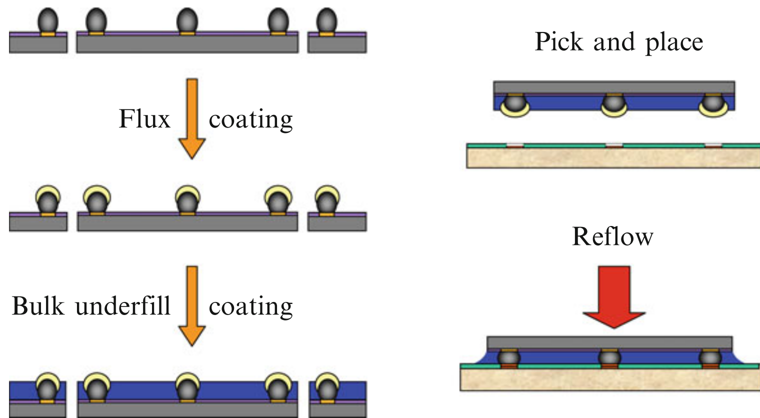


Fig. 5.26 A wafer scale applied reworkable fluxing underfill process

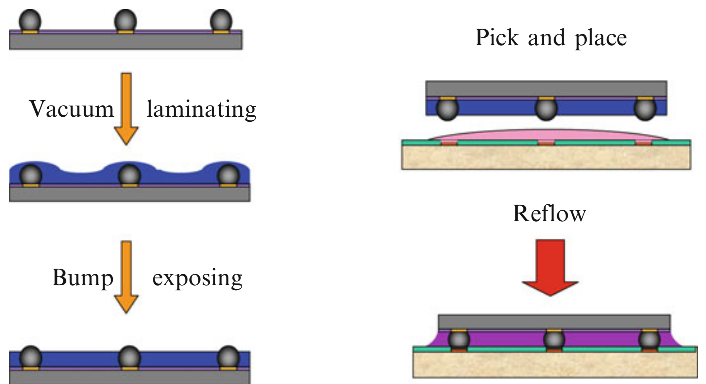


Fig. 5.27 A wafer-applied underfill film laminating process

the saw street clean. The separation of the flux from the bulk underfill material preserves the shelf life of the bulk underfill as well as prevents the deposition of fillers on top of the solder bump so as to ensure the solder joint interconnection in the flip chip assembly. In this process, no additional flux dispensing on board is needed and hence the underfill needs to be tacky in the flip chip bonding process to ensure the attachment of the chip to the board, as discussed in the previous text.

Underfill deposition on wafer using liquid material via coating or printing requires subsequent B-staging, which is often tricky and problematic. The process developed by 3 M and Delphi-Delco circumvents the B-stage step using film lamination [90]. The process steps are shown in Fig. 5.27, in which the solid film comprised of thermoset/thermoplastic composite is laminated onto the bumped wafer in vacuum. Heat is applied under vacuum to ensure the complete wetting of the film over the whole wafer and to exclude any voids. Then a proprietary process is carried out to expose the solder bump without altering the original solder

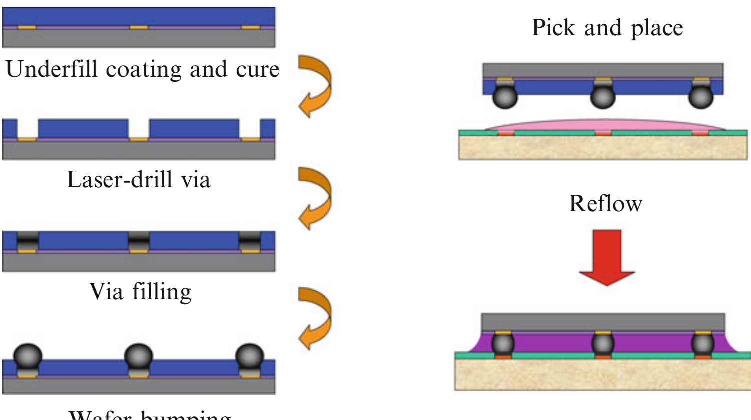


Fig. 5.28 A multilayer wafer-scale underfill process

shape. The subsequent flip chip assembly is carried out in a no-flow underfill-like process in which a curable flux adhesive is applied on the board and then the assembly is reflowed.

Wafer level underfill can also be applied before the bumping process. Figure 5.28 shows a multilayer wafer-scale underfill process developed by Aguila Technologies, Inc. [91]. The highly filled wafer level underfill is screen printed onto an unbumped wafer and then cured. Then this material is laser-ablated to form microvias that expose the bond pads. The vias are filled with solder paste and reflowed. Bumped are formed on top of the filled vias. The flip chip assembly is similar to no-flow underfill process again with a polymer flux dispensed onto the board before chip placement.

One similarity among all these three processes is the separation of flux material from the bulk underfill. Wafer level underfill process provides the convenience of separating different functionalities by using dissimilar materials so that "the one magic material that solves everything" is not required. However, it is likely to create inhomogeneity inside the underfill layer, the impact of which on the reliability is not fully understood.

A novel photo-definable material which acts both as a photoresist and an underfill layer applied on the wafer level was reported by Georgia Tech [92]. In the proposed process as shown in Fig. 5.29, the wafer level underfill is applied on the un-bumped wafer, and then is exposed to the UV light with a mask for cross-linking. After development, the un-exposed material is removed and the bump pads on the wafer are exposed for solder bumping. The fully cured film is left on the wafer and acts as the underfill during the subsequent SMT assembly after device singulation. A polymeric flux is needed during the assembly for holding the device in place on the board and providing fluxing capability, a process similar to the dry film laminated wafer level underfill. In order to enhance the material property, the addition of silica fillers is necessary. In this case, nano-sized silica was used to avoid UV light scattering which hinders the photo-cross-linking process. It also resulted in an optical transparent film

Fig. 5.29 A photo-definable wafer level underfill process

on the wafer to facilitate the vision recognition during dicing and assembly process. The photo-definable nano-composite wafer level underfill presents a cost-effective way of applying wafer level underfill and has potentially fine-pitch capability.

Since wafer level underfill is a relatively new concept and most researches are still in the process and material development stage, there are few reports on the reliability of a flip chip package using wafer level underfill. Although there is no standard process for wafer level underfill yet, the final decision might depend on the wafer and chip size, bump pitch, and package type, etc. Like wafer level CSP, multiple solutions can coexist for wafer level underfill process.

5.9 Summary

Flip chip offers many advantages over other interconnection technologies and is practiced in many applications. Underfill is necessary for a reliable flip chip on organic package but is process-unfriendly and becomes the bottle-neck of a high production flip chip assembly. As silicon technology moves to nanometer nodes with feature size less than 65 nm, shrinking pitch and gap distance, as well as new development in lead-free solder and low-K ILD/Cu interconnect present new challenges to underfill materials and process. Many variation of the conventional underfill have been invented to address the problem, among which, the newly developed no-flow underfill, molded underfill, and wafer level underfill have

attracted much attention. The no-flow underfill process simplifies the conventional flip chip underfill process by integrating flux into the underfill, eliminating capillary flow, and combining solder reflow and underfill cure into one step. However, the pre-deposited underfill cannot contain high level of silica filler due to the interference of filler with solder joint formation. The resulting high CTE of the underfill limits the reliability of the package. Various ways have been explored to enhance the reliability through improved fracture toughness of the underfill material, low T_g and low modulus underfill, and the incorporation of fillers using other process approaches. Molded underfill combines underfill and over-mold together and is especially suitable for flip chip in package to improve the capillary underfill flow and the production efficiency. Careful material selection, mold design and process optimization are required to achieve a robust molded underfill process. Wafer level underfill presents a convergence of front-end and back-end in packaging manufacturing and may provide a solution for low cost and high reliable flip chip process. Various material and process issues including underfill deposition, wafer dicing with underfill, shelf-life, vision recognition, chip placement, solder wetting with underfill, etc. have been addressed through novel material development and different process approaches. Although the research is still in the early stage and there is no standard in the process yet, there have been considerable successes in demonstrating the process, which looks promising for the future packaging manufacturing. All of these three approaches require close cooperation between the material suppliers, package designers, assembly companies, and maybe chip manufacturers. A good understanding in both the materials and the processes and their interrelationship is essential to achieve successful packages.

References

1. Wong CP, Lou S, Zhang Z (2000) Flip the chip. Science 290:2269
2. Davis E, Harding W, Schwartz R, Coring J (1964) Solid logic technology: versatile high performance microelectronics. IBM J Res Dev 8:102
3. Nakano F, Soga T, Amagi S (1987) Resin-insertion effect on thermal cycle resistivity of flip-chip mounted LSI devices. In: The proceedings of the international society of hybrid microelectronics conference, p 536
4. Tsukada Y (1992) Surface laminar circuit and flip-chip attach packaging. In: Proceedings of the 42nd electronic components and technology conference, p 22
5. Han B, Guo Y (1995) Thermal deformation analysis of various electronic packaging products by Moire and microscope Moire interferometry. J Electronic Packaging 117:185
6. Han S, Wang KK (1997) Analysis of the flow of encapsulant during underfill encapsulation of flip-chips. IEEE Trans Compon Packaging Manuf Technol Part B 20(4):424–433
7. Han S, Wang KK, Cho SY (1996) Experimental and analytical study on the flow of encapsulant during underfill encapsulation of flip-chips. In: Proceedings of the 46th electronic components and technology conference, pp 327–334
8. Nguyen L, Quentin C, Fine P, Cobb B, Bayyuk S, Yang H, Bidstrup-Allen SA (1999) Underfill of flip-chip on laminates: simulation and validation. IEEE Trans Compon Packaging Manuf Technol 22(2):168–176

9. Young WB, Yang WL (2006) Underfill of flip-chip: the effect of contact angle and solder bump arrangement. IEEE Trans Adv Packaging 29(3):647–653
10. Rodriguez F (1996) Principles of polymer systems, Chapter 3: physical states and transitions. Taylor & Francis, Washington, DC
11. Ward IM, Hadley DW (1993) An introduction to the mechanical properties of solid polymers, Chapter 4: principles of linear viscoelasticity. John Wiley & Sons Ltd, New York
12. Tobolsky AV (1960) Properties and structure of polymers. Wiley, New York
13. Bressers H, Beris P, Caers J, Wondergerm J (1996) Influence of chemistry and processing of flip-chip underfills on reliability. In: 2nd international conference on adhesive joining and coating technology in electronics manufacturing, Stockholm Sweden
14. Nysaether JB, Lundstrom P, Liu J (1998) Measurements of solder bump lifetime as a function of underfill material properties. IEEE Trans Compon Packaging Manuf Technol Part A 21(2):281–287
15. Dudek R, Schubert A, Michel B (2000) Analyses of flip-chip attach reliability. In: Proceedings of 4th international conference on adhesive joining and coating technology in electronics manufacturing, pp 77–85
16. Palaniappan P, Selman P, Baldwin D, Wu J, Wong CP (1998) Correlation of flip-chip underfill process parameters and material properties with in-process stress generation. In: Proceedings of the 48th electronic components and technology conference, pp 838–847
17. Mercado L, Sarihan V (2003) Evaluation of die edge cracking in flip-chip PBGA packages. IEEE Trans Compon Packaging Technol 26 (4):719–723
18. van Vroonhoven JCW (1993) ASME Trans J Electronic Packaging 115:28–32
19. Luo S, Wong CP (2001) IEEE Trans Compon Packaging Technol 24:38–42
20. Lam DCC, Yang F, Tong P (1999) IEEE Trans Compon Packaging Technol 22:215–220
21. Wong EH, Chan KC, Rajoo R, Lim TB (2000) In: Proceedings of the 50th electronic components and technology conference, pp 576–580
22. Luo S, Wong CP (2004) J Adhes Sci Technol 18(2):275
23. Luo S, Leisen J, Wong CP (2001) In: Proceedings of the 51st electronic components and technology conference, pp 149
24. Andrews EH (1991). In: Lee LH (ed) Adhesive bonding, Plenum Press, New York
25. Mittal KL (ed) (1992) Silanes and other coupling agents. VSP, Utrecht
26. Mittal KL (ed) (2000) Silanes and other coupling agents, vol 2. VSP, Utrecht
27. Welsh DJ, Pearson RA, Luo S, Wong CP (2001) In: Proceedings of the 51st electronic components and technology conference, pp 1502
28. Chen T, Wang J, Lu D (2004) Emerging challenges of underfill for flip-chip application. In: Proceedings of the 54th electronic components and technology conference, pp 175–179
29. Hwang JS (2000) Lead-free solder: the Sn/Ag/Cu system. Surface Mount Technology, p 18
30. Huang B, Lee NC (1999) Prospect of lead free alternatives for reflow soldering. In: Proceedings of SPIE—the international society for optical engineering, vol 3906, p 771
31. Butterfield A, Visintainer V, Goudarzi V (2000) Lead-free solder paste flux evaluation and implementation in personal communication devices. In: Proceedings of the 50th electronic components and technology conference, p 1420
32. Mahalingam S, Goray K, Joshi A (2004) Design of underfill materials for lead free flip-chip application. In: Proceedings of 2004 I.E. international society conference on thermal phenomena, pp 473–479
33. Chee CK, Chin YT, Sterrett T, He Y, Sow HP, Manepali R, Chandran D (2002) Lead-free compatible underfill materials for flip-chip application. In: Proceedings of the 52nd electronic components and technology conference, pp 417–424
34. Tsao P, Huang C, Li M, Su B, Tsai N (2004) Underfill characterization for low-K dielectric/Cu interconnect IC flip-chip package reliability. In: Proceedings of the 54th IEEE electronic components and technology conference, pp 767–769
35. Rajagopalan S, Desai k, Todd M, Carson G (2004) Underfill for low-K silicon technology. In: Proceedings of 2004 IEEE/SEMI international electronics manufacturing technology symposium

36. Pennisi R, Papageorge M (1992) Adhesive and encapsulant material with fluxing properties. US Patent 5,128,746, 7 July 1992
37. Wong CP, Baldwin D (1996) No-flow underfill for flip-chip packages. US Patent Disclosure, April 1996
38. Wong CP, Shi SH (2001) No-flow underfill of epoxy resin, anhydride, fluxing agent and surfactant. US Patent 6,180,696, 30 Jan 2001
39. Wong CP, Shi SH, Jefferson G (1997) High performance no flow underfills for low-cost flip-chip applications. In: Proceedings of the 47th electronic components and technology conference, pp 850
40. Wong CP, Shi SH, Jefferson G (1998) High performance no-flow underfills for flip-chip applications: material characterization. IEEE Trans Compon Packaging Manuf Technol Part A 21(3):450
41. Zhang Z, Shi SH, Wong CP (2000) Development of no-flow underfill materials for lead-free bumped flip-chip applications. IEEE Trans Compon Packaging Manuf Technol 24(1):59–66
42. Zhang Z, Wong CP (2000) Development of no-flow underfill for lead-free bumped flip-chip assemblies. In: Proceedings of electronics packaging technology conference, Singapore, pp 234–240
43. Zhang Z, Wong CP (2002) Study and modeling of the curing behavior of no-flow underfill. In: Proceedings of the 8th international symposium and exhibition on advanced packaging materials processes, properties and interfaces, Stone Mountain, Georgia, pp 194–200
44. Morganelli P, Wheelock B (2001) Viscosity of a no-flow underfill during reflow and its relationship to solder wetting. In: Proceedings of the 51st electronic components and technology conference, pp 163–166
45. Johnson RW, Capote MA, Chu S, Zhou L, Gao B (1998) Reflow-curable polymer fluxes for flip-chip encapsulation. In: Proceedings of international conference on multichip modules and high density packaging, pp 41–46
46. Shi SH, Wong CP (1998) Study of the fluxing agent effects on the properties of no-flow underfill materials for flip-chip applications. In: Proceedings of the 48th electronic components and technology conference, pp 117
47. Shi SH, Lu D, Wong CP (1999) Study on the relationship between the surface composition of copper pads and no-flow underfill fluxing capability. In: Proceedings of the 5th international symposium on advanced packaging materials: processes, properties and interfaces, pp 325
48. Shi SH, Wong CP (1999) Study of the fluxing agent effects on the properties of no-flow underfill materials for flip-chip applications. IEEE Trans Compon Packaging Technol Part A 22(2):141
49. Palm P, Puhakka K, Maattanen J, Heimonen T, Tuominen A (2000) Applicability of no-flow fluxing encapsulants and flip-chip technology in volume production. In: Proceedings of the 4th international conference on adhesive joining and coating technology in electronics manufacturing, pp 163–167
50. Puhakka K, Kivilahti JK (1998) High density flip-chip interconnections produced with in-situ underfills and compatible solder coatings. In: Proceedings of the 3rd international conference on adhesives joining and coating technology in electronics manufacturing, pp 96–100
51. Wang T, Chew TH, Chew YX, Louis Foo (2001) Reliability studies of flip-chip package with reflowable underfill. In: Proceedings of the pan pacific microelectronic symposium, Kauai, Hawaii, 2001, pp 65–70
52. Zhang Z, Wong CP (2002) Assembly of lead-free bumped flip-chip with no-flow underfills. IEEE Trans Compon Packaging Manuf Technol 25(2):113–119
53. Miller D, Baldwin DF (2001) Effects of substrate design on underfill voiding using the low cost, high throughput flip-chip assembly process. Proceedings of the 7th international symposium on advanced packaging materials: processes, properties and interfaces, pp 51–56
54. Zhao R, Johnson RW, Jones G, Yaeger E, Konarski M, Krug P, Crane L (2002) Processing of fluxing underfills for flip-chip-on-laminate assembly. In: Presented at IPC SMEMA Council APEX 2002, Proceeding of APEX, San Diego, CA, pp S18-1-1–S18-1-7

55. Wang T, Lum C, Kee J, Chew TH, Miao P, Foo L, Lin C (2000) Studies on a reflowable underfill for flip-chip application. In: Proceedings of the 50th electronic components and technology conference, pp 323–329
56. Gamota D, Melton CM (1997) The development of reflowable materials systems to integrate the reflow and underfill dispensing processes for DCA/FCOB assembly. IEEE Trans Compon Packaging Manuf Technol Part C 20(3):183
57. Dai X, Brillhart MV, Roesch M, Ho PS (2000) Adhesion and toughening mechanisms at underfill interfaces for flip-chip-on-organic-substrate packaging. IEEE Trans Compon Packaging Technol 23(1):117–127
58. Smith BS, Thorpe R, Baldwin DF (2000) A reliability and failure mode analysis of no flow underfill materials for low cost flip-chip assembly. In: Proceedings of 50th electronic components and technology conference, pp 1719–1730
59. Moon KS, Fan L, Wong CP (2001) Study on the effect of toughening of no-flow underfill on fillet cracking. In: Proceedings of the 51st electronic components and technology conference, pp 167–173
60. Wang H, Tomaso T (2000) Novel single pass reflow encapsulant for flip-chip application. In: Proceedings of the 6th international symposium on advanced packaging materials: process, properties, and interfaces, pp 97–101
61. Zhang Z, Fan L, Wong CP (2002) Development of environmental friendly non-anhydride no-flow underfills. IEEE Trans Compon Packaging Technol 25(1):140–147
62. Shi SH, Yao Q, Qu J, Wong CP (2000) Study on the correlation of flip-chip reliability with mechanical properties of no-flow underfill materials. In: Proceedings of the 6th international symposium on advanced packaging materials: processes, properties and interfaces, pp 271–277
63. Shi SH, Wong CP (1999) Recent advances in the development of no-flow underfill encapsulants—a practical approach towards the actual manufacturing application. In: Proceedings of the 49th electronic components and technology conference, p 770
64. Miao P, Chew Y, Wang T, Foo L (2001) Flip-chip assembly development via modified reflowable underfill process. In: Proceedings of the 51st electronic components and technology conference, pp 174–180
65. Kawamoto S, Suzuki O, Abe Y (2006) The effect of filler on the solder connection for no-flow underfill. In: Proceedings of the 56th electronic components and technology conference, pp 479–484
66. Zhang Z, Lu J, Wong CP (2001) Provisional Patent 60/288,246: "a novel process approach to incorporate silica filler into no-flow underfill"
67. Zhang Z, Lu J, Wong CP (2001) A novel approach for incorporating silica fillers into no-flow underfill. In: Proceedings of the 51st electronic components and technology conference, pp 310–316
68. Zhang Z, Wong CP (2002) Novel filled no-flow underfill materials and process. In: Proceedings of the 8th international symposium and exhibition on advanced packaging materials processes, properties and interfaces, pp 201–209
69. Gross KM, Hackett S, Larkey DG, Scheultz MJ, Thompson W (2002) New materials for high performance no-flow underfill. In: Symposium proceedings of IMAPS 2002, Denvor
70. Gross K, Hackett S, Schultz W, Thompson W, Zhang Z, Fan L, Wong CP (2003) Nanocomposite underfills for flip-chip application. In: Proceedings of the 53rd electronic components and technology conference, pp 951–956
71. Sun Y, Zhang Z, Wong CP (2004) Fundamental research on surface modification of nano-size silica for underfill applications. In: Proceedings of the 54th electronic components and technology conference, pp 754–760
72. Weber PO (2000) Chip package with molded underfill. US Patent 6,038,136, 14 Mar 2000
73. Weber PO (2000) Chip package with transfer mold underfill. US Patent 6, 157,086, 5 Dec 2000
74. Gilleo K, Cotterman B, Chen T (2000) Molded underfill for flip-chip in package. High density interconnection, p 28

75. Braun T, Becker KF, Koch M, Bader V, Aschenbrenner R, Reichl H (2002) Flip-chip molding—recent progress in flip-chip encapsulation. In: Proceedings of 8th international advanced packaging materials symposium, pp 151–159
76. Liu F, Wang YP, Chai K, Her TD (2001) Characterization of molded underfill material for flip-chip ball grid array packages. In: Proceedings of the 51st electronic components and technology conference, pp 288–292
77. Rector LP, Gong S, Miles TR, Gaffney K (2000) Transfer molding encapsulation of flip-chip array packages. In: IMAPS proceedings, pp 760–766
78. Han S, Wang KK (1999) Study on the pressurized underfill encapsulation of flip-chips. IEEE Trans Compon Packaging Manuf Technol Part B 20(4):434–442
79. Rector LP, Gong S, Gaffney K (2001) On the performance of epoxy molding compounds for flip-chip transfer molding encapsulation. In: Proceedings of the 51st electronic components and technology conference, pp 293–297
80. Becker KF, Braun T, Koch M, Ansorge F, Aschenbrenner R, Reichl H (2001) Advanced flip-chip encapsulation: transfer molding process for simultaneous underfilling and postencapsulation. In: Proceedings of the 1st international IEEE conference on polymers and adhesives in microelectronics and photonics, pp 130–139
81. Shi SH, Yamashita T, Wong CP (1999) Development of the wafer-level compressive-flow underfill process and its required materials. In: Proceedings of the 49th electronic components and technology conference, p 961
82. Shi SH, Yamashita T, Wong CP (1999) Development of the wafer-level compressive-flow underfill encapsulant. In: Proceedings of the 5th international symposium on advanced packaging materials: processes, properties and interfaces, p 337
83. Gilleo K, Blumel D (1999) Transforming flip-chip into CSP with reworkable wafer-level underfill. In: Proceedings of the Pan Pacific Microelectronics Symposium, p 159
84. Gilleo K (1999) Flip-chip with integrated flux, mask and underfill. WO Patent 99/56312, 4 Nov 1999
85. Qi J, Kulkarni P, Yala N, Danvir J, Chason M, Johnson RW, Zhao R, Crane L, Konarski M, Yaeger E, Torres A, Tishkoff R, Krug P (2002) Assembly of flip-chips utilizing wafer applied underfill. In: Presented at IPC SMEMA Council APEX 2002, Proceedings of APEX, San Diego, CA, pp S18-3-1–S18-3-7
86. Tong Q, Ma B, Zhang E, Savoca A, Nguyen L, Quentin C, Lou S, Li H, Fan L, Wong CP (2000) Recent advances on a wafer-level flip-chip packaging process. In: Proceedings of the 50th electronic components and technology conference, pp 101–106
87. Charles S, Kropp M, Kinney R, Hackett S, Zenner R, Li FB, Mader R, Hogerton P, Chaudhuri A, Stepniak F, Walsh M (2001) Pre-applied underfill adhesives for flip-chip attachment. In: IMAPS proceedings, international symposium on microelectronics, Baltimore, MD, pp 178–183
88. Zhang Z, Sun Y, Fan L, Wong CP (2004) Study on B-stage properties of wafer level underfill. J Adhes Sci Technol 18(3):361–380
89. Zhang Z, Sun Y, Fan L, Doraiswami R, Wong CP (2003) Development of wafer level underfill material and process. In: Proceedings of 5th electronic packaging technology conference, Singapore, pp 194–198
90. Zenner RLD, Carpenter BS (2002) Wafer-applied underfill film laminating. In: Proceedings of the 8th international symposium on advanced packaging materials, pp 317–325
91. BurressRV, Capote MA, Lee Y-J, Lenos HA, Zamora JF (2001) A practical, flip-chip multilayer pre-encapsulation technology for wafer-scale underfill. In: Proceedings of the 51st electronic components and technology conference, pp 777–781
92. Sun Y, Zhang Z, Wong CP (2005) Photo-definable nanocomposite for wafer level packaging. In: Proceedings of the 55th electronic components and technology conference, p 179

Chapter 6
Conductive Adhesives for Flip-Chip Applications

Daoqiang Daniel Lu and C.P. Wong

Abstract Significant progress has been made to improve electrically conductive adhesive (ECA) and nonconductive adhesive (NCA) technology. Recent material development of various anisotropic conductive adhesives/films (ACAs/ACFs) and their applications in flip chip are reviewed first. Then research achievements in material development and in electrical and mechanical aspects of isotropic conductive adhesives (ICAs), and their applications in flip chip and advanced packages are reviewed in details. In addition, latest advances of NCA technology for flip-chip applications are also reviewed.

6.1 Introduction

Electrically conductive adhesive (ECA) technology is one of the lead-free alternatives to soldering technology. ECAs can offer numerous advantages such as fewer processing steps which reduce processing cost, lower processing temperature which makes the use of heat-sensitive and low-cost substrates possible, and fine pitch capability [1]. There are two main types of conductive adhesives, anisotropic conductive adhesives (ACAs) and isotropic conductive adhesives (ICAs) [2–8]. All of the

D.D. Lu (✉)
Henkel Corporation, Shanghai, China
e-mail: daniel.lu@cn.henkel.com

C.P. Wong
Dean of Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China

on leave from:
Georgia Institute of Technology, School of Materials Science & Engineering,
771 Ferst Drive, Atlanta, GA 30332-0245, USA
e-mail: cp.wong@mse.gatech.edu

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_6, © Springer Science+Business Media New York 2013

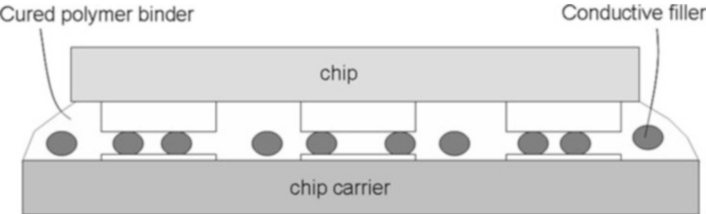


Fig. 6.1 ACA joins formed between two components (chip and chip carrier)

electrically conductive adhesives consist of a polymer binder and a conductive filler. Nonconductive adhesives consist of polymeric matrix without any electrically conductive fillers.

6.2 Anisotropically Conductive Adhesives/Films

6.2.1 Overview of ACAs/ACFs

Anisotropically conductive adhesives (ACAs) provide unidirectional electrical conductivity in the vertical or *Z*-axis. This directional conductivity is achieved by using a relatively low volume loading of conductive filler (5–20 volume percent) [3, 9, 10]. The low volume loading is insufficient for inter-particle contact and prevents conductivity in the *X*–*Y* plane of the adhesive. The ACA, in film or paste form, is interposed between the two surfaces to be connected. Application of heat and pressure to this stack-up causes conductive particles to be trapped between opposing conductor surfaces on the two components. Once electrical continuity is achieved, the polymer binder is hardened by chemical reaction (thermosets) or by cooling (thermoplastics). The hardened polymer binder holds the two components together and helps maintain the contact pressure between component surfaces and conductive particles. Figure 6.1 shows ACA joints formed between a chip and chip carrier. Anisotropically conductive adhesives have been developed for use in electrical interconnection, and various designs, formulations, and processes have been patented in Europe, Japan, and the USA [3].

6.2.2 Categories

Broadly, ACAs fall into two categories: those that are anisotropically conductive before processing and those whose anisotropy arises as a result of processing. Their characteristics can be summarized as follows: (a) Preprocessing anisotropy results from materials characterized by an ordered system of conductor elements interspersed in an adhesive matrix film. They are always in the form of tape or sheet and are complicated to manufacture, requiring an adhesive film to be laser-drilled or etched then filled with conducting materials. They provide predictable

contacts and are typically applied to a substrate as preforms. (b) Post-processing anisotropy results from materials that are a homogeneous mix of conductive fillers and adhesive matrix and which have no internal structure or order prior to processing. All adhesive pastes and some tapes fall into this category.

6.2.3 Adhesive Matrix

The adhesive matrix is used to form a mechanical bond at an interconnection. Both thermosetting and thermoplastic materials are used. Thermoplastic adhesives are rigid materials at temperatures below the glass transition temperature (T_g) of a polymer. Above the T_g, polymers exhibit flow characteristics. Thus, the T_g must be sufficiently high to avoid polymer flow during the application conditions, but the T_g must be low enough to prevent thermal damage to the associated chip carrier and devices during assembly. The principal advantage of thermoplastic adhesives is the relative ease with which interconnections can be disassembled for repair operations [11, 12]. However, thermoplastic ACAs suffer from many disadvantages. One of the most serious issues is that adhesion is not sufficient to hold the conductive particles in position, causing the contact resistance to increase after thermal shocks [11, 12]. Moreover, a phenomenon called "spring back" increases the contact resistance while the adhesive layer recovers from the stress caused by pressing of an ACA onto the components during bonding. This phenomenon, a creep characteristic exhibited by thermoplastic elastomers, occurs much after an ACA film has been heated to create the electrical joints. The contact resistance sometimes increases to more than three times the initial resistance during spring back (i.e., unloading) [11].

Thermosetting adhesives, such epoxies and silicones, form a three-dimensional cross-linked structure when cured under specific conditions. Cure techniques include: heat, UV-light, and added catalysts. As a result of the cure reaction that is irreversible, the initial uncross-linked material is transformed into a rigid solid. The thermosetting ACAs are stable at high temperatures and, more importantly, provide a low contact resistance. This results from a compressive force that maintains the conductive particles in intimate contact after the cure. That is, the shrinking caused by the cure reaction achieves a low contact resistance with long time stability. The ability to maintain strength at high temperature and robust adhesive bonds are the principal advantages of these materials. However, because the cure reaction is not reversible, rework or repair of interconnections is not an option [11, 12]. The choice of adhesive matrix and its formulation is critical to the long-term life properties of a composite. In practice, many options exist for the adhesive matrix. Acrylics can be used in low-temperature applications (under 100 °C), while epoxies are more robust and can be used at higher temperatures (up to 200 °C). Polyimide is used in the harshest environments where the temperature approaches 300 °C [3].

6.2.4 Conductive Fillers

6.2.4.1 Solid Metal Particles

Conductive fillers are used to provide the adhesive with electrical conductivity. The simplest fillers are metal particles such as gold, silver, nickel, indium, copper, chromium, and lead-free solders (SnBi) [3, 11, 13–15]. The particles are usually spherical and range 3–15 μm in size for ACA applications [16]. Needles or whiskers are also quoted in some patents [3].

6.2.4.2 Non-metal Particles with Metal Coating

Some ACA systems employ nonconductive particles with a thin metal coat. The core material is either plastic or glass with a metal coating consisting of gold, silver, nickel, aluminum, or chromium. The basic particle shape of these systems is also spherical. Plastic-cored particles deform when compressed between opposing contact surfaces, thus provide a large contact area. Polystyrene (PS) is often selected as the core material because the coefficient of thermal expansion of metal-coated PS beads is very close thermoset adhesives. The combination of epoxy resin and metal-plated PS beads results in a large improvement in thermal stability [11]. In addition, glass can also be selected as the core material. Glass-cored particles coated with metal lead to a controlled bond-line thickness because the glass core is not deformable. Since the conductive particle size is known, the conductivity of the joint can be predicted.

6.2.4.3 Metal Particles with Insulating Coating

To achieve fine pitch connections, metal spheres or metal-coated plastic spheres coated with insulating resin fillers were developed. The insulating resin layer is only broken under pressure to expose the underlying conductive surfaces, referred to as a microcapsule filler (MCF). Higher filler loading can be achieved with MCFs for fine pitch applications to avoid creating electrical short circuit conditions between printed circuit features [11, 16]. A typical cross section of an ACA interconnection with microcapsule filler material is illustrated in Fig. 6.2.

6.2.5 Flip-Chip Applications Using ACAs/ACFs

In traditional flip-chip packages, solder bumps provide electrical connections between a chip and chip carrier. To achieve high reliability, organic underfill materials are usually required to fill the gap between the chip and chip carrier. The cured underfill creates a monolithic structure that evenly distributes the stress

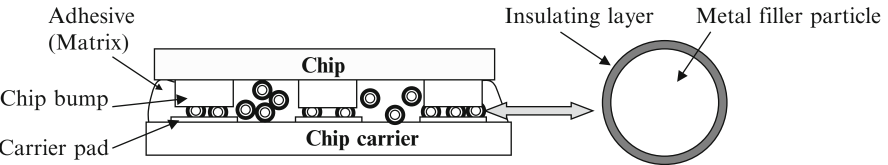


Fig. 6.2 Schematic depicting the cross section of an interconnection using a MCF filled ACA

over all the material in the gap, not just on the solder connections. In the past several years, much research has been conducted to develop flip-chip packages using ACAs in place of solder bumps. The primary advantages of ACA over lead-bearing solder for flip chips include ACAs' fine pitch capability, lead-free, low processing temperature, absence of flux residue, and generally lower cost. Also, ACA flip-chip technology does not require an additional underfilling process because the ACA resin acts as an underfill.

ACA flip-chip technology has been employed in many applications where flip chips are bonded to rigid chip carriers [17]. This includes bare chip assembly of ASICs in transistor radios, personal digital assistants (PDAs), sensor chips in digital cameras, and memory chips in laptop computers. In all the applications, the common feature is that ACA flip-chip technology is used to assemble bare chips where the pitch is extremely fine, normally less than 120 μm. For those fine applications, it is apparent that the use of ACA flip-chips instead of soldering is more cost-effective.

ACA flip-chip bonding exhibits better reliability on flexible chip carriers because the ability of flex provides compliance to relieve stresses. For example, the internal stress generated during resin curing can be absorbed by the deformation of the chip carrier. ACA joint stress analysis conducted by Wu et al. indicated that the residual stress is larger on rigid substrates than on flexible substrates after bonding [18].

6.2.5.1 ACA Flip Chip for Bumped Dies

Two Filler Systems

Y. Kishimoto et al. reported [19] anisotropic conductive adhesive pastes using two different fillers: Au-coated rubber particles (soft) and nickel particles (hard). The ACAs were used to bond a flip chip with Au plated bumps to a board with copper metallization. With the application of pressure, the soft particles were brought into contact with surface pads and were deformed which lowered this contact resistance. The hard particles, however, deformed the bumps and pads, thus were also in intimate contact with the surfaces to help reduce this contact resistance. The study showed that their choice of both hard and soft fillers in ACA materials had similar voltage–current behavior, and both exhibited stable contact resistance values after 1,000 cycles of thermal cycling and 1,200 h of 85 °C/85%RH aging conditions [19].

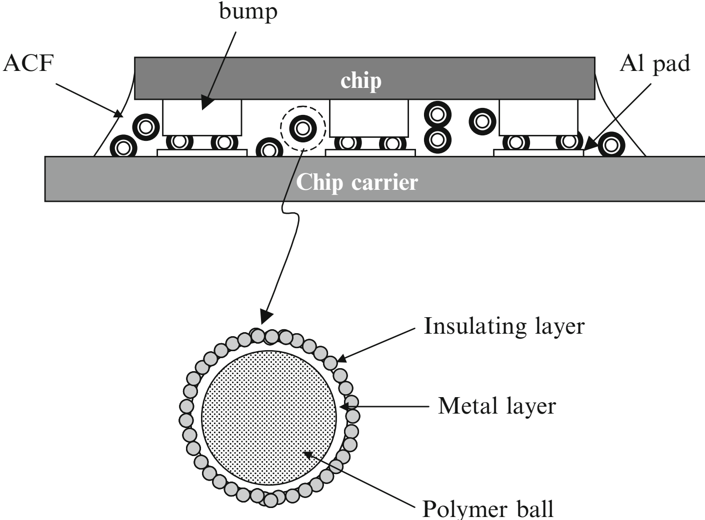


Fig. 6.3 Schematic depicting Casio's ACF technology—microconnector

Coated Plastic Filler

Casio developed an advanced anisotropic conductive adhesive film called the Microconnector (Fig. 6.3) [20–22]. This adhesive contains conductive particles made by coating plastic spheres with a thin layer of metal, followed by an additional 10 nm-thick layer of insulating polymeric material. The insulating layer consists of a large number of insulating micro powder particles that electrically insulate the outer surface of the spheres. The thin insulation layer is formed by causing insulating micro powder particles to adhere to the surface of the metal layer via electrostatic attraction. The base adhesive resin is thermoplastic or thermosetting, producing compressive force when cured. When heat and pressure are applied during bonding, the insulating layer, which is in contact with the bump surface of an IC, is broken. However, the insulating layer remains intact on conductive particles not crushed by the bonding pads, thereby producing only *Z*-axis electrical interconnections and preventing lateral short circuit conditions. With an additional insulating layer, a fine pitch and low contact resistance can be achieved without the risk of lateral short-circuiting by increasing the filler percentage (i.e., amount of particles per unit volume base adhesive resin or film). Casio is manufacturing pocket TV's with a liquid crystal using this material [22].

Solder Filler Systems

Unlike most commercial ACAs, where the electrical conductivity is based on the degree of mechanical contact achieved by pressing conductive particles to contact

pads on board and chip bumps, solder-filled ACAs establish microscopic metallurgical interconnections. The advantage of these joints is that the metallurgical bonds that are established prevent electrical discontinuities from occurring should the adhesive polymeric matrix undergo relaxation during the operational lifetime. Therefore, solder-filled ACAs combine the benefits of both soldering and adhesive joining resulting in more reliable ACA joints. Furthermore, better electrical performance is achieved due to lower contact resistance established through the metallurgical bonds [23].

Joints made with SnBi-filled and Bi-filled ACAs experience brittle intermetallic compound formation and have problems with typical conductor and coating materials such as copper, nickel, gold, and palladium [24]. Bi and SnBi are, however, compatible with tin, lead, zinc, and aluminum. Because Zn and Al are easily oxidized, only Sn and Pb are suitable surface finish materials for SnBi- and Bi-filled ACA applications. High quality interconnections were formed by metallurgically bonding SnPb-bumped chips on SnPb-coated substrates utilizing a Bi particle-filled ACA [25]. The joints once formed at a relatively low temperature could withstand a high temperature. The joint formation process is illustrated in Fig. 6.4. At the bonding temperature, 160 °C, as the Bi particles have locally penetrated thin oxide layers on both SnPb surfaces, the liquid lentil formation occurs immediately. After the Bi particles have dissolved completely into the liquid lentils between the solid SnPb bumps and coating, more Sn and Pb will dissolve into the liquid lentils until the liquid has reached its equilibrium composition at the bonding temperature. After solidification, the dissolved Bi will precipitate out as very fine particles from the saturated solution. Since the melting is transient, the remelting of the solid lentils will happen at higher temperatures than the first melting. The remelting point of the solid lentils can be controlled by the concentration of Bi present in the joint. The formed ACA joints exhibited a stable resistance after 2,000-h 85 °C/85%RH aging or after 1,000-h-temperature cycle testing (−40 to 125 °C). Even though this work is still preliminary, it demonstrates an interesting idea and concept. For lead-free applications, different materials such as pure Sn can be used for chip bumps and surface finish on the substrate [25].

Ni Filler

Toshiba Hino Works developed a flip-chip bonding technology using an anisotropically conductive film (ACF) filled with nickel spheres and LSI chips with gold ball bumps for mobile communications terminals. A resin sealing process at the sides of the LSI chip was added to improve mechanical strength. An FR-5 glass epoxy chip carrier was utilized to improve heat resistance. The assembled pager sets passed qualification consisting of drop, vibration, bending, torsion, and high temperature testing. The process has been demonstrated capable of mass production utilizing full automation of the flip-chip bonding method capable of producing 30,000 pager modules per month [26].

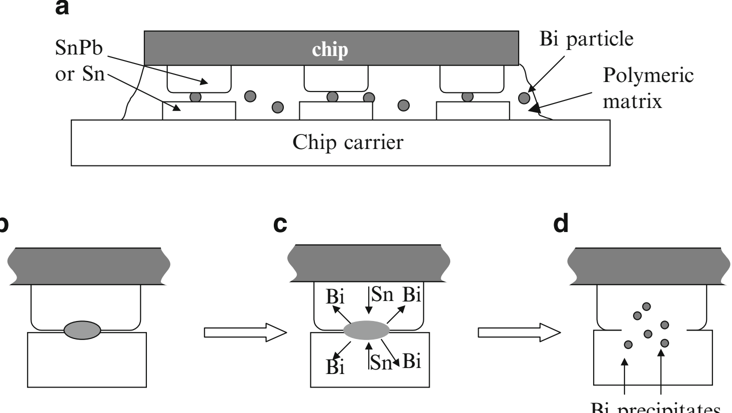


Fig. 6.4 Schematic illustration of the formation of electrical interconnects between a bumped chip and a mating carrier using a Bi-filled ACA. (**a**) The chip is aligned and placed on a chip carrier. (**b**) The Bi-particle is deformed between a chip bump and a carrier pad when a bonding pressure is applied. (**c**) The Bi-particle dissolves into the liquid lentils upon exposure of heat. (**d**) Bi diffuses into the Sn–Pb matrix and form fine solid precipitates

6.2.5.2 ACA Bumped Flip Chips on Glass Chip Carriers

ACAs are probably the most common approach for flip chip on glass applications. The ACA flip chip on glass technology not only provides assemblies with a higher interconnection density and a thinner and smaller size but also has fewer processes and lower costs as compared with TAB (tape automated bonding) technology. Also, bonding IC chips directly to the glass of the LCD panel using ACAs is a better choice when the pitch becomes less than 70–100 µm. Small size and high resolution LCDs such as viewfinders, video-game equipment displays, or light valves for liquid-crystal projectors use flip-chip on glass technology for the IC connections.

Selective Tacky Adhesive Method

Sharp developed a flip-chip bonding approach that utilizes ACA technology depicted in Fig. 6.5 [27, 28]. The novel feature of the Sharp technology is the method of attaching electrical conductive particles onto IC termination pads. This "bumping" procedure consists of coating the wafer with a 1–3 µm thick UV curable adhesive. Coated wafers are irradiated with UV light in a standard photolithographic process while the Al pads on the IC are optically masked. As a result of this process, the thin adhesive film above the Al pads remains uncured and tacky, whereas the adhesive on other chip areas is cured. Due to the tackiness of the adhesive on the Al pads, conductive particles only easily adhere to these sites. The conductive particles utilized by Sharp are gold-coated polymer spheres. UV-curable adhesive is dispensed on LSI chips before being aligned with a glass carrier. While still applying pressure to

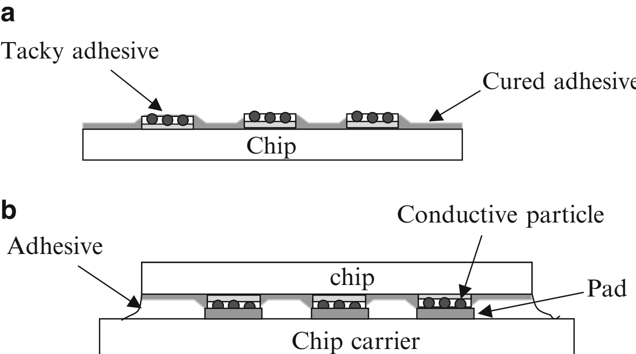


Fig. 6.5 Schematic depicting an ACA flip chip technology scheme utilized by Sharp. (**a**) Conductive particles adhere to uncured tacky adhesive on pad areas of a chip. (**b**) Chip remains in contact with the glass chip carrier since the adhesive exerts a compressive force after it is cured with light (UV)

maintain contact between the LSI chip and glass carrier, a light-setting adhesive is irradiated with UV light. Even upon releasing the pressure, the chip terminations remain electrically connected to their mating carrier pads. This is due to the deformed conductive particles which remain in contact with these termination pads as a result of the compressive force exerted by the cured adhesive. This process has several advantages, among them is that no bump plating is required; and the bonding process can be done by irradiating with UV light at room temperature, thus other materials are not damaged due to the effects of heat. This packaging concept can potentially achieve a high throughput.

The MAPLE Method

Seiko Epson Corporation developed a new flip-chip on glass technology called the “MAPLE” (Metal-Insulator-Metal Active Panel LSI Mount Engineering) method. The MAPLE method is to bond ICs directly to a glass panel substrate using a thermosetting anisotropic conductive film containing uniformly distributed conductive Au particles. While typical flip-chip on glass technologies require several alignment steps, this bonding process is very simple. First an ACA sheet is placed on a glass panel. After aligning the IC bumps with mating glass panel pads and temporarily bonding, proper IC interconnections are established by permanently bonding at high temperature and pressure. It is necessary for the bonding press-tool surface to be flat and parallel to the IC [29]. Comparing Metal-Insulator-Metal (MIM) panel modules made with TAB, MIM panel modules utilizing the MAPLE approach had smaller panel fringe size, thinner panel thickness, fewer assembled sides, fewer processes, and simpler module structure. The panel modules utilizing MAPLE passed all the required reliability tests. The MAPLE approach is being used in mass production of MIM panel modules.

6.2.5.3 ACA Bumped Flip Chips for High Frequency Applications

In many low frequency applications, conductive adhesive joining has proved to be a cost-effective and reliable solution. The high frequency behavior of ACA interconnections has attracted much attention in the past several years. The high frequency behavior of ACAs in flip-chip packages has been reported by several investigators. Rolf Sihlbom et al. demonstrated that ACA-bonded flip chips can provide performance equivalent to solder flip chips in the frequency range of 45 MHz to 2 GHz on FR4 chip carriers and 1–21 GHz on a high-frequency Telfon-based chip carrier. The different particle sizes and materials in the conductive adhesives gave little difference in high frequency behavior of ACA joints [30, 31].

Myung-Jin Yim et al. developed a microwave frequency model for ACF-based flip-chip joints based on microwave network analysis and S-parameter measurements. By using this model, high frequency behavior of ACF flip-chip interconnections with two filler particles, Ni and Au-coated polymer particles, was simulated. It was predicted that Au-coated polymer-particle-filled ACF flip-chip interconnections exhibited comparable transfer and loss characteristics to solder bumped flip chips up to about 13 GHz and thus they can be used for up to 13 GHz, but Ni-filled ACF joints can only be used for up to 8 GHz because the Ni particle has a higher inductance compared the Au-coated particle. Polymeric resins with a low dielectric constant and conductive particles with low inductance are desirable for high-resonance frequency applications [32].

6.2.5.4 ACA for Unbumped Flip Chips

Although ACAs are typically utilized with flip-chip bumped die, they are also used for unbumped flip chips in some cases. For unbumped flip chips, a pressure-engaged contact must be established by bringing the particles to the aluminum chip pads rather than a bump. The pressure must be sufficient to break the oxide on the aluminum pads. A sufficient quantity of particles must be trapped in the contact pad area and remain in place during bonding and curing to achieve a reliable interconnection. In addition to maximizing the number of particles in the contact area, the number of particles located between adjacent pads must be minimized to prevent electrical shorts. An additional factor that must be considered in the case of unbumped flip chips is adhesive flow during bonding and curing. It is essential to control the temperature heating rate to be sufficiently slow when the polymeric resin is curing so the conductive filler particles can migrate from the chip carrier side to the chip side pad [33].

Gold-Coated Nickel Filler

An application utilizing gold-coated nickel particles has been reported to provide reliable connection to unbumped flip chips [34]. Another study showed ACAs containing larger particles could accommodate planarity issues due to surface roughness, non-flat or nonparallel pads, compared to ACAs containing smaller particles. It was very difficult to obtain 100 % consistency in conduction with unbumped flip-chip dice using ACAs with small diameter balls [35].

Ni/Au Coated Silver Filler

A flip-chip technology developed by Toshiba Corporation utilized an ACF to attach bare unbumped chips (with Al pads) onto a PCB with bumps formed from a silver paste screen printed on the PCB [36]. After curing, Ag bumps were formed (70 μm diameter, 20 μm height) which were subsequently over plated with Ni/Au. It was determined that an ACF with a low CTE (28 ppm/ °C), low water absorption rate (1.3 %), and utilizing a Au-plated plastic ball worked best. It was also found that Ni/Au plated Ag paste-formed bumps exhibited a lower initial connection resistance and a lower connection resistance increase as compared to Ag paste-formed bumps which were not overplated with Ni/Au.

Conductive Columns

Nitto Denko Corporation developed an anisotropic conductive film for fine-pitch flip-chip applications [37]. The features of this ACF were as follows: (1) connectable between bumpless chips and a fine-pitch printed circuit board, (2) high electrical conductivity, (3) repairable (easy to peel off chips from a printed circuit board at elevated temperatures), (4) high reliability, and (5) can be stored at room temperature. The other notable features are the following: it is usable at pitches down to 25 μm; the conductive elements are micro-metallic columns as opposed to random-shaped particles; this adhesive matrix consists of a thermoplastic polymer resin; the conductive columns are coated with an insulator; and a high T_g polymer completely separates the columns from the adhesive.

It is easy to change the diameter of the conductive columns in order to make the film compatible with a variety of pitches. Sn/Pb or other solder materials are plated on both the top and bottom of the conductive columns (usually copper). The plated solder on the both ends of the conductive columns melts and forms metallurgical connections between the conductive columns and metal pads on a chip and the mating chip carrier, which ensures a good connection. Fig. 6.6a illustrates the cross section of the film structure. A rough surface, a result of plating, has the advantage of providing a good connection with the mating terminal pads. A typical terminal pad structure of a chip without bumps is shown in Fig. 6.6b. To achieve a good connection, the height of the conductive columns must be larger than the thickness

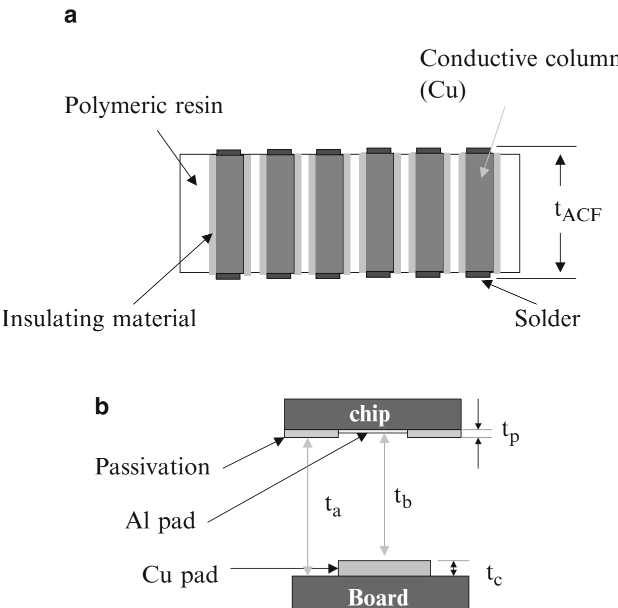


Fig. 6.6 Illustration of a scheme for fine-pitch, flip chip interconnection. (**a**) An anisotropic conductive film filled with micro metallic columns. (**b**) A typical cross-sectional structure of a chip without bumps and the mating chip carrier

of the passivation layer (t_p). Since t_b, the distance from a Cu pad surface of the chip carrier to the passivation layer surface of the chip is usually smaller than t_a (the distance from solder mask surface of the chip carrier to the passivation layer surface of the chip), the conductive columns will assume an inclined position during bonding if the thickness of the conductive columns is larger than the ACF thickness (t_{ACF}). It is important to adjust the thickness of board or chip carrier pads and ACF thickness to achieve good connection and adhesion. Reliability results indicate that the ACF possessing an adhesive matrix with a high T_g (282 °C) exhibits high reliability; the contact resistance remained unchanged after 1,000 cycles of accelerated thermal cycle testing (−25 to 125 °C).

6.2.6 Failure Mechanism of ACA/ACF Interconnections

Since the adhesive matrix is a nonconductive material, interconnection joints rely to some extent on pressure to assure contact for conventional ACAs. Adhesive interconnections therefore exhibit different failure mechanisms compared to soldered connections, where the formation of intermetallic compounds and coarsening of grains are associated with the main mechanisms. Basically there are two main

failure mechanisms that can affect the contacts. The first is the formation of an insulating film on either the contact areas or conductive particle surfaces. The second is the loss of mechanical contact between the conductive elements due to either a loss of adherence, or relaxation of the compressive force.

6.2.6.1 Oxidation of Non-noble Metals

Electrochemical corrosion of non-noble metal bumps, pads, and conductive particles results in the formation of insulating metal oxides and significant increase in contact resistance. Electrochemical corrosion only occurs in the presence of moisture and metals that possess different electrochemical potentials. Humidity generally accelerates oxide formation and so too the increase in contact resistance. Reliability test results for flip chip on flex (FCOF) using gold bumps and ACFs filled with Ni particles indicated that the connection resistance increased with time under elevated temperature and humidity storage conditions [38]. In this case, the gold bump acts as cathode and the Ni particle as an anode. A nickel oxide, which is electrically insulating, eventually forms on the surface of the Ni particles.

6.2.6.2 Loss of Compressive Force

The compressive forces acting to maintain contact among the conductive components are partly due to curing shrinkage achieved when curing the polymeric matrix of ACAs. Both the cohesive strength of the adhesive matrix and the interfacial adhesion strength between the adhesive matrix and the chip and chip carrier must be sufficient to maintain the compressive force. However, the thermal expansion of adhesives, their swelling due to moisture adsorption, and mechanical stresses due to applied loads tend to diminish this compressive force created as a result of curing. Moreover, water not only diffuses into the adhesive layer but also penetrates to the interface between adhesive and chip/chip carrier causing a reduction in adhesion strength. As a result, the contact resistance increases and can even result in a complete loss of electrical contact [39].

6.2.7 Recent Advances in Nano-ACAs/ACFs

6.2.7.1 Low Temperature Sintering of Nano-Ag-Filled ACAs/ACFs

One of the concerns for ACAs/ACFs is the higher joint resistance since interconnection using an ACA/ACF relies on mechanical contact, unlike the metallurgical bonding of soldering. An approach to minimize the joint resistance of an ACA/ACF is to make the conductive fillers fuse to each other and form metallic joints. However, to fuse metal fillers in polymers does not seem feasible, since a typical

organic printed circuit board (T_g ~125 °C), on which the metal filled polymer is applied, cannot withstand such a high temperature; the melting temperature (T_m) of Ag, for example, is around 960 °C. Research showed that the T_m and sintering temperatures of materials could be dramatically reduced by decreasing the size of the materials [40, 41]. It has been reported that the surface pre-melting and sintering processes are a primary mechanism for the T_m depression of the fine nanoparticles (<100 nm). For nano-sized particles, sintering could occur at much lower temperatures and, as such, the use of the fine metal particles in ACAs would be promising for fabricating high electrical performance ACA joints through eliminating the interface between metal fillers. The application of nano-sized particles can also increase the number of conductive particles on each bond pad and result in more contact area between the particles and bond pads. Fig. 6.7 shows SEM photographs of Ag nanoparticles annealed at various temperatures. Although very fine particles (20 nm) were observed for as synthesized (in Fig. 6.7a) and 100 °C treated particles (in Fig. 6.7b), dramatically larger particles were observed after heat treatment at 150 °C and above. With increasing temperature, the particles became larger and appeared as solid matter rather than porous particles or agglomerates. The particles shown in Fig. 6.7c–e were fused through their surface and many dumbbell type particles could be found. The morphology was similar to a typical morphology of the initial stage in the typical sintering process of ceramic, metal and polymer powders. This low temperature sintering behavior of the nanoparticles is attributed to the extremely high inter-diffusivity of the nanoparticle surface atoms, due to the significantly energetically unstable surface condition of the nano-sized particles with large proportion of the surface area to the entire particle volume.

For the sintering reaction in a material system, temperature and duration are the most important parameters, in particular, the sintering temperature. Current–resistance (I–R) relationship of ACAs filled with Ag nanoparticles is shown in Fig. 6.8. As can be seen from the figure, with increasing curing temperature, the resistance of the ACA joints decreased significantly, from 10^{-3} to 5×10^{-5} ohm. Also, higher curing temperature ACA samples exhibited higher current carrying capability than the low temperature samples. This phenomenon suggested that more sintering of Ag nanoparticles and subsequently improved interaction between the nanoparticles and metal bond pads were achieved at higher temperatures [42], yet the *x–y* direction of the ACA maintains an excellent dielectric property for electrical insulation.

6.2.7.2 Self-Assembled Monolayers for Nano-ACAs/ACFs

In order to enhance the electrical performance of ACA/ACF materials, self-assembled monolayers (SAMs) have been introduced into the interface between metal fillers and metal-finished bond pad of ACAs [43, 44]. These organic molecules adsorb to the metal surface and form physicochemical bonds, which allow electrons to flow. As such, they reduce electrical resistance and enable a high current flow.

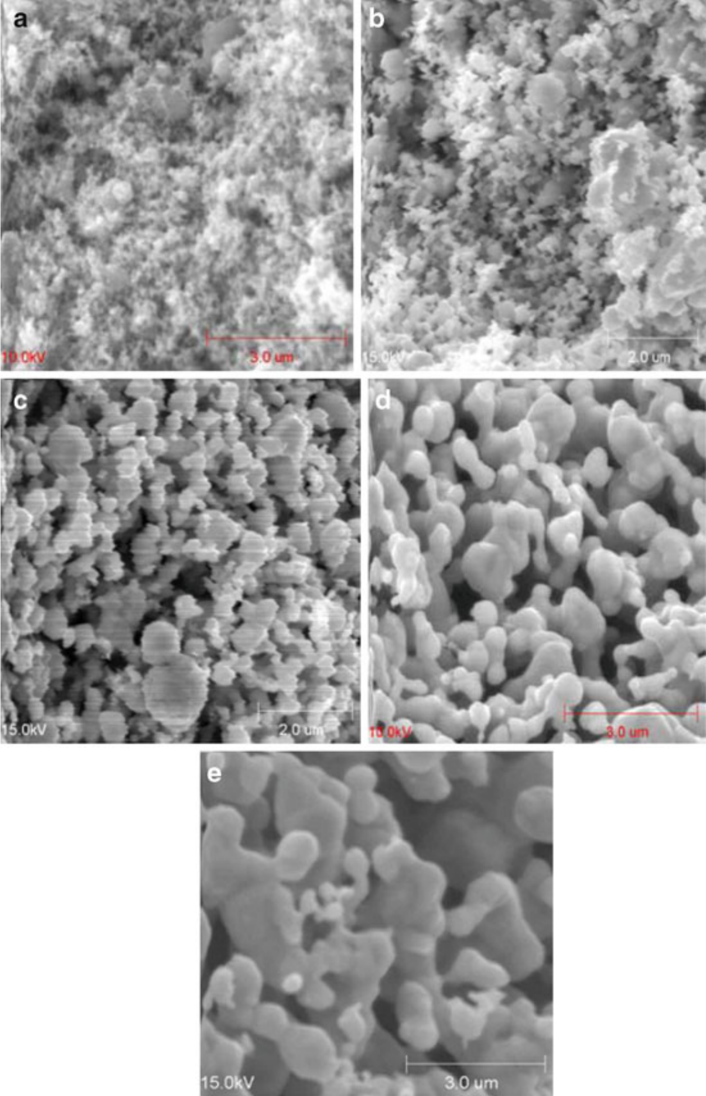


Fig. 6.7 SEM micrographs of 20 nm-sized Ag particles annealed at different temperatures for 30 min: (**a**) room temperature (no annealing); (**b**) annealed at 100 °C; (**c**) annealed at 150 °C; (**d**) annealed at 200 °C; and (**e**) annealed at 250 °C [40]

The unique electrical properties are due to the tuning of metal work-function by these organic monolayers. Metal surfaces can be chemically modified by the organic monoalyers and the reduced work-functions can be achieved by using suitable organic monolayer coatings. An important consideration when examining the

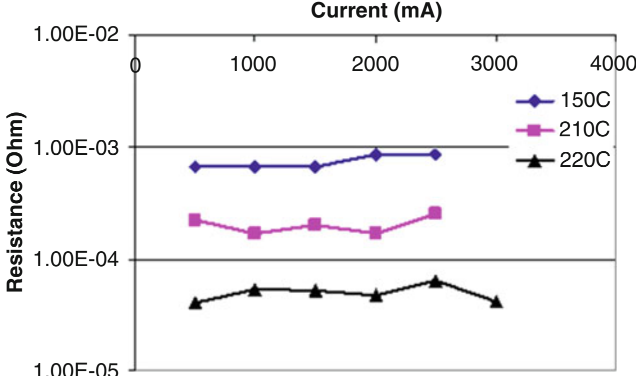


Fig. 6.8 Current–resistance (I–R) relationship of nano-Ag-filled ACAs filled with Ag nanoparticles at different curing temperatures [42]

advantages of organic monolayers pertains to the affinity of organic compounds to specific metal surfaces.

Different organic self-assembled monolayers (SAMs), e.g., dicarboxylic acids and dithiols, have been introduced into ACA/ACF joints. For ACAs with SAMs and micron-sized gold/polymer or gold/nickel fillers, lower joint resistance and higher maximum allowable current (highest current applied without inducing joint failure) were achieved for low temperature curable ACAs (<100 °C). For high curing temperature ACAs (150 °C), however, the improvement was not as significant as low curing temperature ACAs, due to the partial desorption/degradation of organic monolayer coating at the relatively high temperature [45]. However, when dicarboxylic acid or dithiol was introduced into the interface of nano-silver filled ACAs, significantly improved electrical properties could be achieved for a high temperature curable ACA/ACF, suggesting the coated molecular wires did not suffer degradation on silver nanoparticles at the curing temperature (Fig. 6.9). The enhanced bonding could be attributed to the larger surface area and higher surface energy of nanoparticles, which enabled the monolayers to be more readily coated and relatively thermally stable on the metal surfaces [46].

6.2.7.3 Silver Migration Control in Nano-Silver Filled ACAs

Silver is the most widely used conductive filler in ICAs and exhibits exciting potential in nano-ACAs/ACFs due to many unique advantageous properties of silver. Silver has the highest room temperature electrical and thermal conductivities among all the conductive metals. Silver is also unique among all the cost-effective metals because of its conductive oxide (Ag_2O). In addition, silver nanoparticles are relatively easily formed into different sizes (a few nanometers to 100 nm) and shapes (such as spheres, rods, wires, disks, flakes, etc.) and are well dispersed in a variety of polymeric matrix materials. Also the low temperature sintering and high surface

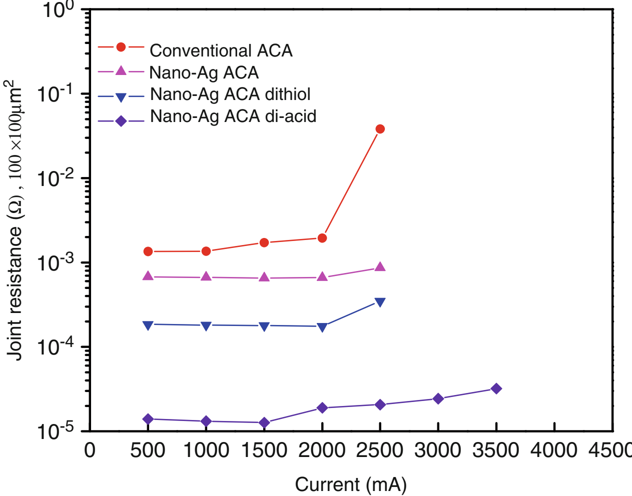


Fig. 6.9 Electrical properties of nano-Ag filled ACAs with dithiol or dicarboxylic acid [46]

energy makes silver one of the promising candidates as a conductive filler in nano-ACAs/ACFs. However, silver migration has long been a reliability concern in the electronics industry. Metal migration is an electrochemical process, whereby, metal (e.g., silver), in contact with an insulating material, in a humid environment and under an applied electric field, leaves its initial location in ionic form and deposits at another location [47]. It is considered that a threshold voltage exists above which the migration starts. Such migration may lead to a reduction in electrical spacing or cause a short circuit between interconnections. The migration process begins when a thin continuous film of water forms on an insulating material between oppositely charged electrodes. When a potential is applied across the electrodes, a chemical reaction takes place at the positively biased electrode where positive metal ions are formed. These ions, through ionic conduction, migrate toward the negatively charged cathode and, over time, they accumulate to form metallic dendrites. As the dendrite growth increases, a reduction of electrical spacing occurs. Eventually, the dendrite silver growth reaches the anode and creates a metal bridge between the electrodes, resulting in an electrical short circuit [48].

Although other metals may also migrate under specific environments, silver is more susceptible to migration, mainly due to the high solubility of silver ions, low activation energy for silver migration, high tendency to form dendrite shape and low possibility to form stable passivation oxide layer [49–51]. The rate of silver migration is increased by: (1) an increase in the applied potential; (2) an increase in the time of the applied potential; (3) an increase in the level of relative humidity; (4) an increase in the presence of ionic and hygroscopic contaminants on the surface of the substrate; and (5) a decrease in the distance between electrodes of opposite polarity.

In order to reduce silver migration and improve the reliability, several methods have been reported. These methods include the following: (1) alloying the silver with an anodically stable metal such as palladium [48] or platinum [52] or even tin [53]; (2) using hydrophobic coating over the printed wiring board (PWB) to shield its surface from humidity and ionic contamination, since water and contaminates can act as a transport medium and increase the rate of migration; (3) plating of silver with metals such as tin, nickel or gold, to protect the silver fillers and reduce migration; (4) coating the substrate with polymer [54]; (5) applying benzotriazole (BTA) and its derivatives in the polymeric matrix [55]; (6) employing siloxane epoxy polymers as diffusion barriers due to their excellent adhesion to conductive metals [56]; and (7) chelating silver fillers in ECAs with molecular monolayers [57]. As an example shown in Fig. 6.10 [58], with carboxylic acids forming chelating compounds with silver ions, the silver migration behavior (leakage current) could be significantly reduced and controlled.

6.2.7.4 ACF with Straight-Chain-Like Nickel Nanoparticles

Sumitomo Electric Industries (SEI) recently developed a new-concept ACF using nickel nanoparticles with a straight-chain-like structure as a conductive filler [59]. They applied the formulated straight-chain-like nickel nanoparticles and solvent in a mixture with epoxy resin on a film substrate. Then the particles were aligned toward the vertical direction of the film surface and fixed in the resin by evaporating the solvent. In the evaluation using 30 μm-pitch IC chips and glass substrates (the area of Au bumps was 2,000 μm^2, the distance between neighboring bumps was 10 μm), this new ACF showed excellent reliability of electrical connection after high-temperature, high-humidity (60°C/90%RH) test and thermal cycle test (between −40 and 85 °C). The samples were also exposed to high-temperature, high-humidity (60°C/90%RH) for surface insulation resistance (SIR) study. Although the distance between two electrodes was only 10 μm, ion-migration did not occur and insulation resistance was maintained at over 1 GΩ for 500 h. This result showed that the new ACF has superior insulation reliability. This indicates that this new ACF has potential for application in very fine interconnections.

6.2.7.5 Nanowire ACFs for Ultrafine Pitch Flip-Chip Interconnection

In order to satisfy the reduced I/O pitch and avoid electric shorting, a possible solution is to use high-aspect-ratio metal posts. Nanowires exhibit high possibilities due to the small size and extremely high aspect-ratio. In the literature, it is shown that nanowires could be applied in FET sensors for gas detection, magnetic hard-disks, nanoelectrodes for electrochemical sensors, thermal–electric devices for thermal dissipation and temperature control, etc. [60–62]. To prepare nanowires, a template for growing nanowires needs to be fabricated. Many expensive methods such as e-beam, X-ray or scanning probe lithographies have been used but micrometer length

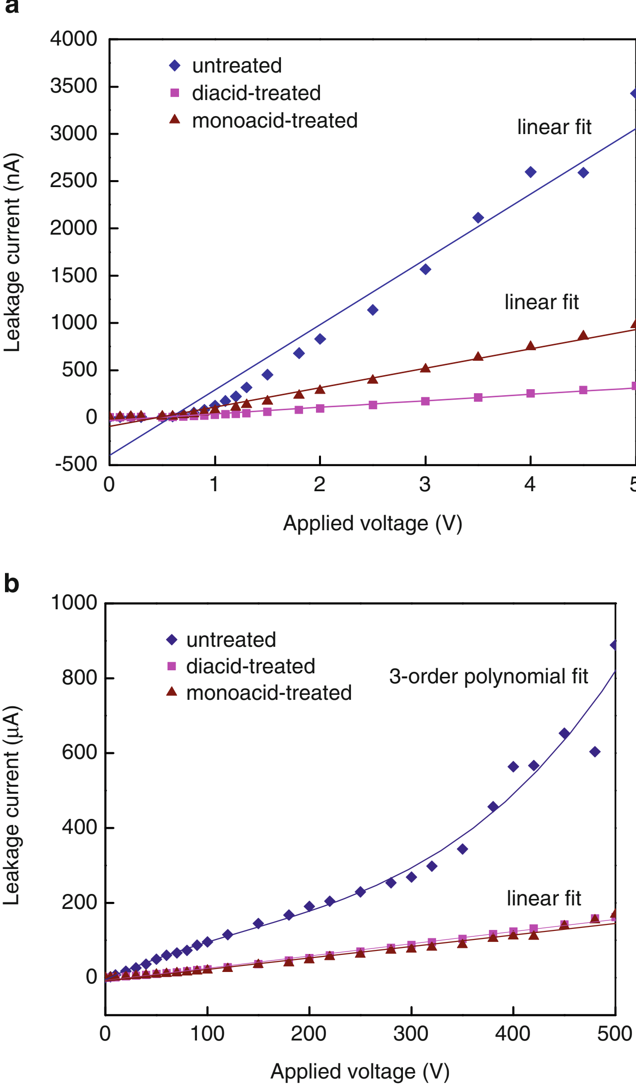


Fig. 6.10 Leakage current–voltage relationships of nano-Ag-filled conductive adhesives at (**a**) low voltages and (**b**) high voltages [58]

nanowires have not been obtained. Another less expensive alternative is electrodeposition of metal into nano-porous template such as anodic aluminum oxide (AAO) [63] or block-copolymer self-assembly template [64]. The disadvantages of a block-copolymer template include its small thickness (which means only short nanowires can be made from the template), nonuniform distribution and poor parallelism of nano-pores. However, AAO has benefits of higher thickness (>10 µm), uniform pore

size and density, larger size and very parallel pores. Lin et al. [65] developed a new ACF with nanowires. They used AAO templates to obtain silver and cobalt nanowire arrays by electrodeposition. And then a low viscosity polyimide (PI) was spread over and filled into the gaps of nanowires after surface treatment. The bi-metallic Ag/Co nanowires could stay parallel during fabrication by magnetic interaction between the cobalt and the applied magnetic field. The silver and cobalt nanowires/polyimide composite films could be obtained with nanowire diameter of about 200 nm and maximum film thickness up to 50 μm. The *X–Y* insulation resistance was about 4–6 GΩ and the *Z*-direction resistance including the trace resistance (3 mm length) was less than 0.2 Ohm.

6.2.7.6 In-Situ Formation of Nano-Conductive Fillers in ACAs/ACFs

One of the challenging issues in the formation of nano-filler ACAs/ACFs is the dispersion of nano-conductive fillers in the ACA/ACF. A considerable amount of research has been conducted in recent years to address the dispersion issue of nano-composite because nano-fillers tend to agglomerate. For the fine pitch electronic interconnects using nano-filler ACAs/ACFs, the dispersion issues need to be solved. The efforts usually include physical approaches such as sonication and chemical approaches such as use of surfactants. Recently, a novel ACA/ACF containing in situ formed conductive nanoparticles was proposed for the next generation high performance fine-pitch electronic packaging applications [66, 67]. This novel interconnect adhesive combines the electrical conduction along the z-direction (ACA-like) and the ultra fine pitch (<100 nm) capability. Instead of adding the conductive nanofillers in the resin, the nanoparticles can be in-situ formed during the curing/assembly process. By using in-situ formation of nanoparticles, during the polymer curing process, the filler concentration and dispersion actually could be better controlled and the drawback of surface oxidation of the nano-fillers could be easily overcome.

6.3 Isotropically Conductive Adhesives

6.3.1 Introduction

6.3.1.1 Percolation Theory of Conduction

Isotropic conductive adhesives are composites of polymer resin and conductive fillers. The conductive fillers provide the composite with electrical conductivity. With increasing filler concentrations, the electrical properties of an ICA transform it from an insulator to a conductor. Percolation theory has been used to explain the electrical properties of ICA composites. At low filler concentrations, the resistivities

of ICAs decrease gradually with increasing filler concentration. However, the resistivity drops dramatically above a critical filler concentration, V_c, called the percolation threshold. It is believed that at this concentration, all the conductive particles contact each other and form a three-dimensional network. The resistivity decreases only slightly with further increases in the filler concentrations [68–70]. A schematic explanation of resistivity change of ICAs based on percolation theory is shown in Fig. 6.11. In order to achieve conductivity, the volume fraction of conductive filler in an ICA must be equal to or higher than the critical volume fraction. A recent study from Agar et al. [71] suggested that the existence of a directly contacted percolated metallic network is infeasible because the strong repulsive forces at the interface will prohibit direct metallic interaction. Experiments on doped epoxies with varying dielectric constants indicated that the dielectric constant and bulk resistivity are directly correlated; implying that tunneling and hopping mechanism dominate electron transport between the metallic flakes.

Similar to solders, ICAs provide the dual functions of electrical connection and mechanical bond in an interconnection joint. In an ICA joint (Fig. 6.12), the polymer resin provides mechanical stability and the conductive filler provides electrical conductivity. Filler loading levels that are too high cause the mechanical integrity of adhesive joints to deteriorate. Therefore, the challenge in formulating an ICA is to maximize conductive filler content to achieve a high electrical conductivity without adversely affecting the mechanical properties. In a typical ICA formulation, the volume fraction of the conductive filler is about 25–30 % [9, 10].

6.3.1.2 Adhesive Matrix

An ideal polymeric matrix for ICAs should exhibit a long shelf life (good room temperature latency), fast cure, relatively high glass transition temperature (T_g), low moisture pickup, and good adhesion [72].

Both thermoplastic and thermoset resins can be used for ICA formulations. The main thermoplastic resin used for ICA formulations is polyimide resin. An attractive advantage of thermoplastic ICAs is that they are reworkable, e.g., can easily be repaired. A major drawback of thermoplastic ICAs, however, is the degradation of adhesion at high temperature. Another drawback of polyimide-based ICAs is that they generally contain solvents. During heating, voids are formed when the solvent evaporates. Most of commercial ICAs are based on thermosetting resins. Epoxy resins are most commonly used in thermoset ICA formulations because they possess superior balanced properties. Silicones, cyanate esters, and cyanoacrylates are also employed in ICA formulations [73–77].

Most commercial ICAs must be kept and shipped at a very low temperature, usually −40 °C, to prevent the ICAs from curing. Pot life is a very important factor for users of the ICAs. In order to achieve desirable latency at room temperature, epoxy hardeners must be carefully selected. In some commercial ICAs, solid curing agents are used, which do not dissolve in the epoxy resin at room temperature. However, these curing agents can dissolve in the epoxy at a higher temperature

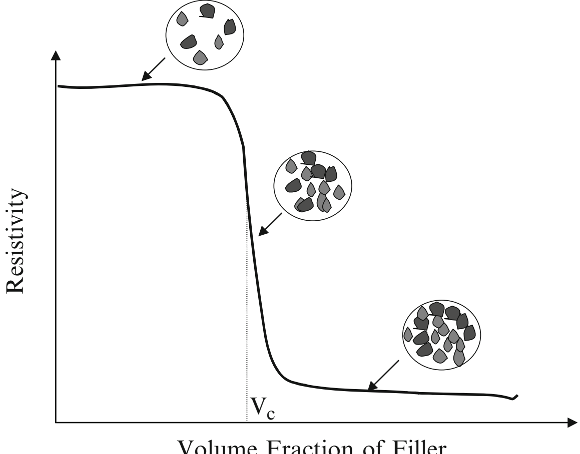


Fig. 6.11 Effect of filler volume fraction on the resistivity of ICA systems

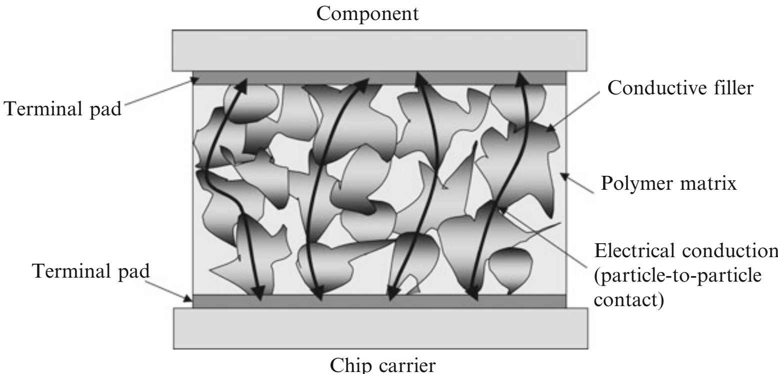


Fig. 6.12 Schematic illustration of how electrical conduction paths are established by uninterrupted particle-to-particle contact between the component and chip carrier terminal pads in an ICA joint

(curing temperature) and react with the epoxy resin. Another approach to achieve latency is to employ an encapsulated imidazole as a curing agent or catalyst. An imidazole is encapsulated inside a very fine polymer sphere. At room temperature, the polymer sphere does not dissolve or react with the epoxy resin. But at a higher temperature, after the polymer shell is broken, the imidazole is released from the sphere to cure the epoxy or catalyze the cure reaction. Fast cure is another attractive property of a desirable ICA. Shorter cure times increase throughput resulting in lower processing cost. In epoxy-based ICA formulations, proper hardeners and catalysts such as imidazoles and tertiary amines can be used to achieve rapid cure.

Conductive adhesives with low T_gs can lose electrical conductivity during thermal cycling aging [78, 79]. Electrical conductivity in metal-powder filled conductive adhesives is achieved through the contact of adjacent metal particles

Table 6.1 Effects of moisture on ICA joints

Degrade bulk mechanical strength
Decrease interfacial adhesion strength and cause delamination
Promote the growth of voids present in joints
Give rise to swelling stress in joints
Induce the formation of metal oxide layers resulting from corrosion

with each other, thus producing a continuous electrical path between a component lead and the metal pad of a chip carrier. When a joint is subjected to thermal cycling conditions, it experiences repeated cyclic shear motion of the lead relative to the chip carrier pad. The amount of shear strain is primarily dependent on the thermal cycling conditions and thermal expansion mismatch between the component and chip carrier. Besides lead-deformation and substrate compliance, the majority of the shear strain produced is accommodated by viscoelastic or visco-plastic deformation of the conductive adhesive. When a conductive adhesive deforms to accommodate the shear strain produced, the metal particles move, thus changing the position of contact point(s) between adjacent metal particles. If the organic matrix is too compliant, it will flow to fill the area left behind the moving metal particles. When the direction of the shear strain is reversed during thermal cycling, adjacent metal particles move back to their original contact locations, which now are partially covered with the compliant, dielectric organic-matrix material. As the number of thermal cycles increases, the contact resistance between adjacent particles increases, thus increasing the interconnection joint resistance [78].

Moisture absorption can influence the reliability of conductive adhesive interconnection joints. Moisture in polymer composites is known to have an adverse effect on both mechanical and electrical properties of epoxy laminates [79, 80]. Studies relating to the reliability and moisture sensitivity of electronic packages indicate similar degrading effects. It was determined that moisture absorption can cause an increase in contact resistance, especially if the metallization on the bond pads and components are not noble metals [81]. Effects of moisture absorption on conductive adhesive joints are summarized in Table 6.1. In order to achieve high reliability, conductive adhesives with low moisture absorption are required. High adhesion strength to pad and component metallization is a necessary property for conductive adhesives used for interconnections in electronic assemblies. Epoxy-based ICAs tend to have better adhesion strength than polyimide and silicone-based ICAs. However, a silicone matrix tends to have lower moisture absorption than epoxy resins [74].

6.3.1.3 Conductive Fillers

While polymer matrices are dielectric materials, conductive fillers in ICA formulations provide the material with electrical conductivity. In order to achieve

high conductivity, the filler concentration must be at least equal or higher than the critical concentration predicted by percolation theory.

Pure Silver vs. Ag-Coated Fillers

Silver (Ag) is by far the most popular conductive filler, although gold (Au), nickel (Ni), copper (Cu), and carbon are also used in ICA formulations. Silver has the highest room temperature electrical and thermal conductivity among all the metals. Silver is unique among all of the cost-effective metals by nature of its conductive oxide (Ag_2O). Oxides of most common metals are good electrical insulators and copper powder, for example, becomes a poor conductor after aging. Nickel and copper-based conductive adhesives generally do not have good conductivity stability because they are easily oxidized. Even with antioxidants, copper-based conductive adhesives show an increase in volume resistivity on aging, especially under high-temperature and humidity conditions. Silver-plated copper has been utilized commercially in conductive inks, and should also be appreciable as a filler in adhesives. While composites filled with pure silver particles often show improved electrical conductivity when exposed to elevated temperature and humidity or thermal cycling, this is not always the case with silver-plated metals, such as copper flake. Presumably, the application of heat and mechanical energy allows the particles to make more intimate contact in the case of pure silver, but silver-plated copper may have coating discontinuities that allow oxidation/corrosion of the underlying copper and thus reduce electrical paths [9].

Particle Shape and Size

The most common morphology of conductive fillers used for ICAs is flake because flakes tend to have a large surface area, and more contact spots and thus more electrical paths than spherical fillers. The particle size of ICA fillers generally ranges from 1 to 20 μm. Larger particles tend to provide the material with a higher electrical conductivity and lower viscosity [82]. A new class of silver particles, porous nano-sized silver particles, has been introduced in ICA formulations [83, 84]. ICAs made with this type of particles exhibited improved mechanical properties, but its electrical conductivity is lower than that of the ICAs filled with silver flakes. In addition, short carbon fibers have been used as conductive fillers in conductive adhesive formulations [85, 86]. However, carbon-based conductive adhesives show much lower electrical conductivity than silver-filled ones.

Nano-Sized Fillers

Recently, nano-sized fillers including nanowires, nanoparticles, graphenes, and carbon nanotubes (CNTs) have also been introduced into ICA formulations. Quite

some research work has been done especially on using silver-based nano-fillers such as Ag nanowires and nanoparticles for ICA applications. The simplest and the most commonly used bulk-solution synthetic method for Ag nanoparticles is the chemical reduction of metal salts (e.g., $AgNO_3$) [87, 91]. Ag nanowires are generally prepared using a physical template [92, 93] or a template-less wet chemical process [94, 95]. As an example, Fig. 6.13 shows SEM images of an Ag nanowire, an Ag nanoparticle, and a micron-sized Ag flake.

Silver-Copper Fillers

Copper can be a promising candidate for conductive filler metal, due to its low resistivity, low cost, and improved electromigration performance, but oxidation causes this metal to lose its conductivity. Two approaches have been reported for surface treatment and oxidation prevention of copper fillers for ICA application. One is inorganic material coating, and the other is organic material coating. For inorganic coating materials, silver, gold and nickel/gold and solder materials, such as Sn and InSn, are some examples which are coated by electro- or electroless deposition.

Yokoyama et al. reported a powder with a specific structure as a filler for conductive adhesives in 1992 [96]. The powder particle consists of two metallic components, copper and silver. Silver is highly concentrated on the particle surface and the concentration gradually decreases from the surface to the inner of the particle, but always contains a small amount of silver. Conductive adhesive paste filled with this powder exhibits excellent oxidation resistance, i.e., can be exposed to 100 ppm of oxygen content in a nitrogen atmosphere without oxidation. It also exhibits higher solderability than commercially available copper pastes, sufficient adhesion strength even after heating and/or cooling test, and the least migration, almost the same degree as pure copper paste [96].

Cu Fillers

Copper has been identified as a promising candidate for conductive filler due to its high electrical conductivity, low cost, and good electromigration performance. However, the major problem for copper fillers is the oxidation of copper surface and the deterioration of the electrical properties for ICAs. There are basic two surface treatment approaches to prevent the oxidation of the copper filler surface. One is inorganic coating and the other is organic coating.

For organic coating materials on copper surfaces for oxidation/corrosion protection, self-assembly monolayer (SAM) formation, such as azole or thiol compounds, and organosilicic compound formation are the representatives; however, their thermal stabilities have been the concerns, because most of the coatings lose their effectiveness when exposed to the curing condition of ICAs. Yim et al. [97] reported a Cu-filled ICA by using novel silane coupling agents for oxidation prevention of Cu fillers for conventional ICA applications. In their study,

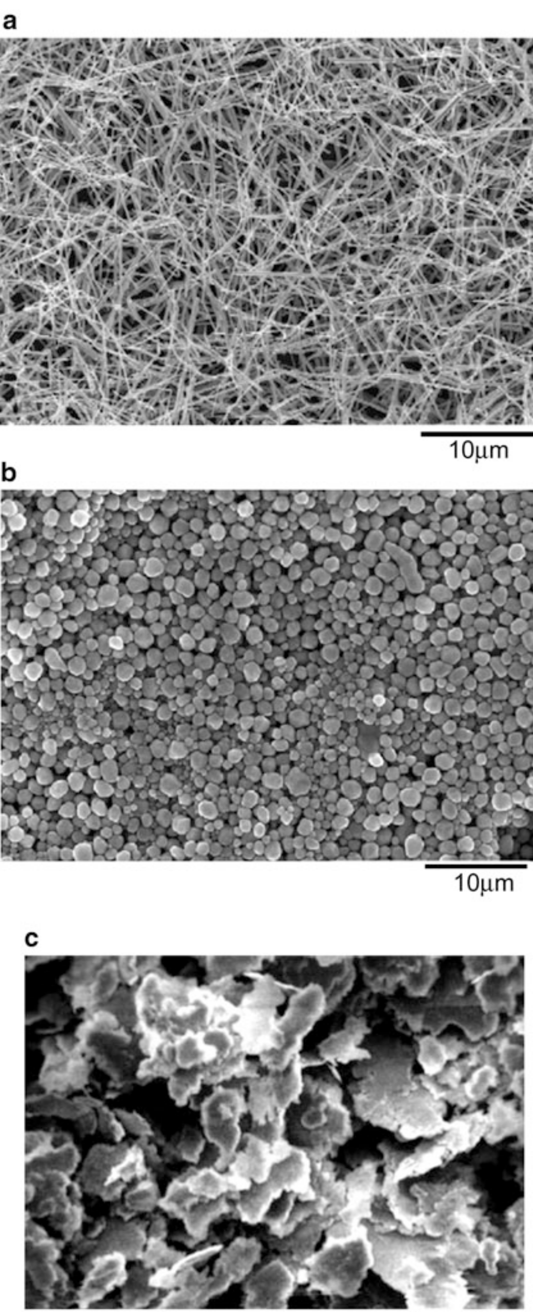


Fig. 6.13 SEM images of a Ag nanowire (**a**), a Ag nanoparticle (**b**), and a micron-sized Ag flake (**c**)

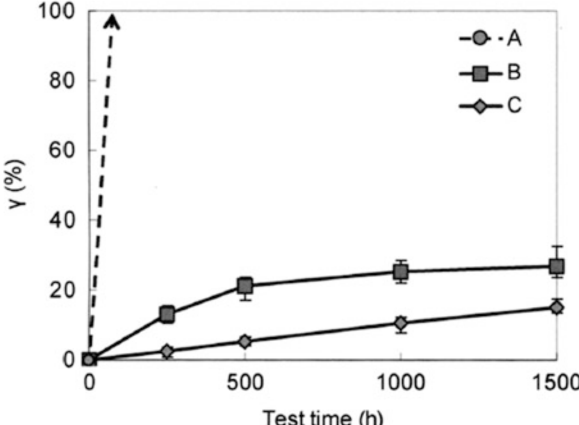


Fig. 6.14 Change rate of electrical resistance of conductive adhesives during 85 °C/85 % RH aging (A—ICA filled with spinous copper filler; B—ICA filled with spinous silver-coated copper filler; C—ICA filled with spherical silver-coated copper filler) [98]

a high-performance, low-cost, Cu-filled ICA was demonstrated through the use of silane coupling agent and the in situ mixing with a matrix resin. The silane coupling agent was shown to be very effective in preventing the copper powder oxidation in ICA, and the improved thermal stability of Cu-filled ICAs was achieved by using the silane coupling agents with aromatic structures. Bulk resistivity of a Cu filled ICA with a silane coupling agent was achieved at 1.28×10^{-3} Ohm cm. They also tried to use of bimodal sized fillers and an optimized concentration of same aromatic silane coupling agent, and a much lower bulk resistivity of 7.5×10^{-4} Ohm cm was obtained.

The other method for preventing copper oxidation is to coat the copper filler surface with silver [98]. Nishikawa et al. investigated the electrical conductivity of ICAs filled with silver-coated copper particles after curing and after reliability tests. It was found that the electrical resistance of ICA with silver-coated copper filler was lower than that of the ICA filled with copper filler. It was also observed that the silver-coated copper filled ICAs showed more stable electrical resistance after high temperature and high humidity aging, and spherical silver-coated copper filler provided ICAs with more stable electrical resistance than the spinous silver-coated copper filler (as shown in Fig. 6.14).

Low-Melt Fillers

In order to improve electrical and mechanical properties, low-melting-point alloy fillers have been used in ICA formulations. A conductive filler powder is coated with a low-melting-point metal. The conductive powder is selected from the group consisting of Au, Cu, Ag, Al, Pd, and Pt. The low-melting-point metal is selected from the group of fusible metals, such as Bi, In, Sn, Sb, and Zn. The filler particles are coated with the low-melting-point metal, which can be fused to achieve

metallurgical bonding between adjacent particles, or between the particles and the bond pads that are joined using the adhesive material [99, 100].

6.3.2 *Flip-Chip Applications Using ICAs*

A key factor in achieving a low-cost, flip-chip technology is the use of isotropic conductive adhesives. In comparison to the classical FC technologies, the use of ICAs for the bumping and joining provide numerous advantages (Table 6.2).

6.3.2.1 ICA Process for Unbumped Chips

Motorola successfully demonstrated an ICA flip-chip bumping process using stencil printing technology both through mathematical modeling and experimentation [101]. Both GaAs and Si flip-chip devices with Au thin-film metallization, and alumina and FR4 chip carriers also with Au metallization were used in this study. The electrical performance of chip and chip carrier combinations (i.e., GaAs/Al_2O_3, GaAs/FR4, and Si/FR4) utilizing conductive adhesive polymer bumps, showed no difference from Au and AuSn bumps (all of the flip-chip dies are mounted onto the chip carriers using an ICA). However, the ICA joints showed premature failure in HAST and thermal shock tests.

The polymer bumping method is a low-cost and efficient process conducted at the wafer-level and suitable for large-scale production. Data of joint resistance stability under accelerated aging conditions such as 85 °C/85 % relative humidity and temperature cycling demonstrates polymer flip-chip interconnections are capable of long-term stability. The polymer flip-chip (PFC) assembly is compatible to a large range of rigid carriers, and heat-sensitive, flexible chip carriers, a key advantage of PFC over solder FC technology. Currently PFC is widely used for flip-chip bonding on low-cost heat-sensitive chip carriers noted in Table 6.3 [102].

Flip Chip with Printed ICA Bumps

The PFC process is a stencil printing technology in which an ICA is printed through a metal stencil to form polymer bumps on bond pads of IC devices subsequent to the under bump metallization deposition on aluminum termination pads. The sequential processes to achieve PFC interconnects are UBM deposition, stencil printing an ICA, bump formation (ICA solidification), flip-chip attach to achieve electrical connections, and underfill for enhanced mechanical and environmental integrity [101, 103].

Table 6.2 Advantages of flip-chip technologies utilizing ICAs

Process simplification and reduction of indexing steps by eliminating activation and purification processes
A smaller temperature load on elements and wiring carriers
The availability of a large spectrum of material combinations
A broad range of applicable adhesive systems allows at the selection of processing parameters and joining characteristics
Few requirements for under bump metallization (UBM) since alloy phase formation does not have to be considered

Table 6.3 Low-cost, heat-sensitive chip carriers utilized in polymer flip-chip applications

Applications of polymer flip-chip bonding
Microcontroller chips on PET substrates
Transponder chips on PVC/ABS/PET/PC/PI foils for smart card inlays
Circuits on flexible systems
Controller and driver circuits on polyester base materials in combination with adhesive-bonded components
Circuits on rigid and multilevel substrates such as FR4 boards or BGA's in combination with SMD components
Temperature sensitive sensors and actuators on most diverse carriers and complex microsystems

UBM Deposition

As with virtually all flip-chip processes, the Al bond pads must be protected from the formation of nonconductive aluminum oxide. This insures a low and stable resistance at bond–bond pad interface. The polymer flip-chip process utilizes an electroless plating technique, Ni/Au or Pd, to cover the Al bond pads prior to polymer bumping. The typical metal thickness is 0.5–1.0 μm for Pd and 3.0–5.0 μm for Ni/Au.

ICA Printing

The PFC process combines high precision stencil printing techniques with highly conductive ICAs. These polymers can be thermosetting or thermoplastic. First, the polymer bumps are formed by deposition of an ICA through the metal mask directly onto the metallized bond pads on a wafer. Printed conductive adhesive bumps can offer an attractive alternative to the other bumping technologies in terms of low cost and manufacturability. The printing process typically involves a screen or stencil with openings through which bumps are deposited. During the printing process, the paste is typically dispensed some distance away from the stencil apertures. Typically, the stencil is separated from a substrate by the snap-off distance. The squeegee is lowered, resulting in contact of the stencil to the substrate or wafer surface. As the squeegee moves across the stencil surface, a stable flow pattern develops in the form of a paste roll. The consequent hydrodynamic pressure developed by the squeegee pushes the paste into the patterned stencil openings. The stencil lifts away from the substrate surface with the paste remaining on the substrate.

ICA Bump Curing

The polymer bumps are then either fully cured or partially cured to the so-called B-stage for thermosetting polymer bumps. For thermoplastic polymer bumps, after stencil printing the solvent is removed to form solid bumps. Bump heights are typically 50–75 μm and process can accommodate pitches down to 5 mils. Bump density of up to 80,000 bumps/wafer has been formed with excellent coplanarity.

Once the bumped wafers are diced, chips are picked from the wafers, flipped over, and then placed on and bonded to chip carriers. Different process procedures are utilized to bond thermosetting polymer bumps to similar thermoplastic bumps as noted in Fig. 6.15. Final processing involves a heat cure for thermosetting bumps, while thermoplastic bump connections only require a few seconds under heat and pressure to melt the thermoplastic.

Underfilling

An underfill is then injected into the gap between the chip and chip carrier and then cured to complete the flip-chip process. The function of the underfill, or encapsulation as it is sometimes referred to, is to provide mechanical integrity and environmental protection to a flip-chip assembly. Studies have demonstrated that both thermoset and thermoplastic ICAs can offer low initial joint resistances of less than 5 milliohms and stable joint resistances (Au-to-Au flip-chip bonding) during all the accelerated reliability testing. The reliability results have indicated that there is no substantial difference in the performance of thermoset and thermoplastic bumps and both types of polymers offer reliable flip-chip electrical interconnections [103].

Flip Chip with Micromachined ICA Bumps

Another polymer flip-chip bumping process is known as micro machined bumping [104]. The bumping procedure is illustrated in Fig. 6.16. Initially Cr/Au contact metal pads for conductive-polymer bumps are deposited on Si wafers, followed by patterning a thick photoresist to create bump holes. A high-aspect ratio and straight sidewall patterns are very important in shaping the conductive-polymer bumps. After the lithography, thermoplastic conductive polymer materials, usually thermoplastic paste filled with Ag flakes, is applied by either dispensing or screen printing the paste into the bump-hole patterns. The wafer is heated in a convection oven to remove the solvent. Due to the difference in curing conditions between the thick photoresist and conductive-polymer, the photoresist can be carefully stripped to expose the dried polymer bumps. Finally, the wafer is diced into individual chips.

Chips with thermoplastic bumps are placed on chip carriers and preheated to approximately 20 °C above the melting point of the polymer causing the bumps to reflow onto the matching chip carrier pads. Mechanical and electrical bonds are established as the chip carrier cools below the polymer melting temperature.

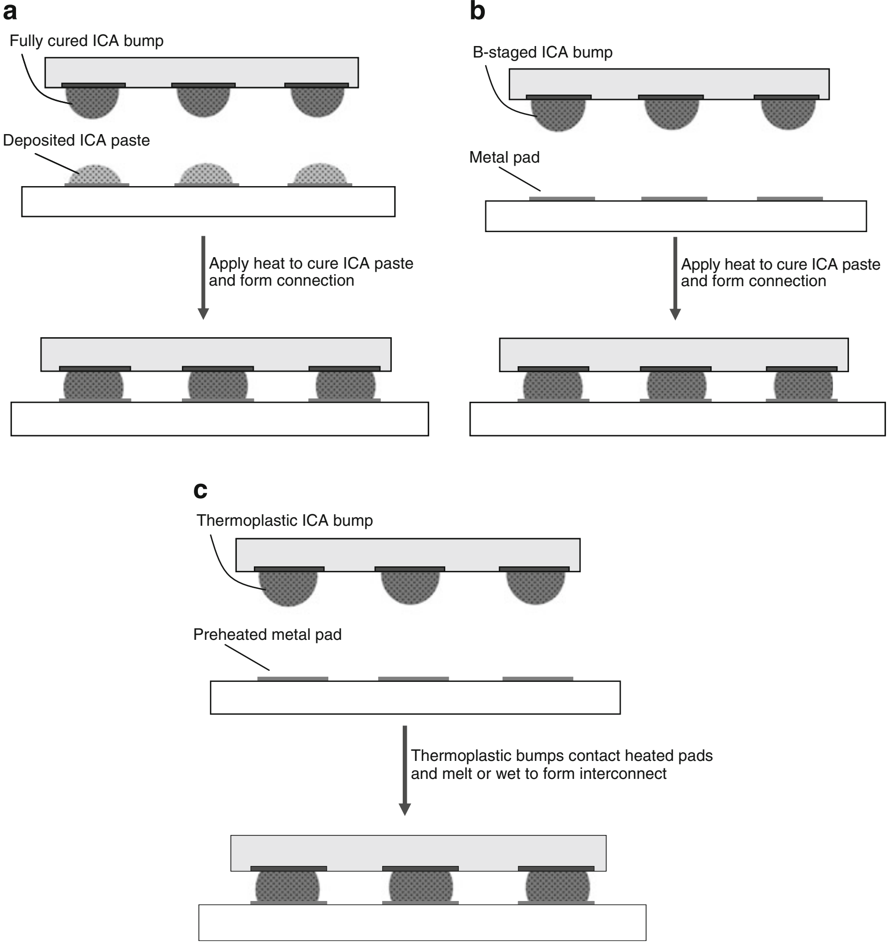


Fig. 6.15 Schematic illustrating various die attachment assembly processes utilizing ICAs. (**a**) chip with cured ICA bumps mated with uncured ICA on carrier pads. (**b**) Chip with partially cured (B-staged) ICA bumps mated with bare carrier pads. (**c**) Chip with thermoplastic ICA bumps mated with bare but preheated carrier pads

To enhance the mechanical bonding strength, a small amount of pressure can be applied by placing a weight on the chip.

This flip-chip bonding technique has high potential to replace conventional solder flip-chip techniques for sensor and actuator systems, optical micro electro-mechanical systems (MEMS), optoelectronic multichip modules (OE-MCMs), and electronic system applications.

Lohokare et al. [105] recently fabricated conductive adhesive bumps using a similar process, but with some extra polishing steps to make the bump surface flat (as shown in Fig. 6.17). Flip-chip interconnects formed using this new process offered

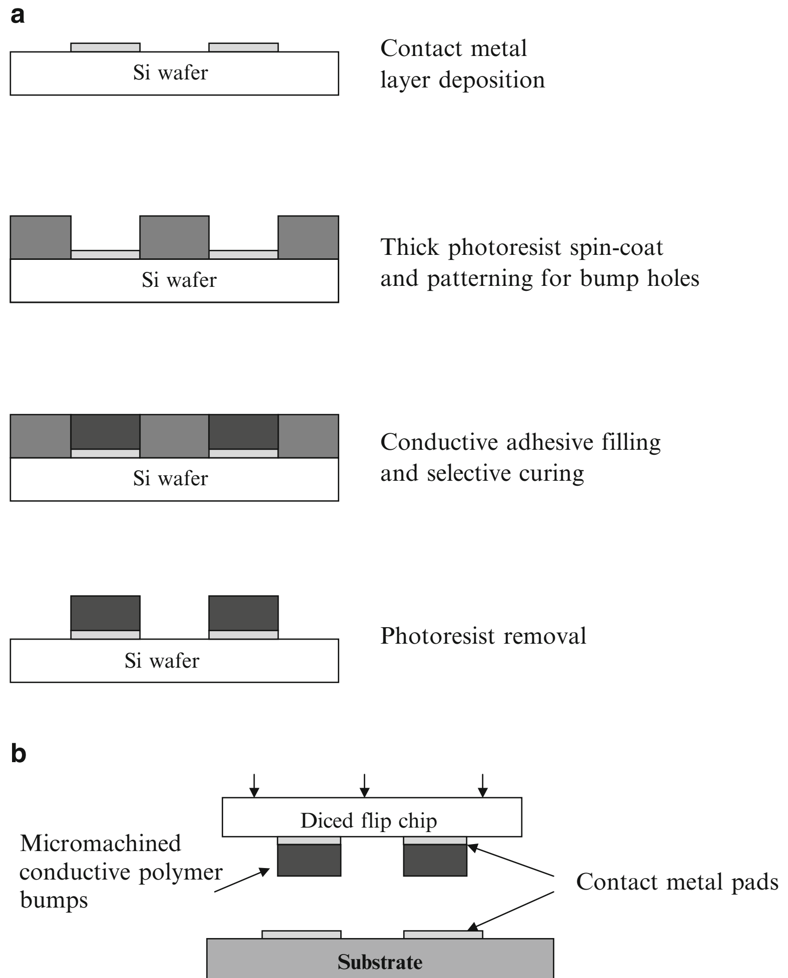


Fig. 6.16 A schematic depicting flip-chip application utilizing chips with micromachined polymer bumps. (**a**) Process flow for creating micromachined polymer bumps in the wafer state. (**b**) Die attachment to a chip carrier

lower contact resistance as compared to those of the printed bumps. Furthermore, in order to study the high-speed electrical performance characteristics of these conductive adhesive bumps, a 10-GHz p-i-n photodetector fabricated in the antimonide material system was used, and it was found that the polymer flip-chip-integrated detector showed electrical performance of the conductive adhesive bumps comparable to metallic interconnects.

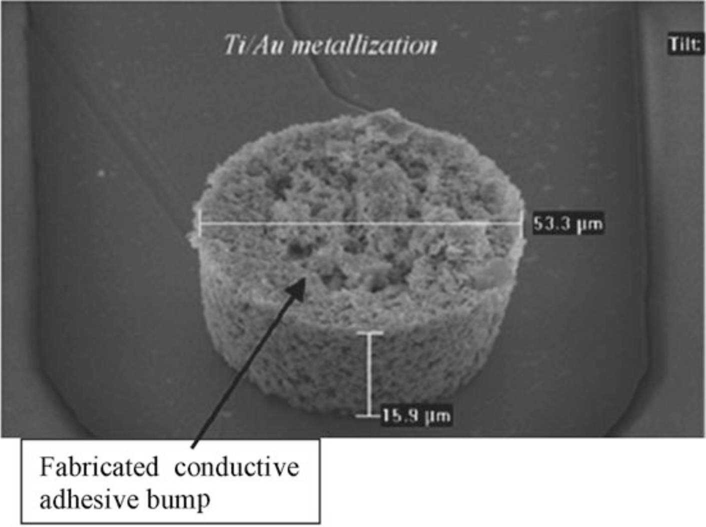


Fig. 6.17 SEM image of a flat surface of conductive adhesive bump [104]

6.3.2.2 Metal-Bumped Flip-Chip Joints

ICAs can also be used to form electrical interconnections with chips that have metal bumps. Isotropic conductive adhesive materials utilize much higher filler loading than ACAs to provide electrical conduction isotropically throughout the material. In order for these materials to be used for flip-chip applications, they must be selectively applied to only those areas that are to be electrically interconnected. Also, the materials are not to spread during placement or curing to avoid creating electrical shorts between circuit features. Screen or stencil printing is most commonly used to precisely deposit the ICA pastes. However to satisfy the scale and accuracy required for flip-chip, bonding requires very accurate pattern alignment. To overcome this difficult requirement, Matsushita developed a dip-transfer method [106].

Raised studs or pillars are required on either the die or chip carrier. Matsushita uses a conventional ball bonder to form Au-stud bumps. Bumping is significantly faster than creating complete wire bonds. A ball bumping process eliminates the need for traditional sputtering and plating processes used for standard bump formation. To prevent the bond area from becoming too large, the bumps are formed in a conical shape. The bumps are pressed level by a flat surface, which adjusts both height and planarity. The ICA is selectively transferred on the bump tips by contacting the face of the die to a flat thin film of the ICA which is produced by screen printing and whose transfer thickness is controlled by changing the printed film thickness. Then the die is picked, aligned, and placed on a chip carrier. The whole assembly is exposed to heat to cure the ICA and form connections between the die and chip carrier. Finally, an underfill (an insulating adhesive) is dispensed between the die and the chip carrier and cured. This method offers the

options of oven curing an assembly since bonding pressure is not required. A specially formulated ICA is used to avoid silver migration, containing 20 % palladium in a silver palladium alloy. A schematic of the process flow of forming joints with stud-bumped flip chips using ICAs is shown in Fig. 6.18. This process is particularly well suited when it becomes necessary to assemble flip-chip devices onto circuit boards already populated with soldered components—so called mixed technology. Regulating the amount of epoxy material transferred onto the Au-stud bumps is critical for ensuring high assembly yields, since electrical shorts may develop between adjacent bumps, either from bridging during the dip process itself, or from epoxy spreading during chip placement.

Another process for bonding a flip chip with metal bumps consists of stencil-printing an ICA on a chip carrier, aligning and placing the chip, curing the ICA to form bonds, and underfilling. A comparison study between an ICA-bonded and solder-bonded flip chips on FR4 chip carrier with Ni/Au metallization was conducted Y. Bessho [107]. The number of thermal cycles (−55 to 125 °C) to failure for both solder and ICA flip-chip circuits was compared. The study showed that stable contacts could be maintained for at least 1,000–2,000 cycles for ICA flip-chip joints. This is comparable to the lifetime for solder flip-chip joints. However, the variation among ICA samples was very high and optimization of assembly processes is needed in order to achieve more reproducible joint resistance [107]. Similar flip-chip interconnection process has been demonstrated for connecting a device with a pin count as high as 8,000 I/O [108]. Figure 6.19 shows the stencil printed ICA bump and interconnection joint formed between ICA and Au stud bump [108].

6.3.2.3 Carbon Nanotube Flip Chip

Flip-chip technology using solder bumps for the LSI interconnect has several issues such as electromigration and thermal stress against bumps [109]. Carbon nanotubes (CNTs) are a promising material for flip-chip interconnection. CNTs offer good electrical properties such as a high conductivity and current density exceeding 10^9 A/cm^2 [110], ballistic transport along the tubes [111]. They have been suggested for use as future internal wiring materials for LSIs [112–116]. In addition, CNTs have a high aspect ratio structure [117–119] and an excellent mechanical flexibility and strength.

A CNT-pattern-transfer process was used for CNT bump fabrication on a die [120, 121]. At the first step, conductive adhesive was printed onto the metal pads on a chip. Next, the chip was pressed on the CNT film blanket which had the vertically aligned CNTs on the flexible substrate. After peeling the CNT film blanket, the CNT bump pattern was left on the die due to the adhesion provided by uncured conductive adhesive dots on the metal pads, and the die with CNT bumps was cured at 180 °C. The CNT bumps were fixed on the metal pads on the chip. With this pattern transfer process, CNT bumps could be fabricated without being exposed to high-temperature carbon precipitating conditions which may potentially deteriorate the materials of the chips. The SEM image of the CNT bumps was shown in Fig. 6.20. The mean diameter and length of the CNT bumps were about 170 μm and 100 μm, respectively.

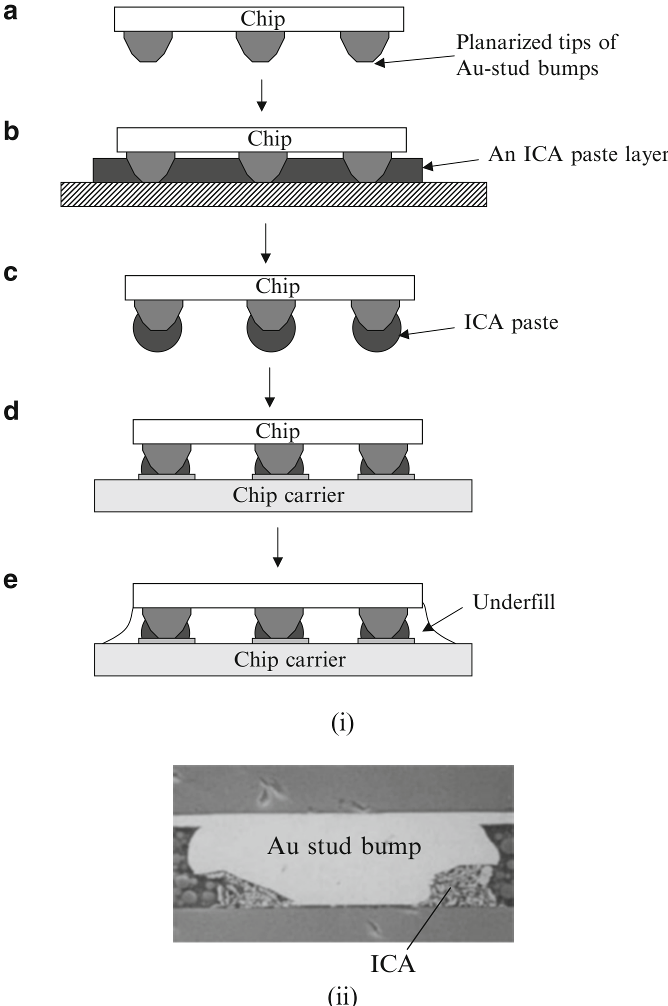


Fig. 6.18 (**a**) A schematic of the process flow of joints formed with stud-bumped flip-chips using ICAs. (a) Tips of the gold stud bumps formed with a wire bond tool are planarized. (b) Planarized bumps are dipped into a thin layer of ICA. (c) The chip is withdrawn, leaving the bumps coated with ICA. (d) The chip is placed on mating pads of a chip carrier with on pressure required during curing. (e) An underfill (an insulating adhesive) is dispensed and cured; (**b**) A SEM image of a ICA joint

The resistance of CNT bumps was measured using CNT bump contact chain. It was found that coating CNT bumps with gold greatly improved the contact resistance of the bumps with the chip and the substrate, resulting in a lower bump resistance of 2.3 Ohm. It was also demonstrated the resilience and flexibility of CNT bumps which can realize thermal stress free flip-chip structure. The CNT

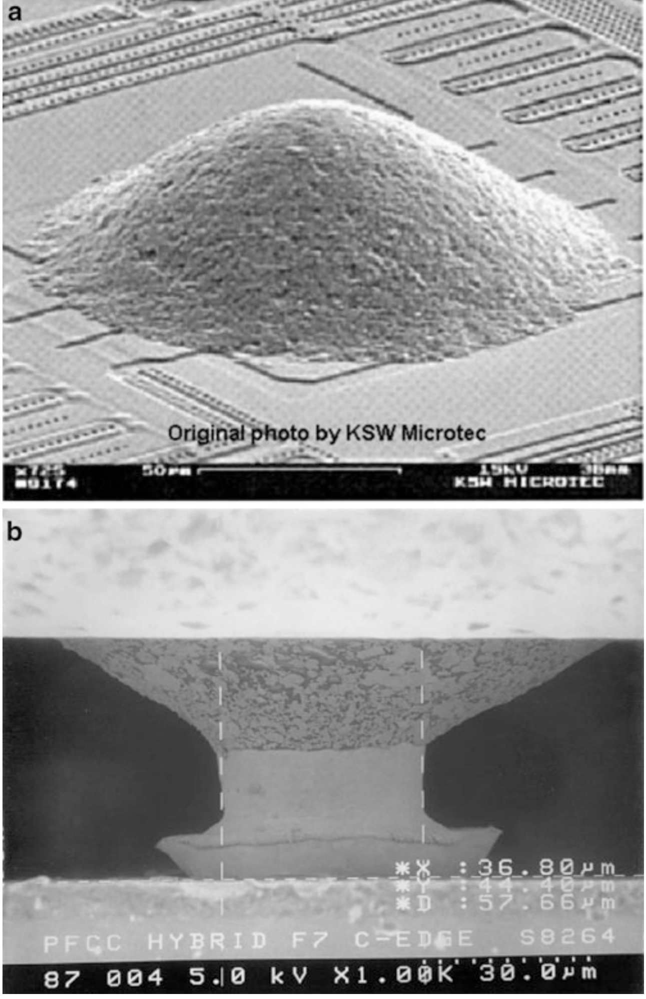


Fig. 6.19 (**a**) SEM image of a stencil printed ICA bump; (**b**) SEM image of a flip-chip joint formed between Au stud bump and printed ICA [108]

bumps were able to absorb a displacement between the chip and substrate up to 10–20 % of the CNT bump height.

6.3.3 ICAs for Advanced Packaging Applications

6.3.3.1 Solar Cell

Thin solar cells are difficult to interconnect with standard soldering techniques. High temperature during soldering introduces stress on the joints and cells. This can

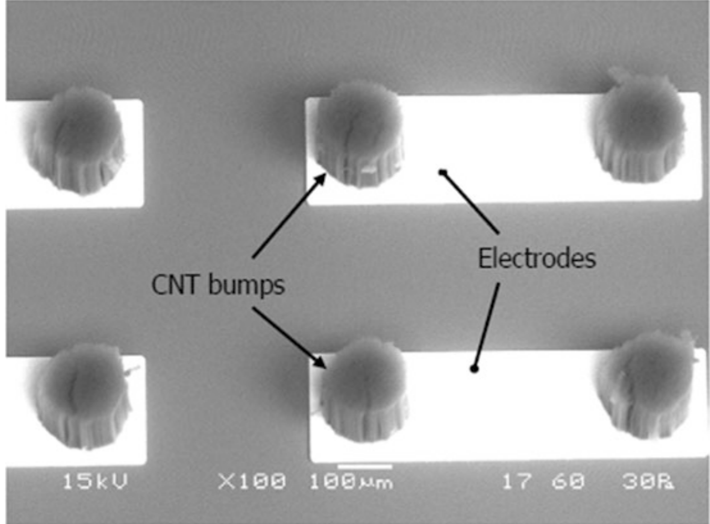


Fig. 6.20 SEM image of CNT bumps fabricated by the transfer process [121]

cause warping and possible breakage of cells. Substituting soldering for a low temperature joining technique would avoid building up of mechanical stress, thus increasing process yield and reliability.

A promising alternative technique is interconnection with conductive adhesives. The processing temperature is much lower, depending on the applied adhesive. D.W.K. Eikelboom et al. [122] studied rear-contacted solar cells interconnected with conductive adhesives. The stress on thin cells and joints between traditional soldering and conductive adhesives was compared.

Solar cells generate high currents, requiring low resistances. Long-term stability in outdoor conditions requires excellent optical and mechanical properties. High volume production demands screen printable adhesives. Contacts made on silver plated substrates with silver plated tabs show excellent electrical properties. Contact resistances are in the milli-ohm range like soldered contacts.

Samples have been damp/heat tested along with soldered references. After 2,500 h at 85 °C/85 % humidity, no degradation occurred. Also temperature cycling, between −40 °C/+80 °C, on the same samples has shown no effects after 200 cycles. No differences are measured between samples that have been encapsulated and plain samples without encapsulation. Adhesives applied to plain copper tabs or applied to screen-printed and fired aluminum give low resistances directly after curing, but show rapid degradation after aging tests due to oxidation and breaking of the adhesive-to-metal bond. Porous screen printed aluminum also soaks up the adhesive component from the paste giving poor bonding.

6.3.3.2 3D Stacking

Wirebonding process applies significant mechanical force [123, 124] in order to produce a reliable friction welded connection, often creating latent defects within a conic region of the silicon underneath bonds in die thinned below 75 μm. Wirebond induced defects can also extend beyond the bonded die itself through underlying die attach materials and into die beneath the originally bonded die within the stack. Reverse loop bonding processes have been developed [125] to mitigate this problem in stacked die packages somewhat, although the same reliability issues often reemerge as die are further thinned below 50 μm. Wirebonding to extremely thin die presents a significant challenge, and an electrical interconnect process that dramatically reduces applied force during application will result in both enhanced yields and superior product reliability over traditional electrical interconnect methodologies.

Andrews et al. [126] demonstrated forming electrical interconnections along the edges of stacked integrated circuits by an extrusion process that utilizes automated needle dispense equipment to form local deposits of conductive adhesive paste. This vertical interconnect process was designed to form three dimensional circuits without the imposition of significant mechanical forces that are known to cause mechanical damage to thin die or fragile substrate materials. The process was demonstrated with productivity rates that exceed 100 interconnections per minute. Figure 6.21 shows an image of a stack die package with conductive adhesive vertical interconnections. The paste extrusion process utilizes automated needle dispense equipment to produce electrically conductive structures from die-to-die, or die-to-substrate, without imposing any significant mechanical forces known to cause damage to extremely thin die or other fragile substrate materials such as GaAs or InGaP. The paste vertical interconnect process is insensitive to die type, die count (up to 128 vertical die have been demonstrated), stacking configuration, or final packaging requirement (e.g., QFN, BGA, WL-CSP, etc.).

Significant processing cost advantages were achieved by cycle-time reduction vs. traditional wirebond stacked-die interconnect, because serial die attach/wirebond process steps are eliminated in favor of a single staging at each process tool regardless of die-count. Additional cost advantages were also found due to the elimination of gold (by eliminating the wirebonding process and reducing the gold thickness upon bondable surfaces). Further cost reduction could be achieved by adopting gang vertical interconnect methods, including but not limited to multi-needle dispense and print based methods.

6.3.3.3 Microspring

Micromachined springs are attracting attentions in microsystem packaging for low-stress interconnection, high density and low-damage probing for device testing. For instance, new probe cards consisting of micromachined springs are in practical use

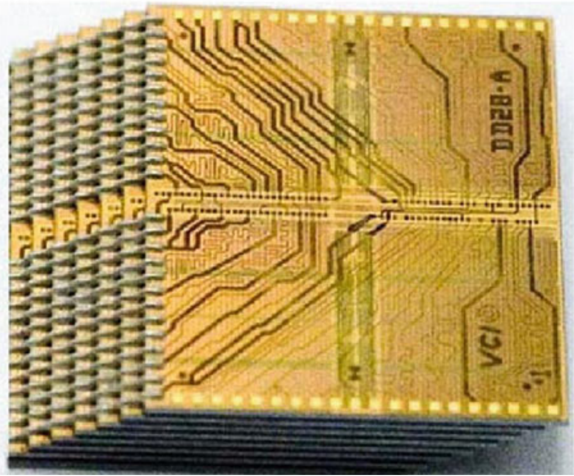


Fig. 6.21 Vertical interconnection formed by a conductive adhesive on Shingle-Tier stack [126]

[127] and essential for high performance device testing, because their dimensions can easily be smaller than a few hundred micrometers. They are therefore applicable to higher pad-density and smaller pad pitch chips, and capable of being effective in a test using high speed signals above 1 GHz. Curl-up microcantilever spring probes [128] and s-type microspring probes using multilayer electroplating [129] for the probe card applications were reported in the literature.

On the other hands, microsprings may also be required for microelectronic packaging. Flip-chip packages, which are utilized for high performance microsystems, are carried out using the standard interconnect technologies, such as solder balls, gold bumps, and conductive adhesives. However, since their interconnect structures have very limited compliance, thermal expansion mismatches within the packages between the silicon integrated circuit (IC) and the package substrate can cause problems such as failures during thermal cycle testing and damages in low-k dielectrics. To overcome these, Chow et al. experimentally demonstrated that a compliant, low force (10 mgf), pressure contact using thin film microsprings [130] could make reliable electrical contact to a gold pad [131]. Furthermore, in 0-level MEMS packaging, a commonly applicable cap wafer may need through-hole interconnect vias and microsprings to interconnect the device electrodes.

These microfabricated springs may not be suitable to microsystem packaging, because multiple photolithography procedures cause high production costs. Aiming to fabricate microspring probes without photolithography steps, T. Itohl et al. utilized an ultra-precise dispenser with three-axis stage system to develop a new three-dimensional microstructure forming method utilizing a continuously repeated dispensing of conductive adhesive pastes [132, 133]. The inner diameter of the nozzle used is 22 μm and the minimum dispense dot size was as small as 22 μm in diameter. High aspect ratio structures were realized by dispensing the paste dots repeatedly. By heating the substrate over 350 K, the organic solvent in the paste dispensed on the substrate vaporizes and the viscosity of the paste increases. Taking

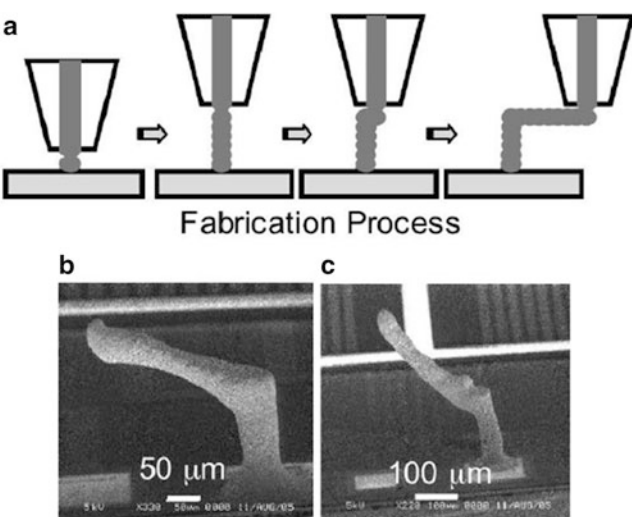


Fig. 6.22 Microcantilever springs fabricated using conductive paste continuous dispensing [133]

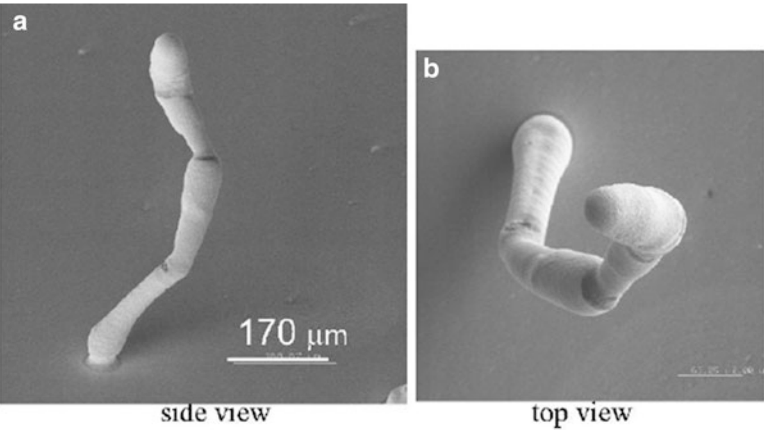


Fig. 6.23 A spiral structure fabricated using conductive adhesive dispensing [133]

advantage of this process, overhanging structures such as cantilever shapes were formed when the nozzle was moved in lateral direction as shown in Fig. 6.22a. For example, in the case of the cantilever shown in Fig. 6.22b, 6.20 times dispensing for the post part and 40 times dispensing for the lever part was carried out. The shapes of cantilevers did not change after curing at 423–523 K for 30 min. It was found that fabricated micro cantilevers have probing resistance lower than 1 Ohm with a low contact force of 1 mN. More complicated shapes such as spiral structure shown in Fig. 6.23 was successfully realized by controlling the dispense conditions and the substrate temperature.

6.3.4 High Frequency Performance of ICA Joints

Only limited work has been conducted to investigate the high frequency behaviors of ICA joints. J. Felba et al. [134] investigated a formulation of ICA that performed well as a solder replacement in microwave applications. The study involved in various different adhesive base materials and several types of main (silver flakes, nickel and graphite) and additional (soot and silver semiflake powder) filler materials. In order to assess the usefulness of a given adhesive formulation, an additional gap in the gold strip of a standard microstrip bandpass filter was made and bridged by an adhesive bonded silver jumper. Both the quality factor (*Q*-factor) and loss factor (*L*) of the filter with the bonded jumper were measured at a frequency of 3.5 GHz in a preliminary experiment and at 3.5 and 14 GHz in a final experiment. It was determined that silver flake powders are the best filler materials for ICA for microwave applications because ICAs filled with the silver flake powders exhibit the highest *Q*-factor and lowest loss factor. Also, addition of soot should be avoided since it decreases the quality factor [134].

A study at Georgia Tech was reported where a flip-chip test vehicle was mounted on a FR4 chip carrier with a gold-plated copper transmission line [135]. The performance of eutectic Sn–Pb and ICAs were evaluated and compared using this test device. Both ICAs and eutectic Sn–Pb solder were determined to exhibit almost the same behavior at a frequency range of 45 MHz to 2 GHz and the measured transmission losses for both materials were minimal. It was also found that the S_{11} characteristics of both Sn–Pb and ICAs after exposure to 85 °C/85 % relative humidity aging for 150 h did not vary from the signals prior to aging, but S_{12} value of the Sn–Pb joints deviated more than that of ICA joints after the aging.

Recently, Kaoru Hashimoto et al. [136] studied feasibility of the conductive adhesive joints for high-speed signal transmission, both transmission characteristics and power supply ability at the conductive adhesive joints using specially designed interconnection models which consisted of a high speed CMOS driver LSI, a model BGA package and a model circuit board. It was found that differential pulse signals could transmit without degradation at 12 Gbps through the 8 conductive adhesive joints in daisy chain configuration, and the conductive adhesive joint exhibits waveform degradation in the case of long transition time. This was considered to be caused by its DC resistance which is about ten times higher than that of the solder joint. However, it showed less waveform degradation in the case of short transition time. This is probably caused by the capacitive coupling interference effect due to proximity arrangement of Ag flakes in the conductive adhesive.

6.3.5 Reliability of ICA Joints

To date, quite some work has been published demonstrating the usefulness and limitations of isotropic adhesives to bond a variety of surface mount compounds

(e.g., QFP) and passive components (R and C) under different aging conditions using various circuit and component-metallizations [72]. Most of the isotropic conductive adhesives require noble metallizations (e.g., Au or AgPd) to survive harsh environmental conditions as for instance 85 °C/85%RH and temperature cycling from −40 to +125 °C. Most of the adhesives give bad results on SnPb surfaces; a few special types show better results in 85 °C/85%RH. Deterioration of the electrical properties is due to an increase in the contact resistance. The bulk resistance of the adhesive, although considerably higher than that of solder, usually remains quite constant. On passivated Cu substrates reasonable good results were obtained.

Jon B. Nysæther et al. [137] compared the failure of flip chip on board (FCOB) with solder or ICA joints and with underfill under thermal cycling. The measurements of solder bump lifetime are compared to a lifetime model based on analytical calculations of solder strain. For two filled types of underfill with CTE nearly matched to that of solder, the measured average lifetimes vary from around 2,700 to 5,500 cycles. Measurements of the lifetime of FCOB's with ICA connections have been carried out for two different material systems. The obtained lifetimes vary between approximately 500 and 4,000 cycles. The lifetime seems to be dependent on the properties of the bump on the chip pad. Delamination, for instance at the ICA/bump interface, was found to be an important failure mechanism. The best results (>4,000 cycles) were obtained for 5 μm high Ni/Au bumps.

Aiming to understand the performance of ICA interconnects under fracture and fatigue loading, J. Constable et al. [138] investigated performance of ICA interconnects under fracture and fatigue loading by monitoring resistance changes (micro-ohm sensitivity) of ICA joints during pull and fatigue testing (cyclic loading up to 1,000 cycles). Observation of the fracture surface suggested that the ICA joint life depended upon the adhesive failure of the bond to the metal surface. It was observed that fracture strains for the ICAs were in the range of 20–38 %, and resistance remained approximately constant in the elastic region, but the resistance started to increase rapidly as soon as the pull-force departed from linear elastic behavior. For fatigue tests, linear displacement was ramped up the preprogrammed maximum displacement and ramped back to the starting position. It was observed that the shear strain for ICA joints surviving 1,000 cyclic loading was typically 10 %, which is about an order of magnitude greater than solders. This suggests that using conductive adhesives may be advantageous for some flip-chip applications. It is believed that since silver filler particles of ICAs cannot accommodate this large strain, the silver filler particles must move relative to one another as the epoxy matrix is strained. The most common pattern of resistance change was only increased to a point corresponding to about a 70 % loss in interface contact resistance before sudden failure. This was an indication that the interface crack slightly propagated into the adhesive [138].

In an effort to gain a fundamental understanding of the fatigue degradation of ICAs, R. Gomatam et al. [139] studied the behavior of ICA joints under temperature and humidity conditions. The fatigue life decreased at elevated temperature and high humidity conditions. It was also observed that the fatigue life of the ICA joints

decreased considerably as the temperature cycle frequency was decreased. This effect was attributed to the fact that as the frequency was decreased, the propagating crack was exposed to higher loads for longer periods of time, effectively resulting in high creep loading [139].

M. Yamashita et al. [140] investigated the interfacial degradation mechanism between Ag-filled ICA and SnPb surface finish at elevated temperatures. It was found that, at 150 °C, the interface degradation was caused by the preferential diffusion of Sn from the plated SnPb layer into the Ag filler in the ICA. Due to this diffusion, Ag–Sn intermetallic compounds were formed on the Ag fillers adjacent to the plating layer, and many large Kirkendall voids were formed in the SnPb plating layer. Also, an interfacial debonding was observed between the ICA and the SnPb plating layer after the heat exposure.

Xu et al. [141] investigated the mechanical behavior of electrically conductive adhesive (ECA) joints exposed to elevated temperature and relative humidity conditions and failure mechanisms of conductive adhesive joints. Three silver-filled, epoxy-based adhesives were used in conjunction with printed circuit board (PCB) substrates with metallizations of Cu/Ni/Au and Cu. Double cantilever beam (DCB) tests was adopted to investigate the effects of environmental aging on ECA joints. This study revealed that conductive adhesives as well as substrate metallizations both play important roles in the durability of conductive adhesive joints. The rate of water attack on the interface of conductive adhesive joints with Cu-plated PCB substrates is faster than for those with Au/Ni/Cu metallization. A possible explanation of this phenomenon was based on considerations of surface free energy and interfacial free energy. The fracture energy of all three ICAs decreased with time under the aging condition. Following drying of the aged conductive adhesive joints, the fracture energy recovered to some extent. This recovery in the fracture energy could be attributed to the reversible effect of plasticization of the bulk adhesives, as well as the regaining of bond strength between the adhesive and the substrate during drying at 150 °C. However, the fracture energy of the adhesive joints showed little recovery after the metal surface was oxidized. For ECA/Cu joints, water attack on the adhesive joint may be divided into three phases: displacing the adhesive from the substrate, oxidizing copper, and weakening the copper oxide. At the end of aging, the three ECA/Cu joints exhibited different modes of failure. ECA1/Cu joints failed interfacially along the adhesive/copper oxide interface, while ECA2/Cu joints exhibited the locus of failure within the copper oxide layer. For ECA3/Cu joints, the failure occurred within the secondary layer of the adhesive, which is adjacent to the interface and is a silver-depleted layer. XPS analysis of DCB failure surfaces suggested that diffusion of Cu to the Au surface might have occurred on the Au/Ni/ Cu-plated PCB substrates during aging. Copper oxide was detected on the substrate surface upon exposure of the conductive adhesive joints to the hot/wet environment [141].

S. Kuusiluoma et al. [142] compared the reliability of isotropically conductive adhesive (ICA) attachments on liquid crystal polymer (LCP) substrate to the reliability of lead-free solder (SnAgCu) attachments on same substrate material. The assembled components were similar surface mount components and the reliability was

assessed through real time measurements of contact resistance of each connection. The devices were subjected to two environmental stress tests, a thermal cycle test and a sinusoidal vibration. Results showed that when using LCP as a substrate material, the reliability of the device under thermal cyclic stress is somewhat inferior to the reliability of SnAgCu solder. No significant differences could however be observed from the vibration test results since none of the assemblies failed in the test. The failure analysis revealed that most failures occurred at the interface between component lead and the attachment material in both cases, but the test methods need to be developed to make further conclusions.

Ales Duraj et al. [143] investigated the influence of dynamic mechanical load (bending of testing boards) on resistance of the adhesive joints. The load was induced by a definite deflection of testing boards (fiberglass laminated PCB assembled with 1,206 SMD resistors). It was found that the applied dynamic load caused changes of basic electrical parameters of the bonds. The more the deflections applied the more changes of resistance were observed. The increase of joint resistance was not linear and was not same for all tested adhesives.

J. Lee et al. [144] studied the junction resistance variation of Ag epoxy-filled ICAs on Cu and immersion-Ag finished PWB during the 85 °C/85 % relative humidity aging. It was found that the junction resistances of immersion-Ag PWB are lower than those of Cu-finished PWB, and the junction resistance shift on the immersion Ag PWB was much smaller than those of Cu-finished PWB during the aging.

6.3.6 *Recent Advances on Nano-ICAs*

To meet the requirements for future fine pitch and high performance interconnects in advanced packaging, ECAs with nano-material or nanotechnology attract more and more interests due to the specific electrical, mechanical, optical, magnetic, and chemical properties. There has been extensive research on nano-conductive adhesives which contain nano-filler such as nanoparticles, nanowires, carbon nanotubes, and graphenes. This chapter provides a comprehensive review of most recent research results on nano-conductive adhesives.

6.3.6.1 ICAs with Silver Nanowires

Wu et al. [145] developed an epoxy-based ICA filled with silver nanowires, and compared the electrical and mechanical properties of this nano-ICA to two other ICAs filled with micrometer-sized (roughly 1 μm and 100 nm) silver particles. The nanowires had a diameter of roughly 30 nm and a length up to 1.5 μm, and the nanowires were polycrystalline in nature. It was found that at a low filler loading (e.g., 56 wt%), the bulk resistivity of ICA filled with the Ag nanowires was significantly lower than the ICAs filled with 1 μm or 100 nm silver particles. The better electrical conductivity of the ICA filled nanowires was contributed by the

lower contact resistance between nanowires and a more significant contribution from the tunneling effect among the nanowires [145]. It was also found that at the same filler loading (e.g., 56 wt%), the ICAs filled with Ag nanowires showed shear strength similar to that of the ICAs filled with the 1 μm and 100 nm silver particles. However, to achieve the same level of electrical conductivity, the filler loading must be increased to at least 75 wt% for the ICA filled with micrometer-sized Ag particles, and the shear strength of these ICAs is then decreased (lower than that of the ICAs filled with 56 wt% nanowires) due to the higher filler loading.

6.3.6.2 Effect of Nano-Sized Silver Particles on the Conductivity of ICAs

Lee et al. [146] studied the effect of nano-sized filler on the conductivity of conductive adhesives by substituting nano-sized Ag colloids for micro-sized Ag particles partially or wholly in a polymeric system (poly(vinyl acetate)—VAc). Electrical resistivity was then measured as a function of silver volume fraction. It was found that when nano-sized silver particles were added into the system at 2.5 wt% increment, the resistivity increased in almost all cases, except when the quantity of micro-sized silver was slightly lower than the percolation threshold value. At that point, the addition of the about 2.5 wt% brought about significant decrease in resistivity. Near the percolation threshold, when the micro-sized silver particles are still not connected, the addition of a small amount of nano-sized sliver particles helps to build the conductive network and thus lowers the resistivity of the composite. However, when the filler loading is above the percolation threshold and all the micro-sized particles are connected, the addition of nanoparticles seems only to increase the relative contribution of contact resistance between the particles. Due to its small size, for a fixed amount of addition, the nano-sized silver colloid contains a larger number of particles when compared with micro-sized particles. This large number of particles should be beneficial to the interconnection between particles. However, it also inevitably increases the contact resistance. As a result, the overall effect is an increase in resistivity upon the addition of nano-sized silver colloids.

Ye et al. [147] also reported a similar phenomenon, i.e., the addition of nanoparticles showed a negative effect on electrical conductivity. They proposed two types of contact resistance, i.e., restriction resistance due to small contact area and tunneling resistance when nanoparticles are included in the system. It was believed that the conductivity of micro-sized Ag particle filled adhesives was dominated by constriction resistance, while that of the nanoparticle-containing conductive adhesives was controlled by tunneling and even thermionic emission. Fan et al. [148] and Mach et al. [149] also observed a similar phenomenon (adding nano-size particles reduced electrical conductivity).

Lee et al. [146] also studied the effect of temperature on the conductivity of ICAs. Heating the composite to a higher temperature can reduce the resistivity quite significantly. This is likely due to the high activity of nano-sized particles. For micro-sized paste, this temperature effect was considered negligible. The interdiffusion of silver atoms among nano-sized particles helped to reduce the contact

resistance quite significantly and the resistivity reached 5×10^{-5} Ohm cm after treatment at 190 °C for 30 min. Jiang et al. [150] showed that when suitable surfactant was used in the nano-Ag containing ICA, the dispersion and inter-diffusion of silver atoms among nano-sized particles could be facilitated and the resistivity of ICA could be reduced to 5×10^{-6} Ohm cm.

6.3.6.3 ICAs Filled with Aggregates of Nano-sized Ag Particles

To improve the mechanical properties under thermal cycling conditions while still maintaining an acceptably high level of electric conductivity, Kotthaus et al. [151] studied an epoxy-based ICA material system filled with aggregates of nano-sized Ag particles. The idea was to develop a new filler material which did not deteriorate the mechanical property of the polymer matrix to a great extent. A highly porous Ag powder was attempted to fulfill this requirement. The Ag powder was produced by the inert gas condensation (IGC) method. The powder consisted of sintered networks of ultrafine particles in the size range from 50 to 150 nm. The mean diameter of these aggregates could be adjusted up to some micrometers. The as-sieved powder was characterized by a low level of impurity content, an internal porosity of about 60 %, and a good ability for resin infiltration. Using the above nano-sized Ag powder instead of Ag flakes is more likely to retain the properties of the resin matrix because of the infiltration of the resin into the pores. Measurements of the shear stress–strain behavior indicated that the thermomechanical properties of bonded joints could be improved by up to a factor of 2, irrespective of the chosen resin matrix.

Resistance measurements on filled adhesives were performed in a temperature range from 10 to 325 K. The specific resistance of the nano-sized Ag powder filled adhesive was about 10^{-2} Ohm cm and did not achieve the typical value of commercially available adhesives of about 10^{-4} Ohm cm. The reason may be that Ag nanoparticles are more or less spherical in shape, which provides fewer conduction paths than Ag flakes, and have the intrinsically lower specific conductivity. For certain applications where mechanical stress plays an important role, this conductivity may be sufficient, and therefore, the porous Ag could be suitable as a new filler material for conductive adhesives.

6.3.6.4 ICAs Filled with Nano-sized Ni Particles

It is generally known that metal powders present properties that are different from those of bulk metals when their particle size is made as small as nanometer size. Powders are classified into particles, microparticles and nanoparticles according to size. Although the classification criterion is not clear, particles with diameter smaller than 100 nm are generally called nanoparticles. This classification is based on the fact that when particle size is smaller than 100 nm, the particle possesses properties that are not found in the microparticles larger than 100 nm. For example, when the particle diameter of such magnetic materials as iron and

nickel is near 100 nm, their magnetic domains change from multiple to single, and their magnetic properties also change [152]. Majima et al. [152] reported an application example of metal nanoparticles to conductive pastes, focusing on the properties of a new conductive adhesive that were not found in conventional ICAs. Sumitomo Electric Industries, Ltd. (SEI) has developed a new liquid-phase deposition process using plating technology [152]. This new nanoparticle fabrication process achieves purity greater than 99.9 % and allows easy control of particle diameter and shape. The particle's crystallite size calculated from the results of X-ray diffraction measurement is 1.7 nm, which leads to an assumption that the size of primary particles is extremely small. When the particle size of nickel and other magnetic metals becomes smaller than 100 nm, they change from multi-domain particles to single-domain particles, and thus their magnetic properties change. That is, if the diameter of nickel particles is around 50 nm, each particle acts like a regular magnet, and magnetically connects with each other to form chainlike clusters. When the chain-like clusters are incorporated into a conductive paste, electrical conduction of the paste is expected to be better than the original paste. The chain-like nickel particles developed were mixed with a predefined amount of poly(vinylidene difluoride) (PVdF) that acted as an adhesive. Then, *n*-methyl-2-pyrrolidone was added to this mixture to make a conductive paste. This paste was applied on a polyimide film and then dried to make a conductive sheet. Specific volume resistivity of the fabricated conductive sheet was measured by the quadrupole method. The same measurement was also conducted on the conductive sheet that used paste made of conventional spherical nickel particles. Measurement of the sheet resistance immediately after paste application indicated that the developed chain-like nickel powder had low resistance of about one-eighth of that of the conventionally available spherical nickel particles. This result showed that when the newly developed chain-like nickel particles were incorporated in the conductive paste, high conductivity could be achieved without pressing the sheet. SEI tested and developed the metal nanoparticles and investigated the possibility of their application in a conductive paste.

6.3.6.5 Nano-ICAs Filled with CNT

Electrical and Mechanical Characterization of CNT-Filled ICAs

The density of commercially available silver-filled conductive adhesives is around 4.5 g/cm^3 after cure. Metal-filled electrically conductive adhesives offer an alternative to typical lead–tin soldering with the advantages of being simple to process at lower temperatures without toxic lead or corrosive flux. The disadvantage of conventional metal-filled conductive adhesives is that a high loading of filler decreases the mechanical impact strength, while a low filler loading results in poor electrical properties. Carbon nanotubes are a new form of carbon, which was first identified in 1991 by Sumio Iijima of NEC, Japan [153]. Nanotubes are sheets of graphene rolled into seamless cylinders. Besides growing single wall

nanotubes (SWNTs), nanotubes can also have multiple walls (MWNTs)—cylinders inside other cylinders. A carbon nanotube can be 1–50 nm in diameter and up to few cm in length, with each end "capped" with half of a fullerene dome consisting of five or six member rings. Along the sidewalls and cap, additional molecules can be attached to functionalize the nanotube to adjust its properties. CNTs are chiral structures with a degree of twist such that the graphene rings join into cylinders. The chirality determines whether a nanotube will conduct in a metallic or semiconducting manner. Carbon nanotubes possess many unique and remarkable properties. The measured electrical conductivity of metallic carbon nanotubes is in the order of 104 S/cm (ballistic transport) [153]. The thermal conductivity of carbon nanotubes at room temperature can be as high as 3,000–6,600 W/mK [154]. The Young's modulus of carbon nanotubes is about 1 TPa. The maximum tensile strength of carbon nanotubes is close to 30 GPa, with some reported at TPa [155]. The density of MWNTs is 2.6 g/cm^3 and the density of SWNTs ranges from 1.33 to 1.40 g/cm^3 depending on the chirality [156]. Since carbon nanotubes have very low density and long aspect ratios, they have the potential of reaching the percolation threshold at very low weight percent loading in the polymer matrix. Wu et al. [157] developed a process to prepare silver-coated carbon nanotubes (SCCNTs), and then used the SCCNTs to formulate an ICA and compared its electrical and mechanical properties to ICAs filled with traditional multi-walled carbon nanotubes (CNTs) and micrometer-sized Ag particles. It was found that the ICA filled with SSCNTs had a lower bulk resistivity (2.21×10^{-4} Ohm cm), and higher shear strength than that of the Ag particle filled ICA at the same filler volume content (28 %).

Experiments conducted by Qian et al. [158] showed 36–42 % and 25 % increases in elastic modulus and tensile strength, respectively, in polystyrene (PS)/CNT composites. The TEM observations in their experiments showed that cracks propagated along weak CNT–polymer interfaces or relatively low CNT density regions and caused failure. If the outer layer of MWNTs can be functionalized to form strong chemical bonds with the polymer matrix, the CNT/polymer composites can be further improved in mechanical strength and have controllable thermal and electrical properties.

Effect of Adding CNTs to the Electrical Properties of ICAs

Lin and Lin [159] studied the effect of adding CNTs on the electrical conductivity of silver-filled conductive adhesives which had epoxy as the base resin and had various filler loadings. It was found that the CNTs could enhance the electrical conductivity of the conductive adhesives greatly when the silver filler loading was still below the percolation threshold. For example, the 66.5 wt% filled silver conductive adhesive without CNTs had a resistivity of 104 Ohm cm, but showed a resistivity of 10^{-3} Ohm cm after adding 0.27 wt% CNTs. Therefore, it is possible to achieve the same level of electrical conductivity by adding a small amount of CNTs instead of the silver fillers.

Composites Filled with Surface Treated CNT

Although CNTs have exceptional physical properties, incorporating them into other materials has been very challenging due to their nonpolar and hydrophobic surfaces. Problems such as phase separation, aggregation, poor dispersion in the matrix, and poor adhesion to the matrix must be overcome. A multifunctional bridge was created between the CNT sidewalls and the host matrix. The power of this bridge was demonstrated by comparing the fracture behavior of the polycarbonate, polystyrene or epoxy composites filled with untreated and surface treated nanotubes. It was observed that the untreated nanotubes interacted poorly with the polymer matrix, and thus left behind voids in the matrix after fracture. However, for the composite filled with treated nanotubes, the nanotubes remained in the matrix even after the fracture, indicating strong interaction with the matrix. Due to their superior dispersion in the polymer matrix, the treated nanotubes achieved the same level of electrical conductivity at much lower loadings than the untreated nanotubes [160–162].

6.4 Nonconductive Adhesive for Flip-Chip Applications

Compared with ECAs, NCAs are materials composed of an adhesive polymer resin and a curing agent without any kind of conductive fillers. NCA technology has significant potential as an alternative for flip chip on organic boards because it offers fine pitch, simplified process, low packaging cost, and a lead-free interconnect system for flip-chip packaging.

Matsushita was the first to use a nonconductive adhesive for flip-chip bonding in 1988 referred as “Micron Bump Bonding Assembly Technology” [163, 164]. Figure 6.24 depicts a cross-sectional view of a LSI chip with Au bumps bonded to a chip carrier (with Cr/Au pads) utilizing this method. The chip bonding accomplished by shrinkage induced in a light-curable resin. That is, the chip is held in place by the adhesive force of the resin and the bonding connection is made by the internal shrinkage stress exerted by the resin. The technology can support pitches down to the micron range. There are virtually no restrictive factors to limit the pitch since an insulating resin is used for the bonding. Due to its flexible structure, the thermal induced stresses in the assembly can be relieved to improve the reliability.

Ferrando et al. demonstrated a low-cost flip-chip process using Au-stud bumps with a NCA paste that does not require underfilling the die [165]. This process was found to be particularly suitable for organic chip carriers where CTE mismatches are critical concerns and is compatible with large-scale production as well as prototyping. The process flow is illustrated in Fig. 6.25. The Au-stud bumping step is preferably done at the wafer level. The gold-stud bumps compensate for chip carrier non-coplanarity by deforming the bonding pads. The PCB must be designed to be compatible with the NCA flip-chip process. For example, metallized vias located in the flip-chip area must be blind to prevent any leakage of the NCA paste

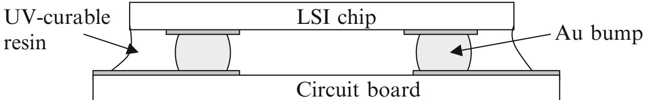


Fig. 6.24 A cross-sectional view of NCA flip-chip joints formed between a LSI chip with Au bumps and a chip carrier

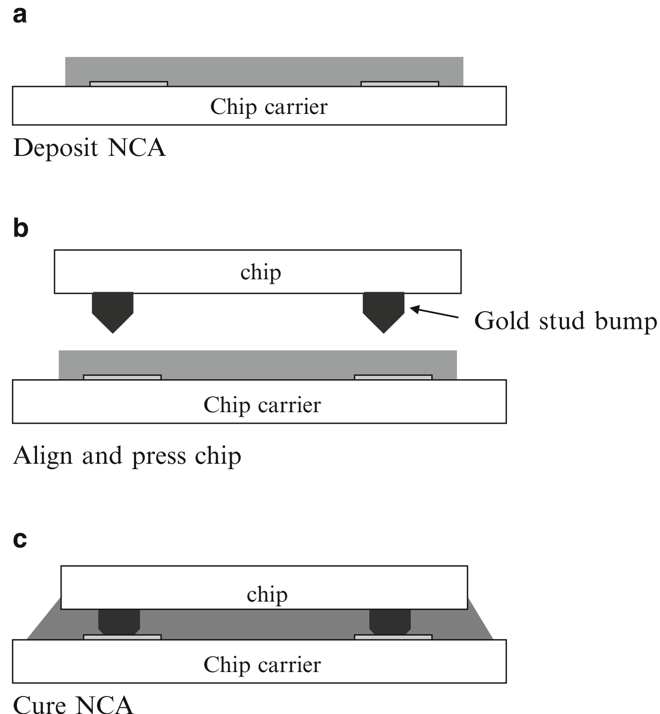


Fig. 6.25 Process steps for forming flip-chip joints between a chip with Au stud bumps to a chip carrier using a NCA

to the other side the substrate. Before dispensing the NCA, the PCB must be dried and receive a surface treatment to reactivate the surface and increase its surface tension. After dispensing the NCA using a screen printing or dispensing machine, a die is aligned and its stud bumps are pressed onto mating NCA deposits on the chip carrier, causing the stud bumps to be deformed when contacted by the PCB pads. In the meantime, both sides of the package are heated to provide the thermal energy to cure the NCA. The flip-chip bonder for this application must be capable of applying a wide range of forces and temperatures while keeping the placement accuracy of the die. Equipment is available that meets the requirements of this process with a

throughput of 100 dies per hour. Devices assembled using several selected NCAs showed acceptable reliability performance after thermal cycling and humidity aging tests. This technology is utilized in manufacturing to assemble sub-miniaturized electronic devices such as hearing aids and heart defibrillators [165].

6.4.1 Recent Developments on NCA Flip Chip

6.4.1.1 NCAs with Low CTE

The most commonly observed flip-chip failure occurred during the thermal cycling test, which is due to the thermal expansion mismatch between chips and substrates. Therefore, the problem of CTE mismatch between chips and substrates becomes serious in the NCAs flip-chip assembly because of high CTE of NCAs materials. For this reason, Yim et al. [166] developed NCAs with reduced CTE by introducing nonconducting low CTE filler to improve reliability of flip-chip assembly on organic substrates. Figure 6.26 shows the SEM of a NCA flip-chip joint. The addition of nonconducting filler noticeably affected the CTE of NCA materials and adhesion on flip-chip assemblies. It was found that the reliability of flip-chip assembly using modified NCA with optimized content of nonconducting filler was significantly improved.

Chang-Kyu Chung et al. [167] investigated the effects of multi-functional epoxy and the addition of silica fillers on thermomechanical properties of cured nonconductive films (NCFs) and thermal cycling reliability of NCFs flip-chip-on-organic board (FCOB) assemblies. It was found that FCOB assemblies using NCFs with multi-functional epoxy had better thermal cycling reliability than those using NCFs with di-functional epoxy. And 10 wt% and 20 wt% silica added NCFs showed the best thermal cycling reliability in the electroplated Au bump application and the stud Au bump application respectively. Consequently, thermomechanical properties of NCFs, especially T_g, should be improved and the amount of added silica fillers should be optimized for high thermal cycling reliability of NCFs FCOB assemblies.

Chuang et al. [168, 169] demonstrated the process feasibility a new NCA flip-chip bonding process—thermosonic bonding, and the bonding strength improvement of a flip-chip-on-flex (FCOF) assembly. Prior to flip-chip bonding, a nonconductive paste was deposited on the surface of copper electrodes over a flex substrate, and a chip with gold bumps was then flipped and thermosonically bonded onto copper pads on the flex substrate. The ultrasonic power plays an important role in removing the nonconductive paste beneath the gold bumps during bonding process, so that gold stud bumps could be directly bonded onto copper electrodes to achieve good electrical connections between chips and the flex substrate. The ultrasonic power provided necessary energy to form a metallurgical bonding between gold bumps and the copper electrodes. The bonding strength of chip on the flex substrate increases with increasing ultrasonic power to an appropriate

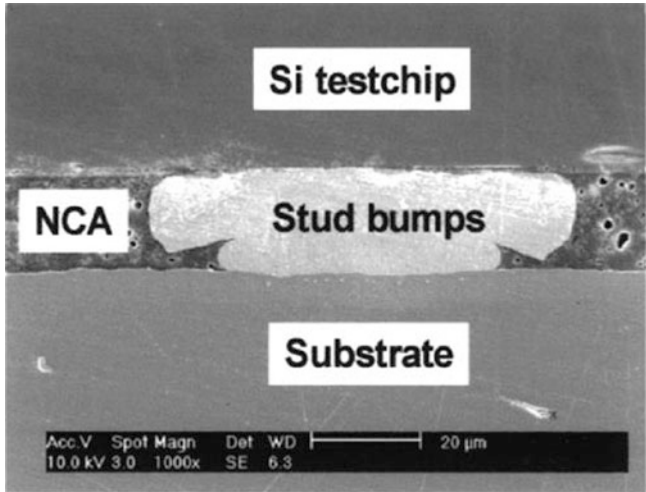


Fig. 6.26 SEM image of NCA flip-chip joint [166]

amount. Selecting an appropriate curing temperature is also essential in improving bonding strength between chips and flex-substrates. This process can be potentially applied to packaging of IC-chips to flex-substrates with copper pads, such as packaging of TFT LCD driver IC. Due to the metallurgical bonding between Au bump and the Cu electrode, the bonding strength was improved. Applying the adequate ultrasonic power to bonding process was not only to improve the bonding strength, but also the bonding strength could be maintained in high level after PCT (pressure cooker test) aging test. There was no significant change in bonding strength for chips bonded on flex substrates after TCT (thermal cycling test). The bonding strength increased with time during HT/HH test. Neither cracks nor defects at boding interface were observed by the authors. The reliability for chips bonded the flex substrate using thermosonic flip-chip bonding process with NCP meets the requirements stated in JEDEC specifications, exception that the adhesion strength of NCP after PCT requires further improvement.

6.4.1.2 Fine Pitch Chip-on-Flex by Pre-applied Wafer Level Adhesives

Chun-Chih Chuang et al. [170] developed a novel process which combined wafer level package technology with ultrafine pitch chip-on-flex (COF) by sidewall-insulated compliant bumps. Two types of adhesives, single-layer nonconductive adhesive (NCA) and double-layer of nonconductive adhesive/anisotropic conductive adhesive (NCA/ACA) were laminated on wafers, respectively. After wafers with laminated adhesives were diced into chips, thermocompression bonding process was executed by using high accuracy flip-chip bonder. Samples that passed electrical tests were then performed to reliability test under 85 °C/85%RH aging.

It was found that the assembled 20 μm-pitch sidewall-insulated compliant-bump-bonded COF packages with pre-applied NCA and NCA/ACA chips showed high reliability.

6.4.1.3 Fast Curing NCA

Frye et al. [171] recently developed a reliable, very fast curing and user-friendly NCA using a thermocompression bonding process. In order to prevent voiding during bonding, a low-viscosity formulation with moderate thixotropy was determined to give the best results and optimum flow. Fine-pitch devices require small, metal interconnects, so particle size control was identified as a critical characteristic. Among a variety of different NCAs with various chemistries, the authors discovered that free radical cure helped to achieve bonding time target—1 s below 300 °C. It was found that minimizing the release of volatiles during cure reduced void formation inside the cured NCA. The authors was successful in developing a new nonconductive paste (NCP) that offers fast bonding time and good MSL L3 reliability. The material passed 1,000 Thermal Cycles from −55 to 125 °C and passed 168 h of HAST (130 °C at 85 % relative humidity) testing.

6.4.1.4 NCAs Versus ACAs in Flex Circuits

Reliability of two different interconnection technologies—ACAs and NCAs, mainly to connect driver electronics to liquid crystal displays, were compared. The connection was made between flexible kapton circuits and rigid glass chip carriers and in one case to FR4 carriers. Tests were performed both on test structures and real displays. Anisotropic conductive film is extensively used in producing high-quality LCDs. Anisotropic conductive films contain a small volume fraction of conductive particles to accomplish electrical conductivity between the two connected contact areas. Nonconductive adhesive uses the surface roughness on the individual mating contact pads to establish electrical contact. In comparison with conventional anisotropic conductive films (ACF) based on thermoset adhesives, the use of nonconductive adhesive (NCA) proved to be superior with respect to reliability. This result was verified by thermal cycling (−30 to 90 °C) and humidity testing (85 °C/85 % RH), conducted on the test structures and real displays [172].

References

1. Cdenhead R, DeCoursey D (1985) Intl J Microelectron 8(3):14
2. Jagt JC (1998) IEEE Trans Compon Packag Manuf Tech A 21:215–225
3. Ogunjimi AO, Boyle O, Whalley DC, Williams DJ (1992) J Electron Manuf 2:109–118
4. Harris PG (1995) Soldering Surf Mount Technol 20:19–21
5. Gilleo K (1995) Soldering Surf Mount Technol 19:12–17

6. Liu J, Lai Z, Kristiansen H, Khoo C (1998) Proceedings of the 3rd international conference on adhesive joining & coating technology in electronics manufacturing. Binghamton, NY, pp 1–17
7. Corbett S, Dominano MJ (1997) Surf Mount Technol 48:48–52
8. Bolger J, Morano S (1984) Adhesives Age 17:17–20
9. Gilleo K (1995) Assembly with conductive adhesives. Soldering Surf Mount Technol (19):12–17
10. Hariss PG (1995) Conductive adhesives: a critical review of progress to date. Soldering Surf Mount Technol (20):19–21
11. Asai S, Saruta U, Tobita M, Takano M, Miyashita Y (1995) Development of an anisotropic conductive adhesive film (ACAF) from epoxy resins. J Appl Polym Sci 56:769–777
12. Chang DD, Crawford PA, Fulton JA, McBride R, Schmidt MB, Sinitski RE, Wong CP (1993) An overview and evaluation of anisotropically conductive adhesive films for fine pitch electronic assembly. IEEE Trans Compon Hybrids Manuf Technol 16(8):320–326
13. Ando H, Kobayashi N, Numao H, Matsubara Y, Suzuki K (1985) Electrically conductive adhesive sheet. European Patent 0,147,856
14. Gilleo K (1987) An isotropic adhesive for bonding electrical components. European Patent 0,265,077
15. Pennisi R, Papageorge M, Urbisch G (1992) Anisotropic conductive adhesive and encapsulant materials. US Patent 5,136,365
16. Date H, Hozumi Y, Tokuhira H, Usui M, Horikoshi E, Sato T (1994) Anisotropic conductive adhesives for fine pitch interconnections. In: Proceedings of ISHM'94, Bologna, Italy, Sept 1994, pp 570–575
17. Liu J (2000) ACA bonding technology for low cost electronics packaging applications-current status and remaining challenges. In: Proceedings of 4th international conference on adhesive joining and coating technology in electronics manufacturing, Helsinki, Finland, Jun 2000, pp 1–15
18. Wu CML, Liu J, Yeung NH (2000) Reliability of ACF in Flip chip with various bump height. In: Proceedings of 4th international conference on adhesive joining and coating technology in electronics manufacturing, Helsinki, Finland, Jun 2000, pp 101–106
19. Kishimoto Y, Hanamura K (1998) Anisotropic conductive paste available for flip chip. In: Proceedings of 3rd international conference on adhesive joining and coating technology in electronics manufacturing, Binghamton, NY, Sept 1998, pp 137–143
20. Sugiyama K, Atsumi Y (1991) Conductive connecting structure. US Patent 4,999,460, 12 Mar 1991
21. Sugiyama K, Atsumi Y (1992) Conductive connecting method. US Patent 5,123,986, 23 Jun 1992
22. Sugiyama K, Atsumi Y (1993) Conductive bonding agent and a conductive connecting method. US Patent 5,180,888, 19 Jan 1993
23. Nagle R (1998) Evaluation of adhesive based flip-chip interconnect techniques. Int J Microelectron Packag 1:187–196
24. Kivilahti JK (1999) Design and modeling of solder-filled ACAs for flip-chip and flexible circuit applications. In: Liu J (ed) Conductive adhesives for electronics packaging. Electrochemical Publications Ltd, Port Erin, British Isles, pp 153–183
25. Vuorela M, Holloway M, Fuchs S, Stam F, Kivilahti J (2000) Bismuth-filled anisotropically conductive adhesive for flip-chip bonding. In: Proceedings of 4th international conference on adhesive joining and coating technology in electronics manufacturing, Helsinki, Finland, Jun 2000, pp 147–152
26. Torii A, Takizawa M, Sawano M (1998) The application of flip chip bonding technology using anisotropic conductive film to the mobile communication terminals. In: Proceedings of international electronics manufacturing technology/international microelectronics conference, Tokyo, Japan, Apr 1998, pp 94–99

27. Atarashi H (1990) Chip-on-glass technology using conductive particles and light-setting adhesives. In: Proceedings of the 1990 Japan international electronics manufacturing technology symposium, Tokyo, Japan, Jun 1990, pp 190–195
28. Matsubara H (1992) Bare-chip face-down bonding technology using conductive particles and light-setting adhesives. In: Proceedings of international microelectronics conference, Yokohama, Japan, pp 81–87
29. Endoh K, Nozawa K, Hashimoto N (1993) Development of 'The Maple Method'. In: Proceedings of Japan international electronics manufacturing technology symposium, Kanazawa, Japan, pp 187–191
30. Sihlbom R, Dernevik M, Lai Z, Starski JP, Liu J (1998) Conductive adhesives for high-frequency applications. IEEE Trans Compon Packag Manuf Tech A 20(3):469–477
31. Dernevik M, Sihlbom R, Axelsson K, Lai Z, Liu J, Starski P (1998) Electrically conductive adhesives at microwave frequencies. In: Proceedings of 48th IEEE electronic components & technology conference, Seattle, Washington, May 1998, pp 1026–1030
32. Yim MJ, Ryu W, Jeon YD, Lee J, Kim J, Paik K (1999) Microwave model of anisotropic conductive adhesive flip-chip interconnections for high frequency applications. In: Proceedings of 49th electronic components and technology conference, San Diego, CA, May 1999, pp 488–492
33. Gustafsson K, Mannan S, Liu J, Lai Z, Whalley D, Williams D (1997) The effect on ramping rate on the flip chip joint quality and reliability using anisotropically conductive adhesive film on FR4 substrate. In: Proceedings of 47th electronic components and technology conference, San Jose, CA, May 1997, pp 561–566
34. Connell G (1997) Condutive adhesive flip chip bonding for bumped and unbumped die. In: Proceedings of 47th electronic components and technology conference, San Jose, CA, May 1997, pp 274–278
35. Oguibe CN, Mannan SH, Whalley DC, Williams DJ (1998) Flip-chip assembly using anisotropic conducting adhesives: experimental and modelling results. In: Proceedings of 3rd international conference on adhesive joining and coating technology in electronics manufacturing, Binghamton, NY, Sept 1998, pp 27–33
36. Hirai H, Motomura T, Shimada O, Fukuoka Y (2000) Development of flip chip attach technology using Ag paste bump which formed on printed wiring board electrodes. In: Proceedings of international symposium on electronic materials & packaging, Hong Kong, China, Nov–Dec 2000, pp 1–6
37. Hotta Y, Maeda M, Asai F, Eriguchi F (1998) Development of 0.025 mm pitch anisotropic conductive film. In: Proceedings of 48th IEEE electronic components & technology conference, Seattle, Washington, May 1998, pp 1042–1046
38. Chan YC, Hung KC, Tang CW, Wu CML (2000) Degradation mechanisms of anisotropic conductive adhesive joints for flip chip on flex applications. In: Proceedings of 4th international conference on adhesive joining and coating technology in electronics manufacturing, Helsinki, Finland, Jun 2000, pp 141–146
39. Kristiansen H, Liu J (1998) Overview of conductive adhesive interconnection technologies for LCDs. IEEE Trans Compon Packag Manuf Tech A 21(2):208–214
40. Moon K, Dong H, Maric R, Pothukuchi S, Hunt A, Li Y, Wong CP (2005) J Electron Mater 34:132–139
41. Efremov MY, Schiettekatte F, Zhang M, Olson EA, Kwan AT, Berry RS, Allen LH (2000) Phys Rev Lett 85:3560–3563
42. Li Y, Moon K, Wong CP (2006) J Appl Polym Sci 99:1665–1673
43. Li Y, Moon K, Wong CP (2004) Proceedings of 54th IEEE electronic components and technology conference, Las Vegas, NV, pp 1968–1974
44. Li Y, Wong CP (2005) Proceedings of 55th IEEE electronic components and technology conference, Lake Buena, FL, pp 1147–1154
45. Li Y, Moon K, Wong CP (2005) J Electron Mater 34:266–271
46. Li Y, Moon K, Wong CP (2005) J Electron Mater 34:1573–1578

47. Davies G, Sandstrom J (1976) Circuits Manufacturing 56–62
48. Harsanyi G, Ripka G (1985) Electrocompon Sci Technol 11:281–290
49. Di Giacomo G (1992) In: McHardy J, Ludwig F (eds) Electrochemistry of semiconductors and electronics: processes and devices. Noyes Publications, Park Ridge, NJ, pp 255–295
50. Manepalli R, Stepniak F, Bidstrup-Allen SA, Kohl P (1999) IEEE Trans Adv Packag 22:4–8
51. Di Giacomo G (1997) Reliability of electronic packages and semi-conductor devices, Chapter 9. McGraw-Hill, New York, NY
52. Klein Wassink RJ (1987) Microelectron Int 9:9–12
53. Shirai Y, Komagata M, Suzuki K (2001) Proceedings of the 1st international IEEE conference on polymers and adhesives in microelectronics and photonics, Potsdam Germany, pp 79–83
54. Schonhorn H, Sharpe LH (1983) US Patent 4,377,619
55. Brusic V, Frankel GS, Roldan J, Saraf R (1995) J Electrochem Soc 142:2591–2594
56. Wang PI, Lu TM, Murarka SP, Ghoshal R (2007) US Patent 7,285,842
57. Li Y, Wong CP (2005) US Patent pending
58. Li Y, Wong CP (2006) Appl Phys Lett 89:112
59. Toshioka H, Kobayashi M, Koyama K, Nakatsugi K, Kuwabara T, Yamamoto M, Kashihara H (2006) SEI Tech Rev 62:58–61
60. Lieber CM (2001) Science 293:1289–1292
61. Prinz GA (1998) Science 282:1660–1663
62. Martin CR, Menon VP (1995) Anal Chem 67:1920–1928
63. Xu JM (2001) Appl Phys Lett 79:1039–1041
64. Russell TP (2000) Science 290:2126–2129
65. Lin R, Hsu Y, Chen Y, Cheng S, Uang R (2005) Proceedings of 55th IEEE electronic components and technology conference, Lake Buena, FL, pp 66–70
66. Li Y, Moon K, Wong CP (2006) Proceedings of 56th IEEE electronic components and technology conference, San Diego, CA, pp 1239–1245
67. Li Y, Zhang Z, Moon K, Wong CP (2006) US Patent pending
68. Jana PB, Chaudhuri S, Pal AK, De SK (1992) Polym Eng Sci 32:448–456
69. Malliaris A, Turner DT (1971) Influence of particle size on the electrical resistivity of compacted mixtures of polymers and metallic powders. J Appl Phys 42:614–618
70. Ruschau GR, Yoshikawa S, Newnham RE (1992) Resistivities of conductive composites. J Appl Phys 73(3):953–959
71. Agar JC, Lin KJ, Zhang R, Durden J, Lawrence K, Moon K-S, Wong CP (2010) ECTC, pp 1713–1718
72. Jagt JC (1998) Reliability of electrically conductive adhesive joints for surface mount applications: a summary of the state of the art. IEEE Trans Compon Packag Manuf Technol A 21(2):215–225
73. Lutz MA, Cole RL (1990) High performance electrically conductive adhesives. Hybrid Circuits (23):27–30
74. Pujol JM, Prud'homme C, Quenneson ME, Cassat R (1989) Electroconductive adhesives: comparison of three different polymer matrices. Epoxy, polyimide, and silicone. J Adhesion 27:213–229
75. Gonzales JIJ, Mena MG (1997) Moisture and thermal degradation of cyanate-ester-based die attach material. In: Proceedings of 47th electronic components and technology conference, San Jose, CA, May 1997, pp 525–535
76. Chien IY, Nguyen MN (1994) Low stress polymer die attach adhesive for plastic packages. In: Proceedings of 1994 electronic components and technology conference, San Diego, May 1994, pp 580–584
77. Galloway DP, Grosse M, Nguyen MN, Burkhart A (1995) Reliability of novel die attach adhesive for snap curing. In: Proceedings of the IEEE/CPMT international electronic manufacturing technology (IEMT) symposium, Austin, TX, Oct 1995, pp 141–147

78. Keusseyan RL, Dilday JL (1994) Electric contact phenomena in conductive adhesive interconnections. Int J Microcircuits Electron Packag 17(3):236–242
79. Antoon MK, Koenig JL (1981) Fourier-transform infrared study of the reversible interaction of water and a crosslinked epoxy matrix. J Polym Sci (Physics) 19:1567–1575
80. Antoon MK, Koenig JL (1981) Irreversible effects of moisture on the epoxy matrix in glass-reinforced composites. J Polym Sci (Physics) 19:197–212
81. Khoo CGL, Liu J (1996) Moisture sorption in some popular conductive adhesives. Circuit World 22(4):9–15
82. Pandiri SM (1987) The behavior of silver flakes in conductive epoxy adhesives. Adhesives Age 30:31–35
83. Günther B, Schäfer H (1996) Porous metal powders for conductive adhesives. In: Proceedings of the 2nd international conference on adhesive joining & coating technology in electronics manufacturing, Stockholm, Sweden, Jun 1996, pp 55–59
84. Kotthaus S, Gunther BH, Haug R, Schafer H (1996) Study of isotropically conductive adhesives filled with aggregates of nano-sized Ag-particles. In: Proceedings of the 2nd international conference on adhesive joining & coating technology in electronics manufacturing, Stockholm, Sweden, Jun 1996, pp 14–17
85. Pramanik PK, Khastgir D, De SK, Saha TN (1990) Pressure-sensitive electrically conductive nitrile rubber composites filled with particulate carbon black and short carbon fibre. J Mater Sci 25:3848–3853
86. Jana PB, Chaudhuri S, Pal AK (1992) Electrical conductivity of short carbon fiber-reinforced poly-chloroprene rubber and mechanism of conduction. Polym Eng Sci 32(6):448–456
87. Chaudhari VR, Haram SK, Kulshreshtha SK (2007) Colloids Surf A 301:475–480
88. Pal A, Shah S, Devi S (2007) Colloids Surf A 302:51–57
89. Chen Z, Gao L (2007) Mater Res Bull 42:1657–1661
90. Kumar A, Joshi H, Pasricha R, Mandale AB, Sastry M (2003) J Colloid Interface Sci 264:396
91. Guzmán MG, Dille J, Godet S (2008) Synthesis of silver nanoparticles by chemical reduction method and their antibacterial activity. Proc World Acad Sci Eng Technol 33:367–374
92. Hu Z, Xu T, Liu R, Li H (2004) Mater Sci Eng A 371:236–240
93. Sun Y, Yin Y, Mayers B, Herricks T, Xia Y (2002) Chem Mater 14:4736–4745
94. Hernandez EA, Posada B, Irizarry R, Castro ME (2004) A new wet chemical approach for selective synthesis of silver nanowires. NSTI Nanotech 2004 3:156–158
95. Korte KE, Skrabalak SE, Xia Y (2008) Rapid synthesis of silver nanowires through a CuCl- or CuCl2-mediated polyol process. J Mater Chem 18:437–441
96. Yokoyama A, Katsumata T, Fujii A, Yoneyama T (1992) New copper paste for CTF applications. In: IMC Proceedings, pp 376–338
97. Yim MJ et al (2007) Proceedings of electronic components and technology conference (ECTC), p 82
98. Nishikawa H (2008) 2nd Electronics system integration technology conference, p 825
99. Kang SK, Rai R, Purushothaman S (1997) Development of high conductivity lead (Pb)-free conducting adhesives. In: Proceedings of 47th electronic components and technology conference, San Jose, CA, May 1997, pp 565–570
100. Kang SK, Rai R, Purushothaman S (1998) Development of high conductivity lead (Pb)-free conducting adhesives. IEEE Trans Compon Packag Manuf Technol A 21(1):18–22
101. Lin JK, Drye J, Lytle W, Scharr T, Subrahmanyan R, Sharma R (1996) Conductive polymer bump interconnects. In: Proceedings of 46th electronic components and technology conference, Orlando, FL, May 1996, pp 1059–1068
102. Seidowski T, Kriebel F, Neumann N (1998) Polymer flip chip technology on flexible substrates-development and applications. Proceedings of 3rd international conference on adhesive joining and coating technology in electronics manufacturing, Binghamton, NY, Sept 1998, pp 240–243
103. Estes RH (1998) Process and reliability characteristics of polymer flip chip assemblies utilizing stencil printed thermosets and thermo-plastics. In: Proceedings of 3rd international

conference on adhesive joining and coating technology in electronics manufacturing, Binghamton, NY, Sept 1998, pp 229–239
104. Oh KE (1999) Flip chip packaging with micromachined conductive polymer bumps. IEEE J Select Topics Quantum Electron 5(1):119–126
105. Lohokare SK, Lu Z, Schuetz CA, Prather DW (2006) Electrical characterization of flip-chip interconnects formed using a novel conductive-adhesive-based process. IEEE Trans Adv Packag 29(3):542–547
106. Gaynes M, Kodnani R, Pierson M, Hoontrakul P, Paquette M (1998) Flip chip attach with thermoplastic electrically conductive adhesive. In: Proceedings of 3rd international conference on adhesive joining and coating technology in electronics manufacturing, Binghamton, NY, Sept 1998, pp 244–251
107. Bessho Y (1990) Chip on glass mounting technology of lysis for LCD module. In: Proceedings of international microelectronics conference, May 1990, pp 183–189
108. Clayton JE (2003) Proceedings of international symposium on microelectronics, Boston, MA, 16–20 Nov 2003, pp 1–7
109. Chang YW, Chiang TH, Chih Chena (2007) Appl Phys Lett 91:132113
110. Kong J, Yenilmez E, Tombler TW, Kim W, Dai H (2001) Phys Rev Lett 87:106801
111. Yao Z, Kane CL, Dekker C (2000) Phys Rev Lett 84:2941–2944
112. Nihei M, Horibe M, Kawabata A, Awano Y (2004) Proceedings of the IEEE interconnect technology conference, pp 251–253
113. Nihei M, Kondo D, Kawabata A, Sato S, Shioya H, Sakaue M, Iwai T, Ohfuti M, Awano Y (2005) Proceedings of the IEEE interconnect technology conference, pp 234–236
114. Kreupl F, Graham AP, Duesberg GS, Steinhögl W, Liebau M, Unger E, Höenlein W (2002) Microelectron Eng 64:399–408
115. Li J, Stevens R, Delzeit L, Ng HT, Cassell A, Han J, Meyyappan M (2002) Appl Phys Lett 81:910–912
116. Nihei M, Horibe M, Kawabata A, Awano Y (2004) Jpn J Appl Phys 43:1856–1859
117. Terrones M, Grobert N, Olivares J, Zhang JP, Terrones H, Kordatos K, Hsu WK, Hare JP, Townsend PD, Prassides K, Cheetham AK, Kroto HW, Walton DRM (1997) Nature 388:52–55
118. Pan ZW, Xie SS, Chang BH, Wang CY, Lu L, Liu W, Zhou WY, Li WZ, Qian LX (1998) Nature 394:631–632
119. Murakami Y, Maruyama S (2004) Chem Phys Lett 385:298–303
120. Fujii T, Someya M (2006) US Patent 7,150,801
121. Ikuo Soga, Daiyu Kondo, Yoshitaka Yamaguchi, Taisuke Iwai, Masataka Mizukoshi, Yuji Awano, Kunio Yube, and Takashi Fujii (2008) Proceedings of electronic components and technology conference, pp 1390–1394
122. Eikelboom DWK, Bultman JH, Schönecker A, Meu-wissen MHH, Van Den Nieuwenhof MAJC, Meier DL (2002) Conductive adhesives for low-stress interconnection of thin back-contact solar cells. In: 29th IEEE photovoltaic specialists conference, May 2002, pp 403–406
123. Prasad SK (2004) Advanced wirebond interconnection technology. Springer, New York
124. Harman GG (1997) Wirebonding in microelectronics: materials processes reliability and yield, 2nd edn. McGraw Hill, New York
125. Carson F (2007) Advanced 3D packaging and interconnect schemes. In: Ku-licke and Soffa symposium at Semicon, San Francisco, CA
126. Andrews LD, Caskey TC, McElrea SJS (2007) 3D Electrical interconnection using extrusion dispensed conductive adhesives. In: International electronics manufacturing technology symposium, pp 96–100
127. http:Hwww.formfactor.com/
128. Kataoka K, Kawamura S, Itoh T, Suga T, Ishikawa K, Honma H (2002) Low contact-force and compliant mems probe card utilizing fritting contact. In: Proceedings 15th International conference on micro electro mechanical systems (MEMS'02), Las Vegas, 20–24 Jan 2002, pp 364–367

129. Kataoka K, Itoh T, Inoue K, Suga T (2004) Multi-layer electro-plated micro-spring array for MEMS probe card. In: Proceedings 17th international conference on micro electro mechanical systems (MEMS'04), Maastricht, 25–29 Jan 2004, pp 733–736
130. Smith DL Alimonda AS (1996) A new flip-chip technology for high-density packaging. In: Proceedings 46th electronic components and technology conference, Orlando, 28–31 May 1996, pp 1069–1073
131. Chow EM, Chua C, Hantschel T, van Schuylenbergh K, Fork DK (2005) Solder-free pressure contact micro-springs in high-density flip-chip packages. In: Proceedings 55th electronic components and technology conference, Lake Buena Vista, FL, May 31–June 3 2005, pp 1119–1126
132. Itoh T, Kataoka K, Suga T (2006) Fabrication of microspring probes using conductive paste dispensing. In: Proceedings 19th international conference on micro electro mechanical systems (MEMS'06), Istanbul, 22–26 Jan 2006, pp 258–261
133. Itoh T, Suga T, Kataoka K (2007) Microstructure fabrication with conductive paste dispensing. In: Proceedings of the 2nd IEEE international conference on nano/micro engineered and molecular systems, Bangkok, Thailand, Jan 2007, pp 1003–1006
134. Felba J, Friedel KP, Moscicki A (2000) Characterization and performance of electrically conductive adhesives for micro-wave applications. In: Proceedings of 4th international conference on adhesive joining and coating technology in electronics manufacturing, Helsinki, Finland, Jun 2000, pp 232–239
135. Liong S, Zhang Z, Wong CP (2001) High frequency measurement for isotropically conductive adhesives. In: Proceedings of 51th electronic components and technology conference, Orlando, FL, May 2001, pp 1236–1240
136. Hashimoto K, Akiyama Y, Otsuka K (2008) Transmission characteristics in GHz region at the conductive adhesive joints. In: Proceedings of electronic components and technology conference, pp 2067–2072
137. Jon B, Lai Z, Liu J (2000) IEEE Tran Adv Packag 23(4):743
138. Constable JH, Kache T, Teichmann H, Muhle S, Gaynes MA (1999) Continuous electrical resistance monitoring, pull strength, and fatigue life of isotropically conductive adhesive joints. IEEE Trans Compon Packag Technol 22(2):191–199
139. Gomatam R, Sancaktar E, Boismier D, Schue D, Malik I (2001) Behavior of electrically conductive adhesive filled adhesive joints under cyclic loading, part I: experimental approach. In: Proceedings of 4th international symposium and exhibition on advanced packaging materials, processes, properties and interfaces, Braselton, GA, Mar 2001, pp 6–12
140. Yamashita M, Suganuma K (2002) Degradation mechanism of conductive adhesive/Sn-Pb plating interface by heat exposure. J Electron Mater 31:551–556
141. Xu S, Dillard DA, Dillard JG (2003) Environmental aging effects on the durability of electrically conductive adhesive joints. Int J Adhesion Adhesives 23:235–250
142. Kuusiluoma S, Kiilunen J (2005) The reliability of isotropically con-ductive adhesive as solder replacement—a case study using LCP substrate. In: Proceedings of electronic packaging and technology conference (EPTC), pp 774–779
143. Duraj, Mach P (2006) Stability of electrical resistance of isotropic conductive adhesives within mechanical stress. In: International conference on applied electronics, Pilsen, Sept 2006
144. Jeahuck L, Cho CS, Morris JE (2007) Proceedings of international conference on electronic materials and packaging, 19–22 Nov 2007, pp 1–4
145. Wu HP, Wu XJ, Liu JF, Zhang GQ, Wang YW (2005) Development of a novel isotropic conductive adhesive filled with silver nanowires. J Composite Mater 40(21):1961–1968
146. Leea HH, Choua KS (2005) Effect of nano-sized silver particles on the resistivity of polymeric conductive adhesives. Int J Adhesion Adhesives 25:437–441
147. Ye L, Lai Z, Johan L, Tholen A (1999) Effect of Ag particle size on electrical conductivity of isotropically conductive adhesives. IEEE Trans Electron Packag Manuf 22(4):299–302

148. Fan L, Su B, Qu J, Wong CP (2004) Electrical and thermal conductivities of polymer composites containing nano-sized particles. In: Proceedings of electronic components and technology conference, Las Vegas, NV, pp 148–154
149. Mach P, Radev R, Pietrikova A. (2008) Electrically conductive adhesive filled with mixture of silver nano and microparticles. In: Proceedings of 2nd electronics system integration technology conference, 2008, pp 1141–1146
150. Jiang H, Moon KS, Li Y, Wong CP (2006) Surface functionalized silver nanoparticles for ultrahigh conductive polymer composites. Chem Mater 18–13:2969–2973
151. Kotthous S, Günther BH, Hang R, Schafer H (1997) Study of isotropically conductive bondings filled with aggregates of nano-sized Ag-particles. IEEE Trans Compon Packag Manuf Technol A 20(1):15–20
152. Majima M, Koyama K, Tani Y, Toshioka H, Osoegawa M, Ka-shihara H, Inazawa S (2002) Development of conductive material using metal nano particles. SEI Tech Rev 54:25–27
153. Iijima S (1991) Helical microtubules of graphitic carbon. Nature 354:56
154. Berber S, Kwon YK, Tománek D (2000) Unusually high thermal conductivity of carbon nanotubes. Phys Rev Lett 84(20):4613–4616
155. Yu MF, Files BS, Arepalli S, Ruoff RS (2000) Tensile loading of ropes of single wall carbon nanotubes and their mechanical prop-erties. Phys Rev Lett 84(24):5552–5555
156. Gao G, Cagin T, Goddard WA (1998) Energetics, structure, mechanical and vibrational properties of single walled carbon nano-tubes (SWNT). Nanotechnology 9:184–191
157. Wu HP, Wu XJ, Ge MY, Zhang GQ, Wang YW, Ji-ang JZ (2007) Properties investigation on isotropical conductive adhesives filled with silver coated carbon nanotubes. Composites Sci Technol 67:1182–1186
158. Qian D, Dickey EC, Andrews R, Rantell T (2000) Load transfer and deformation mechanisms in carbon nanotubepolystyrene composites. Appl Phys Lett 76:2868
159. Lin X, Lin F (2004) Improvement on the properties of silver-containing conductive adhesives by the addition of carbon nanotube. In: Proceedings of high density microsystem design and packaging, Shanghai, China, pp 382–384
160. Rutkofsky M, Banash M, Rajagopal R, Jian C (2005) Using a carbon nanotube additive to make electrically conductive commercial polymer composites. SAMPE Journal 41(2):54–55
161. Lin W, Xiu Y, Jiang H, Zhang R, Hildreth O, Moon K, Wong CP (2008) Self-assembled-monolayer-assisted chemical transfer of in-situ functionalized carbon nanotubes. J Am Chem Soc 130(30):9636–9637
162. Lin W, Moon K, Wong CP (2009) A combined process of in-situ functionalization and microwave treatment to achieve ultra-small thermal expansion of aligned carbon nanotube/polymer nanocomposites: toward applications as thermal interface materials. Adv Mater 21(23):2421–2424
163. Hatada K, Fujimoto H, Kawakita T, Ochi T (1988) A new LSI bonding technology: micron bump bonding assembly technology. In: Proceedings fifth IEEE/CHMT international electronic manufacturing technology symposium, Orlando, FL, pp 45–49
164. Hatada K, Fujimoto K, Ochi T, Ishida Y (1989) Applications of new assembly method ‘Micron Bump Bonding Method’. In: Proceedings of ‘89 IEEE/CHMT Japan international electronic manufacturing technology symposium, Nara, Japan, Apr 1989, pp 45–48
165. Ferrando F, Zeberli J-F, Clot P, Chenuz J-M (2000) Proceedings of 4th international conference on adhesive joining and coating technology in electronics manufacturing, Helsinki, Finland, Jun 2000, pp 205–211
166. Myung-Jin Yim, Jin-Sang Hwang, Woonseong Kwon, Kyung Woon Jang, Kyung-Wook Paik (2003) IEEE Trans Electron Packag Manuf 26(2):150–155
167. Chang-Kyu Chung, Kyung-Wook Paik (2007) Proceedings of 57th electronic components and technology conference, Reno, NV, pp 1831–838
168. Cheng-Li Chuang, Jong-Ning Aoh, Qing-An Liao, Shi-Jie Liao, Guo-Shing Huang (2008), Proceedings of international conference on electronic materials and packaging, Oct 2008, pp 208–211

169. Cheng-Li Chuang, Jong-Ning Aoh, Wei-How Chen (2009) Proceedings of international conference on electronic packaging technology & high density packaging, Aug 2009, pp 725–732
170. Chun-Chih Chuang, Su-Tsai Lu, Tao-Chih Chang, Kyoung-Lim Suk, Kyung-Wook Paik (2009) Proceedings of 4th international microsystems, packaging, assembly and circuits technology conference, Oct 2009, pp 56–59
171. Frye D, Guino R, Gupta S, Sano M, Sato K, Iida K (2010) Proceedings of electronic components and technology conference, pp 427–430
172. Kristiansen H, Bjorneklett A (1992) J Electron Manuf 2(1):7–12

Chapter 7
Substrate Technology

Yutaka Tsukada

Abstract A flip chip package started with a ceramic substrate at the beginning. A Low CTE chip (silicon: 3 ppm/°C) and a relatively low CTE substrate (alumina: 8 ppm/°C) with a ductile solder joint (high lead solder) provided a good reliable system for high density packaging. However, it has disadvantages such as high cost, low electrical property, largeness, and heaviness. In early 1990s, an organic substrate utilizing an epoxy base printed circuit board technology has emerged. An underfill resin fills a gap between a chip and substrate and resolves a stress issue by a large CTE mismatch of a chip and an organic substrate (17 ppm/°C). In this system, the stress at a flip chip joint is dispersed in to the total package entity. This technology opened a door to low cost, high electrical property, small and light package, and has spread throughout semiconductor packages. In this chapter, a major focus is put on the organic substrate technology. The material, process, and reliability influences to a package are described in detail. Particularly, it is emphasized that the basics why the organic substrate is designed in such a way as we see today. Though there are many variations in the organic substrate technology, the basics are common and very important for a productive and reliable package, and the future progress of a flip chip package is strongly dependent on the progress of such basics.

7.1 Introduction

In the beginning, the substrate of flip chip package was a ceramic material. The practical use of it started with SLT (Solid Logic Technology) package that was employed on IBM System/360 announced in 1964. A substrate size was a half inch square with PGA (pin grid array) configuration. A size of substrate has been enlarged

Y. Tsukada (✉)
i-PACKS,
e-mail: yutsukada@aol.com

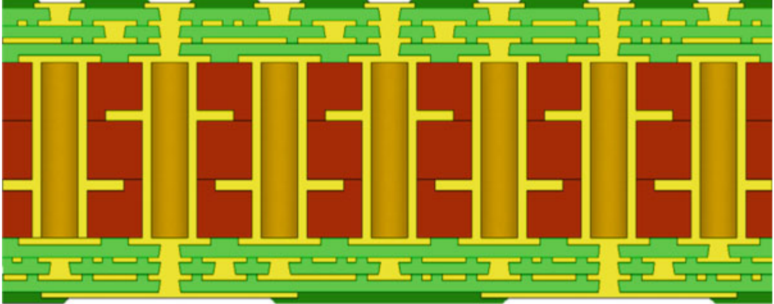
H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_7, © Springer Science+Business Media New York 2013

within the next generation from a half inch square to the most popular size 35 mm square PGA that was MST (Monolithic System Technology) package. Its first version was announced in 1970. There were further larger sizes according to requirements, but the 35 mm PGA was predominantly used from the middle of 1970s and 1980s. Not only a size increase but also a multilayer construction has come in to satisfy a density increase requirement for higher system performance. The major peak of such an innovation was the emergence of TCM (thermal conduction module) employed in IBM System/308X series, 35 layers with 90 mm square substrate at the starting point that carried 131 chips at maximum case. The property and nature of ceramic substrate were well explored and good descriptions were provided in the past. One of the examples was a book "Microelectronics Packaging Handbook" edited by Tummala and Ymaszewski [1]. Due to such assets through history, this chapter does not touch in detail the description of ceramic substrates except some cases for comparison purpose, and puts emphasis on organic substrates.

Within the age of ceramic substrates, a flip chip technology was a valuable asset of packaging but usage was limited mainly in high performance area due to its high cost. In the meantime, a requirement for system usage has been shifted from high end to low end along with spread of network throughout the world, i.e., the so-called wave of computer downsizing. A major event that responded to such stream was the emergence of organic substrate for flip chip packaging technology announced from Yasu Technology Application Laboratory, IBM Japan in 1991 [2]. It consisted of two parts. One was an underfill reinforced flip chip bonding technology that enabled to attach bare chip to epoxy base substrate that has large CTE (*coefficient of thermal expansion*) mismatch to silicon chip [3] and the other was a build-up PCB substrate technology that enabled to accept high I/O of flip chip by the structure similar to a semiconductor with blind via hole, so called micro-via hole, for fine pitch wiring connection [4]. Today, after two decades from the emergence, these technologies are widely spread in low cost, small and light and high performance areas such as supercomputers. Also, significant numbers of derivative technologies are used in many applications. A production of total organic substrate group has reached to 20 % of world PCB production at the end of 2010 in financial basis. However, the technology is still exploring to adopt various situations. This chapter is intended to describe its basics and backgrounds to understand the nature of technology that makes it possible to handle future applications without facing a major reliability issue and with better efficiencies.

7.2 Type of Construction

There are basically two types of structure in organic micro-via substrate. One is a sequential build-up type shown in Fig. 7.1a and the other is Z-stack type shown in Fig. 7.1b.



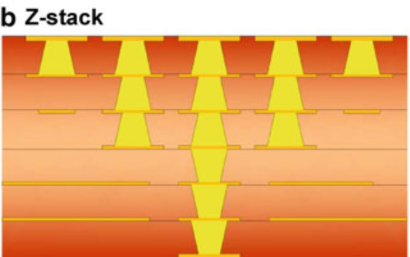


Fig. 7.1 Build-up substrate variations. (**a**) Sequential build-up, (**b**) Z-stack

7.2.1 Sequential Build-Up Structure

A sequential build-up type has a core at the center that employs ordinary PCB technology and build-up layers on both side of the core that employs micro-via hole for layer to layer connection. The role of core is to provide mechanical rigidity and power layers with containing thick Cu conductor plane. Build-up layers are processed sequentially put over on both side the core to provide high density wiring for flip chip bonding. Since the core provides mechanical stability during the fabrication process, a sequential build-up type has higher dimensional stability that is a key to achieve high density wiring. A weak point of sequential build-up type is at through holes that connect front and backside. The pitch of through hole is relatively low density since the bare hole is made by a mechanical drilling process. As the result, utilization of backside wiring layers is relatively low compared with front side though layers cannot be reduced to avoid an unbalanced structure that creates a warpage. A warpage is a common issue in organic substrate technologies due to low modulus of material and becomes an obstacle for assembly to the next level. Another weak point is an increasing process yield impact along with an increasing layer count. It is a predestinate nature of a sequential build-up. The total process yield of build-up layer fabrication follows a power-law degradation of each layer yield as formula (7.1).

$$\text{Through yield} = \text{layer 1 yield} \times \text{layer 2 yield} \times \cdots \times \text{layer } N \text{ yield} \tag{7.1}$$

where each layer field is an average of front side and back side yield.

For example, when the construction is "3 build-up layer + core + 3 build-up layer," the total yield is core yield × (average build-up layer yield)^3. By this nature, the yield degrades exponentially as increase of layer count. Therefore, reducing layer count by utilizing high wiring density per layer is a key to reduce the a cost of sequential build-up substrate.

7.2.2 Z-Stack Structure

Z-stack type is called as "any layer via" substrate. Each layer is fabricated with micro-via and a conductor pattern beforehand and stacked necessary layer numbers by press process to complete a substrate. The structure is uniform that can provide a good electrical property. The process is rather simple and possible to achieve low cost. However due to stacking process by high pressure mechanical press required, the alignment capability of circuit pattern and via hole is worse compared with a sequential build-up type. As the result, wiring density per plane is lower than that of sequential build-up type and layer numbers are increased. Other difficulty is that there is a difference in numbers of stacking count of via hole in a substrate. It creates stacking process window narrower. Since a dielectric layer is soft compared with metal in a via hole, a pressure at interface of low numbers stacking via is less compared with high numbers stacking via where a via to via contact force reaches earlier to the required level. It requires more pressure to low numbers stacking via reaches to the required level that impacts the registration degradation further.

The basic natures of fabrication process of these structures are a semiconductor wiring like structure in case of a sequential build-up type and an extension of conventional PCB but using micro-via in case of a Z-stack type. Such natures deliver a difference in a primary application such as a sequential build-up type for high density processor and ASIC chips and a Z-stack type mostly for consumer products.

7.3 Sequential Build-Up Substrate

Figure 7.2 shows a cross section photograph of a typical flip bonding on organic substrate. The substrate has a core at the center and high density build-up layers with micro-via hole are on both side of the core. The core provides power planes and the build-up layers provide flip chip fan-out wirings. On the surfaces of substrate, there are terminal pads for flip chip bonding on the front side and BGA

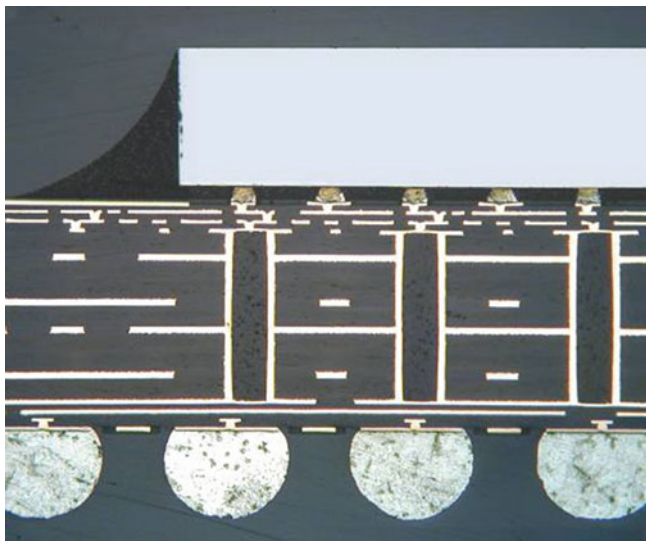


Fig. 7.2 Typical organic package cross section. After Tsukada, [5]

terminal pads on the back side. A gap between chip and substrate is filled with underfill resin to protect flip chip joints.

7.3.1 Process Flow

Figure 7.3 describes a fabrication process step of sequential build-up substrate. The figure shows a single side only but actual processes are implemented on front and back side simultaneously.

(a) The process starts from making a core with an ordinary PCB structure. Plated through holes are filled with resin contains some silica filler and capped by copper plating so that a via hole of the next layer can be placed directly on through hole.

(b) On both side of the core, a dielectric resin is applied by vacuum laminator and cured in a half way. Via hole are drilled by a laser. The surface of resin and inside of via hole is treated by permanganate to provide necessary roughness for copper plating and, at the same time, clean a resin smear in a via hole after laser drilling.

(c) A seed layer is formed with an electro-less copper plating on the surface and inside of via hole. A Pattern resist is applied, expose and develop to form a circuit pattern negatively. Then, an electro-copper plating is applied to make a conductor pattern. Via holes are plated at the same time. Pattern resist is removed after forming a conductor pattern. The dielectric resin is finally cured.

(d)–(g) Process is repeated to make necessary numbers of build-up layer.

(h) A solder mask resin is applied to cover entire surface and cured in a half way.

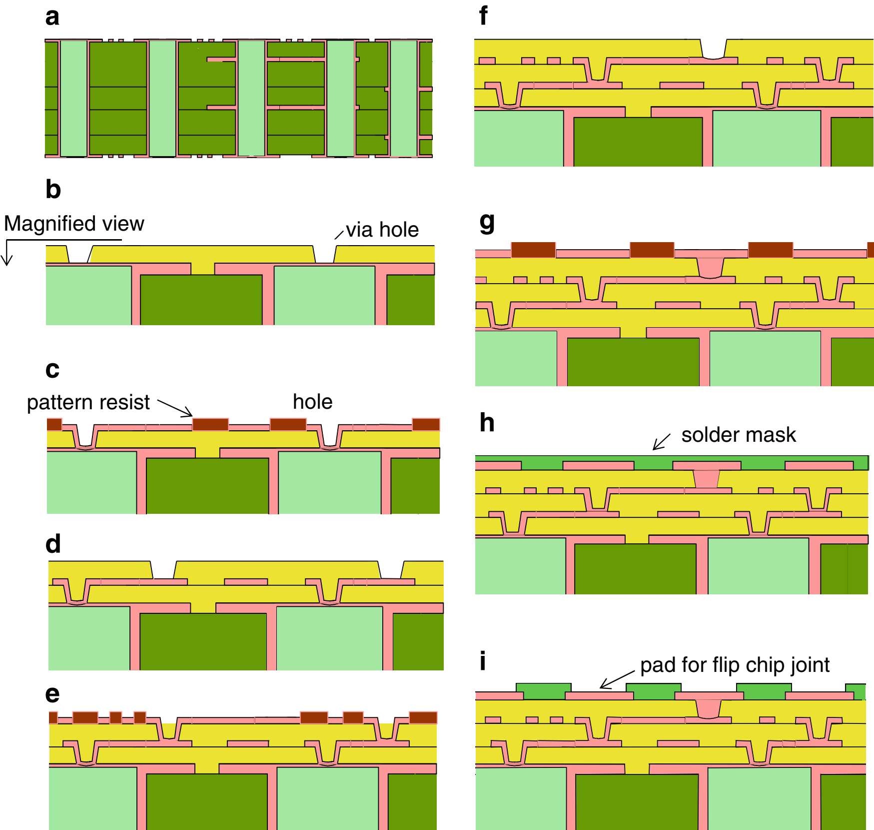


Fig. 7.3 Process flow of sequential build-up

(j) Pads for flip chip joint are formed by photo-etching of solder mask resin. A solder mask resin is cured completely.

7.3.2 Conductor Line

Since an organic substrate employs high definition photo-circuitizing process with a plated copper conductor, wiring density of organic substrate per layer is higher than ceramic substrate that employs a paste printing for conductor patterning. Figure 7.4 shows a concept of escape line design from flip chip bonding area to the area outside of a chip. The figure is assuming 150 μm pitch of flip chip joint and shows a substrate terminal and conductor wiring design in the case. Upper side of the figure is toward for chip center and lower side for chip perimeter. In 150 μm flip chip joint pitch, the dimension is shared 75 μm to a terminal pad and 75 μm to a space

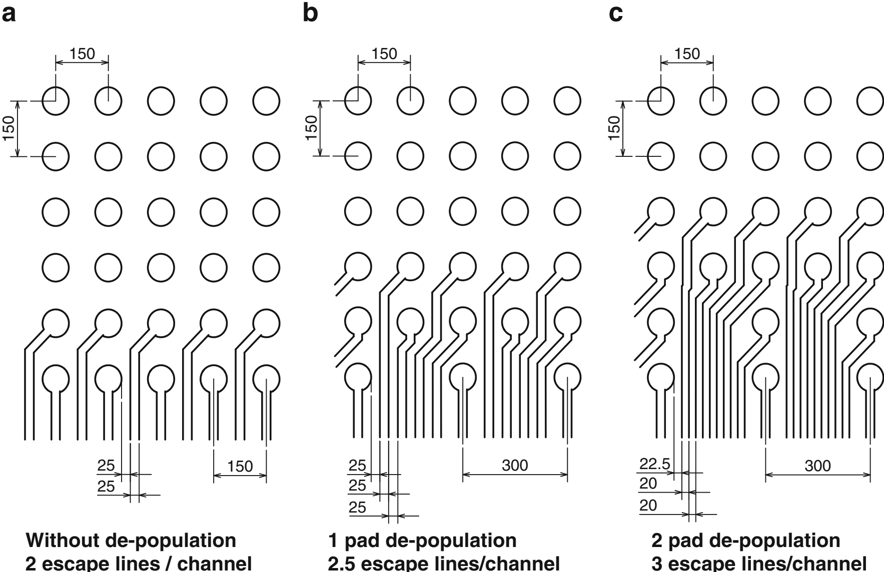


Fig. 7.4 Escape line design. (**a**) Without depopulation 2 escape lines/channel, (**b**) 1 pad depopulation 2.5 escape lines/channel, (**c**) 2 pad depopulation 3 escape lines/channel. After Tsukada [5], 2002

between pads. When there is a 75 μm space between pads, it can be divided to 25 μm space, 25 μm line and 25 μm space. i.e., one 25 μm line can escape between terminal pads. As the result, two lines can be drawn out simply from the first and second row as shown in (a) of the figure. Escape lines can reach to two rows inside from the chip perimeter. Taking some optional arrangement, following cases can be considered. If one pad is deleted from the first row that is the lowest row in the figure, four lines can escape between the first row pads with same 25 μm line and 25 μm space rule. As the result, five lines can escape from two pitch of flip chip joint. i.e., 2.5 lines escape per 150 μm is achieved. The deletion of one pad increased numbers of escape line and escape lines can reach deeper to the third row from the chip perimeter. If the capability of fabrication can provide a finer line width, more lines and more rows can be managed with repeating the same way. 20 μm line/20 μm space capability can provide three escape lines per 150 μm pitch and reach to four rows inside. Since this figure shows signal lines only, an actual design is much more complicated because many power pads must be placed among signal pads. But, this concept indicates an importance of design collaboration between a chip and a substrate designer. If the collaboration between a chip and a substrate is not maintained, a design may end up with higher cost and/or performance degradation.

High density organic substrate for flip chip bonding uses a pattern plating (semi-additive plating) to form copper conductor lines. Figure 7.5 shows pattern resist after development (a) and actual lines made by copper pattern plating (b). Lines are 30 μm

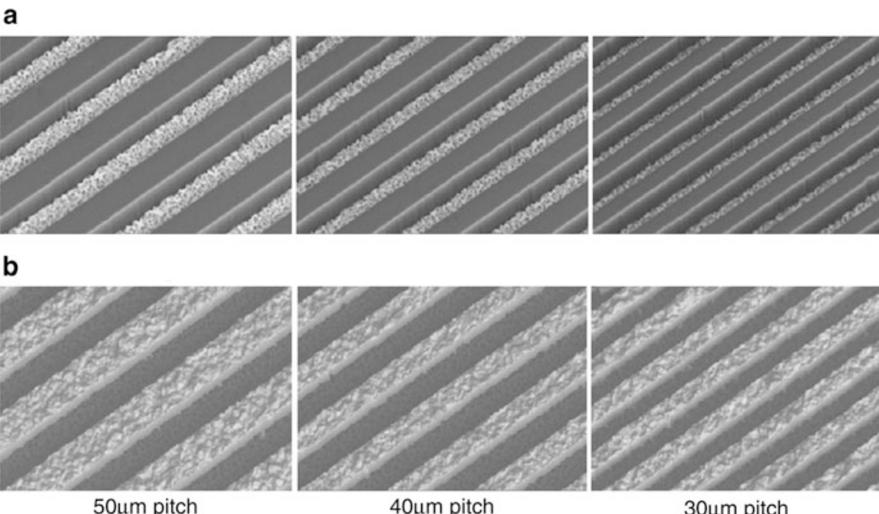


Fig. 7.5 Pattern plating. (**a**) After plating resist development, (**b**) after Cu plating. After Tsukada [5], 2002

pitch (15 μm line/15 μm space), 40 μm pitch (20 μm line/20 μm space), and 50 μm pitch (25 μm line/25 μm space). In between pattern resist lines, surface asperity to sustain an adhesion of copper to a dielectric material. It is actually seen an electro-less copper surface plated over a dielectric resin. In the meantime, a surface of copper conductor line after plating also shows asperity that is formed by granularity of copper plating.

Figure 7.6 shows a test result of copper adhesion to a dielectric material. The adhesion test is set by peel strength of 1 cm copper strip that is pulled perpendicular to the plated surface along with bake time of samples up to 240 h (10 days). In the figure, a peel strength trend of samples from four different lots (1 through 4) with $n = 30$ and 90 is shown. A peel strength is down about 20 % at the first reading but not degraded further, and slightly increased toward the end of test by aging.

The bake temperature is defined with a following formula.

$$t2 = 1.076(t1 + 288) - 273 \tag{7.2}$$

Where $t1$: maximum temperature of application.

$t2$: bake temperature of test.

The result of test is evaluated to a certain minimum peel strength of copper strip defined in a specification. This procedure is described in UL specification for organic material degradation by environmental aging. Though there is a formal aging test, this 10 days bake test that is a simplified version of the formal test is used

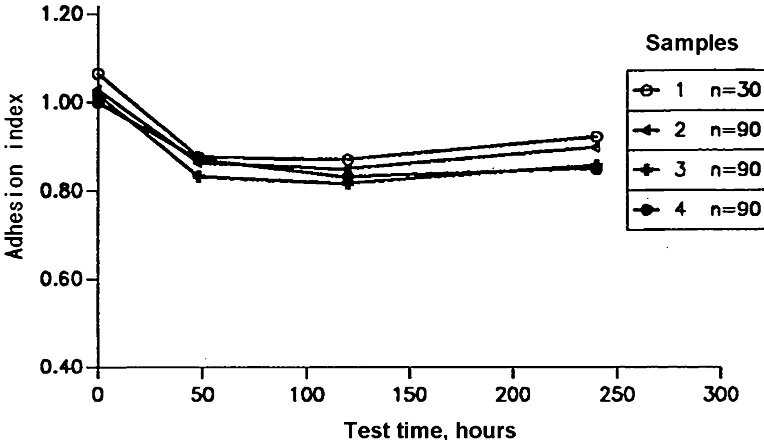


Fig. 7.6 Copper peel strength. After Tsukada [6], 1998

commonly. In case of ordinary PCB material (FR4), a minimum adhesion strength is normally more than 1 kgf per 1 cm copper strip (38 μm in thickness) peel at the initial value. In case of build-up substrate material, an adhesion strength is lower than that of ordinary PCB but enough to meet a degradation limit of the test. The difference in initial adhesion strength is due to a difference in anchor mechanism of copper. In ordinary PCB, a dielectric resin is pressed to copper foil that has an asperity for adhesion. In the meantime, in build-up substrate, copper is plated to asperity of resin surface created by an etching of the resin. There are two major requirements of UL to an organic substrate. One is this adhesion degradation test and the other is a flammability of dielectric material.

Figure 7.7 shows a cross sectional photograph of lines immediately after a pattern resist removal. There is a seed layer on dielectric material for providing a cathode in pattern plating. It is formed by electro-less copper plating that is adhered to roughened surface by permanganate etching. To maintain necessary adhesion of a line, a formation of good anchor mechanism is a critical point of pattern plating. In the photo, it looks the line width is a little wider than the space between the lines. This is intentionally designed such a way to keep a line dimension as designed after a seed layer etching. When a seed layer is etched, a surface of line is also etched and the line width becomes narrower to be a designed nominal value. The next figure describes the situation in more detail.

Figure 7.8 shows a magnified view of “50 μm pitch with 25 μm line and 25 μm space” cross section before a seed layer etching process. A surface of dielectric has irregular asperity formed by permanganate etching to provide mechanical anchor for copper plating adhesion. Electro-less copper covers the asperity surface. This photograph is a result of following processes such as permanganate etching of dielectric, electro-less copper plating, pattern resist apply, expose, develop, and electro-copper plating in sequence. In the middle of photograph, there is a pattern

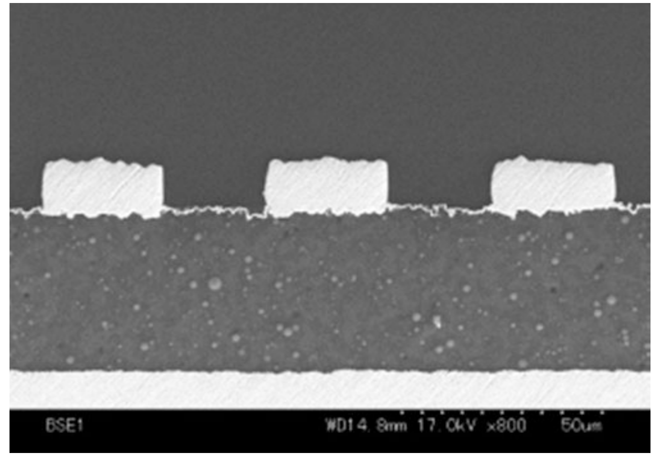


Fig. 7.7 Cross section of seed layer. After Tsukada [5], 2002

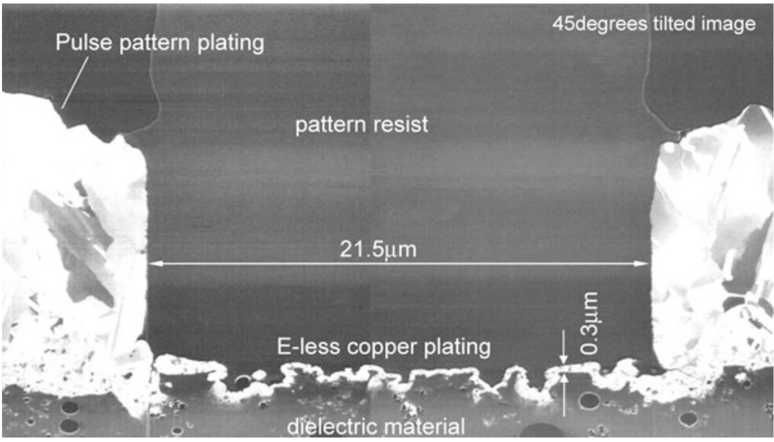


Fig. 7.8 Pattern plating cross section. After Tsukada [5], 2002

resist that occupies an area of space between lines. There are lines on both side of resist that are formed by an electro-copper plating. A space between copper lines is 21.5 µm in width. After resist removal, a seed layer etching removes electro-less copper plating on a dielectric material. At the same time, the surface of lines is etched and the space width becomes wider from 21.5 to 25 µm that is a nominal dimension of the design. In contrast, the line becomes narrower from 28.5 to 25 µm and also reaches to a design nominal. As this photograph, a line always must be wider for seed layer etching margin, a pattern resist in a space is always narrower for the size of etching margin. Therefore, the asperity of seed layer must be smaller when the line pitch is reduced. Otherwise, a yield of resist patterning is significantly deteriorated due to form too narrow resist with high aspect ratio. By this

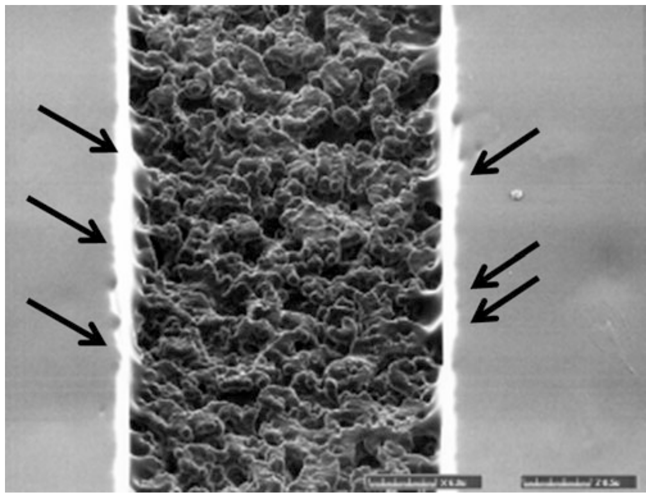


Fig. 7.9 Residual copper whisker at line edge. After Tsukada [5], 2002

dimensional issue, there is a strong demand to reduce a surface asperity of dielectric material with maintaining adhesion of copper plating to necessary level for processability and reliability.

There are other reasons that request a surface asperity of dielectric resin smaller. Figure 7.9 shows a residual copper whisker after seed layer etching. When a seed layer is etched, it takes longer time to etch out a copper in a large concave in surface asperity. In addition, since an adhesion of pattern resist to such asperity is not perfect, there is a narrow extrusion of copper under the edge of pattern resist occasionally remained as indicated by arrows in the photo. When a line pitch is reduced and a space becomes narrower, such phenomenon becomes more critical and sometimes violates a minimum space limitation. Also such extrusion becomes a starting point of copper migration due to a concentration of electrical field that eventually cause a reliability problem.

Another reason is that a skin effect describes later brings a major requirement to reduce such asperity. An original dielectric material shows 1–3 μm anchor depth and it is reduced to less than 1 μm these days by changing the resin formulation and filler size reduction. Other than such reduction of asperity, there are other activities to secure adhesion of copper without asperity by modifying surface property of resin [7] or providing molecular interface between copper and resin [8].

In addition to adhesion of copper on a dielectric resin, adhesion of a dielectric resin on copper is equally important. No matter where bottom or top of line, losing adhesion causes stress concentration at somewhere inside of a substrate and cause a crack during thermal cycle when in use. Conventional PCB uses oxidation of copper to form microscopic structure as shown in the figure (a). However, internal stress of flip chip substrate is much higher and other mechanism is required. Since a copper oxide is easily dissolved by acid, a defect called “pink ring” that is a separation of

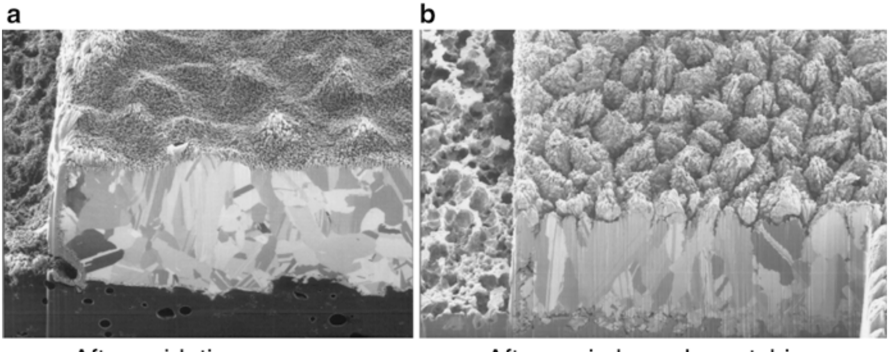


Fig. 7.10 Copper surface asperity. (**a**) After oxidation process, (**b**) after grain boundary etching. After Tsukada [5], 2002

dielectric from copper surface appears around laser drilled micro-via hole and propagate to area around by temperature excursion of thermal cycle. Figure 7.10 shows a view of anchor mechanism provided by etching of copper grain boundary of conductor line surface. The anchor size is a few microns. Eventually shown in the photo, there are asperity for anchor at the bottom and top of line as shown in the figure. There are activities also to reduce this side asperity of copper conductor that are to create finer asperity by a chemical reaction with maintaining required adhesion or add interface material or molecule layer similar described in the previous figure.

Besides pattern plating, a subtractive etching process is used for patterning a copper. It has its own improvement by anisotropic etching. Figure 7.11 shows an effect of etching chemical for this purpose. In the figure, line and space geometry of etching resist is 30/30 μm. An etching resist thickness is 10 μm and copper thickness is 20 μm. (a) Shows a result of copper geometry by a conventional etching solution. From top to bottom picture, etching of copper progresses isotropically that shows entire etching front is moving in semi-ellipse shape. As the result, top of the line becomes quite narrow compared with the bottom. In the meantime, (b) shows a progress in the case of anisotropic solution. From the beginning of the step, line wall is formed vertically and only an etching front at the bottom is showing semi-ellipse shape. It is noticeable that this is the result of anisotropic etching chemical used. The result shows almost complete rectangular shape of lines in cross section even though this is a subtractive etching.

Figure 7.12 shows how a chemical works for this anisotropic etching by steps (a) through (d). A chemical contains an etching inhibitor that selectively deposits on a side wall because a working solution flow is weak there compared with a bottom. As its progress, the inhibitor forms a film on the wall and a direction of etching is controlled by the inhibitor film. Finally, very high etching factor is obtained with forming vertical side wall.

Table 7.1 describes a property of dielectric material “ABF GX13” with other innovative versions from Ajinomoto Fine-techno Co., Inc. GX13 is a de facto

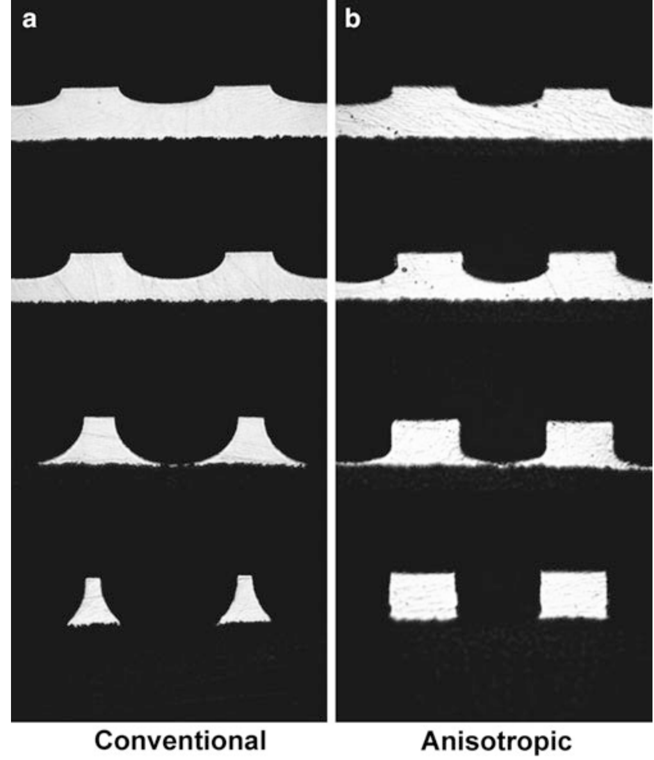


Fig. 7.11 Anisotropic etching. (**a**) Conventional, (**b**) anisotropic. Courtesy of MEC Co., LTD

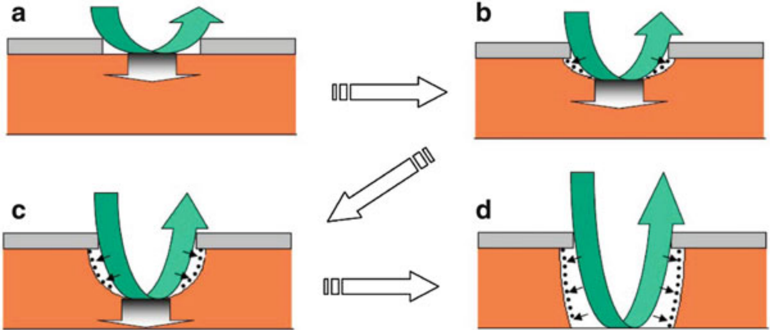


Fig. 7.12 Mechanism of anisotropic etching. After Toda, MEC Co., LTD. [9], 2011

standard material of dielectric resin for an organic substrate today. It is an improved version of the predecessor "GX3" that has a higher CTE as 60 ppm/°C. CTE is lowered to 46 ppm/°C by mainly increasing filler amount in the resin. Lowering CTE supports not only to reduce *X–Y* direction CTE but also, more importantly, to reduce

Table 7.1 Dielectric material, after Mago [10], 2011

Material name			G X 13	G X 92	G Z41
Features			Low CTE	Low profile	Low tan δ
CTE	ppm/°C	x–y, 30–150 °C	46	39	20
CTE	ppm/°C	x–y, 150–240 °C	120	117	67
Tg	°C	TMA	156	153	176
Tg	°C	TMA	177	168	198
Young's modulus	GPa	23 °C	4.0	5.0	9.0
Tensile strength	GPa	23 °C	93	98	120
Elongation	%	23 °C	5.0	5.6	1.7
Dielectric constant		Cavity resonance at 5.8 GHz	3.1	3.2	3.3
Dielectric loss		Cavity resonance at 5.8 GHz	0.019	0.017	0.0074
Water absorption	wt%	100 °C, 1 h	1.1	1.0	0.5

Z-direction CTE of the substrate as previously described. In the electrical property, a dielectric constant of GX13 is low as 3.1 because there is no glass in the film for electrical property and smaller via hole size capability. It is not provided in the table, but glass epoxy layer in an ordinary core has a dielectric constant around 4.2. A dielectric loss is 0.019 that is far higher than a ceramic substrate but it is offset by low resistivity of plated copper wiring in case of an organic substrate compared to a ceramic substrate with a paste wiring. Water absorption is 1.1 % but epoxy is essentially a hydrophobic material and there is no issue for electro chemical property as far as a material is cured appropriately. Instead, an attention must be paid a solder reflow in high humidity environment that may cause a delamination between a resin and copper conductor and sometimes with other resin material.

There is a low profile version in the table that covers a requirement for surface asperity lower to maintain a skin depth control and geometry issues that are also already described. Figure 7.13 shows a comparison of surface asperity profile of ABF by 3500× magnified SEM images. (a) Shows a regular profile of GX13 that is 600–700 nm in Ra with copper strength 0.7–0.8 kgf per cm peel in a test described previously. (b) Shows a profile of improved version reduced to 300–400 nm in Ra with slightly lower peel strength but maintained enough for required level.

In Table 7.1, there is a low tan δ version of ABF. It is very important to reduce a dielectric loss in an organic substrate for high performance area since a dielectric loss of epoxy is an order of magnitude higher than that of a ceramic (Alumina). Figure 7.14 shows a simulation of signal attenuation in an organic substrate. In this figure, a signal loss is calculated with a sum of resistive loss of copper conductor and dielectric loss. Two different dimensions of signal line that are 25 and 50 μm with 10 μm in thickness and three levels of dielectric loss are compared. If is clearly shown that a signal attenuation is improved when a line width is wider that lower the resistive loss and a dielectric loss of material lower. It is rather difficult to improve a dielectric loss than a change of line dimension since it requires a change of material composition that impacts an entire part of substrate manufacturing process and reliability. Since the figure shows signal attenuation per unit length of conductor

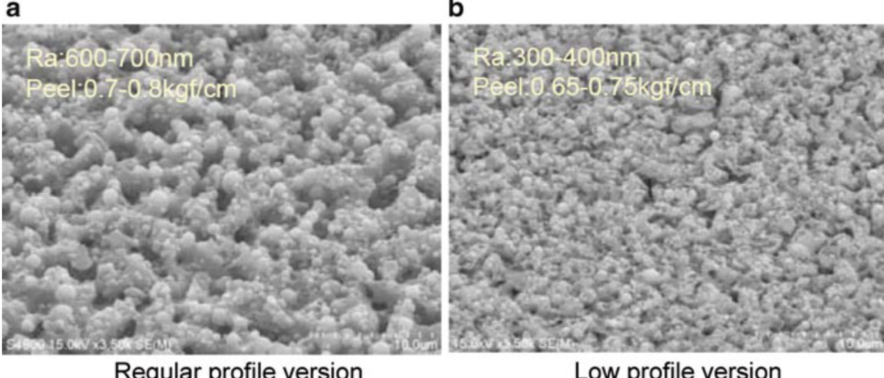


Fig. 7.13 Comparison of surface asperity profile. (**a**) Regular profile version, (**b**) low profile version. Courtesy of Mago, Ajinomoto Fine-Techno Co., Inc. [10], 2011

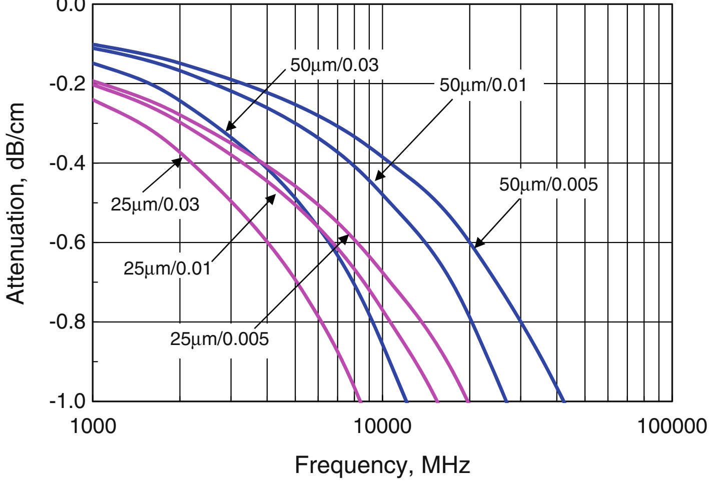


Fig. 7.14 Signal attenuation. After Tsukada, ref. [5], 2002

line, one may recognize that reducing a length of line has a direct effect to improve attenuation. Making substrate smaller directly improves situation even though a resistive loss increases due to a change in line width in the case.

There is other important property of a dielectric resin. Since a dielectric resin has to flow and fill a space between lines and micro-via holes after plating, the resin has to have a good flow nature when laminated to the previous layer. When a dielectric material is applied, there are copper conductor lines as shown in Fig. 7.7 and any other irregular space within a conductor pattern. Also any irregular asperity of dielectric surface and via holes on the previous layer must be filled. In the meantime, it has to be achieved with a low pressure since there is no glass fiber cloth inside of a dielectric

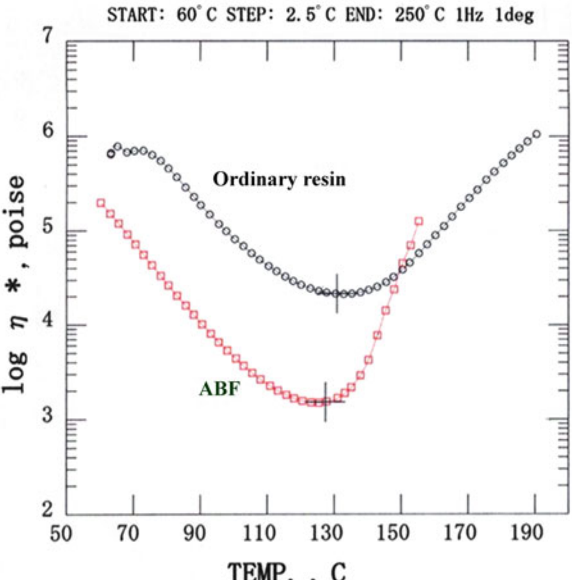


Fig. 7.15 Melt viscosity of resin. Courtesy of Mago, Ajinomoto Fine-Techno Co., Inc. [10], 2011

resin film. A high pressure used for ordinary PCB material like a glass epoxy prepreg flows a resin excessively and difficult to maintain a thickness, or flow out lines on the previous layer. Therefore, low pressure laminator with vacuum must be used. With such conditions, a good flow of resin becomes one of a key property of build-up dielectric resin. In Fig. 7.15, ABF reaches to a quite low viscosity compared with ordinary PCB resin (generally FR4).

7.3.3 Micro-Via Hole

After laminating a film of dielectric resin with low pressure vacuum laminator, the resin is cured in a half way for laser drilling. The reason that a resin is not cured completely is to provide an easier etching of surface at desmear process by permanganate for copper adhesion. In laser drilling, mainly two kinds of laser, CO_2 and UV-YAG, were used for micro-via drilling in a dielectric layer. Table 7.2 describes properties of each laser. In the case of UV-YAG, third harmonic laser, wavelength 355 nm, is listed. The nature of CO_2 laser is high power, low frequency, large beam, and small focus depth. In the meantime, UV-YAG is low power, high frequency, small beam diameter, and large focus depth compared with CO_2. Because of the size of beam diameter and power characteristics, CO_2 laser is used to drill a micro-via larger than 60 μm and UV-YAG for lesser diameters. In the case of CO_2 laser, a micro-via hole with a diameter larger than 60 μm is drilled with a few shots (2–5 shots for 35 μm dielectric thickness) in punching mode. In the meantime, YAG laser

Table 7.2 Laser comparison, after Tsukada [5], 2002

Parameters	UV-YAG	CO_2
Wavelength	355 nm	9.3 μm
Oscillator type	YAG solid	CO_2 gas
Power range	3–7 W	100–300 W
Pulse frequency	40–60 kHz	4–5 kHz
Pulse width	20–200 ns	10–200 μs
Beam diameter	20–80 μm	50–200 μm
Focus depth	100 μm	10 μm
Shot/via	40–70	2–5
Throughput	200–300 vias/s	200–300 vias/s
Work size (typical)	530 × 630 mm	530 × 630 mm
Shot numbers/via (typical)	40–70	2–5
Throughput	200–300 vias/s	200–300 vias/s
Focus depth	100 μm	10 μm

takes several, say ten, shots per via hole in normally tray-pan mode that is a circular shot with small beam to form a necessary via diameter. Due to each frequency, throughput is similar level eventually. In a production tool, multi-head is provided to achieve higher throughput for lowering the cost of drilling.

Figure 7.16 shows cross section photographs of laser drilled micro-via hole. Photo (a) is a drilled hole on a standard epoxy film for build-up layer that has no glass inside. YAG laser with 355 nm wavelength with Tray-pan mode is used in this case. A resulted via diameter is 48 μm at the top. Photo(b) is a drilled hole on a standard glass epoxy layer used ordinary PCB. CO_2 laser with punching mode is used in this case. Since standard glass epoxy layer is thick due to woven glass inside and CO_2 laser that has a larger beam diameter is used, a resulted via diameter is large around 90 μm at the top of via hole. In case of drilling a glass epoxy layer, CO_2 is preferred because an energy absorption to glass is low in case of 355 nm wavelength UV-YAG. Of course 355 nm UV-YAG can drill through a glass but the energy to do that level is high and a risk to punch out copper plane at via bottom increases.

A glass cross in prepreg has been innovated for laser micro-via hole. Figure 7.17 shows glass cross photographs. In case of regular glass cross, since the number of mono filament is large and its diameter is large that make the thread wide and thick, a woven cloth is observed as Photo(a) in the figure when it viewed from top. There are thick bundle area fiber and no fiber area clearly that make a major difference in laser abrasion. In side of drilled hole is rough in shape and diameter is not stable. Photo(b) in the figure shows a flat cross version. Threads are spread and there is no area where threads are vacant and views open like (a). Laser energy is absorbed uniformly and a hole diameter becomes stable. It allows more finer via hole size as a result. Photo(c) shows a cross section of ultrathin glass. The thickness is about 10 μm. Not only thin but threads are completely spread as a single layer of mono filament. This cloth can provide ultrathin substrate or thin build-up layer with a glass cross inside to reinforce a dimensional stability of substrate.

After laser drilling, a bare hole is cleaned with permanganate to remove a carbonated resin smear caused by laser heat. At the same time, the permanganate

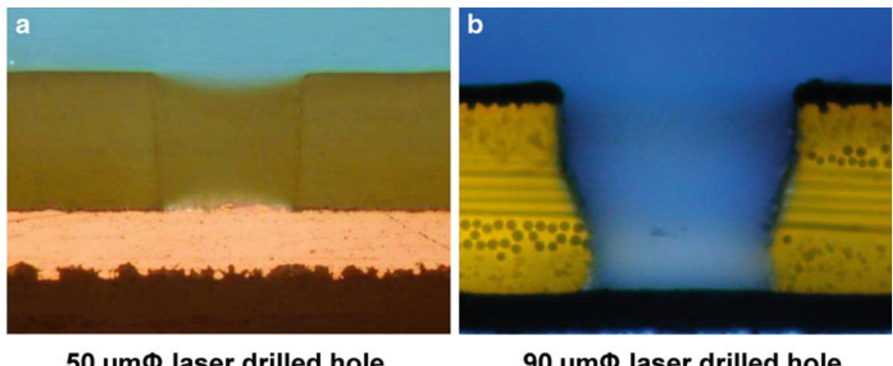


Fig. 7.16 Comparison of drilled hole. (**a**) 50 µm Φ laser drilled hole, (**b**) 90 µm Φ laser drilled hole. After Tsukada [5], 2002

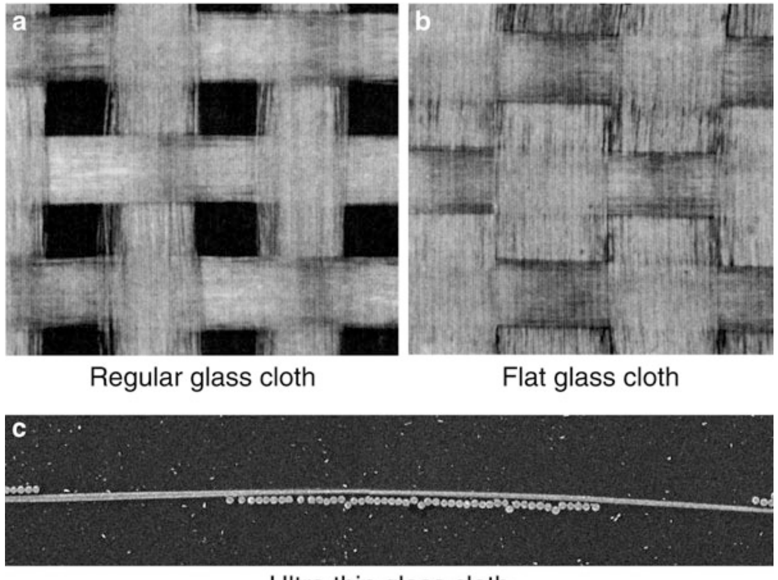


Fig. 7.17 Glass cloth for laser drilling. (**a**) Regular glass cloth, (**b**) flat glass cloth, (**c**) ultrathin glass cloth. Courtesy of Gondoh, Asahi Kasei E-materials Corp.

process works to etch the surface of a dielectric resin to form necessary anchor asperity for adhesion of copper plating. Figure 7.18a shows a bare hole on a standard build-up film after laser drilling by CO_2 laser. Photo(b) shows a hole after permanganate process. In a view of drilled hole Photo(a), there is a heavy smear at the bottom of hole. The smear is cleaned and removed in Photo(b). A surface granularity of copper is seen at the bottom of via hole. In photo(b), rough asperity of dielectric resin that is

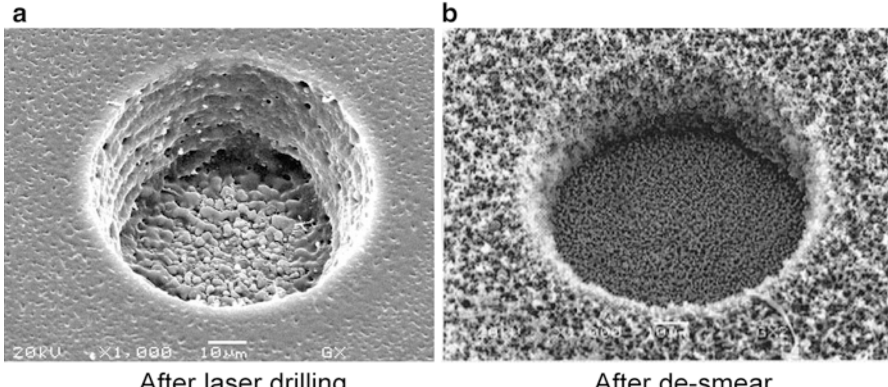


Fig. 7.18 Micro-via hole drilled by CO_2 laser. (**a**) After laser drilling, (**b**) After desmear. Courtesy of Mago, Ajinomoto Fine-Techno Co., Inc. [10], 2011

formed by a permanganate etching is shown at entire surface and via wall as well. When UV-YAG laser is used, a smear at the bottom is not heavy like CO_2 case, but it still remains at the bottom of base hole.

After laser drilling, via hole is cleaned with acid to remove resin smear in side of via, particularly, at the bottom. Permanganate for surface roughing also works to remove a smear. The next process is electro-less copper plating to provide a seed layer for subsequent pattern plating as described before. Electro-less copper is plated not only the surface but also inside of via hole. It is very important that electro-less copper plating securely cover the boundary between resin of via wall and copper at bottom of hole. Otherwise, there will be a trench at the boundary that may be separated with a stress in later. Pattern plating forms lines on the surface and plate inside of via hole at the same time. When via hole is plated in conformal mode, a pulse plating is used. Pulse plating can deposit the same thickness copper to surface and via hole by changing polarity of current with different pulse rates. Figure 7.19 shows an effect of pulse plating for making conformal plated via hole. In Photo(a), by initial plating, copper is deposited thicker at the surface than the bottom of via hole since less copper ions reached to the bottom of via hole. The point where electrical field is concentrated has more deposition. When polarity changed with denser pulse rate, the place where electrical field is higher, copper dissolve in higher rate. As the result, copper remains more at the bottom of via hole. By repeating such sequence, copper thickness at surface and via bottom becomes even shown in Photo(b) thorough (d).

Figure 7.20 shows via hole types used in an organic substrate. Photo(a) shows a conformal via hole. "conformal" means copper is plated in the same thickness along with a shape of via hole. Opening of inside of via hole after plating is eventually filled with next layer dielectric resin or solder mask resin if in the most outside layer. A conformal via hole is the most standard type of via hole. In recent days, a filled via hole is frequently used. A filled via hole is named because inside of via

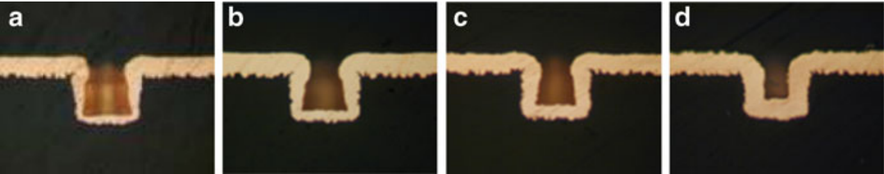


Fig. 7.19 Pulse plating of via hole. After Tsukada [5], 2002

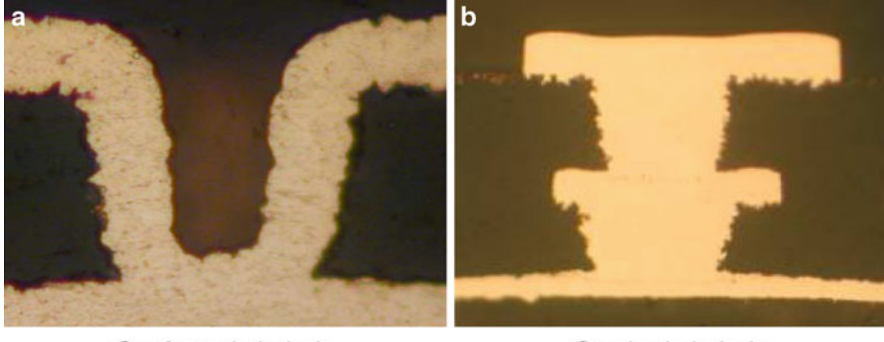


Fig. 7.20 Type of via hole. (**a**) Conformal via hole, after Tsukada [5], 2002. (**b**) Stacked via hole, Courtesy of Okuno Chemical Industries Co., LTD.

hole is filled with copper plating. A filled via hole is required to stack via holes as shown in Photo(b) of the figure. Stacked via hole is inevitably required for high density wiring of a substrate. To have via holes stacked, it is not impossible to use a conformal via hole. Upper via hole is simply just located on a previous layer via hole. Making laser drilling of upper layer deep is not difficult but a plating depth deeper takes a time much longer than ordinary depth. If a lower via is a filled via hole, next layer via hole can be processed with exactly the same process and total control is much easier.

Organic sulfuric compound (Brightener) is generally used in plating to make plated membrane shinny and give ductility by making deposited grain size finer. It also has a leveling effect as shown in Fig. 7.21 (a). Brightener is initially deposited uniformly on flat portion at surface and concave portion of via hole inside. After plating is started, it is not captured into the plating membrane and remains in the surface. As the result, the brightener density increases where the area reduces along with growth of plating thickness. Since a brightener has an acceleratory effect, plating of narrow concave area is promoted for resulting leveling effect. Since absorption activity of leveler is controlled by diffusion, the deposited density becomes low at low agitation area such as concave portion of via hole and high at high agitation area like flat surface as shown in (b). Since leveler has a suppression effect for electro-deposition, copper plating at surface is suppressed. Leveler is a quaternary amine compound and the nature is cation that is preferentially

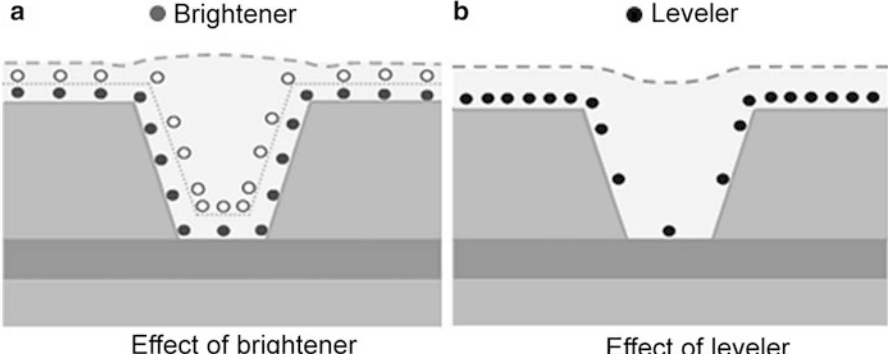


Fig. 7.21 Chemical effect for filled via hole. (**a**) Effect of brightener, (**b**) effect of leveler. Courtecy of Nishiki, Okuno Chemical Industries [11], 2009

absorbed at higher current portion to bring suppression effect there for leveling effect. With such basic leveling effect, chemical solution for filled via hole has its composition of solution and additives adjusted to obtain optimum filling effect for via hole.

A concept of component design of circuit in build-up layer is shown in Fig. 7.22. A required line width is defined as an average line width on the substrate global wiring with considering escape line design at the narrowest part as described previously and process capability for entire conductor layer of a substrate. When line width is defined, dielectric thickness is calculated with formula (7.3) to keep the characteristics impedance at 50 Ω. Before emergence of build-up PCB, ordinary PCB has higher characteristics impedance like 75–90 Ω. When build-up PCB is introduced as a flip chip substrate, it was decided to defined the structure as a substrate for direct chip attach and not as a PCB. Then, 50 Ω characteristics impedance was selected to match it with ordinary semiconductor output impedance with a consideration to expect the transmission line shorter by increasing the density for wiring in an organic substrate. As calculating by formula (7.3), a dielectric thickness is given with strip line structure as shown in upper right of the figure. When a dielectric thickness is defined, a via diameter is calculated with maintaining process capability of via hole plating. In this case, an aspect ratio is given as 0.7 that is considered with required via quality and process time shorter as possible. Aspect ratio is shown as lower right corner of the figure. Then, the base hole diameter is defined as 50 μm in this case. By expecting 2 μm removed by permanganate, a hole diameter to be delivered by laser drill becomes 48 μm.

$$Z_0 = \frac{377}{E_\mathrm{r}} \frac{H}{W_\mathrm{eff}} \frac{1}{2 + 2.8\ (H/W_\mathrm{eff})^{3/4}} \tag{7.3}$$

where; Z_0: Characteristics impedance; E_r: Dielectric constant; H: Conductor height, μm; W_eff: Effective conductor width, μm.

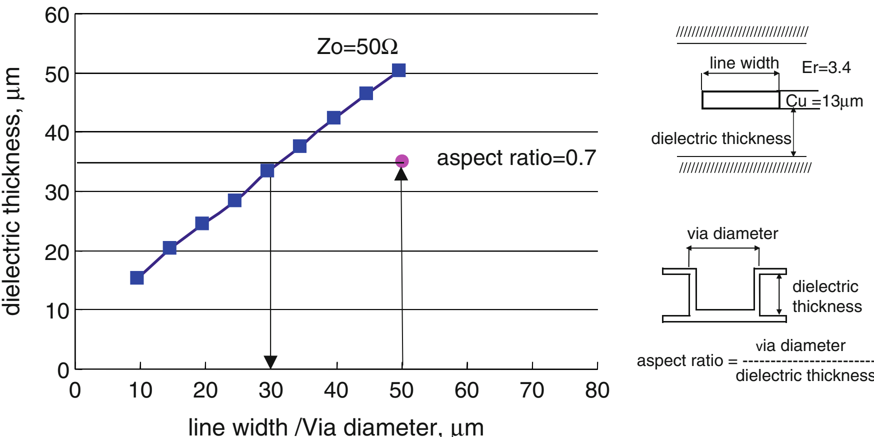


Fig. 7.22 Via hole geometry design. After Tsukada [5], 2002

Figure 7.23 shows an extraction image of copper grain boundary for a plated via hole by EBSP (electron backscatter diffraction pattern). The via hole diameter is about 60 µm. Copper plating is expected to connect upper circuit to lower circuit land by filled copper. When plating is perfectly done, copper grain grows on 111 plane of metallic lattice continued from lower land copper. Then, lower land copper and plated copper becomes to be unified metallic connection. However, in the figure, there is an obvious horizontal line of grain boundary between lower land copper and plated copper. In such connection, copper is not metallically connected but just fit a grain to grain mechanically. In this condition, when plated copper is pulled, it is separated from lower land relatively easy. And separated surfaces of upper side and lower side show concavo-convex shapes that perfectly matched to fit. If one make stacked via holes under such condition, there will be a high risk for separation due to *Z*-direction stress caused by CTE difference between stacked copper and surrounding resin. This level of via hole technology will have a high risk for future smaller via hole for higher wiring density of an organic substrate.

Figure 7.24 shows a cross section image of other via hole. In this image, copper grains grow across a top line of lower land where there are lined up micro voids (indicated by arrows) at a location of electro-less copper. It is an evidence that copper grains grow from copper grains at lower land. It means that a lower land and plated copper for via hole is metallically connected. This particular via hole in the photo has a diameter of about 25 µm and a good demonstration showing via hole plating integrity. Removal of oxide film and residue of organic material by chemical via clean process with maintaining enough turnaround of cleaning to get rid of dissolved chemicals inside of the hole. Other cleaning means, for example, a plasma cleaning, may also help to remove such materials.

A weak point of a standard sequential build-up substrate compared with other method is a through hole featured in a core. A plated through hole at a core connects front build-up layers and back build-up layers. A requirement of electrical connection

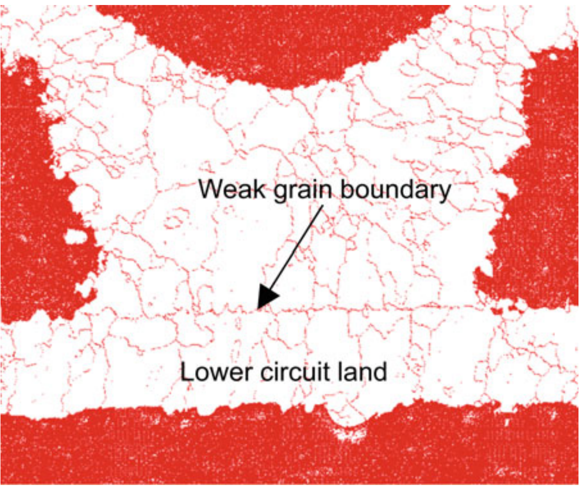


Fig. 7.23 Poor via hole connection. After Tsukada [12], 2008

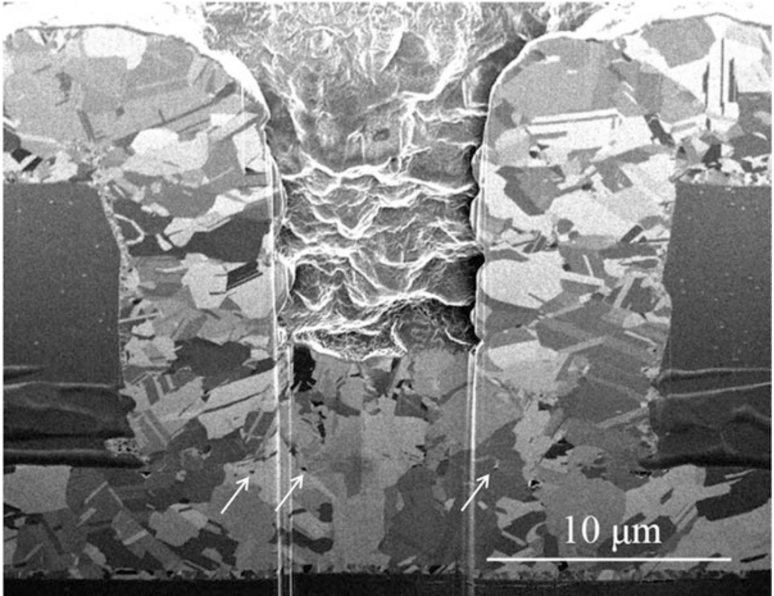


Fig. 7.24 Good via hole connection. Courtesy of KYOCERA SLC Technology Corp.

path from front to back of a flip chip substrate is, preferably, to have the same pitch with a flip chip joint pitch so that signal and power can be connected to a board level in straight through stacked via holes and through hole. However, current through hole drilling is using a mechanical drill. Though there is a small diameter mechanical drill like 60 µm, drill bit cost is significantly high and not practically used widely.

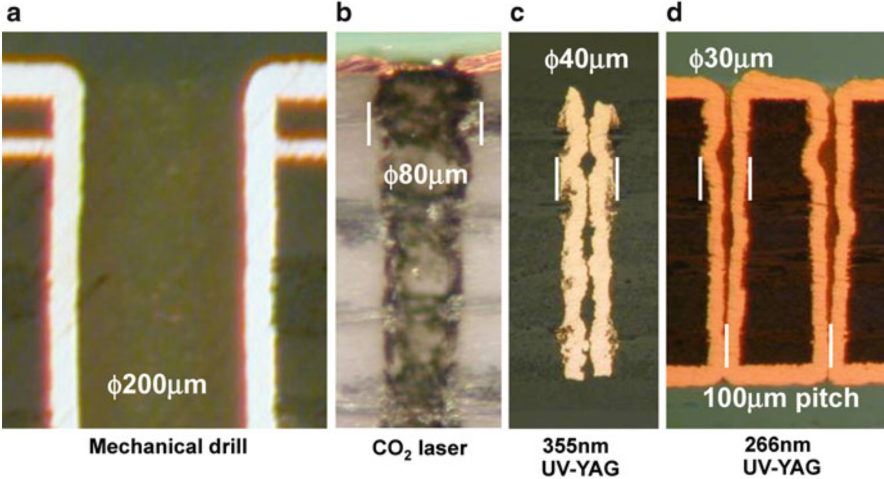


Fig. 7.25 Variation of through hole drilling. (**a**) Mechanical drill, (**b**) CO_2 laser, (**c**) 355nm UV-YAG, (**d**) 266nm UV-YAG. After Tsukada, ref. [13], 2005

Generally 100–200 μm diameter drill bit is used and not achieved the same pitch with flip chip joint pitch. For improving density of substrate and respond to a requirement for narrower flip chip joint pitch, laser drilling is a mandatory selection for the next step. Figure 7.25 shows a comparison of through holes made by various drilling method. In case of CO_2 laser, a diameter of drilled hole is large due to a laser beam size and is not be small enough for future requirement. And even though the energy absorption to glass is high but glass fiber is not sharply penetrated and showing very rough surface inside of the hole. Energy concentration of a beam is not sufficient. By using UV-YAG laser, a condition is improved significantly. However, using 355 nm UV-YAG is still not sufficient to have a clean drilled hole for a portion of glass fiber because an energy absorption rate to glass is low in case of 355 nm. This particular hole is drilled from both side of the core to try to obtain uniform diameter from front to back. With 266 nm UV-YAG, through hole drilling shows a good condition through a core with 0.4 mm in thickness. A drilled diameter is 30 μm and achieves 100 μm pitch through holes. There is a slight irregularity at the point of glass but no issue on quality and reliability. The result is enough. Though the through hole diameter varies from top to bottom, there is no sign of problem for copper plating. The integrity of through hole quality and pitch reduction for future is enough showed in this photograph.

Figure 7.26 shows energy absorption of related materials by a wavelength of laser. A dielectric resin has even absorption rate along wavelength. CO_2 and UV-YAG show a similar rate. In case of copper, an absorption rate is highly depending on a wavelength. UV-YAG is absorbed relatively high rate but not in CO_2 case. It is significantly low in case of CO_2. An absorption rate to glass is quite opposite. It is very high in case of CO_2 and not so high in case of 355 nm UV-YAG.

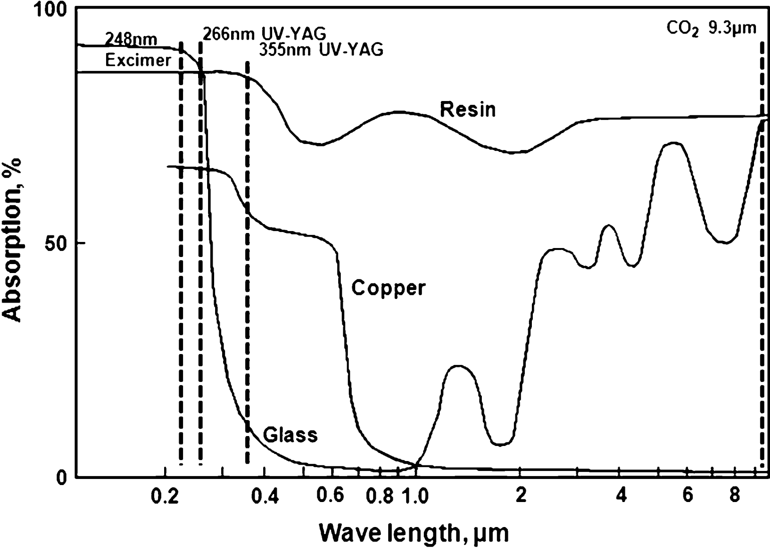


Fig. 7.26 Laser energy absorption by material. After Tsukada [13], 2005

Therefore, both CO_2 and UV-YAG have a similar effectiveness to drill a hole through a dielectric resin. If there is a glass, CO_2 works effectively. CO_2 has high absorption rate to resin and glass but not to copper. According to the figure, if one is intending to drill through a glass epoxy panel without an internal copper plane, CO_2 can work effectively and if there is an internal copper plane, 266 nm 4th harmonic of UV-YAG is the best selection considering absorption rate to resin, glass and copper. 248 nm KrF Excimer laser has a result similar with 266 nm UV-YAG. More importantly, as in the previous figure, a drilled through hole diameter that is different in each case must be primarily considered.

7.3.4 Pad Finish

Solder mask of substrate surface has a significant role to form flip chip joint. In PCB technology, solder mask is used to cover the surface of substrate to prevent solder at soldering process touches to conductor pattern and bridges a narrow space to cause a short defect. Since an outside of conductor layer of organic substrate for flip chip has lines and terminals with far finer pitch than regular PCB, the role of solder mask is significantly important. Figure 7.27 shows two types of solder mask design for flip chip terminal on a substrate and resulted flip chip joint shapes. In this particular design, the pitch of joint is 225 μm and terminal pad size is 125 μm. Lower side photographs are showing a view of solder mask opening from a chip side. Photographs up above show a cross-sectional view of resulted joint shape. (a) is

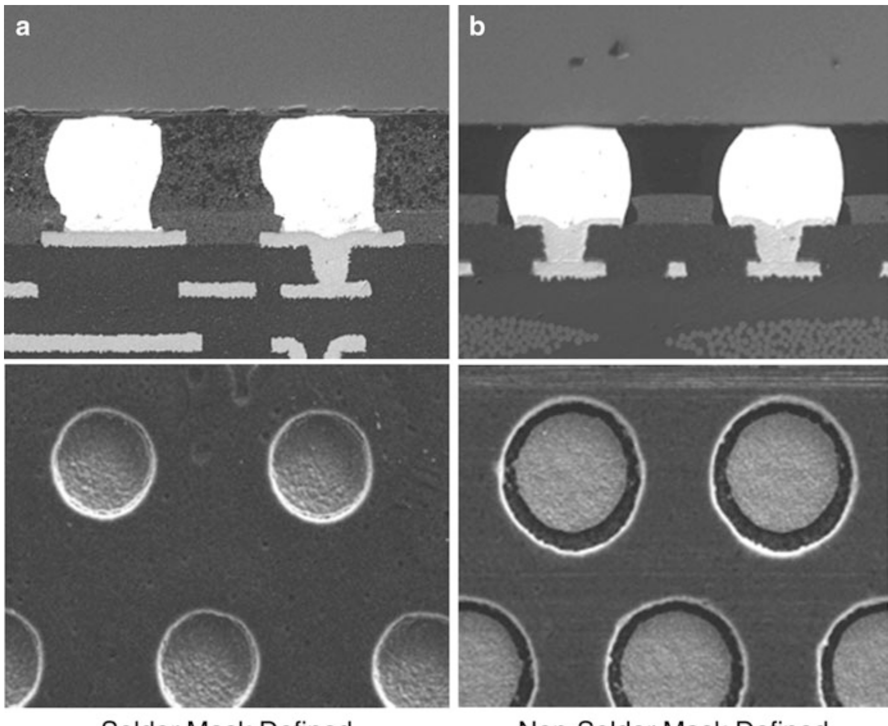


Fig. 7.27 Flip chip joint cross section. (**a**) Solder mask defined, (**b**) non-solder mask defined. After Tsukadaf [5], 2002

named as solder mask defined (SMD). The edge of terminal pad for flip chip joint is covered by solder mask and the size of opening processed by photo-etching is smaller than an actual pad. (b) is named as non solder mask defined (Non SMD). The opening is wider and a terminal pad is completely inside of it. Each type has pros and cons with relating this opening size. Non-SMD type has a large opening and a solder bump of chip is fitting easily inside of opening. In SMD type, the opening is much smaller and a bump of chip is sometime difficult to fit. Since the opening of solder mask is formed by etching, smaller is more difficult to form a opening completely without residue of resin remained inside. Occasionally invisible thin layer of resin is remained and cause a defect for gold plating which eventually cause non wet solder to the pad. When flip chip joint is formed, the joint height of Non-SMD is lower compared with SMD as appeared in upside photographs when solder volume is the same. Low joint height has disadvantages for the gap space cleaning and underfill flow. Both are the same level in terms of reliability. SMD has stress concentration due to narrow down the bottom of joint, but it is not an issue since flip chip joint life with underfill has a large safety margin to required product life. In Non-SMD case, a small dielectric resin crack appears

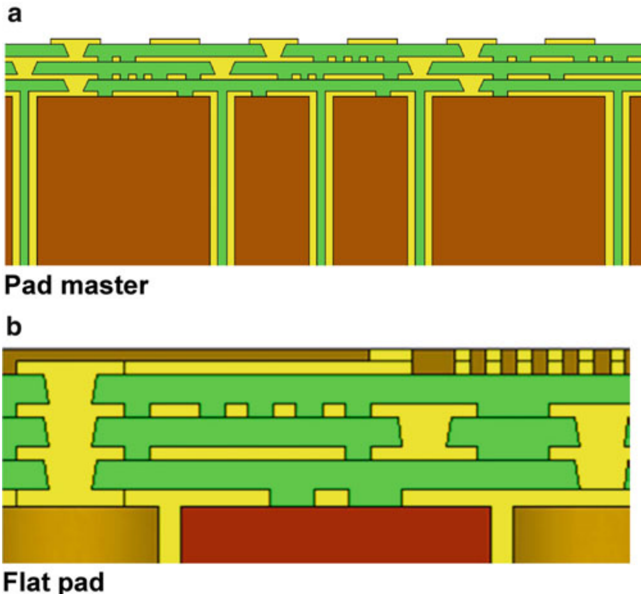


Fig. 7.28 Substrate terminal pad variation. (**a**) Pad master, (**b**) Flat pad

occasionally at the edge of terminal pad caused by *Z*-direction CTE mismatch of resin and solder but it does not extend after the stress is released by causing an un-harmful crack.

Figure 7.28 shows an alternative design for flip chip pad on an organic substrate. Currently flip chip terminal design of substrate is SMD and Non-SMD as described. For future ground rule improvement, both types have a barrier to reduce a pitch. SMD will have a difficulty to form clear opening for the smaller diameter and a narrow area between opening will be break easily in case of Non-SMD. The surface conductor layer will be pad only eventually like a ceramic substrate. It is called as pad master that share one conductor layer for pad only as shown in (a) of the figure. In case of substrate for wire-bonding, flat pad structure that the top surface of terminal is leveled to the same with substrate surface is already implemented.

Various terminal finish is used for organic substrate. Au/Ni plating, Au/Pd/Ni plating, OSP (organic solderability preservative) and solder are generally used. In case of solder, plating, paste and ball attach are used. The most popular process is a paste bumping. The size of solder ball in paste varies to raise filling rate as shown in Fig 7.29. 150–200 μm pitch terminal is a processable range and 100 μm on challenge. To achieve fine paste printing, tool is improved such as to provide one direction squeeze by a rotary squeeze and air bag chamber that pressurize a mask surface after squeezing for clean release of printed solder from mask as shown in Fig. 7.30. Electroless AuNi plating is used with Ni thickness around 5–7 μm and Au thickness 0.4–0.5 μm. AuPdNi is also used with higher shear test result of solder ball compared with AuNi because a surface of Ni is oxidized while a substrate go

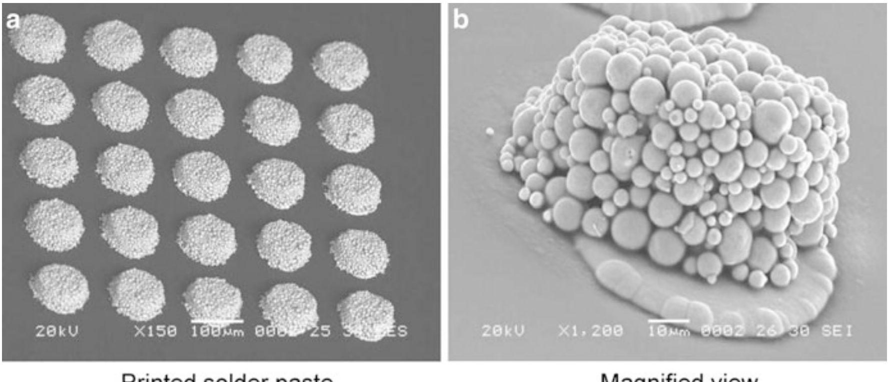


Fig. 7.29 Solder paste printing. (**a**) Printed solder paste, (**b**) Magnified view. Courtesy of Murakami, Minami Co., LTD.

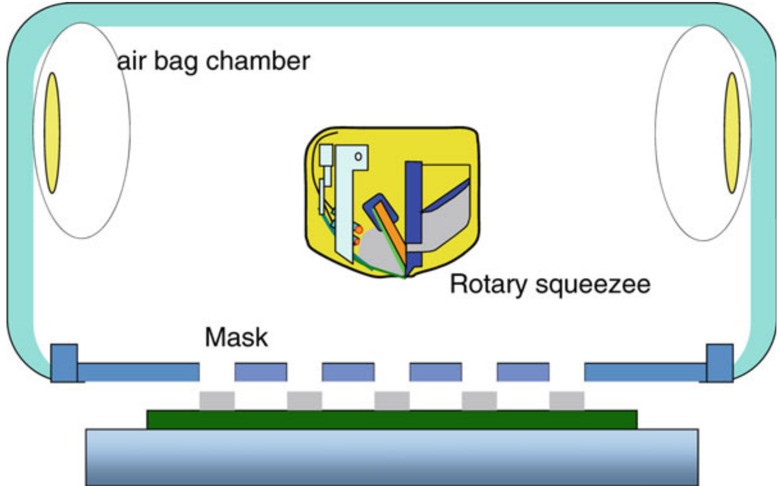


Fig. 7.30 Improved screen printer. Courtesy of Murakami, Minami Co., LTD.

through multi-reflow cycle and a interface with solder ball is occasionally separated. OSP is popular in PCB manufacturing and an enhanced version for Pb free solder is used, so called as pre-flux.

Figure 7.31 shows a cross section of solder, applied by paste, after reflow on a terminal pad where Ni/Au plated as pad finish. Before flip chip bump is placed, solder is flattened so that a placed bump do not move easily. The height of this carrier terminal bump is about 30 μm from a copper terminal surface and protruding 20–25 μm from a surface of solder mask. When the bump is flattened, the top is wider than the terminal pad opening and height from solder mask surface is depressed to 5–10 μm. However, it moves back to the original height when it is put to solder

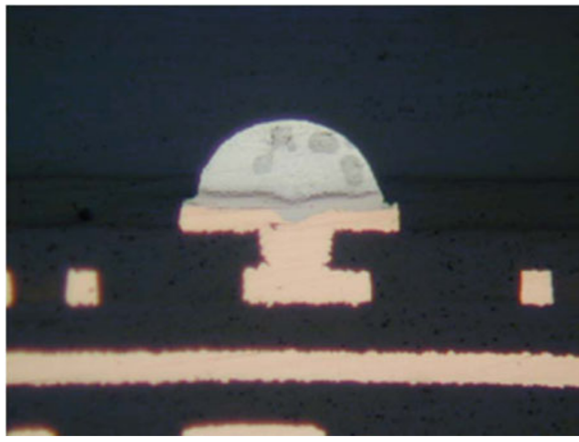


Fig. 7.31 Substrate terminal finishing. After Tsukada [13], 2005

reflow temperature. This step of solder height change creates 10–20 μm dimensional margin in *Z*-direction to formation of flip chip solder joint. In another words, the margin fills a space gap that is caused by height variations exists in parts of flip chip bonding. Possible variations are as follows.

Chip bump height variation that is a sum of a chip terminal height variation (up to 1 μm maximum) and chip bumping height variation (1–5 μm).

Equation (7.2) substrate terminal height variation (5–20 μm) that is caused by a remained topology of copper pattern. The topology is significantly depends on the substrate wiring design.

Since a solder finish on terminal pad compensates height variation of parts consist of flip chip joint, it is the most safe process in terminal finish in terms of dimension and solder wettability as well. One has to consider this issue of *Z*-direction height variation precisely when plans to squeeze a dimension and/or introduce other method and material in a flip chip joint construction. Otherwise, it will degrade chip join yield and remain a high risk in reliability of the joint.

Normally when a copper pad is wet with solder, the height of solder from copper pad surface is limited around 10 μm when the pad size is 100 μm. It is due to a wetting angle of solder to copper (23.5°) in case of Eutectic SnPb solder. The amount is far less than to form optimum solder joint with chip side bump since solder bump height at chip side varies by its applied process nature or deformation during transfer after bumping process. Also a substrate surface topology due to circuit pattern under a dielectric layer makes a pad height variation around 10, 20 μm in the worst case. Therefore a special effort is required to deposit required solder amount to form the bump height higher at certain level. Typical method is utilize “wet back” nature of solder that is to apply solder to large area than a pad by plating, screen printing and other deposit process with mask and form the bump height higher when solder is reflowed and excess solder applied wider than a pad area is wet back to the pad area to form higher bump eventually. However “wet back” required space around a pad so that to deposit excessive

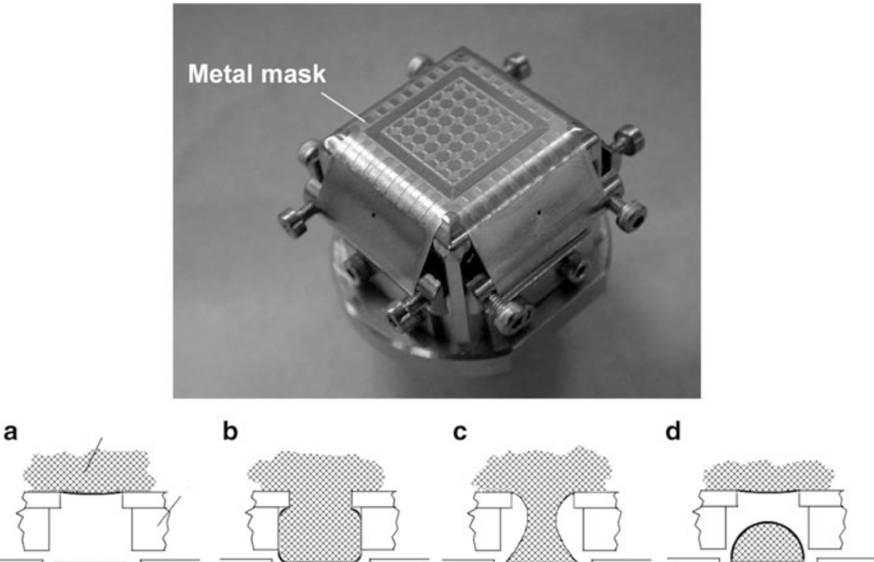


Fig. 7.32 Solder injection tool. After Tsukada, ref. [14], 2000

amount of solder. It is possible to apply a pad for low I/O chip case but not possible to use for high density logic chip where the bump pitch is tight and no room to deposit solder around.

Figure 7.32 shows a tool that was used to resolve this issue in an early stage of implementing a flip chip bonding on an organic substrate. Solder injection tool is used for bumping solder on a flip chip pad of substrate for high density logic chip. The head features a metal mask that has a hole pattern matching with required pad pattern of chip site on a substrate. A molten solder is in a reservoir of head. Operation sequence is as follows.

(a) Align a hole pattern of mask to identical pad pattern on a substrate.
(b) Pressurize the reservoir so that a molten solder is extruding from the hole of mask and a solder touch to a pad below.
(c) After a few seconds to allow solder wet the pad surface, apply negative pressure to the reservoir to draw back molten solder into it.
(d) Molten solder is pinched off and a good solder amount is remained on the pad.

The volume of solder can be designed by a pad size, mask hole size and height of the mask from a pad. One of major benefit of this type method is to be able to use any kind of solder as far as it melts. This tool can deposit a solder around 50 μm on 100 μm size pad. A disadvantage is that has to process each chip site on the substrate one by one.

Figure 7.33 shows scan solder tool that is an improved version of solder injection to use in continuous operation. Instead of mask with hole pattern, molten solder from reservoir is pushed out by pressure from a slit on the head to touch to a substrate

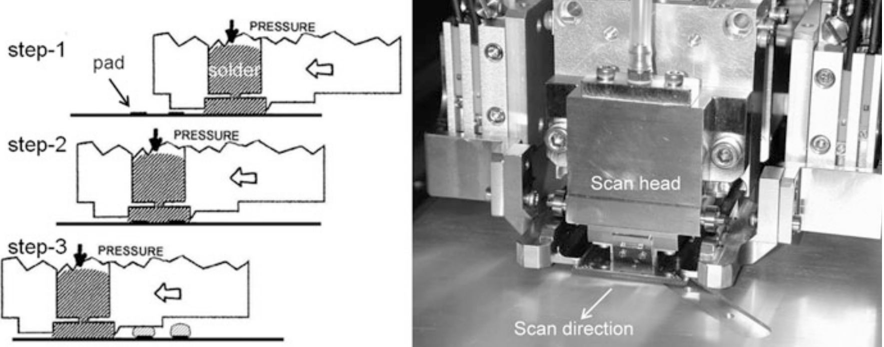


Fig. 7.33 Scan solder tool. After Tsukada [14], 2000

surface (step-1). When the head scans, solder is depressed between a head and substrate surface and touch to pads on the substrate to let solder wet to copper (step-2). As the result of head moving on a substrate surface, solder is split to each pad while still depressing by a part of head. When solder released to free from the depression, some wet back action works to form a round solder bump on a pad (step-3). In this method, the advantage is still to be able to use any kind of solder and quick scan to apply solder by continuous operation. There is some sensitivity to split solder to each pad that depends on the design of pad pattern and pressure control so that molten solder do not spilt out from the space between the head and a substrate.

Figure 7.34 shows an advanced tool that handles a molten solder. A mask made of organic film that has same pattern with pad pattern on a substrate is aligned to the substrate pad. A diameter of hole on a mask is smaller at top and wider at bottom so that extruded solder from the top does not move back to a solder reservoir. The head scans on the mask which has a hole pattern and solder is filled into mask holes to wet pads on a substrate. After cooling and solder freezes, mask is removed. A major advantage of this method is that the height of bumps is controlled by a thickness of mask. Therefore, height of solder is even if pad sizes are different on a mask pattern. 150 μm bump on 250 μm pitch is under production and 75 μm on 150 μm under qualification at the end of 2010. It is planned to move 25 μm on 50 μm for future.

Figure 7.35 shows a process flow of ball bumping that is both used for wafer bumping and substrate bumping. Operation steps are as follows.

1. Sucking balls with an arrangement plate by vacuum.
2. Removing excess balls by applying ultrasonic vibration.
3. Check balls sucked on the arrangement plate.
4. Align to position.
5. Ball placement.
6. Check balls on a substrate.
7. Repair ball placement errors.
8. Reflow and cleaning.

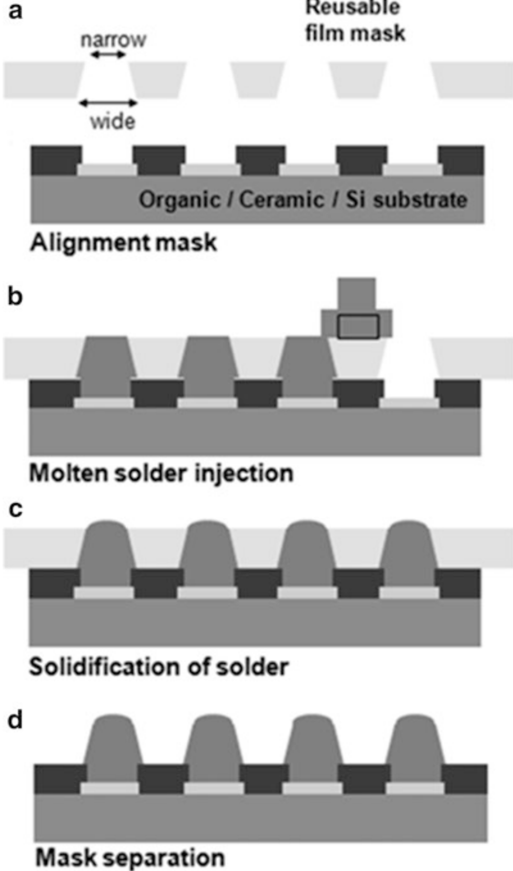


Fig. 7.34 Injection molded solder tool. After Orii [15], 2011

In case of wafer bumping, huge amount of balls is handled in one time. The process is set to allow a rework of missing and irregular bumps after a inspection so that 100 % of bumps is formed completely. If this process is used for a substrate bumping, a rework process may not be necessary since a number of bump may not be large numbers as on a wafer since the process may work on a singulated substrate or a sheet that contains a limited number of substrates.

Figure 7.36 shows a solder deposition by "Super Juffit" process from Showa Denko. Photo(a) shows a SEM photograph after solder powder is applied to fine pitch pads on a substrate. Pitch of pads reached to 50 μm that can accept a very fine pitch wire bond chip. In a magnified view at the corner of photo, it appears fine solder powder is deposited on a pad. Photo(b) shows a view after a reflow of deposited solder powder.

Figure 7.37 shows process steps of "Super Juffit."

(a) Clean a parts so that to remove any contamination and oxide layer of copper.
(b) A chemical with a proprietary composition is applied to entire surface of substrate. After a cleaning and drying, a tacky material is remained on only a copper surface and washing away from other area.

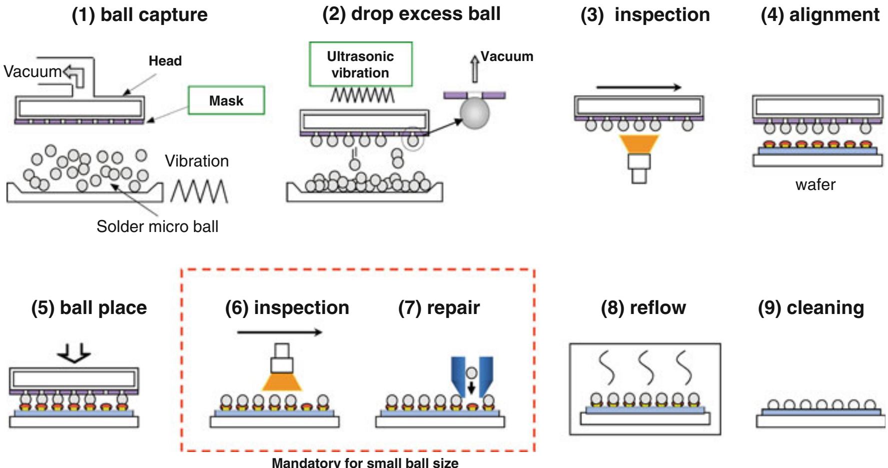


Fig. 7.35 Ball bumping process. (1) ball capture, (2) drop excess ball, (3) inspection, (4) alignment, (5) ball place, (6) inspection, (7) repair, (8) reflow, (9) cleaning. Courtesy of Ishikawa, Nippon Streel Corporation

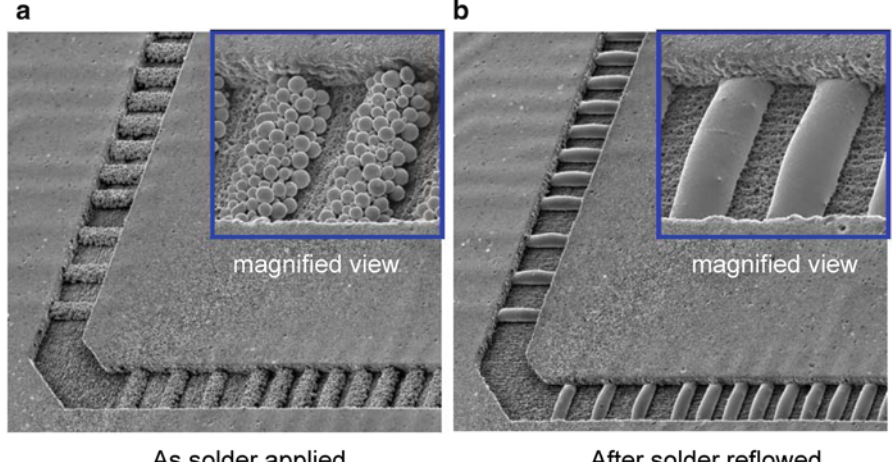


Fig. 7.36 Super Juffit process. (**a**) As solder applied, (**b**) after solder reflowed. Courtesy of Syoji, Showa Denko

(c) Solder powder is applied. Since a pad is covered by the tacky layer, powder sticks only a pad and others are easily put off from surface of a substrate.
(d) Apply a flux for solder wetting.
(e) Reflow solder powder.
(f) Washing out a flux.

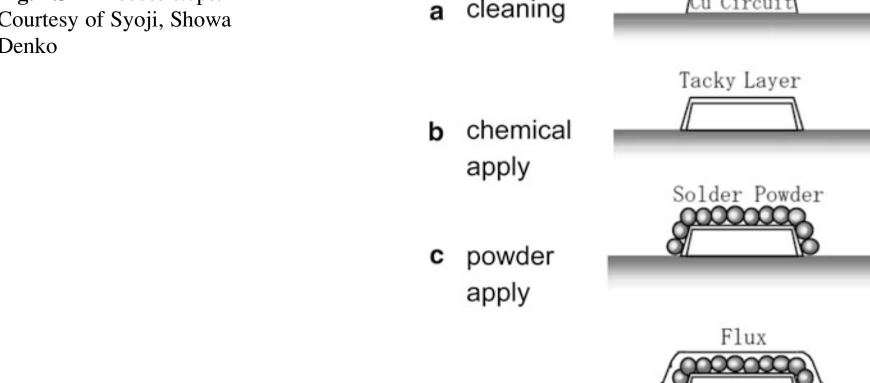


Fig. 7.37 Process steps. Courtesy of Syoji, Showa Denko

Since a process is utilizing the chemical property that reacts with copper only, there is no need of mask and other supplemental mean to deposit solder to very fine pitch pads on a substrate.

Figure 7.38 shows a chip bumping and a matched substrate bumping. By utilizing a solder bumping process on a substrate, a substrate can accept a hard metal pillar bump on a chip with a good yield at a chip bonding process.

7.3.5 Chip Package Interaction

One of major issue of flip chip bonding on an organic substrate is a stress at flip chip joint due to CTE mismatch between chip and substrate. Figure 7.39 shows a schematic of damaged wiring layer cross section of chip that uses low-k material as a dielectric material of chip wiring layer. When a chip is join to substrate, above melting point of solder, solder form a joint between chip and substrate terminal pads. While the temperature is above a melting point, there is no stress even though with CTE mismatch since solder is melting as (a) in the figure. When cooling down and after passing a melting point, solder becomes solidified. Since a substrate CTE is much higher than chip, a substrate shrinks faster than chip. Then a joint is deformed by shear stress due to displacement between terminal pads of chip and substrate as shown (b). By this deformation, a tensile stress is generated and low-k

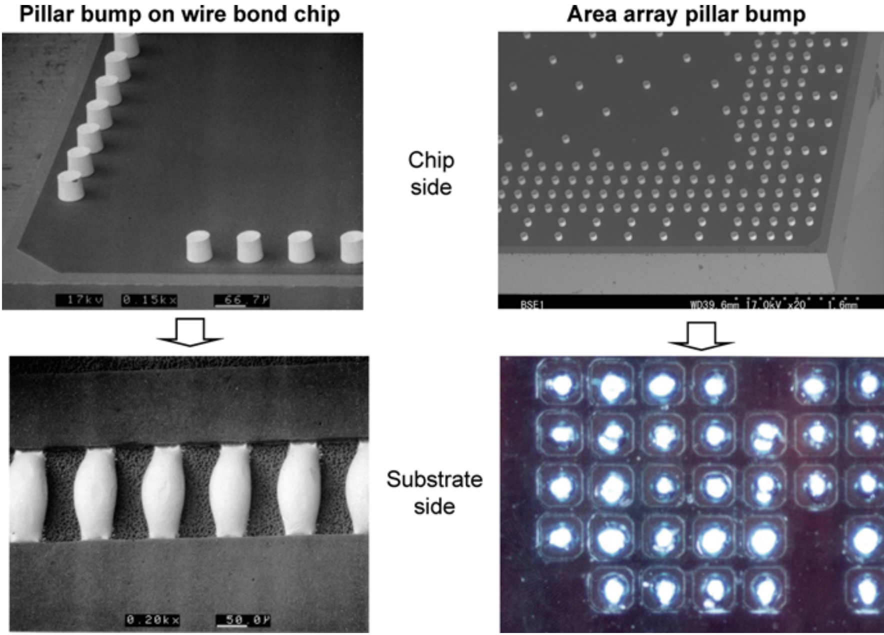


Fig. 7.38 Substrate bumping for chip side pillar. After Tsukada [14], 2000

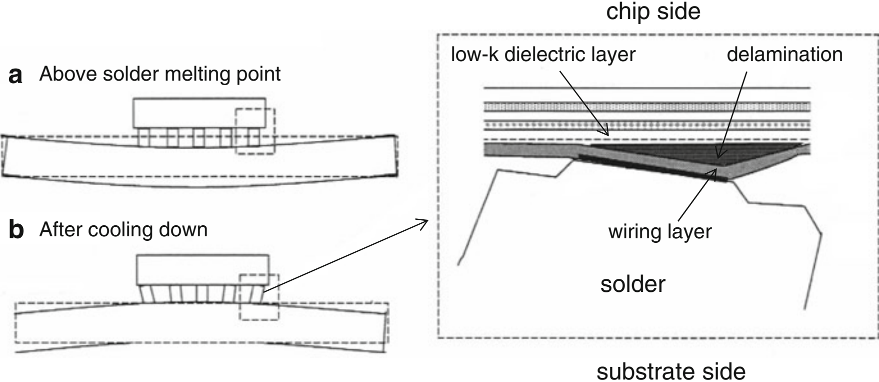


Fig. 7.39 Chip wiring layer damage. (**a**) Above solder melting point, (**b**) after cooling down. After Nishio [16] (2008)

dielectric layer is delaminated to resulted breaking of wiring as shown in the right side of figure.

This problem is essentially caused by CTE mismatch of chip and substrate. Normally chip CTE is 3–3.5 ppm/°C because a base material is a silicon. A CTE may vary depends on how thick a wiring layer is. In the meantime, CTE of organic substrate is 17 ppm/°C at standard, around 15 ppm/°C at low CTE version. It is

because a base core of organic substrate is essentially same with a printed circuit board that has CTE adjusted to 17 ppm/°C. The reason to adjust this level is to protect copper wiring inside of laminate that is established for printed circuit board technology. This adjustment is done by CTE balance of glass fiber woven cloth that employs E-glass with CTE 5–6 ppm/°C and epoxy resin with CTE 70–80 ppm/°C as described before. This technical issue is named as "white bump" since it viewed white by an analysis of SAM (scanning acoustic microscope) and the cause is named as "Chip Package Interaction" and being the most critical issue for low-k chip packaging in these years and also a major barrier for a finer pitch flip chip joint in future.

Figure 7.40 shows schematics of underfill effect when a chip is attached to an organic substrate. (a) in the figure shows when cooling down from a curing temperature of underfill resin to room temperature after a gap between a chip and substrate is filled with the resin. Since a chip and substrate are tightly adhered, displacement between chip and substrate terminal pads are very small. It made possible to extend a flip chip joint life significantly longer. Instead, a chip and substrate entity is bent due to CTE mismatch causes an effect like a bi-metal structure. This bending causes a dominant failure mode in case BGA package that is a failure of BGA joint at the center of package. It is important that CTE mismatch creates a stress but 30 % is converted to this bending when thickness geometries of parts are in a standard configuration, i.e., chip thickness is about 50 % of substrate thickness. Important thing is that remaining 70 % of stress is absorbed into a substrate since material modulus is quite low in case of an organic substrate compared with a ceramic substrate. This distributed stress in a substrate cause various failure inside of substrate when a package is not properly designed. Detail design rule that do not cause a local concentration of stress is very important.

Paring with (a) in the figure, (b) was disclosed when this technology was announced from IBM Yasu Laboratory in 1991. Within a technical disclosure after announcement, a following warning was made about a possible chip damage when a process is not properly handled. After applying underfill, a joint of this package is perfectly protected from the CTE mismatch, however the point after chip join to substrate and cooling down to a room temperature, a joint suffers a heavy deformation due to a displacement between chip and substrate terminal pads that is caused by the CTE mismatch. Since this deformation is very heavy, it was known that the point indicated by "A" has very high tensile stress in a chip wiring layer.

Figure 7.41 is photographs used with the technical warning at the announcement. (a) is a chip wiring crack that was located under the edge of barrier metal. Chip passivation had a crack and the crack extended to damage wiring. When passivation had a crack even if wiring was not broken and chip function does not fail immediately, moisture comes into chip wiring layer and cause a corrosion or metal migration in later day. It was recommended that chip should have a polyimide cushion layer on which a barrier metal can sit on. The technical description is closed by a statement that whether the chip is durable to a stress is the first thing to be confirmed in this technology. This failure mechanism is the same with one described today as chip package interaction. The warning has been made at the

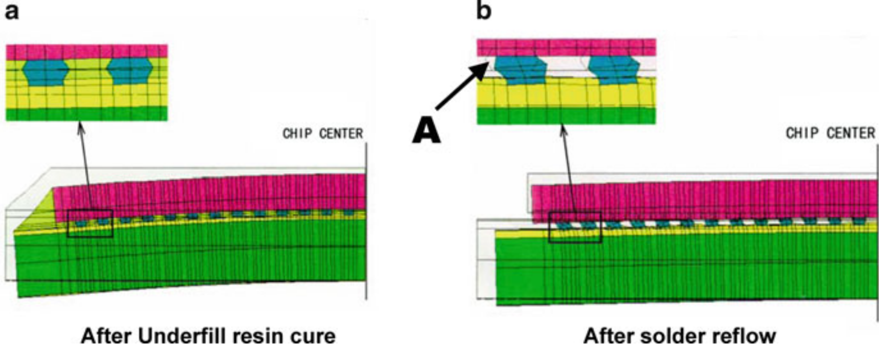


Fig. 7.40 Underfill effect on organic substrate. (**a**) After underfill resin cure, (**b**) after solder reflow. After Tsukada [6], 1998

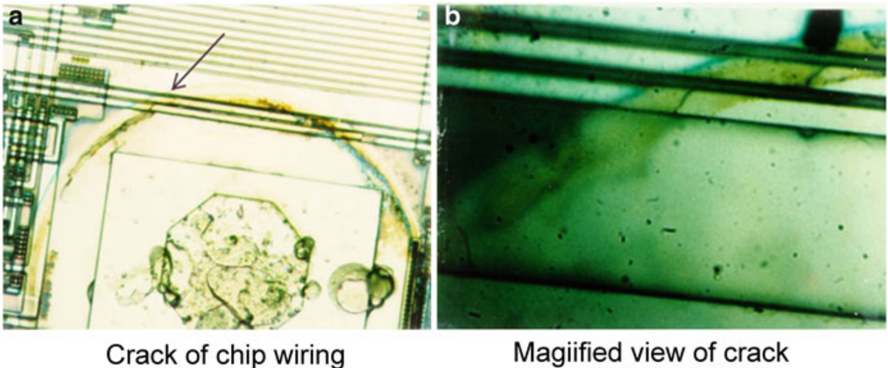


Fig. 7.41 Passivation crack on chip. (**a**) Crack of chip wiring, (**b**) magiified view of crack. After Tsukada [6], 1998

beginning of this technology introduction. It is important one should aware the basic of technology is not changed.

Early stage of flip chip with a ceramic substrate, Goldman issued a paper titled "geometric optimization of controlled collapse interconnections" in 1969. There described in the paper, a flip chip joint is deformed by shear force caused by CTE mismatch between a chip and a ceramic substrate. Due to deformation of joint solder, tensile stress is generated in a chip wiring layer at the edge of chip side terminal. And he indicated there as a critical point shown in Fig. 7.42. This means that the issue of chip wiring layer damage is not only by a usage of organic substrate but ceramic substrate either. A possibility of chip wiring layer damage is an essential thermal stress issue of flip chip joint. One should remind that a dielectric material of chip wiring layer is always under a risk of damage by looking for a density and performance increase. When a technology becomes mature, an issue

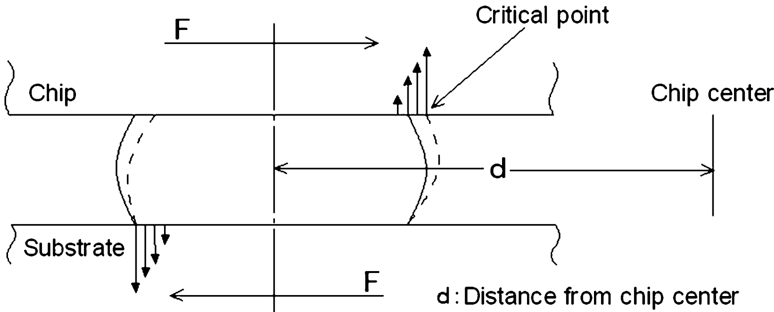


Fig. 7.42 Risk point at flip chip joint. After Goldman [17], 1969

seems to be settled down, but an issue in new step is revealed repeatedly when phase-in to a new technology.

As previously described, a chip and substrate entity bends when ambient temperature changes. Since solder of flip chip joint creeps, a chip and a substrate are both almost flat at the temperature for curing underfill. After cooling down from a curing temperature, the entity has an almost permanent bending. as shown in Fig. 7.43. It shows a bending mode when a 20 mm square chip is attached to 30 mm square substrate and cooling down to room temperature from a resin curing temperature around 130 °C. It shows smooth uniform concave shape. Bending of chip and substrate, i.e., a package bending, is repeated when a chip is heated by a system in operation. The figure shows a bending of just only a package not assembled to the next level board and a stress condition is more complicated when assembled.

When a chip and substrate entity forms BGA package, a state of BGA joint failure in thermal cycle test is shown in Fig. 7.44. (a) Shows a BGA joint pattern and dotted rectangular is the area a chip is attached. “A” through “J” are nets with BGA joints electrically measured a failure in this thermal cycle test. (b) Shows a log-normal chart to show cumulative failures of a measuring point each net. As the test result, Net-A that was located around the center of chip failed first. Then Net-B that was located at the edge of chip failed next. Net-C that was located at the corner of the chip failed third. At the point of time, all other NET distributed along with the substrate diagonal line had no failure. The result that BGA joint under a chip area failed earlier while others were safe means a bending stress was locally in the chip area and other area of substrate remained almost no bending since BGA joints supported each other. This result clearly indicated that an internal stress of substrate increases when the package is attached to a board while uninstalled package is bending freely showing natural concave shape. Also this test result is a main reason there are many packages today that are designed to remove BGA joints from a area under a chip.

A force of bending is significantly strong. It is bending a mother board when BGA package is installed. Figure 7.45 shows a test result when BGA packages are installed on both side of mother board symmetrically. The test result showed a BGA joint life

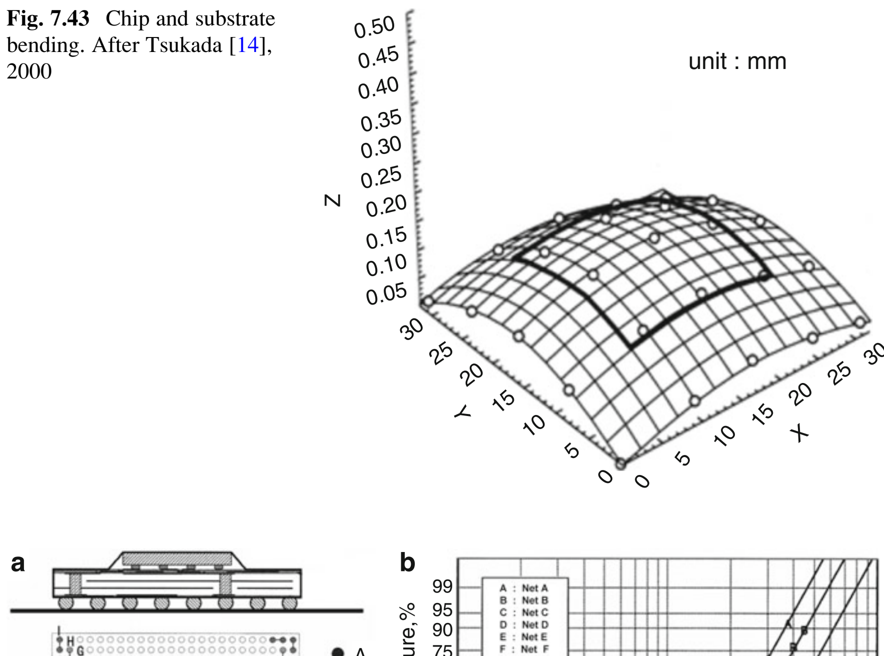


Fig. 7.43 Chip and substrate bending. After Tsukada [14], 2000

Fig. 7.44 BGA joint life with organic substrate. (**a**) BGA joint tested, (**b**) Test result. After Tsukada [14], 2000

in double side loading was shorter about 45 % than single side loading. Earliest fail of BGA joint was also a joint at the center area. This result tells us a bending of chip and substrate entity actually bends a mother board below when a BGA loading is on a single side. When BGA package loaded on double side, a mother board in between BGA packages on front and back is constrained and is not bend since there is a tension caused by a chip and substrate bending from both side. Therefore a BGA joint life becomes significantly shorter in case of double side loading that may violate a requirement of product. In case of flip chip loading on double side on a substrate, the life of the joint also became shorter but the rate was about 35 %. But a joint life has a large safety margin and the life though it is shorter is not the level to deteriorate a product life. In addition in this test, measurement was done with 10 mΩ threshold just

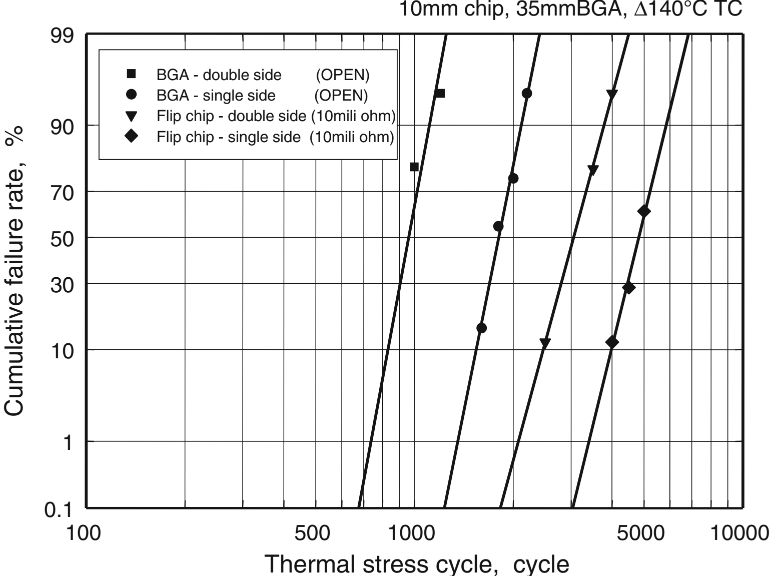


Fig. 7.45 Joint life in double side loading. After Tsukada [14], 2000

for technical comparison purpose. A series of these results tells us a stress condition of flip chip on organic substrate is significantly complicated and one has to analyze and test in a cautious manner when implementing an application.

Figure 7.46 shows examples of chip crack caused by a stress due to CTE mismatch. (a) Shows a crack runs across a chip indicated by arrow. Crack runs horizontally in the photo. An initiation point of crack is a small chipping at the edge of chip. A stress is concentrated at the chipping. The chip in the photo is originally a wirebond chip and attached to a substrate by flip chip bonding with using a gold stud bump. Normally a wirebond chip is singulated from a wafer by a single side dicing from functional side with leaving thin remained thickness of wafer and is braked the thin part finally. Therefore it is not abnormal for wirebond chip to have a chipping at a backside edge. A chip for flip chip is normally uses a double side dicing so that there is not chipping at any edge of chip where a stress is easily concentrated. (b) Shows a very unique crack. A crack propagated in a lateral direction and a silicon base was separated with the functional area of chip remained on the substrate. Since a silicon cracks along with a crystal interface rather easily, too hard substrate and underfill caused this crack by extremely high stress.

A thermal stress condition around a flip chip joint in an organic substrate is more complicated. Figure 7.47 shows a close up schematic of flip chip joint that is sitting on a stacked via placed onto a capped plated through hole in an organic substrate. It is a preferable structure regarding electrical performance since a straight vertical conductive structure from a chip terminal to other side of a substrate potentially a BGA joint. CTE of materials in this schematic are as follows. A chip CTE varies

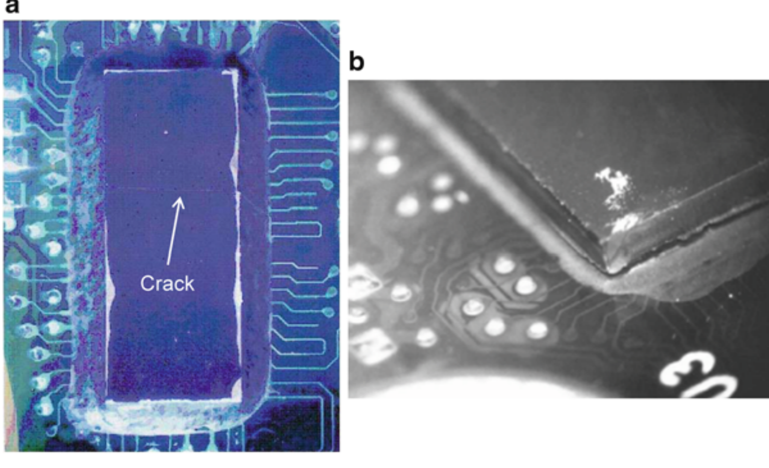


Fig. 7.46 Chip crack caused by stress. (**a**) Crack across a chip, (**b**) Crack laterally separate a silicon. After Tsukada [14], 2000

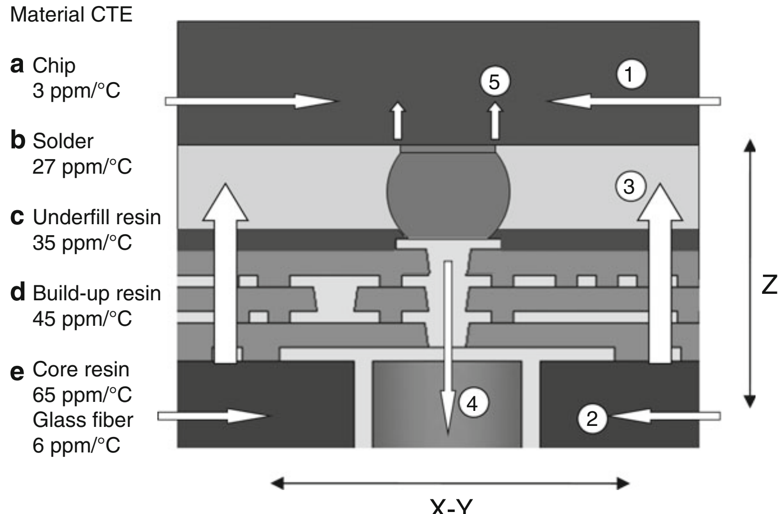


Fig. 7.47 Thermal stress around joint. Material CTE. (**a**) Chip 3 ppm/°C, (**b**) solder 27 ppm/°C, (**c**) underfill resin 35 ppm/°C, (**d**) build-up resin 45 ppm/°C, (**e**) Core resin 65 ppm/°C, glass fiber 6 ppm/°C. After Tsukada [12] 2008

with numbers of wiring layer but basically around 3 ppm/°C that is a CTE of silicon and isotropic. A solder CTE also varies with its composition and about 27 ppm/°C in case of eutectic solder with isotropic nature. Underfill resin is a mixture of epoxy and silica with a combined CTE around 35 ppm/°C and isotropic. Build-up resin is

also a mixture of epoxy and filler material with CTE around 35 ppm/°C and isotropic. A core of substrate is a composite of epoxy resin with CTE 65 ppm/°C and a glass with CTE 6 ppm/°C. Copper of conductor has CTE 17 ppm/°C. Most of materials in a substrate are isotropic material but a glass is a form of woven fabric and laminated laying in *X–Y* direction. Due to a very low CTE with very high modulus of chip and woven glass fiber with low CTE and high modulus compared with all resins that is low modulus with high CTE, a combined CTE in *X–Y* direction of this structure is dominated by a chip and woven glass fiber with force indicated by block arrows ① and ②. As the result, when temperature rises, majority of isotropic expansion of resins is converted to *Z*-direction as ③ in this schematic since X–Y is constrained with a chip and glass fabric. Since a stacked via and a through hole are copper that are relatively low CTE with high modulus, this part of structure connected to flip chip joint does not expand in *Z*-direction as other area between plated through holes. Consequently, a tensile stress is concentrated in stacked via that may bring a separation of stacked structure when metallurgical integrity of plating is not appropriate. Though, a solder of flip chip joint creeps to relax the stress, there is still a high risk of via separation exists. Therefore, to remain a bulk solder in a flip chip joint is important. If in case a solder of joint becomes intermetallic compound, the stress at ④ increases significantly. It will damage a via to via metallic interface and also create tensile stress in chip wiring layer as shown by block arrow ⑤. It is easily understood that the stress increases as a material of joint becomes harder. One has to be cautious when introduces a harder metal at flip chip joint like copper pillar. To avoid this stress concentration, an ideal solder for joint is a low stress and low melting point solder. One of the candidate is Indium that is very ductile to avoid stress concentration and the melting point is low as 151 °C that lowers a thermal stress when cooling from chip joining temperature.

Figure 7.48 shows overall CTE conditions in organic substrate applications. Current issues with a thermal stress are basically caused by a high CTE of a organic substrate that employs an ordinary PCB material set though there are some improvement done after an emergence of technology. No matter how the final figure of package is, there is a high internal stress distributed in a package. It creates risks during fabrication of a package with a concentrated stress at a part of component of the package. It is further complicated if we employ a stacked chip on an organic substrate. To improve this situation, essential CTE reduction of organic substrate materials are required. Activities are on-going in almost all area of organic substrate and its packaging components. Ideally, there are two cases. One is to use a substrate that has a CTE same as silicon. The other is an utilization of low and adjustable CTE construction for a substrate and a board so that one can design a thermal stress of entire package to minimize a stress concentration in any part of the package.

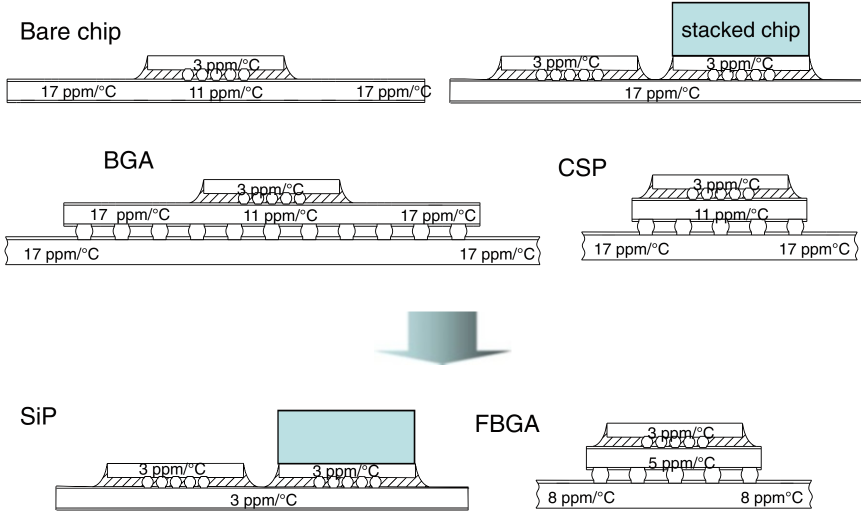


Fig. 7.48 Substrate CTE improvement. After Tsukada [12] 2008

7.3.6 Reliability

Table 7.3 describes typical stress test applied to an organic substrate when target application is cost/performance system such as work station, PC and high end consumer product. There are three categories as applied to other kind of packages. However, enough cautious has to be taken since an acceleration factor of a failure is different by a material. Particularly, some test that has been done for a ceramic substrate is too stringent for an organic substrate because it creates a stress that makes the test as destructive mode rather than a acceleration test. The most important thing is that a failure mode of acceleration test must be the same as presumed failure mode of an application in filed. One of typical example is $\Delta T = 180\ ^\circ\mathrm{C}$ thermal cycle test. Traditionally $\Delta T = 180\ ^\circ\mathrm{C}$ is used for a ceramic substrate since it is based on a inorganic hard material. In case of organic substrate, epoxy is a main material and eventually causes a crack by fatigue from aging effect. In $\Delta T = 180\ ^\circ\mathrm{C}$ thermal cycle test, a crack appears earlier in an organic substrate and cause a conductor wire break down before 1,000 cycles that has been a target cycle in case of a ceramic substrate. However the acceleration factor of cracking is pretty high and extrapolated field life from acceleration test proved has no problem. It is preferable to use lower delta temperature in considering proper accuracy of an acceleration test, $\Delta T = 165\ ^\circ\mathrm{C}$, 1,000 cycle is used in case of organic substrate. If $\Delta T = 180\ ^\circ\mathrm{C}$ is used by some reason, 750 cycles will be the most reasonable target cycle. In a lower delta temperature test as $\Delta T = 100\ ^\circ\mathrm{C}$, target cycle may be set to 3,000 cycles that is the same temperature range widely used for an application package assembly where an ordinary PCB is tested together. Occasionally, wet thermal cycle test is used to apply a stress to a metal portion of a substrate such as

Table 7.3 Stress test, after Tsukada [14], 2000

Stress	Condition	Target
Dry TC (Δ180 °C)	−55–125 °C	750 cycles
Dry TC (Δ165 °C)	−40–125 °C	1,000 cycles
Dry TC (Δ100 °C)	0–100 °C	3,000 cycles
Wet TC (ref.)	0–100 °C	3,000 cycles
THB	85 °C, 85 %, bias	1,000 h
PCBT (HAST)	109.8 °C, 85 %, bias, 1.2 atm.	264 h
PCT (ref.)	121 °C, 100 %, 2 atm.	96 h
HTS	150 °C	1,000 h

conductor wires, micro-via holes and plated through holes that are composed of copper. But it should not be applied to assess an epoxy material related failure mode because it too quick cycle and lose a duration for a time of creep of resin that is brought in low cycle fatigue. Since a creep of copper is low compared with epoxy, wet thermal cycle test can be used for a copper conductor related failure mode, it should be handled as reference test due to its accuracy. In a category of temperature and humidity influences, 85 °C, 85 % with bias voltage is a typical test. Since the target hours of this THB test is 1,000 h that takes about 2.5 months including reading time, HAST (high acceleration stress test) is used frequently. 109.8 °C, 85 %, bias and 1.2 atm. condition that is unsaturated water vapor is one of a typical setting. Occasionally higher temperature is used but the condition has to be unsaturated water vapor, otherwise the test lose a correlation to a failure in field. Even with 109.8 °C, 85 %, bias and 1.2 atm. condition, a corrosion type failure sometime loses correlation. For example, Chloride ion in an organic substrate is driven out and crystallized on surface. PCT (pressure cooker test) is also handled as reference since a saturated water vapor condition is a destructive test in case of epoxy base organic substrate. In HTS (high temperature storage test) with $\Delta T = 150$ °C 1,000 h, a substrate is burned out. But this test is to test a semiconductor related diffusion on a substrate and the burning is not a problem. It may be used to test electro migration or diffusion of any metal structure in a organic substrate in near future along with a density increase for miniaturization of design component. These stress tests are applied after a preconditioning treatment level three defined by JEDEC (joint electron device council).

Figure 7.49 shows a substantial nature of package test that employed an organic substrate for flip chip application. A suite of thermal cycle test was done with different delta temperature is applied to a group of BGA package joined to motherboard and N50 life in between $\Delta T = 180$ °C and $\Delta T = 100$ °C is obtained by extending the stress test until reaching to a failure. The data was extrapolated to temperature below $\Delta T = 100$ °C that is an operation range in field of majority applications.

In the figure, a failure of BGA joint life is indicated by "BGA joint life." As described before, it is caused by a chip and substrate entity bending and the first failure occurs at the joint located in the center of BGA package that became open. It

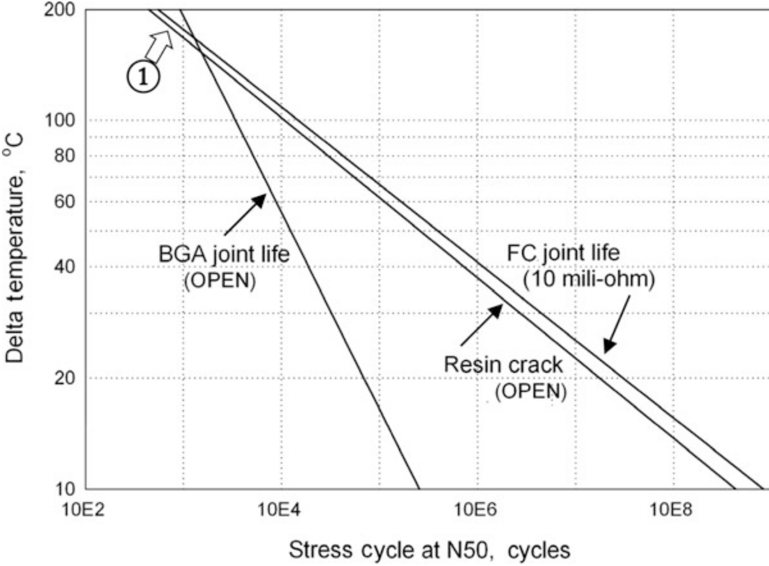


Fig. 7.49 Life estimation by failure mode. After Tsukada [14], 2000

is a fatigue failure of solder, eutectic Sn-Pb in this case, at BGA joint following the modified Coffin-Manson's formula that a life of joint follows inversely proportional to a strain at the point by square-root law. In the meantime, a failure by a line open in a substrate due to a dielectric resin crack indicated as "Resin crack" had significantly higher acceleration compared with a BGA joint and showed following approximately fifth-root law if applied Coffin-Manson's formula. These results were obtained by a different fatigue nature of solder and epoxy resin. A failure by resin crack brings a shorter life compared with BGA joint life above around $\Delta T = 160$ °C pointed by block arrow ①, but the life under $\Delta T = 100$ °C where most of application operates is much longer compared with a BGA joint life. Such result tells two important things.

1. A package life must be tested with loading to mother board, i.e., in an application condition.
2. When a material changed, a stress test to confirm an acceleration factor by failure mode has to be redone with a package loaded to mother board.

In the figure, flip chip joint life is shown as FC joint life. It was not actually a failure by reaching to the failure criteria but 10 mΩ resistance change was used by a way of experiment since it took too long test period if applied a formal failure criteria. Interestingly, as shown in the figure, FC joint life was matching with a curve of Resin crack. This indicates a flip chip joint is protected with underfill resin that is also epoxy and a degradation of joint occurs along with a degradation of resin by losing its protection.

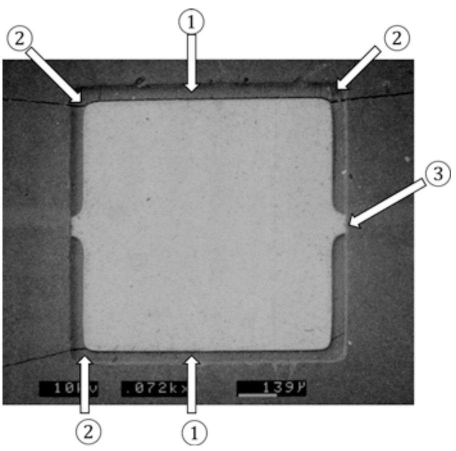


Fig. 7.50 Resin crack around pad. After Tsukada [6], 1998

One of typical failure mode of organic substrate is cracking of a epoxy resin. Even though the acceleration factor of cracking on epoxy is high as described before, it is an intrinsic failure mode and dominates a product life when stress design of a substrate is not properly performed in relation with acceleration test result. Figure 7.50 shows a typical nature of this failure mode that is revealed at certain point when a thermal cycle test is extended more than required for product life assurance. There is a crack along with a edge of large copper pad 1 mm square in size on a substrate surface, It is extended to solder mask from corners ②. Crack does not occur on an edge where there is an entry line to the pad that possibly relax the stress along the edge. This indicates a large copper structure in organic substrate should have a design feature to deconcentrate a stress around it.

Figure 7.51 shows a crack around internal conductor structure. It appears white lines around conductor pattern. When a thermal cycle test is extended more than a product life assurance, not only on surface of a substrate, a crack is possible to be found at internal dielectric layer along with a edge of copper structure such as line and land. As previously described, there is a high *Z*-direction movement in internal build-up layers that create a stress at the edge of internal copper structure. This indicates that there is a risk of such stress concentration anywhere in a substrate and one has to follow a detail design rule to avoid the stress concentration as much as possible. In addition, adding a filler to dielectric material and increase a modulus with making material harder has a certain level of effect to reduce *Z*-direction movement. It is the second benefit of adding a filler to dielectric material other than lowering the CTE. However, there must be an enough caution for adhesion of filler surface to dielectric material. Otherwise a separation of filler and dielectric becomes a starting point of crack and makes a life shorter in contrary.

Figure 7.52 shows a typical failure mode under temperature, humidity, and bias stress test. There is a clear growth of copper dendrite from cathode to anode. However, a properly fabricated organic substrate will not show this failure mode by a harmful timing to product life. Copper migration in 35 μm thickness dielectric

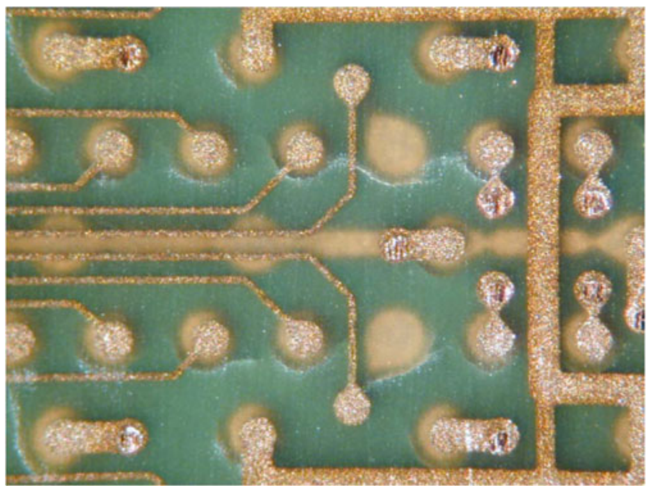


Fig. 7.51 Resin crack around internal pattern. After Tsukada [14] 2008

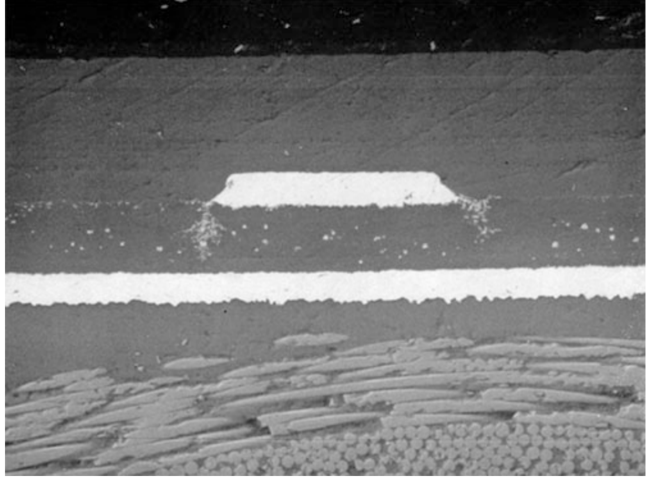


Fig. 7.52 Copper migration. After Tsukada [6], 1998

layer shown in the figure was obtained by more than 2,500 h under 85 °C, 85 %, 5 V stress condition. Since a radius of copper ion is less than an angstrom, a cross-link density of epoxy is not able to completely shut out from movement under electrical filed. Eventually copper metal migration occurs in epoxy based substrate and a consideration is just a matter of timing to a product life. Therefore, one has to have a cautious that insufficient cross-linking may cause a failure in a critical range to a product life and an acceleration factor of this failure will also vary by a bias voltage and change of electrical filed in a build-up layer. In some cases, stack up lines in

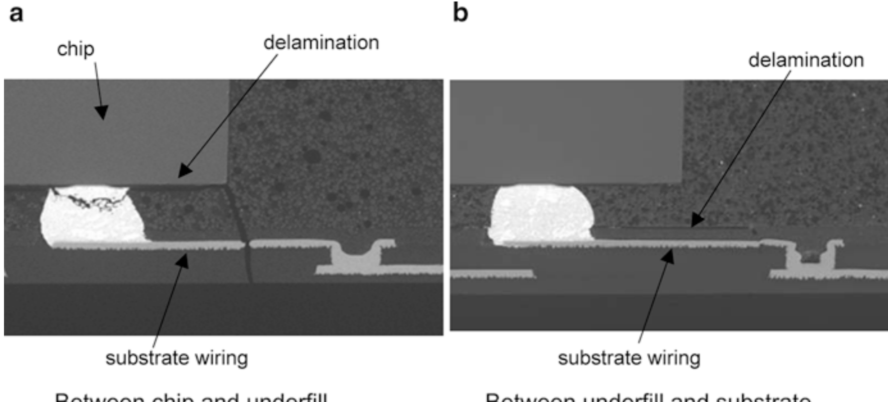


Fig. 7.53 Delamination between materials. (**a**) Between chip and underfill, (**b**) between underfill and substrate, after Tsukada [18] 2004

vertical direction by multi-layer build-up may create a stronger concentration of electrical field at the edge of line and cause a shorter life in copper metal migration.

One of typical failure of plastic material is a loss of adhesion. Since stress is distributed along with each material in an organic package as described previously, adhesion loss causes a stress concentration and ends up a breakdown of substrate wiring in most of the case. Figure 7.53 shows typical cases of such failure mode. In (a) photograph, there is a delamination between a chip and underfill. The separation cause a crack at the end of chip and the crack propagates to a direction of substrate along with a stress gradient at the point. The crack is extended into solder mask and dielectric layer. When a dielectric layer cracks, a wiring line breaks easily. In this case, there is a clear crack at the top of joint that was a result of the separation between a chip and substrate. In (b) photograph, there is a delamination between underfill and solder mask. The separation causes a stress concentration at a certain point and a dielectric resin cracks with breaking down a wiring line. As these examples, a separation of materials causes a fatal failure of a substrate and sometimes a joint failure at the same time.

Figure 7.54 shows a crack came out at the edge of underfill. Due to a nature of package bending caused by CTE mismatch of a chip and substrate, underfill fillet is under a high tensile stress since a chip and substrate tends to be separated at the edge of a chip. Normally, a modulus of underfill is higher compared with substrate materials since substrate materials have to flow and fill a space and corners between wires and/or pads. When a modulus of underfill is excessively higher than substrate materials such as solder mask and dielectric resin, substrate materials are broken down and crack is initiated at the fillet edge that propagates into the substrate as in the figure.

A substrate is the largest part of a flip chip package in its size. Figures 7.55 and 7.56 show a brief comparison of influence of substrate size on a system package size and performance. Figure 7.54 shows SiP package sizes with different size of BGA

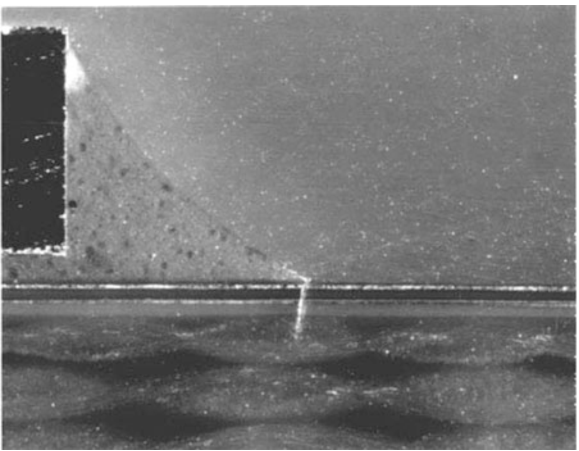


Fig. 7.54 Substrate crack at fillet edge. After Tsukada [18] 2004

Fig. 7.55 Influence of package size for SiP size. After Tsukada [12] 2008

packages. In the figure, there are three cases depends on a substrate size of BGA to construct a SiP with four chips, i.e., four single chip packages. They are Case42.5 with 42.5 mm square BGA substrate, Case35 with 35 mm square BGA substrate and Case10 with 10 mm square BGA substrate as a case that is a several times higher density than other two cases. Four chips are lay out with these substrate size to form a SiP. To assume each substrate can contain 6,000 net for an application, each wiring rule and BGA pitch is defined.

For Case42.5; chip size: 10 mm square, substrate wiring line/space: 25/25 μm, BGA pitch: 1.0 mm.

For Case35: chip size: 10 mm square, substrate wiring line/space: 20/20 μm, BGA pitch: 0.8 mm.

For Case10: chip size: 5 mm square, substrate wiring line/space: 7.5/7.5 μm, BGA pitch: 0.2 mm.

A performance comparison is made with the worst case signal transmission with a pass right upper chip to left lower chip as a critical bus length. Resulted geometry in each case is as follows. In Case42.5, SiP size is 100 mm square and critical bus length is 98.2 mm. In Case35, SiP size is 85 mm square and critical bus length is 82.5 mm. In Case10, SiP size is 29 mm square and critical bus length is 22.8 mm.

With a signal transmission along a critical path, eye pattern of each case is shown in Fig. 7.55. Since a substrate size is large in Case42.5, eye pattern shows an possible operational frequency as 5 GHz. It improves in Case35, however the margin is not significantly increased. 10 GHz is the maximum before an eye pattern is deteriorated. In contrary to these two cases, Case10 shows a remarkable improvement on eye pattern with good margin for 20 GHz. It is obvious that a substrate size has significant influence on a total performance of package and, in addition, a unit cost of substrate becomes lower when the substrate size shrinks since a number of substrate per fabrication work size increases in reverse to the rate of shrink. Then, it allows to use higher cost material and higher cost processes with even an end cost becomes lower by such result of scaling action.

7.3.7 Historical Milestone

Figure 7.57 is a photograph of the first product in the world with a flip chip on an organic substrate that was shipped in the beginning of 1990. It was a 16 MB SIMM card for PC application with 18 chips attached directly on front and back of ordinary FR4 substrate. Chips were reinforced with underfill epoxy to protect a joint from stress caused by CTE mismatch. Since chips were memory that is low I/O, ordinary FR4 PCB was enough to work as a substrate for wiring I/O’s necessary for a function of SIMM card.

Figure 7.58 shows the first build-up substrate product. Two logic chips 12 and 8.7 mm square in size at locations indicated by block arrows were attached directly to two layers build-up PCB as a substrate. (b) Shows a cross section of chip and (c) shows a cross section of build-up layers carrying micro-via holes. A dielectric material was a commercial solder mask and a micro-via hole was formed by photo etching process. Bare chips were processed through “burn-in” before installed to a build-up PCB substrate with using temporary carrier by chip attach and remove processes taking flip chip joint advantage of chip replacement technique. Since the technology was designed to handle a chip as one of a surface mount component, chips were attached in the same assembly line with other surface mount components. A function of the card was character recognition for Kanji letters.

Figure 7.59 shows the first BGA using organic substrate. 17 mm square ASIC chip with 1918 I/O’s were attached on 33.5 mm square build-up substrate with four layers build-up on four layers FR4 (i.e., 4 + 4 + 4 construction). Micro-via holes were formed by photo process. There are four studs at four corners of substrate for securing heat sink. BGA joints are on a full grid array.

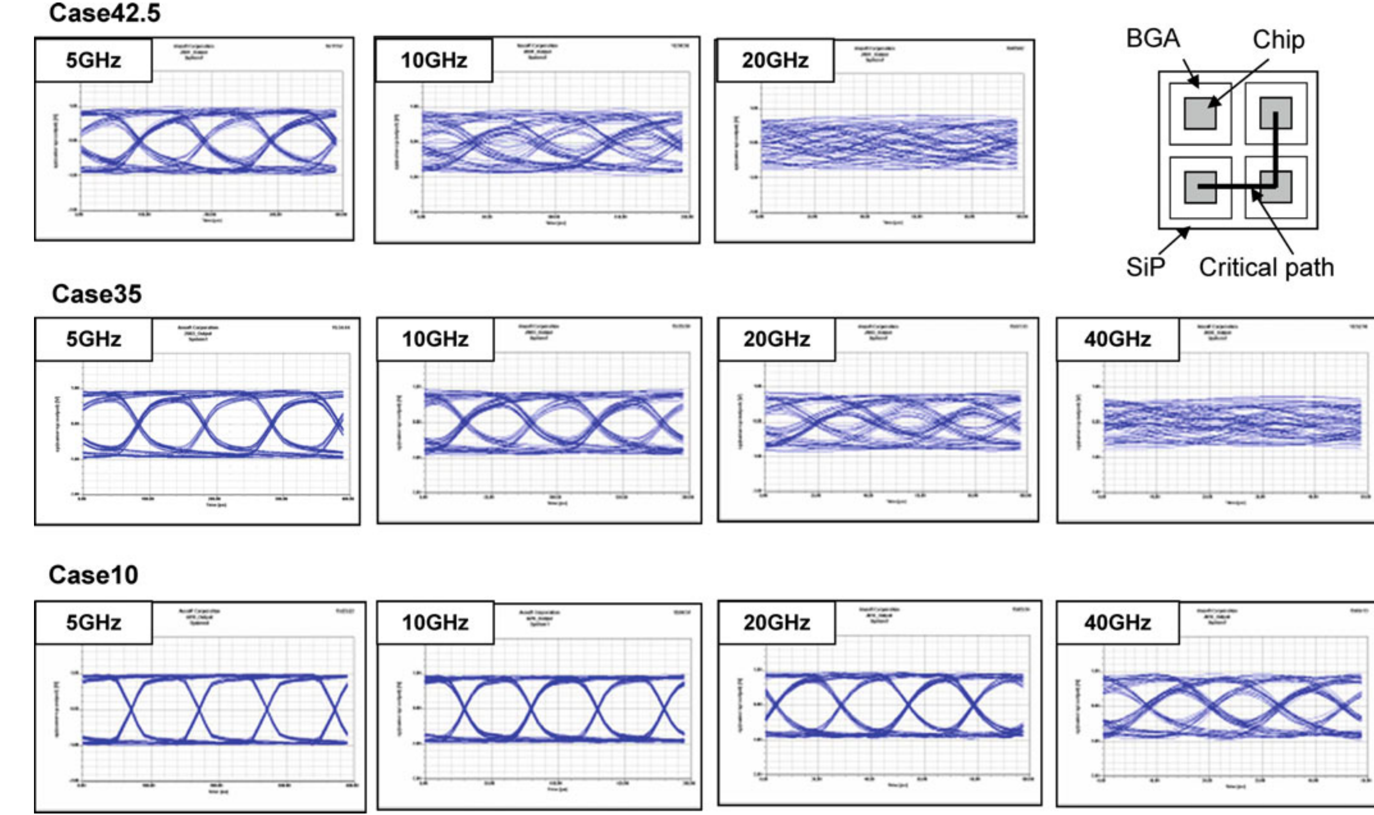


Fig. 7.56 Effect of size for performance. After Tsukada [12] 2008

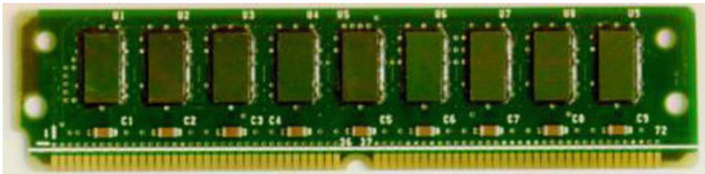


Fig. 7.57 First organic substrate product. After Tsukada [6] 1998

Figure 7.60 shows a cross-sectional photograph of the first organic substrate microprocessor for PC application. Pentium II from Intel Corporation started to use organic substrate in 1998. As shown in a photograph, a chip was attached to organic substrate with BGA package format. The BGA package was loaded onto pinned interposer to make a final format as PGA. The organic substrate employed a laser micro-via hole instead of photo via hole. A laser micro-via hole became a dominant process for organic substrate since this product.

Figure 7.61 shows photographs of the first supercomputer that employed an organic substrate package for main processor. A supercomputer named "Earth Simulator" located in Yokohama-city Japan was the first place of "Super computer TOP500" from June 2002 through November 2004. Table 7.4 describes a specification of package. A main processor package has an organic substrate that has four build-up on both sides of eight layer core ($4 + 8 + 4$ construction) with 140×112.5 mm in size. It carries a large processor chip with 19.84×21.04 mm in size that has 8960 I/O's. At the end of 2010, "Earth Simulator" is still running at 4th in supercomputer ranking with improvement in architecture and linked severs.

7.4 Z-Stack Type Substrate

Since a sequential build-up substrate is leading an organic substrate technology in cost/performance area such as CPU, GPU, and game processor substrates where a superior performance is required as well as low cost, there is not so many variations in its structure other than detail design elements. In the meantime, Z-stack type has more varieties in its application because it has been mainly used in low-end area such as mobile devices and consumer products. Not only applications, there are some unique features in each fabrication method.

7.4.1 Z-Stack Substrate With Pattern Transfer

In this section, Z-stack type substrate "CPCore" from KYOCERA SLC Technology Corporation is introduced as one of such examples. Figure 7.62 shows a process flow of this method. (1) A sheet of copper is attached to a polymer carrier film and a dry

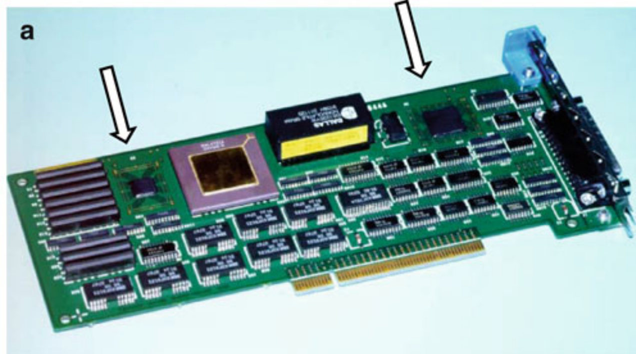


Fig. 7.58 First logic chip on build-up substrate. (**a**) First build-up PCB package, (**b**) first logic chip directly attached, (**c**) Cross section of build-up PCB. After Tsukada [6] 1998

film resist is laminated for circuit patterning. The resist is exposed and developed to form a wiring pattern of copper on the polymer carrier film. (2) In parallel, uncured (green) dielectric sheet is drilled by laser to form via holes. Then via holes are filled with conductive paste made of copper particles and resin binder. (3) The wiring pattern is transferred onto an uncured dielectric sheet. The uncured dielectric sheet with transferred circuit for each layer of substrate design are stacked, then, pressed

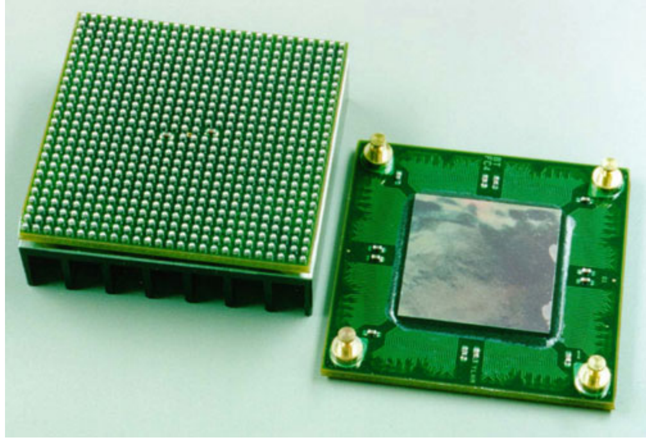


Fig. 7.59 First BGA using organic substrate. After Tsukada [14], 2000

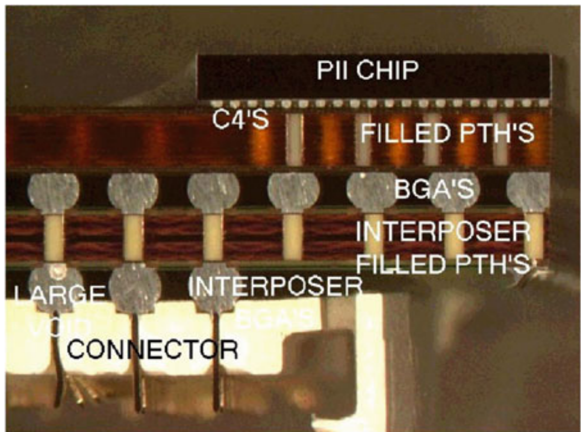


Fig. 7.60 First microprocessor on organic substrate. After Tsukada [13] 2008

and cured by a hot press for completing a substrate. The flow of the simultaneous curing process is similar to that of the green sheet process of ceramic substrate.

Figure 7.63 shows a connection of conductor copper to paste in via hole. In wiring pattern transfer process, a copper conductor pattern is buried into a prepreg by pressure. At the same time, conductive paste that fills a via hole is depressed by a thickness of buried copper conductor. The conductive paste that fills a via hole and copper conductor is firmly adhered by the depression action to secure the connection. Finally, CuSn intermetallic compound is formed in the interface of copper conductor and conductive paste to secure the reliability of connection.

Figure 7.64 shows a cross section of completed seven metal layers substrate. A prepreg is using a thermo setting PPE (Polyphenylene ether) for better processability and electrical characteristics.

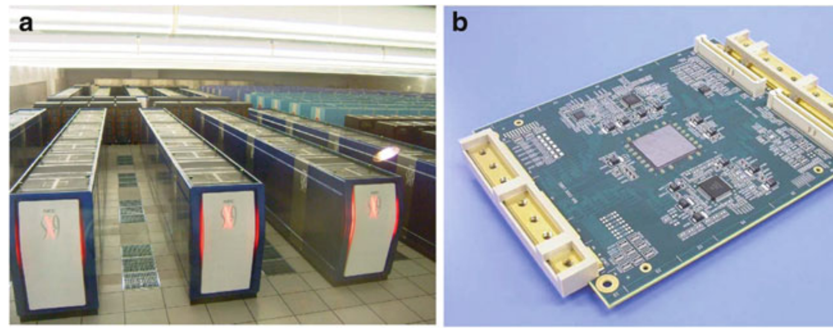


Fig. 7.61 First supercomputer with organic substrate package. (**a**) Total view of the system with 160 Node, 1280 CPU's at original configuration. (**b**) Processor package, chip at the center, power connectors on both end. Courtesy by Inasaka, NEC Corporation

Table 7.4 Package specification, courtesy by Inasaka, NEC Corporation

LSI I/O (Sig.)		8,960 (1,791)
LSI size (mm)		19.84 × 21.04
PKG size (mm)		140 × 112.5
FC joint		Sn/Ag/Cu
Cap.	On chip (nF)	1,020
	On PKG (uF)	103
Layer count		4Bu-8core-4Bu
PKG connection		Cable connector
PKG pins (Sig.)		2,628 (1,521)
Wire length (m)		80
CLK (GHz)		3.2
Power(W)		250

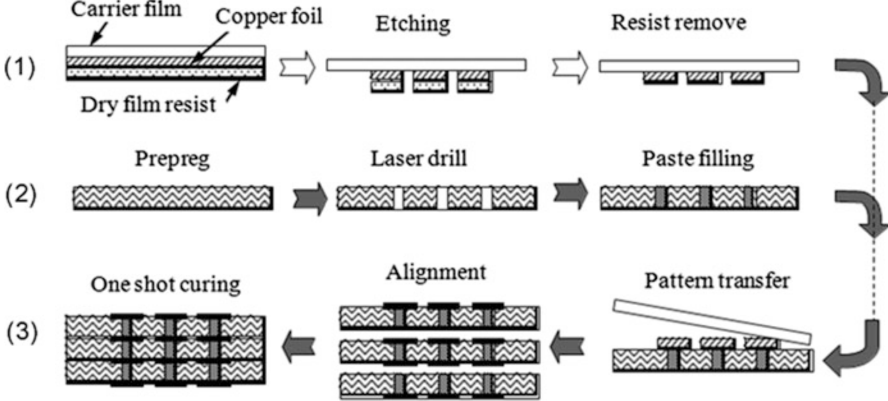


Fig. 7.62 Z-stack and pattern transfer. After Tsukada [18], 2004

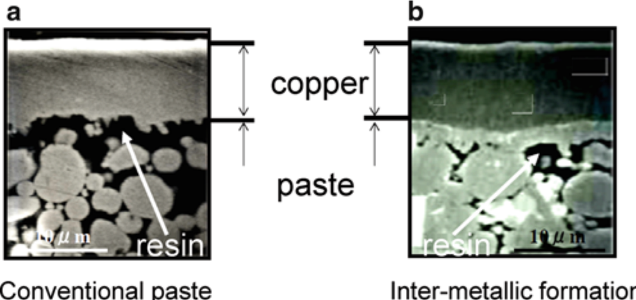


Fig. 7.63 Via hole connection. (**a**) Conventional paste, (**b**) intermetallic formation. After Tsukada [18], 2004

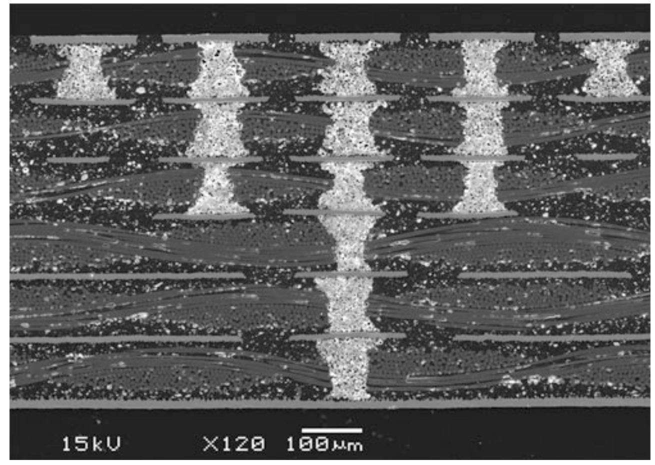


Fig. 7.64 Completed Z-stack substrate. After Tsukada [18], 2004

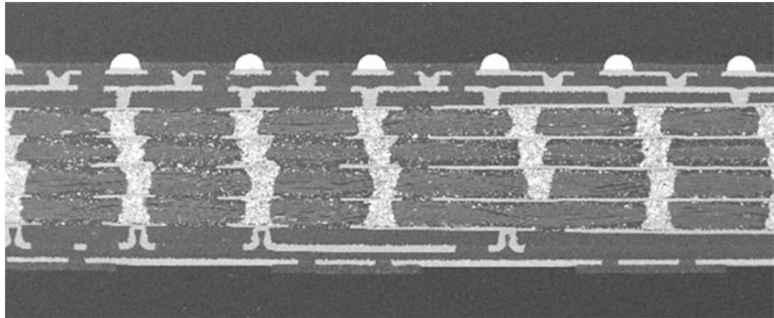


Fig. 7.65 Z-stack + build-up combined substrate. After Tsukada [18], 2004

Figure 7.65 shows a version of Z-stack and sequential build-up combined substrate for a higher density application.

7.4.2 Any Layer Via Substrate

There is a substrate called as "Any layer via." Originally "Any layer via" is the name put on a Z-stack type substrate. It was started with "ALIVH" from Panasonic. The name of "ALIVH" itself is a shorten version of "Any Layer Interstitial Via Hole." It does not have a core and all layer is a stack of same structure. Each layer of substrate is prefabricated with a laser drilled micro-via hole fill with conductive paste and stacked by press to complete a substrate. However, there is another type of substrate called "any layer via." It is essentially a sequential build-up substrate with a core at the center, but the core is very thin as it looks the same structure with build-up layers. Therefore, its cross section looks like coreless substrate. But the design rule is quite different since each layer is using a prepreg that has a glass fiber cloth inside. Though a very thin glass cross that is introduced also in this chapter is used, line and micro-via hole size are far coarse from the level of a regular build-up substrate. High yield with very low level wiring ground rule, a number of sequential build-up count is high as a several layers on each side of a core.

7.4.3 Embedded Component Substrate

An ordinally embedded component substrate is burying components at a core part of a sequential build-up type substrate. A core with buried components is fabricated first and the build-up layer is applied later. Since an embedded component substrate is an application oriented package, there are wide variations of its design in detail. In this section, as a unique example of embedded component substrate, "B^2it" (buried bump interconnection technology) from Dai Nippon Printing with embedded component is introduced. B^2it is one of a Z-stack type substrate with employing a unique bump via construction method. Figure 7.66 shows a basic process flow.

1. Printing: Ag paste is printed on Cu foil to form a conductive bump. A paste is cured and a bump is formed like conically shaped protrusion as shown in a photograph.
2. Piercing: A prepreg sheet is laid up and pressed so that a bump comes through prepreg by breaking and pushing aside a glass fiber.
3. A Cu foil is laminated and pressed to complete a via connection. Then, a laminate is forwarded to a patterning process.

After completing a patterning, proceed to a press process to complete a substrate as shown in Figure 7.67.

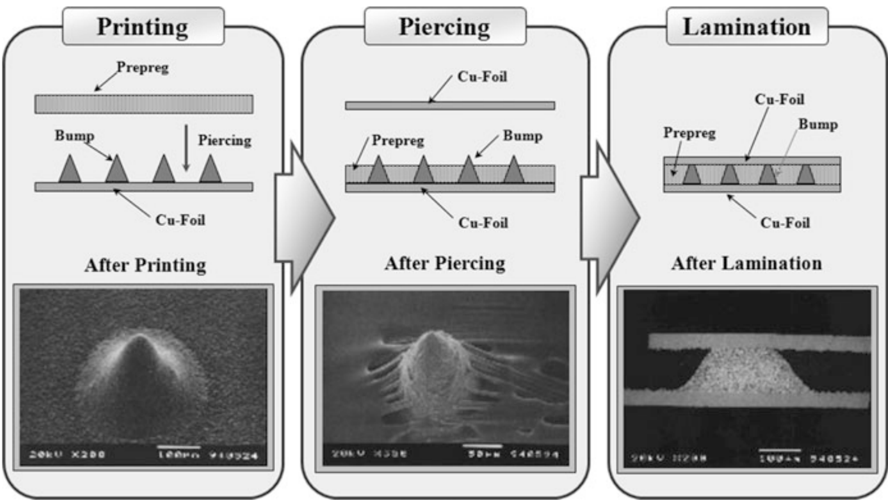


Fig. 7.66 Basic process flow. Courtesy of Dainippon Screen MFG. CO., LTD.

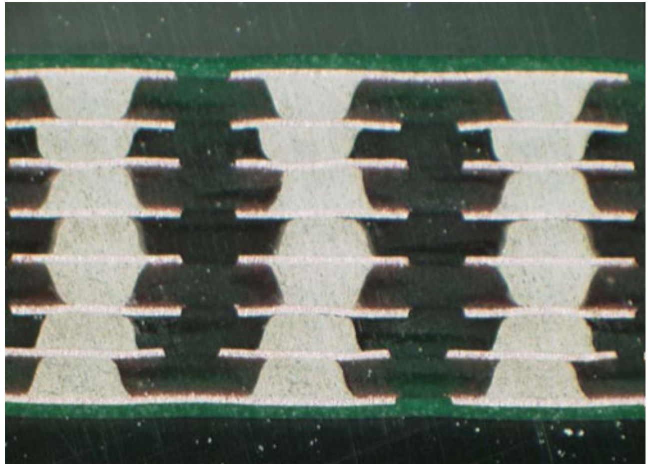
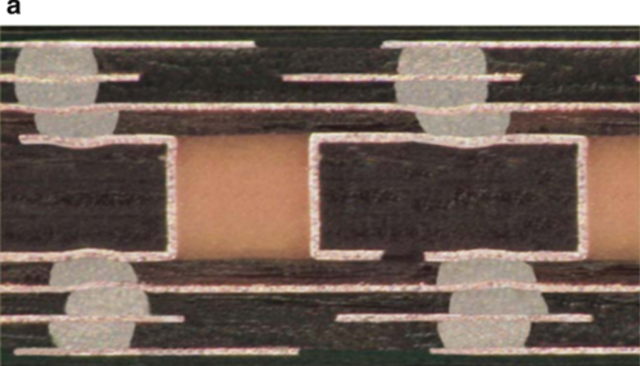


Fig. 7.67 Completed B^2 it substrate. Courtesy of Dainippon Screen MFG. CO., LTD.

Figure 7.68 shows a variation of substrate that is fabricated with other method. (a) Shows a photograph of combination with an ordinary thorough hole core. (b) Shows a combination with build-up layer with a laser micro-via hole.

Figure 7.69 shows an example of component embedded substrate. Buried components are three wafer level CSPs and total 18 passive components that are 1005C and 0603C. The substrate size is 9.2 mm square with 0.65 mm in thickness and constructed with six layer B^2it. A Halogen free dielectric material is used.



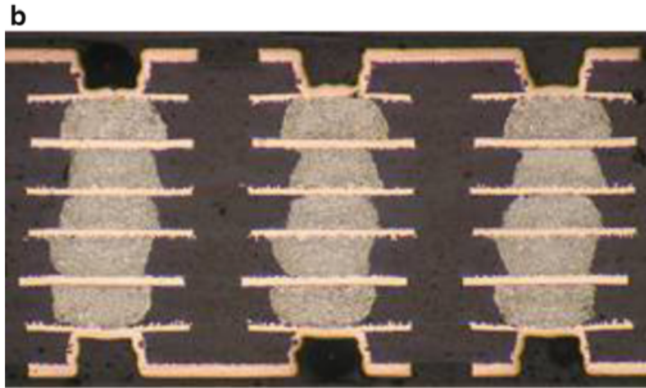


Fig. 7.68 Combination with other method. (**a**) With through hole core, (**b**) with laser via build-up layer. Courtesy of Dainippon Screen MFG. CO., LTD.

Figure 7.70 shows an example of active component embedded substrate. Buried components are a bare chip with 3.1 mm in size, 1005c and 0402C passives in total ten pieces. The substrate is 8.5 mm square with 0.48 mm in thickness with seven layers B^2it. Overall package thickness is 1.0 mm with 46 total components including a quartz. A function of the module is NFC (near field communication) that a requirement is getting high in recent days.

Figure 7.71 shows a process flow of embedded component substrate introduced in a previous figure. As the first step, front side layer and base layer by B^2it process are provided. Also a core part that carries internal wiring plane is provided. Assemble passive components onto a base layered by regular SMT process and a bare chip is attached with flip chip bonding. Bumping and prepreg lamination on one side of front layer and core part are done. Cavities for chip and passives are formed. Lay up all parts and lamination press process completes a stacking of layers. Solder mask is applied and assemble other necessary components to finish a product. Most of

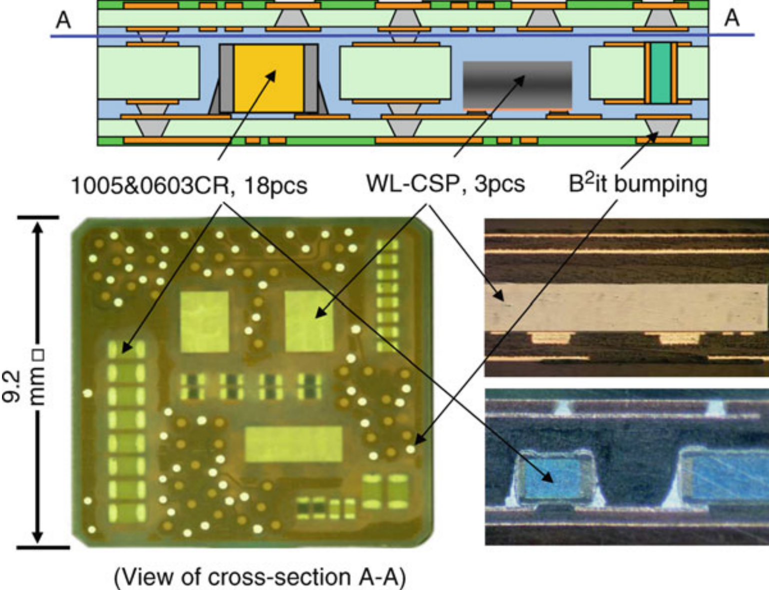


Fig. 7.69 Example of WL-CSP and passives imbedded. Courtesy of Dainippon Screen MFG. CO., LTD. [19] 2010

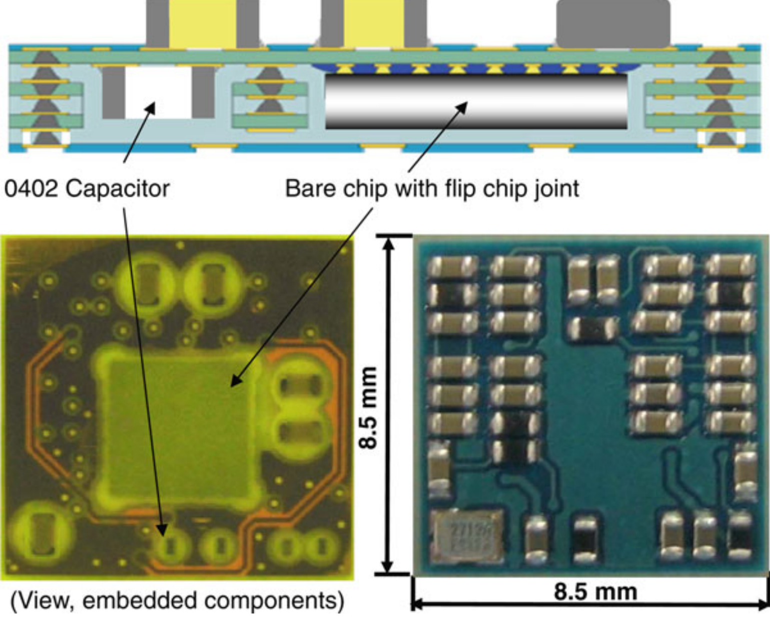


Fig. 7.70 Example of embedded active component. Courtesy of Dainippon Screen MFG. CO., LTD. [20] 2007

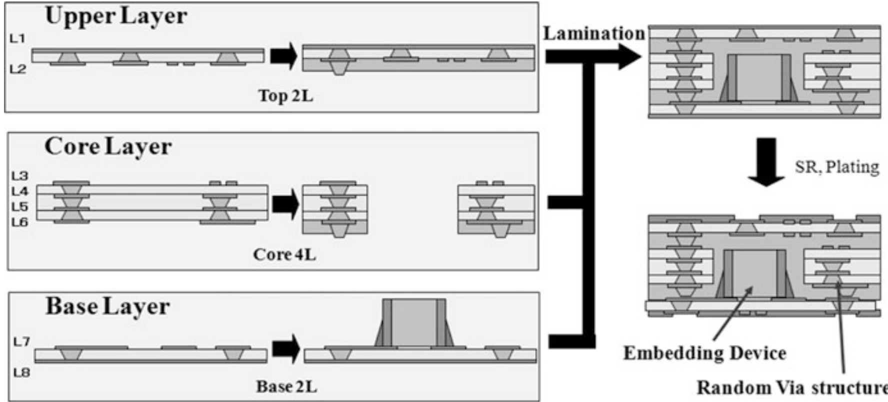


Fig. 7.71 Process flow of embed component substrate. Courtesy of Dainippon Screen MFG. CO., LTD. [20] 2007

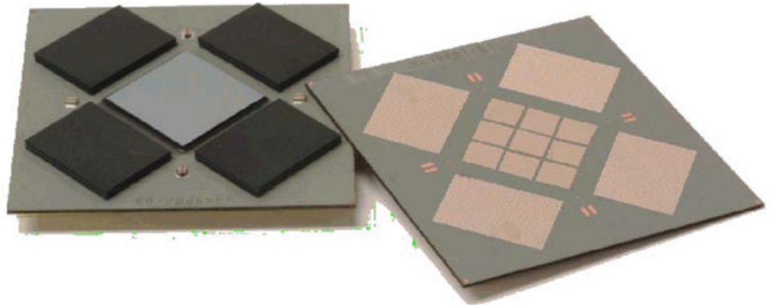


Fig. 7.72 Substrate with PTFE for high speed signal. Courtesy of Endicott Interconnect Technologies, Inc.

embedded component substrates are in peripherally of portable phone product in primacy of thin and small advantages.

7.4.4 Substrate With PTFE Material

Figure 7.72 shows a substrate that employs an unique material for high speed signal transmission. “HyperBGA” from Endicott Inter Connect Technologies, Inc. is a flouropolymer-based coreless semiconductor package that allows run a signal at extremely high rates of speed. The combination of the low loss, low dielectric constant material and strip line cross sections enable signal speeds surpassing 12 Gb/s. The material compliance of the PTFE, combined with the dimensional stability of a copper-invar-copper center plane, enables long field life, with none of the BGA joint wear out, die cracking, delamination or flip chip bump fatigue.

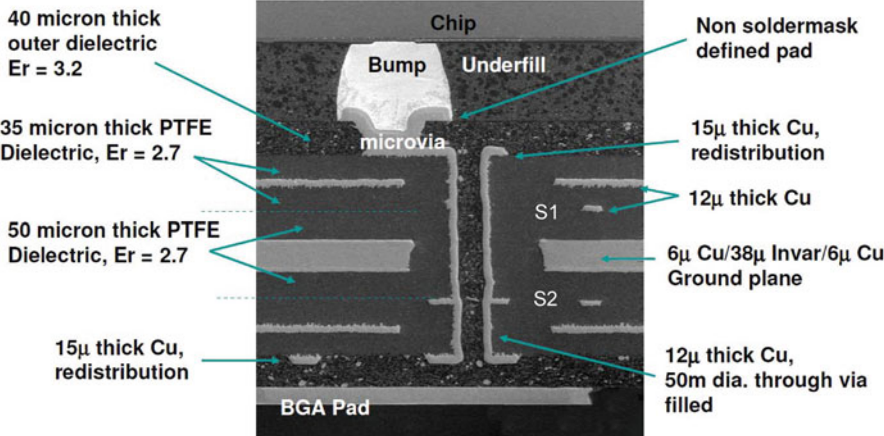


Fig. 7.73 Substrate cross section. Courtesy of Endicott Interconnect Technologies, Inc. [21], 2003

Table 7.5 Material property

PTFE dielectric constant	2.7 at 10 GHz
PTFE loss	0.003 at 1 MHz
Outer layer dielectric constant	3.2 at 1 GHz
Outer layer loss	0.0027 at 1 MHz
Composite flexural modulus	1.2 Mpsi
Coefficient of thermal expansion	10–12 ppm/°C

Figure 7.73 shows a unique coreless structure in cross section. At the center, a thick copper-invar-copper is set for controlling CTE and dimensional stability.

There are two signal layers S1 and S2 that are embedded in a strip line environment sandwiched between either voltage plane and center core (ground plane). Two redistribution planes located outer side of voltage planes support escape lines from flip chip joint to signal wiring. Top and bottom planes only provide a flip chip pad and a BGA pad, respectively. Interlayer connections are accomplished with micro-via holes and plated through-holes (PTHs).

Table 7.5 describes material property of this substrate. PTFE provides a low dielectric constant and low dielectric loss. And, overall low CTE and low modulus lower the stress at flip chip joints.

7.5 Challenges

7.5.1 Coreless Structure

Figure 7.74 shows a cross section of a coreless structure. One of weak points of sequential build-up type substrate is a construction of core, particularly a through

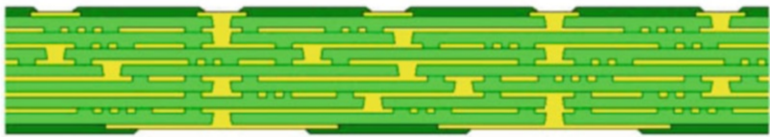


Fig. 7.74 Coreless substrate

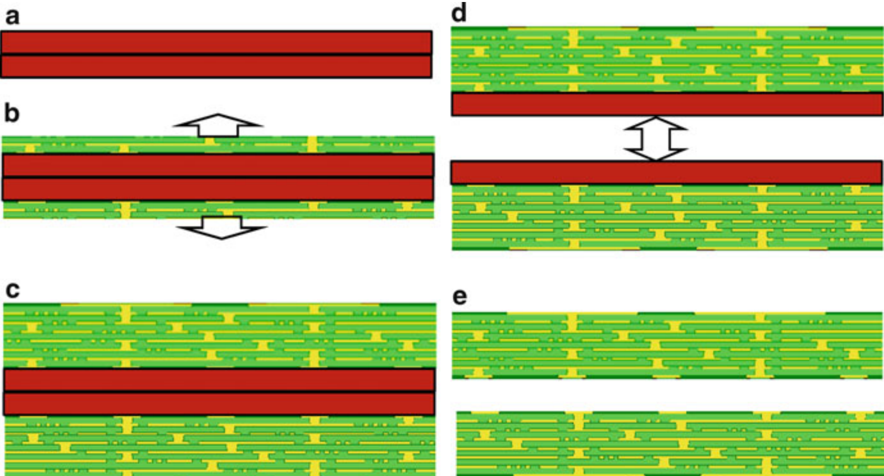


Fig. 7.75 Coreless process steps

hole. Since a core is carrying a role to provide a mechanical rigidity, it is normally thicker than build-up layers. A path from front side to back side is a through hole that is currently processed by a mechanical drilling process. It is larger in dimension compared with a micro-via hole and becomes an obstacle in terms of electrical performance. The idea is to delete a core portion from a standard sequential build-up structure.

However, problems are a low modulus of laminate that causes dimensional stability worse and a cost if it processed by sequential build-up in one side since an ordinary sequential build-up is processed on double side. To resolve a cost issue, as shown in Fig. 7.75, the development is focused to prepare a core made of two parts adhered using a dummy panel and process build-ups on both sides of the core. After finishing the fabrication, the work panel is separated and a remaining part of core that has been with a coreless part is removed. Finally, two sequential build-up coreless panel is made for a process to substrate pieces.

Even though taking such effort, there are fundamental issues on this type of substrate. One is a low modulus and dimensional stability of a finished part since build-up laminate material is a plastic that is essentially very low modulus like

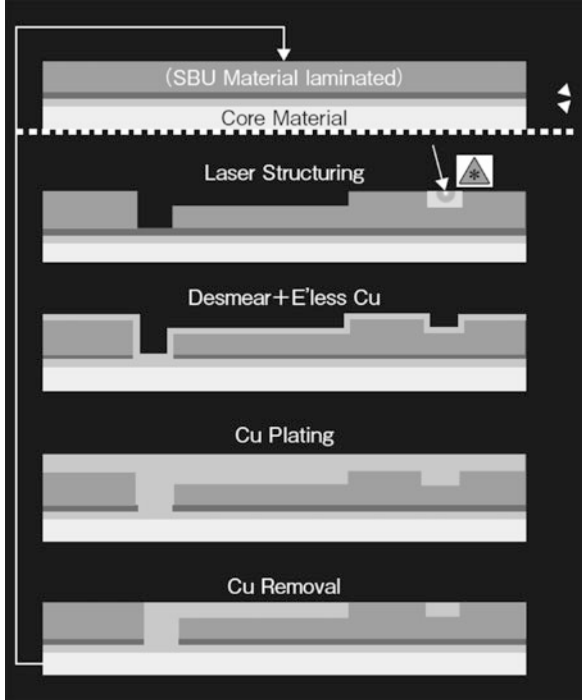


Fig. 7.76 Process flow of trench structure. Courtesy of Baron, Atotech Deutschland GmbH [23], 2011

5–7 GPa. Yet a copper in the laminate is high modulus compared with such resins. Therefore a finished laminate shows a bend, twist and distortion that are more in local area of substrate rather than a global uniform warpage compared with standard construction with a core. This nature will vary by each design since it depends on a conductor pattern. In addition, theoretically it is higher cost since a number of sequence to make a build-up is about twice of making a standard sequential build-up substrate that creates significant impact on yield loss since a lager build-up count becomes one of major weak point of sequential build-up process that is described in early part of this chapter. In standard sequential build-up substrate, a core is shared for power planes that is coarse in wiring rule with utilizing ordinary PCB process to lower the cost. In addition, it is said as disadvantage of core because a lower (back) side build-up wiring plane is not fully utilized since through hole density is not reach to the density of micro-via plane in build-up. However, particularly in BGA substrate, BGA side wiring density is not necessary high compared with front side that is high density since it contains chip area escape lines.

In recent disclosures, it was described as superior in performance and user company of a substrate took an effort to assemble a retainer for warpage counter-measure and implement an improvement on a positioning system in assembly process to avoid tipping of substrate [23].

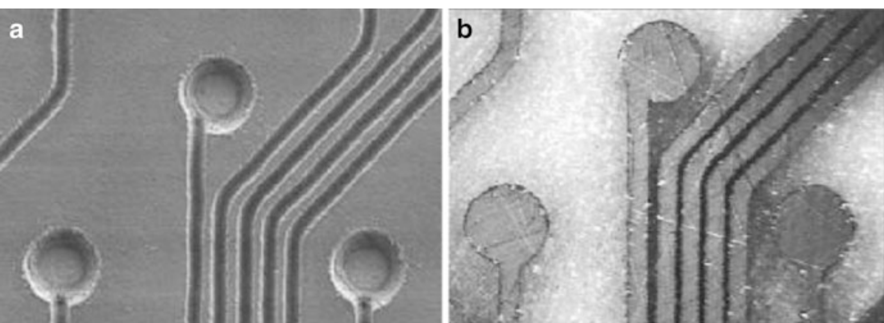


Fig. 7.77 Trench and finished circuit pattern. (**a**) After trench formation by laser, (**b**) after completing circuits. Courtesy of Baron, Atotech Deutschland GmbH [23], 2011

7.5.2 Trench Structure

Figure 7.76 shows a process flow of a substrate fabrication method named laser trench method named as “V2” from Atotech Deutschland GmbH. The process consists of laser curving and plating with planarization, and eliminating photo circuitization process that is carrying a major yield detractor in an organic substrate fabrication process. The first step is a laser abrasion after a lamination of dielectric material. Laser grooves lines and via holes. Since a depth control of UV-YAG laser can be done far finer than CO_2 laser due to its low energy per shot with high frequency, conductor and via depth are controlled. After a surface roughening of light desmear process, seed layer is formed by electro-less copper plating. Lines and via holes are by electroplating. Each circuit element is completed by planarization process and a substrate moves to next layer step.

Figure 7.77 shows grooved trenches. (a) Shows a trench pattern formed by laser grooving and (b) shows completed conductor lines and via holes. One of advantages of is method is to form lines and via holes by a single registration at laser tool compared with two registrations are involved that are laser process and photo process in case of ordinary build-up process steps.

Figure 7.78 shows a cross section of a pattern and via hole. By utilizing a plating additives specially designed for this purpose, filling of all pattern is possible with minimum dimple height with less than 10 μm plating thickness remained on the general area of dielectric surface.

Figure 7.79 shows a result of dense fine pitch lines and cross section of product substrate. It is said in the disclosure that key points for success are plating thickness control and planarization. Filler particle size in dielectric material relative to design rule is also important to achieve uniform geometry definition by laser abrasion. 355 nm UV-YAG and 248 nm excimer laser are capable below 10 μm wiring features.

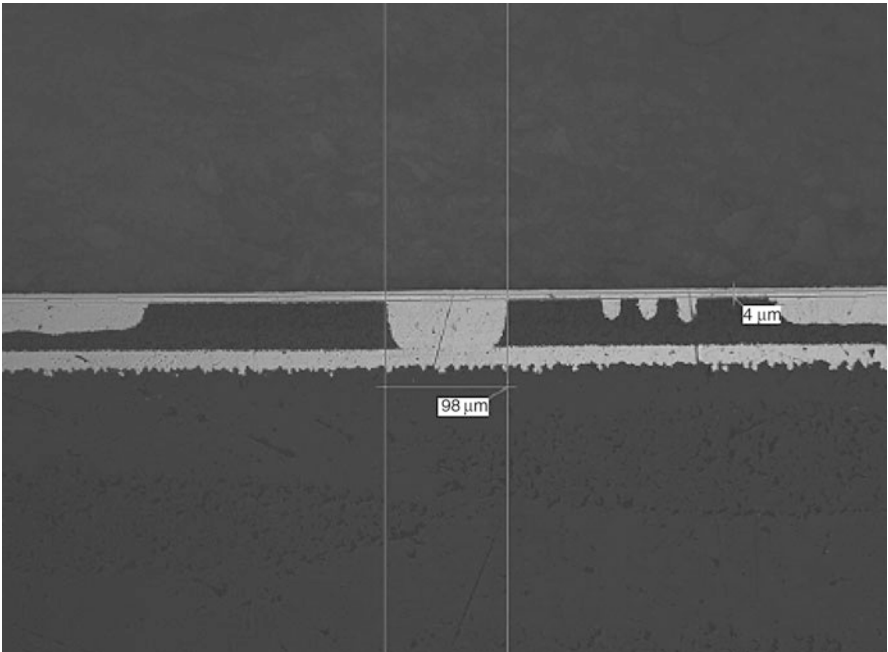
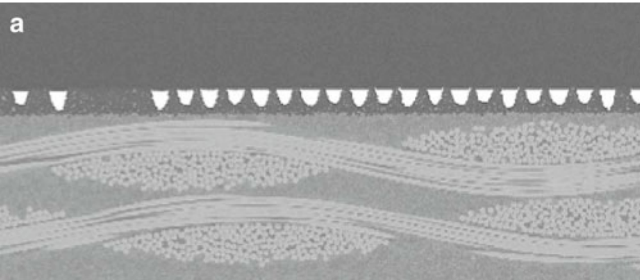


Fig. 7.78 Cross section after copper plating. Courtesy of Baron, Atotech Deutschland GmbH [23], 2011

7.5.3 Ultralow CTE

A basic issue of a current organic substrate is its high CTE due to carrying a conventional PCB material set. It has been a common understanding that an organic substrate CTE has to be lower to reduce a global miss-match of CTE with a silicon. A question has been how low the CTE is possible to be. Figure 7.80a shows a cross section of construction to reduce a CTE with a new material set. The configuration of three build-up layers on both side of two metal layer core. The core has organic fiber made of a poly-p-phenylenebenzobisoxazole (PBO) with a CTE as low as −6 ppm/°C and a high Young's modulus of 270 GPa. The fiber is impregnated with a polyamide resin. The uniqueness of a core is not only a material set but the way of impregnation. The fiber is featured in an unidirectional construction that achieves a resin content of the prepreg as low as 40 % to raise an effectiveness of fiber reinforcement on the CTE. The core CTE is shown as "0 + 0 cu %" in Fig. 7.79b that reaches to −1 ppm/°C. In (b) of the figure, the CTE of 3 + 2 + 3 composite is shown approximately 3.5 ppm/°C. Another point that needs attention in the figure is an influence of copper. Beside a composite CTE measurement that is done with 50 % copper ratio in wiring plane, a case with 100 % copper ratio is shown and tells us a copper has a significant influence to a global CTE of substrate due to high CTE as 17 ppm/°C with high modulus around 100 Gpa. It is a point to be emphasized in low CTE challenge in an organic substrate.



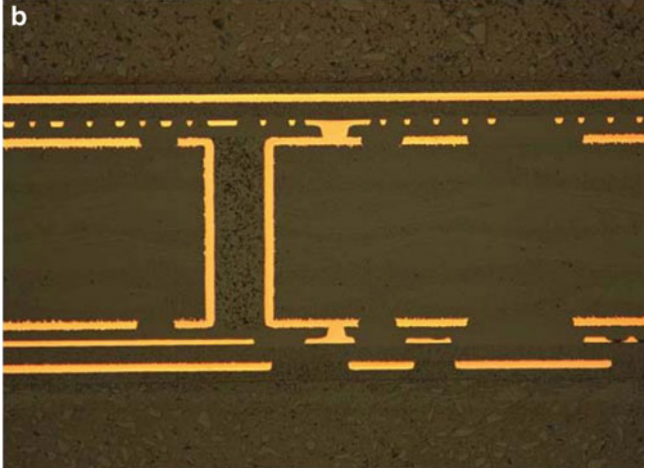


Fig. 7.79 Completed structure. (**a**) High pitch conductor trace, (**b**) substrate cross section. Courtesy of Baron, Atotech Deutschland GmbH [23], 2011

With an ultralow CTE construction, a product prototype is built as shown in Fig. 7.81 and cleared required stress tests target. The substrate size is 10 mm square with 100 μm pitch through holes in the core and 8–10 μm pitch lines in build-up layers with 100 μm pitch flip chip joint. These features accommodate the density of a chip I/O of 104 cm^{-2}, which is about ten times greater than that achieved in a current organic package and expected to satisfy the next generation requirement.

7.5.4 Substrate for Stacked Chip

In the last decade, there have been heavy research and development activities for stacking a semiconductor chip since the density increase has been getting to face higher wall to future generations. Figure 7.82 describes various issues and concerns

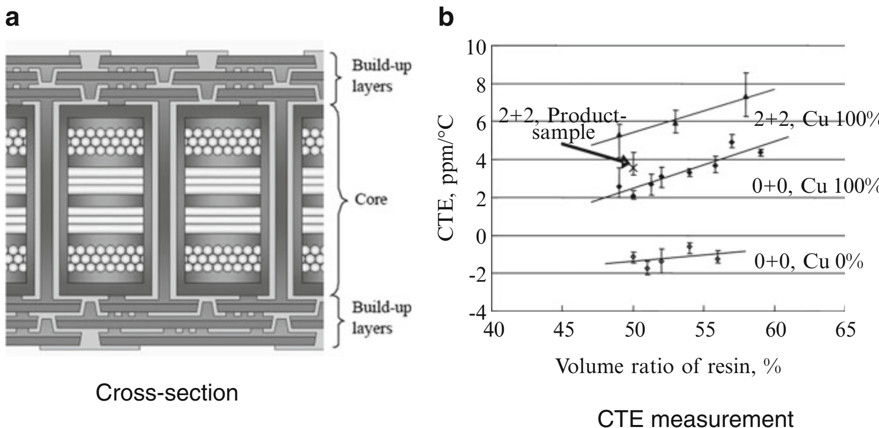


Fig. 7.80 Ultralow CTE substrate. (**a**) Cross section, (**b**) CTE measurement. Courtesy of KYOCERA SLC Technology corp. [24] 2009

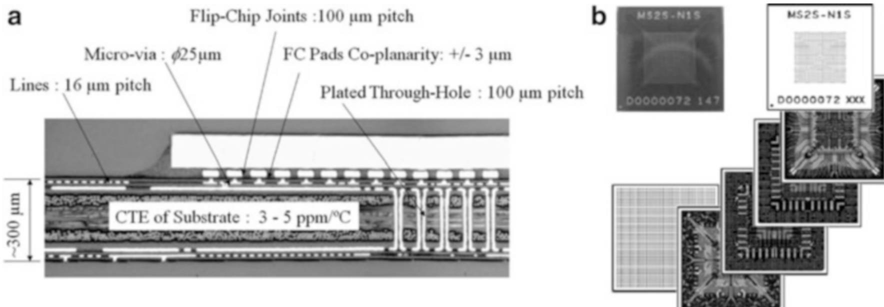


Fig. 7.81 Prototype of next level package. (**a**) Cross section, (**b**) layer patterns. Courtesy of KYOCERA SLC Technology corp. [24] 2009

relating to chip 3D stacking. Most of such efforts have been focused TSV (through silicon via) and the technology seems to become viable in these years. The work has been moving to other issues. Within others, issues connected by an curved arrow in the figure are to be grouped in terms of thermal stress management, i.e., “Chip Package Interaction.” The cause is the same as described in the earlier section of this chapter for 2D packaging but more severe condition by a dimension compared with 2D. The key is that a reliability of package today is not just a reliability of packaged part but must be confirmed by loading the package on an application condition. If the package is used on an application PCB board, the reliability must be assessed with loading to the board. Within CPI related items, silicon substrate has been emphasized and significant numbers of papers and reports were disclosed,

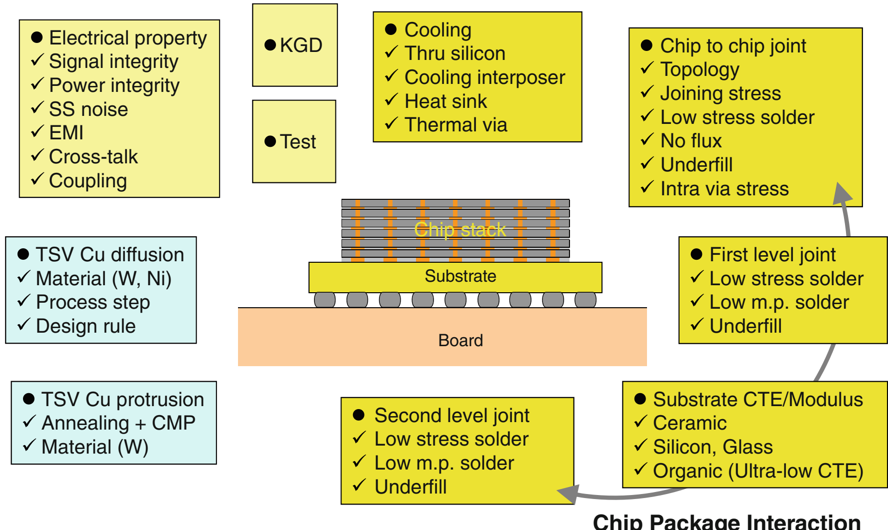


Fig. 7.82 Issues in 3D chip stacking

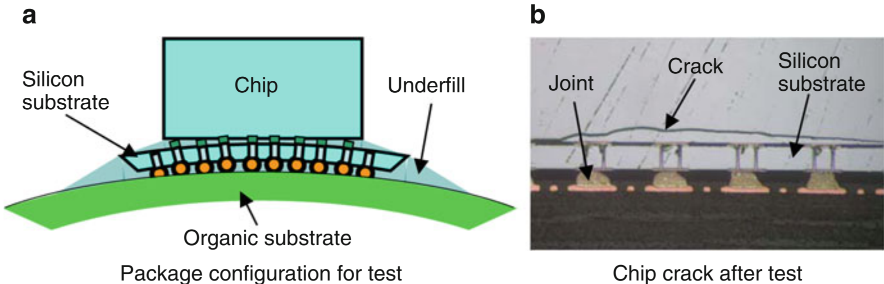


Fig. 7.83 Silicon substrate with IMC flip chip joint. (**a**) Package configuration for test, (**b**) chip crack after test. Courtesy of Orii, IBM Tokyo Research [25], 2010

but there found least reports regarding comprehensive reliability assessment including in an application condition by the time of this manuscript written.

Figure 7.83 is one of least disclosure of silicon substrate assembled on an organic substrate. In the disclosure, a chip was assembled on 150 μm thick silicon substrate. Gaps between chip to silicon substrate and silicon substrate to organic substrate were filled with underfill resin. An organic substrate can be regarded as an application board when a silicon substrate is directly used in an application. If there is no underfill between silicon substrate and organic substrate, the condition is almost the same with a case that bare chip is attached to organic substrate with no underfill and a joint does not withstand to a stress by thermal cycle. (b) Shows a result that the package was stressed under −55 to 125 °C ($\Delta T = 180$ °C) thermal cycle test. (b) Shows a resulted chip crack by the stress when a joint material was CuSn intermetallic. It is also

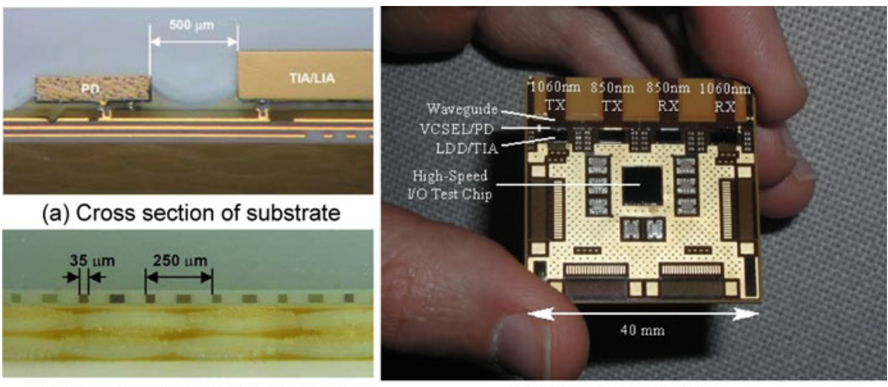


Fig. 7.84 Optical wave guide substrate. (**a**) Cross section of substrate, (**b**) cross section of wave guide, (**c**) recent prototype. After Nakagawa [26] 2008, After Tokunari [27] 2010

reported that there was no crack when a joint material is in that is quite soft compared with the intermetallic. This result indicates that if flip chip joints are harder than certain level, a CTE mismatch between silicon chip and organic substrate is enough to cause a damage to the chip even if there is underfill that tightly adheres the chip with an organic substrate and causes least displacement in *X*–*Y* direction. Hence, an attention to *Z*-direction stress has to be paid. This result clearly indicates that a reliability of 3D chip stack package must be assessed with an application condition, not just a package level reliability. It is the same nature that has been described about reliability in this chapter.

7.5.5 Optical Wave Guide

"Heat" is one of a major issue of today's high performance computer. Suppressing a clock frequency by a parallel processing with multi-core technology is on-going. However, since multi-core requires high bandwidth communication between processor and memory. A large number of electrical connection still require high power and generate heat. It is ideal if such communication is converted to optical technology that is already available to system gate and reaching to board level. Therefore, optical technology on a package level is required in near future. One of such development activity is to provide optical wave guide on substrate and prepare a MCM by flip chip technology. Figure 7.84 shows one of such activity. (a) shows a cross section of a prototype package with wave guide layer put on a build-up layer of a substrate, and a cross section photograph shows a dimension of wave guide tested that contains 35 µm core and 55 µm dummy pattern on 250 µm pitch [26]. (c) shows a latest prototype package that has demonstrated 20 Gb/s data rate per transmitter channel and the 40 mm square module can support 2 Tb/s bandwidth [27].

7.6 Ceramic Substrate

Figure 7.85 shows a typical ceramic substrate of flip chip bonding. Photo(a) shows the outside appearance of a substrate with front side (left) and back side (right). Photo(b) shows a cross section of substrate at a chip side. Vertical via connections with high number of layers are clearly seen.

Table 7.6 with Fig. 7.86 describes design dimensions for a flip chip substrate. Dimensions are in a set with via hole size as a primary parameter. Stacking layer count can be provided up to 35 layers as maximum. Majority is in 18–26 layers for a flip chip package.

Table 7.7 describes material properties available. There are a variety of materials for a substrate. Alumina has several types with detail arrangement. AlN (aluminum nitride) provides lowest CTE. LTCC (low temperature co-fired ceramic) provides low resistivity conductor such as Ag and Cu with sintering low temperature compared with other ceramics. There are unique items in LTCC. Than is a high CTE ceramic substrate material. An underfill effect is available on ceramic substrate. It is not the same magnitude as an organic PCB but enough to extend a flip chip joint life than no underfill flip chip bonding. Since a flip chip joint is protected by underfill, a high CTE substrate material is designed to share a protection more on BGA joint side.

7.7 Roadmap

7.7.1 JEITA

JEITA (Japan Electronics and Information Technology Industries Association) is an industry association for electronics and information industry and issues Japan Jisso Technology Roadmap in every other year that is subjected to fees. The latest was issued on May 2011. Within six working group in total, WG5 with 17 member firms is in charge for printed wiring board including substrate technology. Roadmap consists of ten sections partitioned with four groups and is made with referencing a questionnaire to Japanese material and PWB makers. In rigid PWB section includes build-up, multilayer, double side, and single side PCBs. FPC section includes multilayer, double side, and single side. TAB/COF is a short shingle section. Substrate section consists of tape, rigid, build-up, and ceramic in categories. There are common and difficult challenge parts in addition. In each section, roadmap parameters are described with 10 years outlook and categorized in three lines of numbers that are “class A: high volume production,” “ class B: advanced with limited makers” and “class C: challenge with no high volume production.” For a semiconductor package substrate, a product range is stated as, class-A for a low end product like Memory, class-B for a mid-range like CPU and class-C for a high-

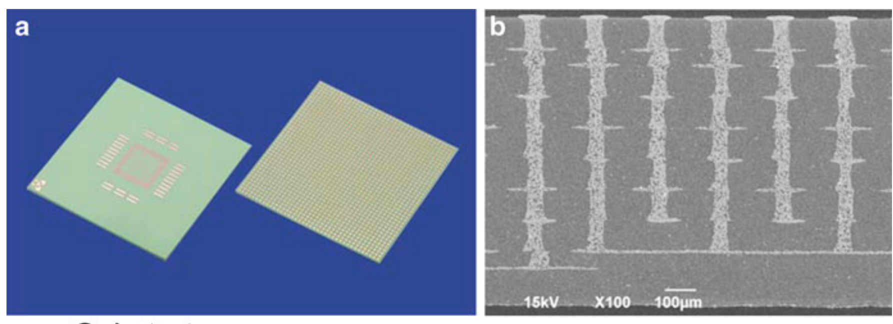


Fig. 7.85 Ceramic substrate. (**a**) Substrate appearance, (**b**) cross section of chip site. Courtesy of KYOCERA Corporation

Table 7.6 Design dimensions, courtesy of KYOCERA Corporation

Item	CC200 (mm)	CC100 (mm)	CC075 (mm)	CC050 (mm)
(a) Via hole dia	0.200	0.100	0.075	0.050
(b) Via cover dot dia	0.381	0.150	0.127	0.100
(c) Via center spacing	0.635	0.254	0.200	0.150
(d) Via cover dot to line clearance	0.254	0.127	0.075	0.050
(e) Line width	0.127	0.100	0.075	0.050
(f) Line to line clearance	0.127	0.100	0.075	0.050
(g) Via cover dot dia (on plane)	0.381	0.150	0.127	0.100
(h) Line width (on plane)	0.200	0.100	0.075	0.050
(i) Clearance dia	0.990	0.455	0.280	0.200
(j) Via pitch (on plane)	1.120	0.560	0.356	0.254
(k) Layer thickness	0.05–0.55	0.05–0.20	0.05–0.15	0.10

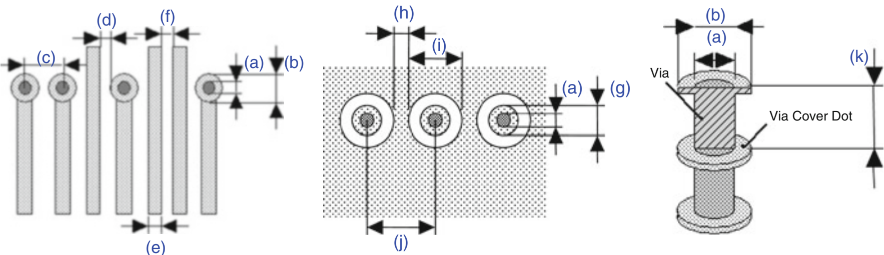


Fig. 7.86 Design element. Courtesy of KYOCERA Corporation

Table 7.7 Material properties, courtesy of KYOCERA Corporation

Ceramic material options		Electrical				Thermal		Mechanical		Conductor material
		Dielectric constant		Dissipation factor (×10E−4)						
		1 MHz	2 GHz	1 MHz	2 GHz	CTE (ppm/K) (RT-400 °C)	Thermal conductivity (W/mK)	Flexural strength (MPa)	Young's modulus of elasticity (GPa)	
Alumina	A473	9.1	8.5	5	10	6.9	18	400	270	W, Mo
(Al_2O_3)	A440	9.8	–	24	–	7.1	14	400	310	W, Mo
	A443	9.6	–	5	–	6.9	18	460	310	W, Mo
	AO600	9.0	8.8	10	21	7.2	15	400	260	CuW
	AO700	9.4	9.2	6	6	7.2	21	620	315	Mo
AlN	AN242	8.7	8.6	1	170	4.7	150	400	320	W
LTCC	GL940	–	18.7	–	2.5	10.7	3.5	220	188	Ag
	GL950	–	9.4	–	14	8.5	4.1	400	173	Ag
	GL330	7.8	7.7	4	5	8.2	4.3	400	178	Cu
	GL570	5.6	5.7	3	7	3.4	2.8	200	128	Cu
	GL771	5.3	5.2	8	36	12.3	2.0	170	74	Cu

Table 7.8 JEITA roadmap, substrate line/space (unit: μm)

Item	Class	2010	2012	2014	2016	2018	2020
Minimum width	A	15	12	10	7	7	7
	B	12	7	7	5	5	5
	C	7	5	5	3	3	3
Tolerance for width	A	±5	±4	±3	±3	±2	±2
	B	±3	±2	±2	±2	±2	±2
	C	±2	±1	±1	±1	±1	±1
Minimum space	A	15	13	10	8	8	8
	B	13	8	8	5	5	5
	C	8	5	5	3	3	3

Table 7.9 JEITA, hole and land diameter, pitch (unit: μm)

Item	Class	2010	2012	2014	2016	2018	2020
Mechanical drill min. hole diameter	A	100	100	100	75	75	75
	B	100	100	75	75	75	75
	C	75	75	75	50	50	50
Mechanical drill min. land diameter	A	220	205	205	180	180	180
	B	200	200	180	160	160	160
	C	140	140	120	120	120	120
Mechanical drill min. hole pitch	A	305	285	280	270	270	260
	B	225	215	210	205	200	190
	C	190	175	175	155	150	145
Laser drill min. hole diameter	A	100	75	75	75	50	50
	B	50	50	50	50	30	30
	C	50	50	50	30	30	30
Laser drill min. land diameter	A	200	180	170	160	150	150
	B	120	120	110	110	90	90
	C	100	100	90	80	80	80
Laser drill min. hole pitch	A	305	285	270	240	240	240
	B	185	175	160	155	140	140
	C	150	135	120	115	105	105

end like FPGA. In the substrate section, roadmap items are T_g, ε, tan δ, CTE x–y, CTE z, warpage, minimum line/space, minimum PTH minimum via/land, etc.

Table 7.8 is an example of the roadmap and describes a roadmap for a build-up substrate wiring line width and space. Shaded area is defined as “No solution in a current development scope.” Table 7.9 is an example of the roadmap for hole parameters on a rigid substrate.

Table 7.10 Substrate related parameters in ITRS

Parameter	Year of production								
	2009	2010	2012	2014	2016	2018	2020	2022	2024
Chip to substrate interconnect land pitch (μm)	150	135	110	100	100	95	95	95	95
Min. finished substrate thickness (mm)	1.1	1.1	1.1	1.1	1.1	0.8	0.8	0.6	0.6
Min. line width/space (μm)	18/18	15/15	12/10	10/10	8/8	5/5	3/3	2/2	1/1
Min. conductor thickness (μm)	25	25	20	15	12	10	5	4	3
Min. through via diameter (μm)	100	100	80	80	70	70	70	60	60
Min. through via land diameter (μm)	250	250	200	200	150	150	150	120	120
Min. micro via diameter (μm)	60	60	60	50	50	30	30	20	20
Min. micro via land diameter (μm)	150	130	120	100	100	70	70	70	70
Min. through via pitch (μm)	300	300	275	275	275	250	250	250	250
Min. solder mask opening (μm)	80	80	60	60	50	50	40	30	30
Min. solder mask opening tolerance (μm)	20	20	18	18	15	15	10	8	8

7.7.2 ITRS

In ITRS (International Technology Roadmap for Semiconductor), there is an Assembly & Packaging chapter. Within the chapter, substrate related parts are sectioned application range as Low Cost (PBGAs), Handheld (FBGA), Mobile Products (SiP, PoP), Cost performance (CPU, GPU, Game Processor), High Performance (High End), and High Performance (LTCC). Table 7.10 describes a roadmap of dimensional parameters for Cost performance (CPU, GPU, Game Processor) as an example with relating to one of JEITA. Other than dimensional parameters, T_g, CTE (X–Y), CTE (Z), Dk@1 GHz, Df@1 GHz, Young's modulus, and water absorption for core material and build-up material are listed. ITRS roadmap is open in WEB.

In parameters, there is a minimum micro-via diameter that is 60 μm in these years. In fact, a micro-via hole diameter for processors of PC and game machines that are primary drivers of application of organic substrate has not been changed actually before year 2000 due to a technical fear in the reliability as previously described in this chapter. It means that a design ground rule of organic substrate has not progressed more than a decade because a true wiring capability of substrate is dominated by via hole density in a unit area. Fabricating a finer pitch line is rather easier compared with

reducing a via hole diameter. Technical establishment for a smaller via diameter has to be achieved for advancement of an organic substrate technology.

7.8 Summary

Flip chip is a superior technology in terms of performance, manufacturability, size and cost with scaling. Substrate material has been innovated from ceramic to organic. By the reduction of cost, applications are widely spreading in any part of electronic products today and will be growing in future. Yet, because of the high performance and density, one has to follow a sound engineering work to achieve its original advantages. In this chapter, it is tried to emphasize its growing element of technologies and applications, and, at the same time, basic technologies and backgrounds as well.

Acknowledgment The author thanks all firms and individuals who offered technical information and acceptance for its usage in this chapter.

References

1. Tummala R, Rymaszewski E (eds) (1989) Microelectronics packaging handbook. Van Nostrand Rheinhold, New York
2. Tsukada Y (1991) Low cost multi-layer thin film substrate, filling a gap to semiconductor. Nikkei Micro Device 73:61–67
3. Tsukada Y, Mashimoto Y, Nishio T, Mii N (1992) Reliability and stress analysis of encapsulated flip chip joint on epoxy base printed circuit board. In: Proceedings of ASEM/JSME joint conference for advanced in electronics packaging-milpitas CA, vol 2, pp 827–835
4. Tsukada Y, Tsuchida S, Mashimoto Y (1992) Surface laminar circuit packaging. In: Proceeding of IEEE 42nd electronics components & technology conference, San Diego, CA, pp 22–27
5. Tsukada Y, Yamanaka K, Kodama Y, Kobayashi K (2002) Features of new laser micro-via organic substrate for semiconductor package. In: Proceeding of ISE 27th international electronics manufacturing technology conference, Dusseldorf
6. Tsukada Y (1998) Introduction of build-up PCB technology. Nikkan kogyo newspaper publication
7. Watanabe K, Fujimura T, Nishiwaki T, Tashiro K, Honma H (2004) Surface modification of insulation resin for build-up process using TiO_2 as a photo catalist and its application to the metallization. J JIEP 7(2):136–140
8. Tsukada Y, Kido Y (2011) Bonding of heterogeneous material using molecular interface technology, application to organic substrate. Electronics Packaging Technol, 27(1)
9. Toda K (2011) Anisotropic etching for ultra fine pitch pattern formation using a subtractive method. In: 12th PWB EXPO technical conference, Tokyo
10. Mago G (2011) Text of technical seminar. 40th Iinter NEPCON Japan, Tokyo
11. Nishiki S (2009) Advanced copper plating wiring technology. CMC Publication, pp 198–207
12. Tsukada Y (2008) Issues on flip chip bonding technology and future direction. In: Presentation at 4th JIEP-west technical lecture, Osaka

13. Tsukada Y (2005) The packaging, 10 years from now. Presentation at JIEP International Conference of Electronics Packaging, Tokyo
14. Tsukada Y (2000) High density, high performance and low cost flip chip technology. Nikkan kogyo newspaper publication
15. Orii Y (2011) C4NP solder bump build technology. In: 481st Technical seminar, electric Journal, Tokyo
16. Nishio T (2008) In: Text of technical seminar, 37th Iinter NEPCON Japan, Tokyo
17. Goldman L (1969) Geometric optimization of controlled collapse interconnections. IBM J Res Dev 13:251–265
18. Tsukada Y (2004) A consideration for total mechanical stress in flip chip packaging utilizing build up substrate technology. In: Presentation at TC6, IEEE 54th electronics component and technology conference, Las Vegas
19. Sasaoka K, Yoshimura H, Takeuchi K, Terauchi T, Tsuchiko M (2010) Development of new buildup PWB with embedded both active devices and chip passive components at the same time. In: Proceedings of 16th symposium on microjoining and assembly technology in electronics, pp 369–374
20. Sasaoka K, Motomura T, Morioka N, Fukuoka Y (2007) Development of substrate with embedded bare chip and passives. In: Proceedings of 17th micro electronics symposium, pp 159–162
21. MaCbride R, Rosser S, Nowak R (2003) Modeling and simulation of 12.5Gb/s on a HyperBGA package. In: Proceedings of 28th IEMT symposium, IEEE
22. (2011) Nikkei Electronics, 5–2(1055):61–68 and pp 87–95
23. Baron D (2011) A Comprehensive packaging solution for advanced IC substrates using novel composite materials. In: Text of technical seminar, 40th Iinter NEPCON Japan, Tokyo
24. Yamanaka K, Kobayashi K, Hayashi K, and Fului M (2009) Advanced surface laminar circuit packaging with low coefficient of thermal expansion and high wiring density. In: Proceedings of 59th ECTC, IEEE, pp 325–332
25. Sakuma K, Sueoka K, Kohara S, Matsumoto K, Noma H, Aoki T, Oyama Y, Nishiwaki H, Andry P, Tsang C, Knickerbocker J, and Orii Y (2010) IMC bonding for 3D interconnection. In: Proceedings of 59th ECTC, IEEE, pp 864–871
26. Nakagawa S, Taira Y, Numata H, Kobayashi K, Terada K, Tsukada Y (2008) High-density optical interconnect exploiting build-up waveguide-on-SLC board. In: Proceedings of 58th ECTC, IEEE, pp 256–260
27. Tokunari M, Tsukada Y, Toriyama K, Noma H, Nakagawa S (2011) High-bandwidth density optical I/O for high-speed logic chip on waveguide-integrated organic carrier. Proceedings of 60th ECTC, IEEE, pp 819–822
28. Japan Packaging Roadmap, JEITA (2011)
29. ITRS roadmap, WEB

Chapter 8
IC-Package-System Integration Design

Yiyu Shi, Yang Shang, Hao Yu, and Shauki Elassaad

Abstract Miniature is massive when it comes to electronics. While there exists a continuous effort in industry to integrate more functionalities into the same area, the prohibitive scaling cost at 45 nm and beyond makes it difficult to continue the trend. Towards this, More-than-Moore techniques have been proposed, which explore new dimensionality of integration by creating and integrating non-digital functionality to semiconductor products (Zhang and Roosmalen. More than Moore—creating hnanoelectronics systems/nanoelectronics systems. Springer, New York, 2009). They motivate new technological possibilities and unlimited application potential.

In order to achieve integration of complete electrical functional blocks within a package, integrated chip-package-system (ICPS) solution has emerged as a cost-effective and flexible solution for More-than-Moore. High-volume commercial and consumer applications are benefiting from the integration of multiple semiconductor devices with other passive and active components within a conventional IC package. The near future of IC packaging includes the integration of MEMS, optical, and photonic devices into subsystems including semiconductors. Package proliferation is the direct result of the demands being placed on the traditional IC package by evolution of device and system technology. As applications become more demanding, packaging technology expands to deliver the optimal solution at the lowest cost.

For integrated device manufacturer (IDM), while system on a chip (SOC) will continue to be a focus area, more applications are taking advantage of package-level integration to deliver complete subsystems to their customers. For example, wafer-fabrication mask costs, short product lives, and relatively low-yielding,

Y. Shi (✉)
Missouri University of Science and Technology (formerly University of Missouri, Rolla), Rolla, USA
e-mail: yshi@mst.edu

Y. Shang • H. Yu
Nanyang Technological University, Singapore, Singapore

S. Elassaad
Stanford University, Stanford, CA 94305, USA

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_8, © Springer Science+Business Media New York 2013

mixed-technology wafer-fabrication processes make SOC undesirable for some applications. The emerging ICPS applications solve these problems by assembling multiple devices into a single IC package. Logic and memory combinations are prevalent. The logic device can be combined with various memory capacities to customize the packaged product for different applications.

In radio frequency (RF) applications, where passive network design is critical to complete the subsystem design, the ICPS solution can move this complexity off the system board and into the package. This approach is becoming common for wireless applications, where standard radio components can be used to deliver wireless connectivity to a variety of end-products without the need to have extensive RF design capability.

Three-dimensional packaging is an approach that is gaining wide acceptance for space-constrained ICPS applications. Stacking silicon dies inside a package allows multiple device types to be integrated into the same space as a single die. The vast majority of mobile phones produced today employ this technology. FLASH and SRAM memories are commonly stacked inside a single package. Going forward, the move to 2.5 and 3 G cell-phone functionality will requires a higher level of integration. Many companies will integrate the digital base band processor device and potentially other ASICs for functionality, such as MP3 decoding and GPS processing, into stacked configurations with increased capacity memory device.

As for semiconductor assembly and test services (SATS), they are following common solutions: Chip Scale Packaging (CSP) provides chip manufacturers with design options to address the various physical form factor and performance needs of their products. Enabling a variety of package types including lead bonded, wire bonded, and various multi-chip package solutions, the current CSP technology satisfies performance, size, and reliability requirements across a wide range of applications. Multi-Chip Packaging includes technologies for die stacking, mixed die System-in-Package (SiP), and Package-on-Package (PoP) stacking through substrate folding and by ball stacking. Flip-Chip Packaging provides a solution for high yield and reliability for high-performance high I/O devices. The flip-chip interconnect offers inherent advantages for high performance from increased interconnect density in a smaller area. Flip-Chip Interconnect is a scalable, copper interconnect technology addressing the industries need for ever-finer pitch flip-chip solutions with larger die, ultra low-*K* dielectrics, and lead-free materials.

Increased package-level integration creates a number of new challenges since various silicon devices in a single package may not come from the same semiconductor company. New levels of cooperation are required between semiconductor companies, packaging companies, and end users to develop an efficient design flow that allows efficient implementation of integrated designs and maintains a fair yield. Design automation and verification of various components within a packaged system needs to be considered and the appropriate level of capability designed into the system.

Current gaps in the ICPS tool flow include multilevel constraint management, multi-technology tool support, multi-domain simulation support, multilevel models and verifications. In addition, a standard data format between IC package and

printed circuit board (PCB) design tools is also required. For example, there is no single design flow with a tool to handle electrical, mechanical, and thermal concerns across the chip, package, and board. Moreover, the flow requires the consideration of electrostatic discharge (ESD), wire-loop optimization for multiple die stacks, as well as new tools to support emerging through-silicon-via (TSV) technologies in three-dimensional packaging.

The chapter is organized as follows. Section 8.1 offers a general overview of the design explorations and considerations to tackle the above problems. Section 8.2 zooms into the topic of decoupling capacitor insertion for noise suppression, and Section 8.3 discusses the three-dimensional integrated systems, an emerging technique for More-than-Moore.

8.1 Integrated Chip-Package-System: What, How, and Why?

8.1.1 Introduction

8.1.1.1 Overview

Miniature is massive when it comes to electronics. While there exists a continuous effort in industry to integrate more functionalities into the same area, the prohibitive scaling cost at 45 nm and beyond makes it difficult to continue the trend. Towards this, More-than-Moore techniques have been proposed, which explore new dimensionality of integration by creating and integrating non-digital functionality to semiconductor products [1]. They motivate new technological possibilities and unlimited application potential.

In order to achieve integration of complete electrical functional blocks within a package, integrated chip-package-system (ICPS) solution has emerged as a cost-effective and flexible solution for More-than-Moore. High-volume commercial and consumer applications are benefiting from the integration of multiple semiconductor devices with other passive and active components within a conventional IC package. The near future of IC packaging includes the integration of MEMS, optical, and photonic devices into subsystems including semiconductors. Package proliferation is the direct result of the demands being placed on the traditional IC package by evolution of device and system technology. As applications become more demanding, packaging technology expands to deliver the optimal solution at the lowest cost.

For integrated device manufacturer (IDM), while system on a chip (SOC) will continue to be a focus area, more applications are taking advantage of package-level integration to deliver complete subsystems to their customers. For example, wafer-fabrication mask costs, short product lives, and relatively low-yielding, mixed-technology wafer-fabrication processes make SOC undesirable for some applications. The emerging ICPS applications solve these problems by assembling multiple devices into a single IC package. Logic and memory combinations are prevalent. The logic device can be combined with various memory capacities to customize the packaged product for different applications.

In radio frequency (RF) applications, where passive network design is critical to complete the subsystem design, the ICPS solution can move this complexity off the system board and into the package. This approach is becoming common for wireless applications, where standard radio components can be used to deliver wireless connectivity to a variety of end-products without the need to have extensive RF design capability.

Three-dimensional packaging is an approach that is gaining wide acceptance for space-constrained ICPS applications. Stacking silicon dies inside a package allows multiple device types to be integrated into the same space as a single die. The vast majority of mobile phones produced today employ this technology. FLASH and SRAM memories are commonly stacked inside a single package. Going forward, the move to 2.5 and 3 G cell-phone functionality will requires a higher level of integration. Many companies will integrate the digital base band processor device and potentially other ASICs for functionality, such as MP3 decoding and GPS processing, into stacked configurations with increased capacity memory device.

As for semiconductor assembly and test services (SATS), they are following common solutions: Chip Scale Packaging (CSP) provides chip manufacturers with design options to address the various physical form factor and performance needs of their products. Enabling a variety of package types including lead bonded, wire bonded, and various multi-chip package solutions, the current CSP technology satisfies performance, size, and reliability requirements across a wide range of applications. Multi-Chip Packaging includes technologies for die stacking, mixed die System-in-Package (SiP), and Package-on-Package (PoP) stacking through substrate folding and by ball stacking. Flip-Chip Packaging provides a solution for high yield and reliability for high-performance high I/O devices. The flip-chip interconnect offers inherent advantages for high performance from increased interconnect density in a smaller area. Flip-Chip Interconnect is a scalable, copper interconnect technology addressing the industries need for ever-finer pitch flip-chip solutions with larger die, ultra low-K dielectrics, and lead-free materials.

Increased package-level integration creates a number of new challenges since various silicon devices in a single package may not come from the same semiconductor company. New levels of cooperation are required between semiconductor companies, packaging companies, and end users to develop an efficient design flow that allows efficient implementation of integrated designs and maintains a fair yield. Design automation and verification of various components within a packaged system needs to be considered and the appropriate level of capability designed into the system.

Current gaps in the ICPS tool flow include multilevel constraint management, multi-technology tool support, multi-domain simulation support, multilevel models and verifications. In addition, a standard data format between IC package and printed circuit board (PCB) design tools is also required. For example, there is no single design flow with a tool to handle electrical, mechanical, and thermal concerns across the chip, package, and board. Moreover, the flow requires the consideration of electrostatic discharge (ESD), wire-loop optimization for multiple die stacks, as well as new tools to support emerging through-silicon-via (TSV) technologies in three-dimensional packaging.

The chapter is organized as follows. The remainder of Sect. 8.1 offers a general overview of the design explorations and considerations to tackle the above problems. Section 8.2 zooms into the topic of decoupling capacitor insertion for noise suppression, and Sect. 8.3 discusses the three-dimensional integrated systems, an emerging technique for More-than-Moore.

8.1.2 Design Explorations

Design exploration for ICPS should try to balance the needs and constraints of the various components that make up the ICPS 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 ICPS, it is equivalently important to choose a packaging substrate that satisfies the design constraints while still minimizing the overall cost and complexity of the design cycle.

In this section, we highlight the design process as well as the design factors associated with ICPS. 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 discuss how to tackle the problems associated with the chip-package co-design early in the ICPS design cycle. In particular, we 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 objective is to concurrently design the chip(s) and package to meet the design constraints in a fast time-to-market fashion. The emphasis 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 are necessarily done. Thus, it is not 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.

Also, we address the constraints associated with the physical design and implementation of ICPS and their impact on technology selection and exploration. In particular, we 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 methodology that addresses all these constraints simultaneously in order to successfully implement the design and meet its constraints.

Figure 8.1 shows the system design flow of ICPS. 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 ICPS to communicate with

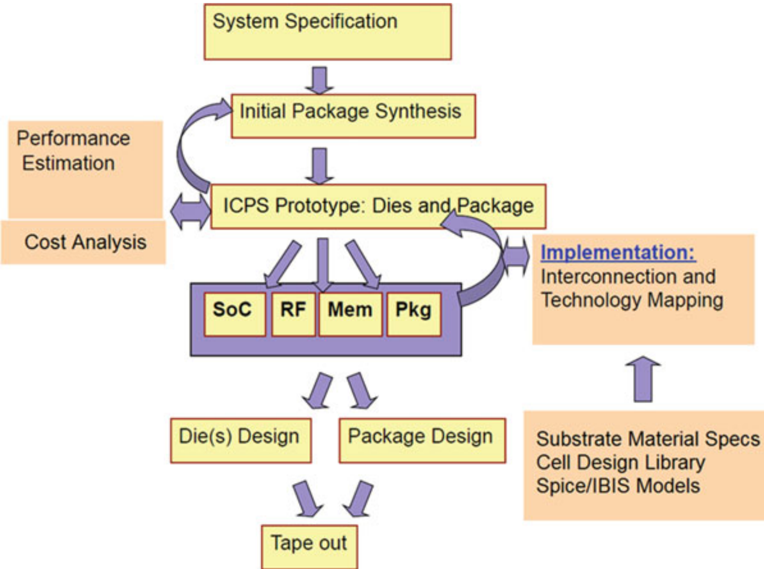


Fig. 8.1 A coherent ICPS design methodology

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 estimates of 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 can 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 on 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 in regards to the number of package layers needed and the design of the power delivery system. 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 result in adequate power sources as needed by the ICPS 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 target 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.

8.1.2.1 On-Chip Design Decisions

During the early stages of ICPS 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, and 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 or 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.

I/O Circuit Design

The proper design of the interface channel (Fig. 8.2) 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 [2–6].

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 like 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 8.3 shows the modeling components needed to estimate the performance and noise metrics of the signal interface in ICPS. 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 loss transmission 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

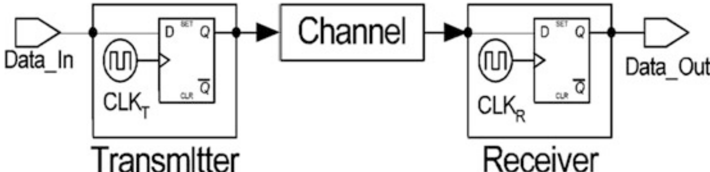


Fig. 8.2 Block diagram of a chip-to-chip communication channel. The main components are: the transmitter, the receiver, and the lossy channel that is composed of the RDL routes, the package traces, and the PCB traces

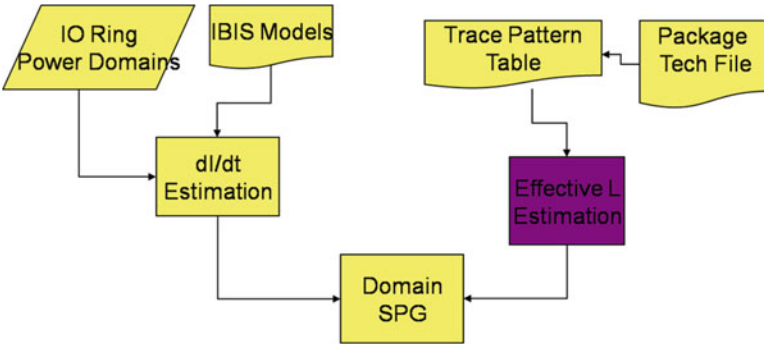


Fig. 8.3 Quick simultaneous switching noise estimation model that uses the IBIS models of the drivers, the package layer parameters, and the trace model of the signal routes from the driver to the package pins

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 [7]. There are 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's 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 which aids in reducing noise and cross talk as well as keeping the clock and data jitter under control [8].

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 [9–13]. 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 [14, 15]. There are power and signal integrity constraints that need to be honored, chief amongst those is the simultaneous switching noise (SSN) constraint [16]. 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 any more to wait until after routing the package to do this analysis and verify that the signal and power integrity constraints are met [17, 18].

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 power delivery system 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.

8.1.2.2 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 over designed 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. Due 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 successful co-design. This is the primary reason making package design difficult.

Package Stack-Up Order

After 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. This means there is a core layer sandwiched between a pair of routing layers. 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 still meeting the design constraints and performance targets [19, 20].

Substrate Layer Assignment

The decision of layer assignment is very crucial 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 [21–25].

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 Z_0 (characteristic impedance), insertion loss, and crosstalk as functions of key substrate design dimensions, DOE (Design of Experiment) methods are employed by high-end design houses. As for Z_0, 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 impedance and a wider trace inordinately constrains wiring density [26]. Most often, the design rules are over designed and are provided by the substrate manufacturers as templates to be followed by the package design teams.

Voltage Domain Planning

The different technologies in ICPS 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

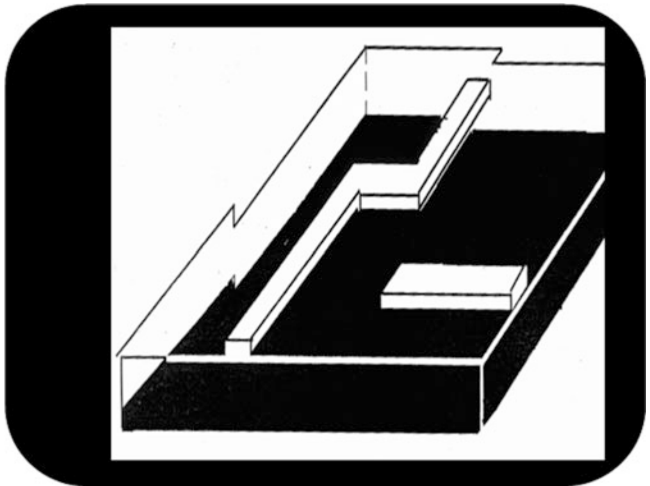


Fig. 8.4 A solid plane that provides a stable reference for the traces above and below it. This design provides a tighter return path loop which renders the effective loop inductance smaller

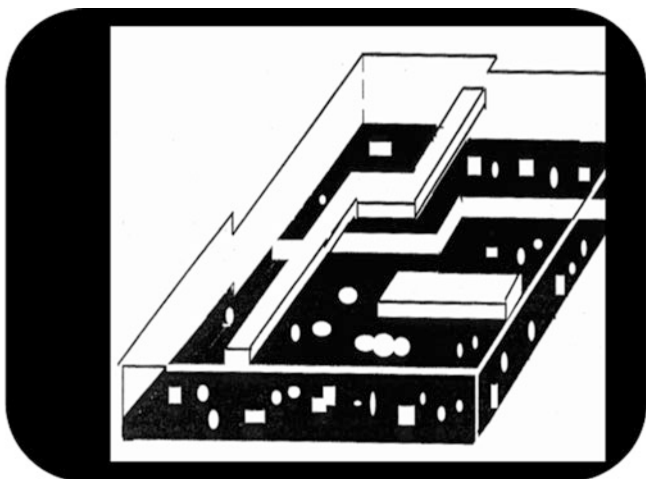


Fig. 8.5 A perforated plane does not provide a stable reference and causes a large inductance loop and noise problems in the package

domain is composed of one or more power domains. These power domains constitute logical partitioning of the different signals and buses in the system. Typically, a solid plane is necessary to provide a stable reference for the traces above and below it, as contrasted by Figs. 8.4 and 8.5.

8.1.3 Modeling and Analysis Decisions

For pre-layout characterization of the package, lumped or transmission line models are utilized. Most 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 entire channel, including PCB and 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 such 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/O. 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.

In regards to the 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 they exist come at a high 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, design of experiment (DOE) technique and statistical analysis are used. Both 2D and 3D electromagnetic analysis and circuit model simulation are carried out in design optimization process. Passive characterizations are also carried out to get the basic parameters used in the package design.

8.1.4 ICPS Design Problems

In this section, we expound on 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.

8.1.4.1 Design and Planning of Power Delivery System

The design of the power delivery system requires time- and frequency-domain modeling and analysis 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 is an integral part of the power design of ICPS. 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 processor voltage rails [28].

8.1.4.2 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 ICPS including the package layers, and onto the receiving end. If the transmitter and the receiver are not in the same ICPS, then the PCB traces should be modeled as well as the receiving ICPS. 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.

Signal Interface Characterization

Signal interface characterization in terms of bandwidth, jitter, and skew has to be carried out before an ICPS is taped out. For design and modeling of the channel, Fig. 8.6 shows a basic model that can be constructed as soon as the chip and package decisions outlined above have been implemented.

Figure 8.6 shows a detailed model for estimating the channel noise as well as the signal timing, skew, and jitter. The accuracy of this model is reasonable in the design planning stage before the chip is fully implemented and the package is

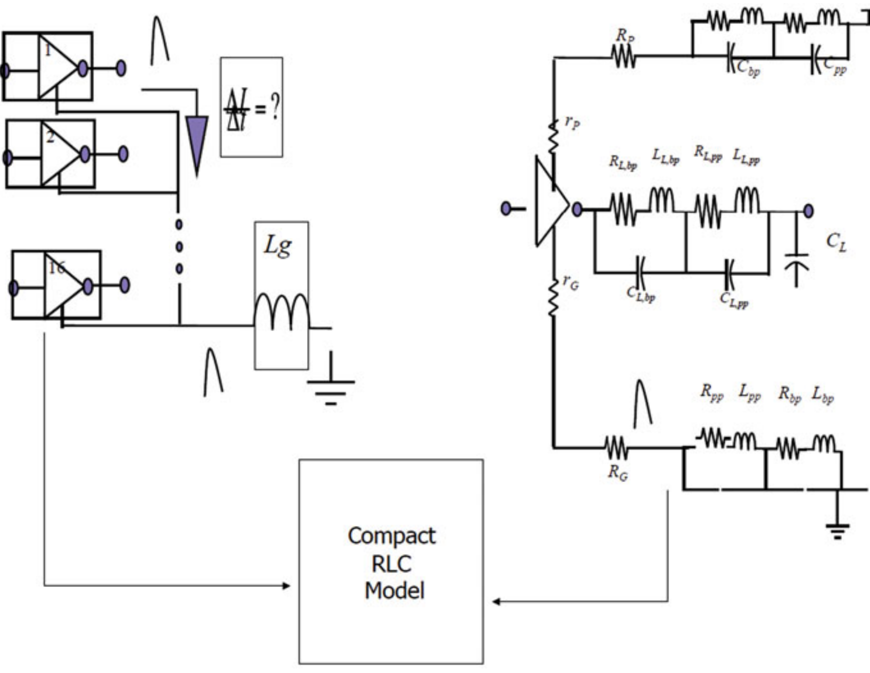


Fig. 8.6 RLC Model for estimating the noise, and timing of the signal interface early in the ICPS design stage

routed. The parasitic 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 [2–6].

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.

In the next section, we discuss some detailed design problems to tackle the power integrity problems in ICPS.

8.2 Decoupling Capacitor Insertion

8.2.1 Introduction

Power integrity is very important for the performance of ICPS. Compromising it may lead to logic errors and slow transition. Nowadays, chips operate at very high frequencies and consume a large amount of power. The number of I/O's is ever increasing. High power leads to large current flows in the power delivery system (PDS), which causes large IR drop and d*i*/d*t* noise. High frequencies cause inductive effects and may trigger resonance, which presents large impedance in PDS. A large number of I/O's lead to serious SSN. All of these may lead to power rail collapse and affect the operation of circuits. Power integrity has to be guaranteed in the entire PDS from voltage regulator module (VRM) to on-chip power grid. We focus on decoupling capacitor (decap) optimization for the power integrity of IC package, especially SSN problem. However, our method can be also used for decoupling capacitor optimization in other part of the power delivery system.

Decaps, which act as temporary current sources and low pass filters for AC signals, are essential to reduce the voltage fluctuation in the PDS. For package decoupling purpose, discrete decaps are used. These decaps are not perfect. Their frequency responses can be modeled with an equivalent serial capacitance (ESC), an equivalent inductance (ESL), and an equivalent resistance (ESR). With different prices, different types of decaps have different ESC, ESL, and ESR, and therefore different effective frequency ranges. As pointed out in [29], the expensive decaps may not be the best choice for electrical performance. Also the effectiveness of the decaps depends on its electrical environment and therefore varies with locations. Therefore, the types and locations of the decaps have to be optimized for most effective design with minimal cost.

The problem of decap optimization has already been presented in the literature. In [30–35], on-chip decap optimization problem has been studied for different objective functions. However, on-chip decaps normally have negligible ESL and ESR and can take continuous values. Unfortunately, these are not true for in-package decaps.

In-package and on-board decap optimization has also been studied, but majority of existing work is trial-and-error methods, such as [36] and [37], both of which are manual processes. Automatic optimization methods also exist. For example, the authors of [38] use the partial-element equivalent circuit (PEEC) model and model order reduction techniques to compute the input impedance and then search for the optimal locations of the decaps to minimize the impedance by gradient-based search. In [39] the authors use FDTD and Fast Fourier Transform (FFT) to obtain frequency dependent Poynting vector 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.

The most comprehensive work on automatic optimization of package decaps is [29]. In this paper, the authors model the inductive effect of packages with

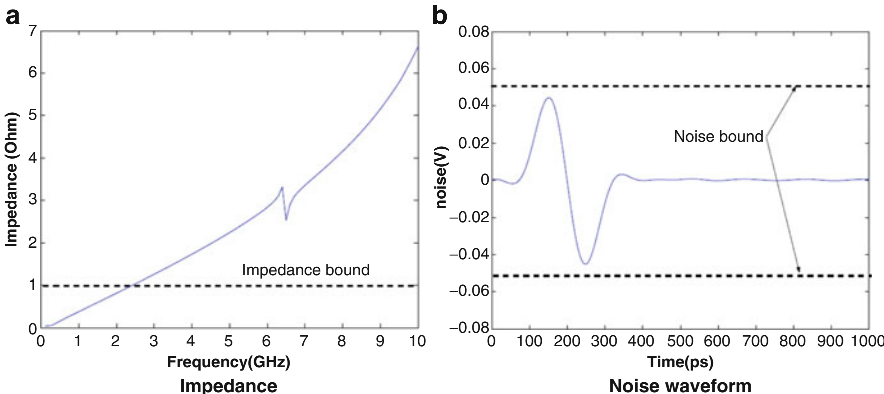


Fig. 8.7 Impedance and noise waveform

susceptance (inverse of inductance) instead of inductance, and extract a resistance–capacitance–susceptance (RCS) 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 algorithm is developed to search for the optimal types of decaps at given 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, the approach is based on impedance metrics, which will lead to significant overdesign. For example, in Fig. 8.7 we show a case where the noise bound is met but impedance bound is not for a signal with effective frequency range up to 10 GHz. Figure 8.7a shows that the target impedance is not met in most part of the frequency band but the noise bound has been met as shown in Fig. 8.7b. It is clear that the target impedance cannot capture the noise accurately and may cause overdesign.

Accordingly, we propose to directly deploy noise as the metric of SSN and develop an efficient noise model to optimize the location and types of decaps. We consider a large number of ports to search for the optimal location for decap insertion. We assume that the impedance matrix is given and develop an efficient model to compute the new impedance matrix with one decap inserted or removed. The time complexity of our algorithm is $O(n_2)$ compared to $O(n_3)$ in the state-of-the-art existing work [40]. With impedance matrix and pre-characterized switching current waveform, we use FFT to compute the noise waveform and obtain the worst-case noise. Based on these models, a simulated annealing algorithm can be employed to minimize the cost subject to the maximum noise constraint. Experimental results demonstrate $3\times$ cost reduction and more than $10\times$ speedup over the previous impedance-based methods. The results have been published at [41]. Although simple, the efficiency of the SA-based approach is still limited in the presence of large designs.

To address the efficiency issue, we have developed a scalable sensitivity-based algorithm using ring-based decap allocation followed by the legalization to complete detailed placement of decaps. The primary contributions of this method are twofolds. First, to generate an effective macromodel considering large numbers of input ports, we propose spectral clustering to find a small amount of principal I/Os based on the I/O correlation. This enables an effective model order reduction. In addition, the information of clustered I/Os can be further used to partition the large RLC-network for power supply. By further incorporating the structure macromodeling [42, 43].we can perform a localized reduction and analysis for each partitioned block. Compared to the macromodel used in [29], our method is 3.04× more accurate and 25× more efficient. Second, given a large number of legal positions, we introduce a ring-based decap allocation to avoid trying every legal position as in SA. To systematically allocate decaps, the map of legal positions is first decomposed into multiple rings. By parametrically describing those rings in the state equations, the nominal responses and the sensitivities of I/Os with respect to the ring can be efficiently generated from a structured and parameterized macromodel for each partitioned block. Then, the decaps can be allocated according to the incrementally calculated sensitivity. Compared to the decap allocation in [29, 41], experiments show that our allocation is 97× faster, and reduces the decap cost by up to 16 %.

8.2.2 Electrical Models

8.2.2.1 Package Model

As shown in Fig. 8.8, packages for semiconductor chips often consist of multiple signal layers, power planes, and ground planes with dielectric in between. Metal signal traces connecting the chip I/O cells to the PCB traces are routed between planes, and package planes are stapled together with vias, and connected to PCB by balls. We assume the locations of chip I/O ports are known and the possible locations for the decoupling capacitors (decaps) are predefined. We can pre-build the macromodel of the package with the specified ports for I/O's and decaps before the optimization process. This macromodel not only includes the power or ground planes, but can also include vias and traces. Other part of PDS such as on-chip power grid, PCB and VRM can also be included. Specifically, for the macromodel we obtain the impedance matrix $Z(f_k)$ for the specified ports at a number of sample frequencies f_k beforehand. The matrix element $Z_{ij}(f_k)$ of $Z(f_k)$ is the transfer impedance from port j to port i at frequency f_k. The frequency dependent impedance Z can be obtained by various methods, such as 3D field solvers, model reduction, or measurement, depending on the time and accuracy requirement and design stages. Our method can be used with any of these methods. With the macromodel, the efficiency of following optimization process no longer depends on the size of the original circuits, but only depends on the number of ports defined. This allows a very complex package to be optimized in a very short time.

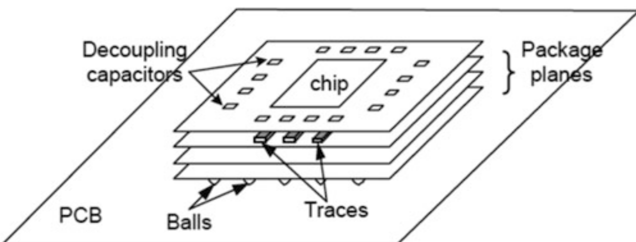


Fig. 8.8 IC package

To start with, we extract a detailed RLCK circuit of the package, and then use a model order reduction technique to obtain the impedance matrix. For the detailed RLCK circuit, the planes are partitioned into grids and the traces are divided into small segments. Then, we extract the resistance, self inductance, and grounding capacitance of each segment, the coupling inductance between each pair of segments and the coupling capacitance between adjacent segments.

8.2.2.2 Decoupling Capacitor Model

As discussed in the introduction, the decaps for the package are discrete elements. Each type of decaps has different frequency-domain response and can be characterized by ESC, ESL, and ESR as shown in Fig. 8.9. We assume there are multiple types of decaps and their ESC, ESL, and ESR are given. For efficient optimization, we pre-compute the frequency dependent impedance of each type at the sample frequencies as

$$Z_{\mathrm{d}}(\omega) = \mathrm{ESR} + \frac{1}{j\omega\mathrm{ESC}} + j\omega\mathrm{ESL} \tag{8.1}$$

8.2.2.3 Model of I/O Cells

Normally each I/O cell drives a transmission line in the package. When an I/O cell switches, it draws a large current from the power delivery system and causes voltage fluctuation (SSN). The electrical behavior of I/O cells can be modeled by various models, for example, a physical model such as the BSIM model [44] or a behavioral model such as the IBIS model [45]. With a given load, we can pre-characterize the I/O cell and obtain the time dependent current waveform similar to IBIS model by simulation. We further transfer the obtained time-domain waveform to frequency-domain and obtain the frequency component of the current to be used in the following optimization process.

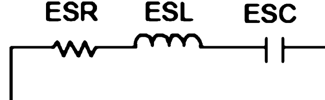


Fig. 8.9 Model of decaps

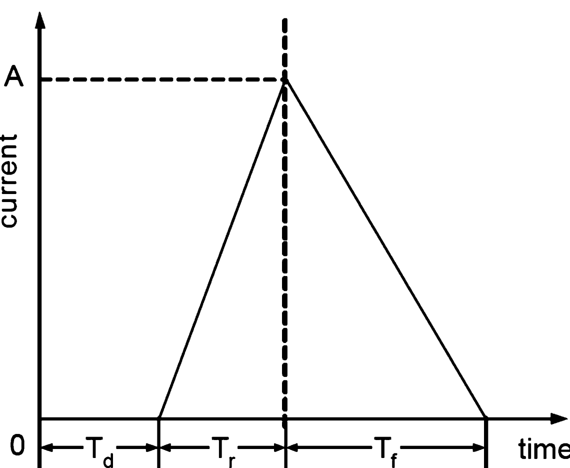


Fig. 8.10 Switching current model

Similarly to [31], for simplicity we model the current waveform as a two-segment piecewise-linear waveform (triangular waveform) as shown in Fig. 8.10. With the parameters T_{d}, T_{r}, and T_{f} defined in the figure, the frequency components are computed as

$$I(\omega) = \frac{a}{s^2}\mathrm{e}^{-sT_{\mathrm{d}}} + \frac{b-a}{s^2}\mathrm{e}^{-s(T_{\mathrm{d}}+T_{\mathrm{r}})} - \frac{b}{s^2}\mathrm{e}^{-s(T_{\mathrm{d}}+T_{\mathrm{r}}+T_{\mathrm{f}})}, \tag{8.2}$$

where

$$\omega = 2\pi f \tag{8.3}$$

$$s = j\omega \tag{8.4}$$

$$a = \frac{A}{T_{\mathrm{r}}} \tag{8.5}$$

$$b = -\frac{A}{T_{\mathrm{f}}} \tag{8.6}$$

In this model, each I/O cells can have different amplitude, rise time, and fall time. Note that our methods discussed in the rest of this chapter are not limited to such waveform but can be applied to any waveform. More accurate and complex current models can be used and the frequency components can be obtained either numerically or analytically beforehand without affecting the optimization process.

8.2.3 Impedance Metric and Its Incremental Computation

With a given current injection, the noise at a port depends on the impedance. With the insertion or removal of decaps, the impedance matrix of the system will change and affect the noise value. Therefore, the impedance matrix has to be updated with changes of decap distribution. In [29], this is done by n_{io} AC sweeps, where n_{io} is the number of I/O ports. Another method is presented in [40], which assumes that the macromodel without decap is given in terms of admittance matrix $Y(\omega)$. And the impedance with decap is computed as

$$Z(\omega) = \left(Y(\omega) + \tilde{Y}(\omega)\right)^{-1}, \tag{8.7}$$

where $\tilde{Y}(\omega)$ is a diagonal matrix with $\tilde{Y}_{ii}$ equal to the admittance of the decap at port i at frequency ω. Both of these methods need at least one matrix inversion, on which the computation time of this operation mainly depends. Because Y is a macromodel, it is usually a dense matrix and the time complexity of the matrix inversion is roughly $O\left(n_{\mathrm{p}}^3\right)$, where n_{p} is the number of ports including the I/O ports and the ports for the decaps.

The approach above is good for computing impedance when simultaneously inserting or removing a large number of decaps. However, in iterative optimization process such as the one to be presented later in this chapter, we normally add or remove one or a small number of decaps each time. In this case, full matrix inversion is not necessary for impedance computation, and an incremental method is needed for efficiency.

Assuming at a certain frequency the impedance matrix before inserting the decap is Z and we insert one decap at port k as shown in Fig. 8.11, we need to solve the new impedance $\widehat{Z}$. $\widehat{Z}_{ij}$, which is the transfer impedance from port j to port i, is equal to the voltage at i when applying an 1 A current source at port j. Because the system is linear, we can replace the rest of the package except the decap with a Thevenin equivalent circuit as shown in Fig. 8.12. The voltage source is equal to Z_{kj} and the source impedance is equal to Z_{kk}. Therefore the current running through the decap is $Z_{kj}/(Z_{kk} + Z_{\mathrm{d}})$, where Z_{d} is the impedance of the decap. Replacing the capacitor with a current source of the same current as shown in Fig. 8.13 will not change the voltage or current in the rest of the circuit. According to the superposition principle, the change of Z_{ij} is equal to $-Z_{ik}Z_{kj}/(Z_{kk} + Z_{\mathrm{d}})$ and

$$\hat{Z}_{ij} = Z_{ij} - \frac{Z_{ik}Z_{kj}}{Z_{kk} + Z_{\mathrm{d}}}, \tag{8.8}$$

where Z_{ij} is the transfer impedance from port j to port i before inserting the decap. We can see that the change of Z_{ij} only depends on Z_{ik}, Z_{jk}, Z_{kk} and Z_{d}. Therefore, the overall impedance matrix with the decap added at port k at a given frequency is

Fig. 8.11 Connection of a decap

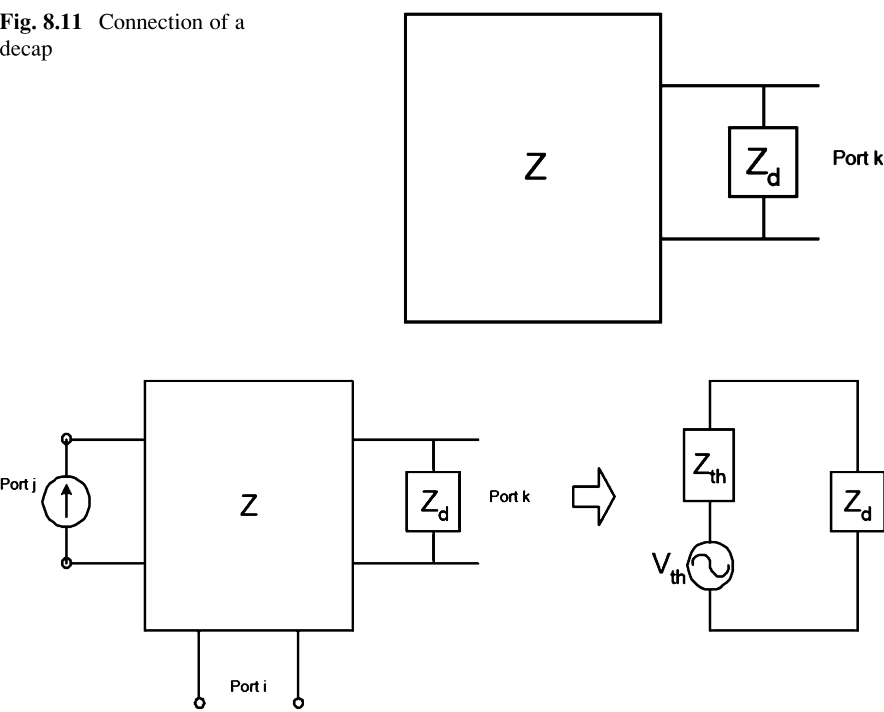


Fig. 8.12 Thevenin equivalent circuit

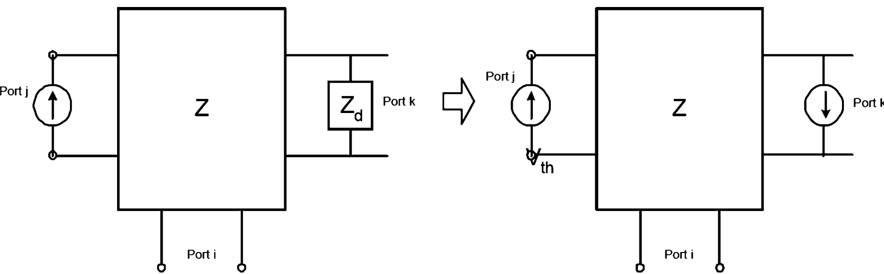


Fig. 8.13 Equivalent current source

$$\hat{Z} = Z - \frac{b_k a_k}{Z_{kk} + Z_\mathrm{d}} \tag{8.9}$$

where a_k is the kth row of Z and b_k is the kth column of Z. The computation time of this process is mainly determined by computing $b_k a_k$ which is an $O(n_\mathrm{p}^2)$ process. Removing a decap from port k is equivalent to adding a negative admittance of the same value at port k. Therefore, the overall impedance matrix with the decap removed from port k at a given frequency is

$$\hat{Z} = Z - \frac{b_k a_k}{Z_{kk} - Z_{\mathrm{d}}} \tag{8.10}$$

Compared to (8.7), this method is obviously more efficient and scalable with the number of ports, when only one decap is added or removed. This is especially suitable for iterative optimization process or trial-and-error process, in which one or a small number of decaps are changed and the impedance matrix is needed to be reevaluated in each iteration. Another advantage of this method is that to obtain certain ports' impedance we only need to selectively compute them with (8.8) without computing the impedance of other ports. This again is good for trial-and-error method. For example, in simulated annealing method, we can first only compute the impedance of I/O ports. If the solution has been accepted, we further compute the impedance of other ports. Otherwise, we can move to the next iteration without any further computation. Since I/O ports are only a fraction of the total ports, we can save significant computation time.

If n decaps are changed, the computation in (8.9) needs to be repeated n times. When $n \ll n_{\mathrm{p}}$, it will still be more efficient than (8.7). The worst case is when $n = n_{\mathrm{p}}$, which means that the distribution of decaps changes at all the ports, and the complexity becomes $O(n_{\mathrm{p}}^3)$, same as [12]. Fortunately, this case will never happen in one iteration.

8.2.4 Noise Metric

One potential issue with the impedance metric is that it is not directly proportional to the noise and may be pessimistic. Physically, it actually 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 needs not to be very small in the entire frequency band. In Fig. 8.14 we show an excitation current waveform and its spectrum up to 10 GHz. It is a triangular waveform with rising and falling time both equal to 100 ps and the amplitude is 50 mA. We can see that the current is mostly distributed from 0 to 10 GHz, but the amplitude of the frequency component gradually decreases with the frequency increasing. The time-domain noise is the convolution of current and impedance in frequency-domain. Therefore, a large impedance at a lower frequency may cause large time-domain noise, but may not cause the problem at a higher frequency. One such case is shown in Fig. 8.7.

Instead of the impedance metric, we can directly consider the noise in the power delivery system at each port of interest. We can easily compute the impedance at different sampling frequencies and also pre-compute the spectrum of the switching

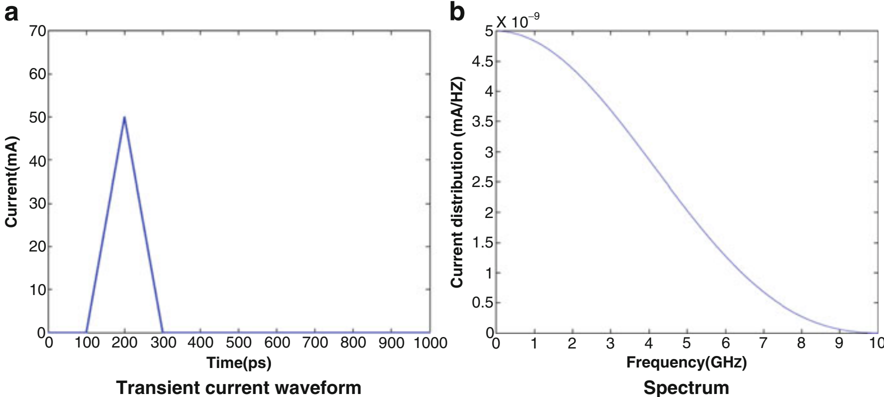


Fig. 8.14 Transient current waveform and its spectrum

current of each port. For the noise at port j induced by the switching activity at port, the noise component at the kth frequency sampling point can be easily computed as,

$$V_{ij}(f_k) = Z_{ij}(f_k)I_j(f_k) \tag{8.11}$$

We then use FFT to compute the time-domain waveform, which is the noise waveform induced by port j at port i. The time complexity of FFT is $O(n\log n)$ where n is the number of the sampling points. For the signal shown in Fig. 8.14, 512 sampling points from 0 to 50 GHz are used. For a signal with shorter rising time or falling time, more sampling points in higher frequencies are needed.

At a given port, we consider both the noise induced by the I/O cells connected to the port and the noise induced by the switching activity of other I/O cells connected at other ports. Because the switching of the I/O cells are random and the system is linear, the worst-case noise at one port is the sum of the maximum noise induced by all the cells. Each maximum noise can be computed with the proposed method.

8.2.5 Simulated-Annealing-Based Decap Insertion

8.2.5.1 Settings

In this section, we use the developed impedance and noise models to minimize the cost of the decoupling capacitors (decaps) in a package under the constraint of noise in the power delivery system. Figure 8.15 shows a sketch of the IC package we consider with I/O cells located on a ring structure along the chip boundary and decaps located around the chip. The package is often cut into different domains for different supply voltages. Each voltage domain can be optimized separately. As an example, in this chapter we only consider one side of the package, which can be considered as one voltage domain of the package.

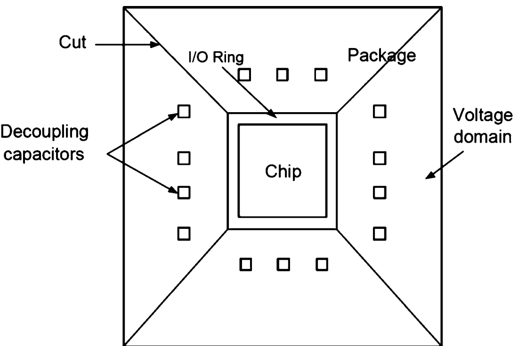


Fig. 8.15 An IC package

Table 8.1 Decoupling capacitors [29]

Type	1	2	3	4
ESC (nF)	50	100	50	100
ESR ( )	0.06	0.06	0.03	0.03
ESL (pH)	100	100	40	40
Price	1	2	2	4

Similar to [29], we also try to minimize the total decap cost. We consider different types of decaps with different prices and assume the same set of decaps as in [29], which are summarized in Table 8.1.

However, different from [29], we do not apply the target impedance constraint. Instead, we directly require the worst-case noise to be less than the given noise bound. We assume that V_{dd} is 2.5 V and require the noise to be less than 15 % of V_{dd}, which is 0.35 V at each port.

8.2.5.2 Algorithm

We use the simulated annealing (SA) algorithm to optimize the types and locations of the decaps so that the total cost is minimized and the noise in the power/ground plane is smaller than a given bound. The objective function is defined as

$$F(p_{\mathrm{i}}, c_j) = \alpha \sum_i p_{\mathrm{i}} + \beta \sum_j c_j \tag{8.12}$$

where α and β are weights for the noise and cost respectively. α is chosen to be much larger than β so that the noise constraint can be achieved. The first summation is over all ports, and the second is over all decaps. p_{i} is the penalty function for violation of the noise constraint and is defined as

$$p_{\mathrm{i}} = \begin{cases} 0 & (V_i < \bar{V}) \\ V_i - \bar{V} & (V_i > \bar{V}) \end{cases} \tag{8.13}$$

where $\bar{V}$ is the noise upper bound and V_i is the worst-case noise at port i, which is computed by the method proposed in Sect. 8.2.4. c_j are the respective cost for each decap as shown in Table 8.1.

There are two types of moves in our simulated annealing (SA) scheme: (1) adding a decap of a random type at a randomly picked port. (2) removing a decap capacitor. At most one decap is allowed at one port. After each move, we compute the new impedance according to (8.9) and the noise according to (8.11). We start the SA with initial temperature of 20 and terminate it at 0.001. The temperature is decreased by a factor of 0.95 and the number of moves at a particular temperature is 100.

8.2.5.3 Results

Case 1

Our model and algorithm can be applied to any package configurations with any number of layers. In this case, we assume 1 cm $\times$ 2 cm rectangular planes with a power plane and a ground plane. I/O cells are located at one edge of the structure. We assume that there are 30 I/O cells. Each of them will draw the current shown in Fig. 8.14. Since cells close to each other have similar impedance and strongly couple to each other, we partition the 30 I/O cells into 3 groups and define 3 I/O ports. Each cell is connected to the closest I/O port and each of the ports is connected with ten I/O cells. Note for higher accuracy, more ports can be defined if necessary. We allow the decaps to be distributed across the plane, and therefore define 90 uniformly distributed ports on the package. Totally, there are 93 ports in our macromodel.

Our noise-based algorithm finds a valid solution where all the ports meet the noise constraint. The worst-case noise of each port is listed in Table 8.2. The total cost of the decaps is 20 based on the price listed in Table 8.1. In Fig. 8.16, we show the distribution of the decaps in a uniform grid. In this figure, the numbers stand for the type of decap as defined in Table 8.1, and “0” means no decap.

We further compare our results with an impedance-based approach. In this approach, for the objective function we substitute the noise with the maximum impedance and replace the noise bound with the target impedance. Because we require the noise to be less than 0.35 V and the total peak current of 10 I/O’s connected to one port is 500 mA, the target impedance for each port is calculated to be 0.7. The distribution of the decaps in the best solution given by the impedance approach is shown in Fig. 8.17. We can see that though the decaps still concentrate around the chip but spread more across the planes than noise driven approach. The total cost is 72, which is more than 3$\times$ larger than the results of noise driven approach. Table 8.3 shows the maximum impedance and the noise at each port. We can see that the target impedance cannot be reached but the noise is already well below the noise bound. Accordingly, if we continue to use target impedance to guide the design, we need to insert additional decaps which will lead to large overdesign.

Table 8.2 Worst-case noise at ports

Port	1	2	3
Before optimization (V)	2.52	2.49	2.48
After optimization (V)	0.344	0.343	0.344

0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0
0 0 0 **1** 0 0 0 0 0 0 0
0 ` 0 0 0 **3** 0 0 0 0 **3**
1 0 0 **1** 0 **4** 0 **2** **3** 0 **1**
0 0 0 0 0 0 0 0 0 0 0

Chip

Fig. 8.16 Optimal distribution of decaps from noise driven approach

0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 4 0 0 0 0 0 0
0 0 0 0 0 0 0 **4** 0 0 0
0 0 0 0 0 0 0 **4** 0 0 0
1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 0 **4** **1**
0 0 0 0 0 2 0 0 **2** **3** **4**
0 0 **2** **4** **1** **2** 0 **4** **2** **2** **1**
2 **4** **3** **3** **1** **1** **0** **1** **4** **1** **4**
0 0 0 0 0 0 0 0 0 0 0

Chip

Fig. 8.17 Optimal distribution of decaps from impedance driven approach

Table 8.3 Impedance and noise at ports

Port	1	2	3	Bound
Maximum impedance	5.31	5.59	7.12	0.7
Worst-case noise (V)	0.256	0.302	0.284	0.35

Case 2

In case 2, we assume a domain on one side of the chip as shown in Fig. 8.18. The package is assumed to have four layers of power or ground planes. All the planes are stapled together with uniformly distributed vias and the bottom power/ground plane is grounded at several locations in the plane. We define 70 ports for the decaps and 3 ports for noise optimization. The capacitor distribution of the best solution is shown in Fig. 8.19. We can see that the capacitor is distributed around the chip and also across the planes. This is because the vias and grounding connections change the electrical environment at different locations. The best location for decaps may not be just closest to the chip.

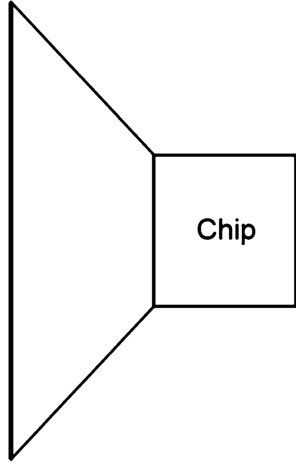


Fig. 8.18 One package domain

Fig. 8.19 Optimal distribution of decaps from noise driven approach

0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	**2**	**3**	0
0	0	0	0	0	0	**1**	0	0	0	**1**	0	0
0	0	**3**	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	**1**	0	0	0	0	**2**	0
0	**2**	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	**4**	0	**3**	0	0	**1**	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	**2**	**2**	0	**1**	**2**	**1**	0	**2**	**2**	**1**
0	0	0	0	0	**4**	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

Chip

8.2.5.4 Runtime

We implement the algorithms in Matlab and conduct experiments on a 2.8 GHz Xeon system. For comparison, we also implement the method of [40]. The runtime of case 1 for different methods listed in Table 8.4 is shown in Table 8.5.[1] In the table, method 1 is the proposed method using the proposed incremental impedance computation and FFT for noise computation. Method 2 uses the impedance computation method from [40] and FFT for noise computation. Method 3 is from [29]. Comparing methods 1 and 2, we can see that the incremental computation of impedance is 11× faster than the matrix inversion-based approach. Comparing with method 3, our method is significantly faster than even considering the speed difference of the computing platforms and with more ports.

[1] The runtime of method 3 in Table 8.5 is taken from [1]. The computation platform is 1 GHz Pentium 3, and the computing language is unknown.

Table 8.4 Approaches

1	Incremental impedance + noise objective
2	Matrix inversion [40] + noise objective
3	Ref. [29]

Table 8.5 Runtime

Approach	1	2	3
Ports	93	93	20
Iterations	5,881	5,403	920
Run time(s)	389.5	4,156.1	2,916.0
Average run time(s)	0.0662	0.7692	0.519

8.2.6 Sensitivity-Based Decap Insertion

The biggest disadvantage of SA-based method discuss above is that it virtually tries every possible location, and accordingly, it can only handle small designs. In this section, we discuss an alternative approach through localized model reduction, parameterized problem formulation, and sensitivity-based optimization.

8.2.6.1 Improved Model and Problem Formulation

Recall that packages often consist of multiple signal planes, power planes, and ground planes with dielectric in between. Metal signal traces connecting the chip I/O cells to the PCB traces are routed between planes, and package planes are stapled together with vias, and connected to PCB by balls. Similar to the SA-based method, we assume that the locations of chip I/O ports are known, and the allowed number of possible locations called legal positions for decaps are predefined for each region in a multi-layer package with consideration of congestion due to package routing and ball assignment. The legal positions are slots to connect the terminals of decaps, but are not necessarily where decaps are located. As shown in Fig. 8.20, the I/Os are located in the center of the package, and the legal positions to allocate decaps surround the chip with consideration of reserved routing area. After one decap is assigned to one legal location the decap is then called legally placed.

Note that a complete RLC model is required for accurate representations of interactions among package layers, C4 bumps, vias, on-chip power grids, and all other signal traces. The power/ground plane can be uniformly discretized into N_{v} tiles, and each tile is modeled by RLC element under the PEEC model [46]. However, a detailed 3D extraction using PEEC [46] introduces densely coupled inductances (L) that increase the model complexity. This problem can be solved by stamping a sparsified L^{-1} element as discussed in Sect. 8.2.6.2.

Same as the SA-based method, the design freedoms are still the legal locations and decap types. Brute-force examination of every possible combination is computationally expensive if not impossible. To allocate decaps in a manageable way, we propose a ring-based decomposition of all legal positions. This is based on the

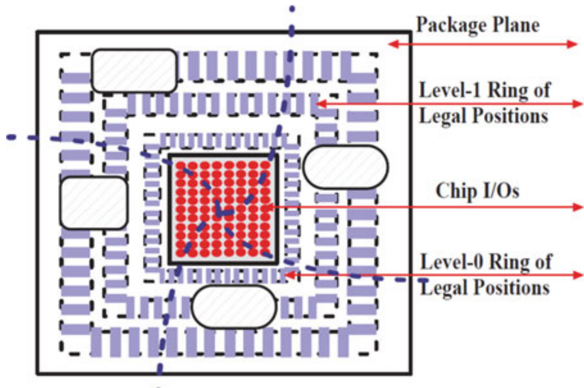


Fig. 8.20 The typical diagram of package plane, chip I/Os, and legal positions for decaps. The legal positions can be decomposed into multiple rings. The entire system can be further partitioned

observation that the impact of decaps to I/O power integrity can be distinguished by the distance to the center of the chip. The legal positions are decomposed into a total number of M_1 rings. Each ring is composed by a group of legal positions, and is levelized according to its distance to the center of the chip. The illegal positions due to package routing are not included in each level of rings. In addition, legal positions are assumed to be uniformly distributed on one ring, and there are mi legal positions at ith-level ring. The level-0 ring is closest to the I/O but has smallest numbers of legal positions, and the level-M_0 is farthest to the I/O but has largest numbers of legal positions.

Moreover, because of non-uniformly distributed I/Os in space, in addition to the distance, the orientation of legal positions can have different impact to I/O power integrity as well. This can be compensated by allocating different types of decaps on one levelized ring. In other words, one level of ring can be allocated non-uniformly with different types of decaps. Assuming that there are M_2 types of decaps, each levelized ring is duplicated by M_2 copies, called template in this chapter. Note that the legal positions on one template have the same decap type (cost), but only one template at one level is selected to allocate decaps.

As a result, there are total $M = M_1 \cdot M_2$ templates, and a vector of templates can be defined by

$$T = [T_1, T_2, \ldots, T_M]. \tag{8.14}$$

Usually, there are less than ten types of decaps [29, 41] to choose during the realistic design. Therefore, the number of M is still manageable.

Moreover, we need to define an accurate figure of metric to describe the power integrity at each I/O. The power integrity, i.e., the voltage bounce at each I/O, is time and space variant during a sufficient long time-period t_p. The noise amplitude (worst-case noise) metric used by the SA-based method can still be used here. However, a very narrow noise with a large amplitude may not lead to noise violation.

To avoid over-design, we improve the metric by using the noise integral with consideration of the noise pulse width. The noise integral above one target voltage Vd_i for ith I/O is

$$f_i = \int_{t_o}^{t_p} \max[y_i(T,t), Vd_i]\mathrm{d}t = \int_{t_s}^{t_e} [y_i(T,t) - Vd_i]\,\mathrm{d}t, \tag{8.15}$$

with a pulse width (t_s, t_e) during which the voltage is above the target voltage, and transient noise waveform $y_i(T,t)$ at ith I/O. We require this noise metric to be smaller than some bound Vc_j at the jth I/O cell, i.e.,

$$f_j \leq Vc_j \ \ (j = 1, .., p). \tag{8.16}$$

Recall that our design freedoms are twofolds: one is the level of ring, and the other is the decap type. Accordingly, our problem formulation becomes: Given the allowed noise (Vc), legal positions (M_1) and decap types (M_2), the decap allocation problem is to decide and minimize the total cost of decap under a given bound of decap number (M), such that the voltage violations at each I/O is smaller than the allowed noise.

This problem can be mathematically represented by

$$\begin{aligned} &\min \sum_{i=1}^{M} n_i T_i, \quad (i = 1, \ldots, M) \\ &\text{s.t. } Uf \leq Vc \text{ and } \sum_{j}^{M_1} m_j \leq M. \end{aligned} \tag{8.17}$$

where $f = [f_1, \ldots, f_N]^{\mathrm{T}}$, $U = I_{N \times N}$, $Vc = [Vc_1, \ldots, Vc_N]^{\mathrm{T}}$. In addition, n_i is the cost for i th template $(i = 1, \ldots, M)$, and m_j is the legal position number of j th level $(j = 1, \ldots, M_1)$. As discussed later, this problem can be efficiently solved by an allocation according to sensitivity. The key is to calculate the parameterized sensitivity from a localized integrity analysis.

8.2.6.2 Parameterized Circuit Equation

Because the partial inductance in PEEC introduces massive magnetic couplings, it would slow down the simulation. As shown by [42], the inverse of L ($L - 1$) [47] described by VPEC model can be sparsified and stamped stably and passively in the circuit matrix by a vector-potential nodal analysis (VNA). In this chapter, the nominal RLC-network for package planes is modeled by VPEC model and is stamped by VNA in frequency (s) domain:

$$(G_0 + sC_0)x(s) = BI(s), \quad y(s) = B^{\mathrm{T}}x(s) \tag{8.18}$$

with

$$x(s) = \begin{bmatrix} v_n \\ a_l \end{bmatrix}, \quad B = \begin{bmatrix} E_i \\ 0 \end{bmatrix}$$

and

$$G_0 = \begin{bmatrix} G & E_l L^{-1} \\ -E_l^T L^{-1} & 0 \end{bmatrix}, \quad C_0 = \begin{bmatrix} C & 0 \\ 0 & L^{-1} \end{bmatrix} \tag{8.19}$$

All notations in (8.19) are summarized in Table 8.6. Note that B is the adjacency matrix to describe P identical inputs and outputs, where the inputs $J = BI(s)$ are I/O current sources, and outputs $y(s)$ are the voltage responses (noise) at those I/Os. As discussed in Sect. 8.2.6.3, studying such an I/O map can guide the network partition.

To obtain the sensitivity, we need to first parameterize the system. Each template T_i is described by a pair of topology matrices T_i^g and T_i^c, where T_i^g describes how to connect the nodal equivalent conductance, and T_i^c defines how to connect the nodal capacitance and the branch equivalent susceptance (inverse of inductance). For an ith template, adding decaps between tiles m and n results in:

$$T_i^g(k,l) = T_i^g(l,k) = \begin{pmatrix} -g_i & \text{if } k = m, \quad l = n \quad \text{and} \quad k \neq l \\ \sum_l \left|T_i^1(k,l)\right| & \text{if } k = l \\ 0 & \text{otherwise} \end{pmatrix} \tag{8.20}$$

where $k, l \in 1, 2, \ldots, N$, and g_i is the equivalent conductance of one decap. $T_i^c(k,l)$ can be given similarly to add the equivalent capacitance and susceptance c_i and s_i. This decomposition enables us to apply an efficient decap allocation discussed with details later.

Accordingly, the decaps can be parametrically added into the nominal state matrix

$$\left[G_0 + sC_0 + \sum_{i=1}^{M}\left(T_i^g + sT_i^c\right)\right]x(T,s) = BI(s),$$
$$y(T,s) = B^{\mathrm{T}}x(T,s). \tag{8.21}$$

Recall that $x(T_M, s)$ is the total voltage response. For the purpose of design optimization, similar to the way of handling variations in [48], the state variable $x(T,s)$ is first expanded into Taylor series with respect to T_i, and then reconstructed into a new state variable using the nominal values and the first-order sensitivities

Table 8.6 Notations for system equation (8.18)

$x(y)\ (\in R^{N\times 1})$	State variable (at output)
$v_n\ (\in R^{N_v\times 1})$	Nodal voltage variables
$a_l\ (\in R^{N_l\times 1})$	Branch vector-potential variables
$G\ (\in R^{N_v\times N_v})$	Nominal conductance matrix
$C\ (\in R^{N_v\times N_v})$	Nominal capacitance matrix
$L^{-1}\ (\in R^{N_l\times N_l})$	Nominal inverse-inductance matrix
$E_l\ (\in R^{N_v\times N_l})$	Inductive incident matrices
$B\ (\in R^{N\times p})$	Input/output incidence matrix

Note that $N = N_V + N_L$

$$x_{ap} = \left[x_0^{(0)}, x_1^{(1)}, \ldots, x_M^{(1)}\right]^{\mathrm{T}}. \tag{8.22}$$

A dimension-augmented system can be reorganized according to the expansion order

$$(G_{ap} + sC_{ap})x_{ap} = B_{ap}I(s), \quad y_{ap} = B_{ap}^{\mathrm{T}}x_{ap}, \tag{8.23}$$

where

$$G_{ap}=\begin{bmatrix} G_0 & 0 & 0 \\ T_1^g & G_0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ T_M^g & 0 & G_0 \end{bmatrix} \text{and} C_{ap}=\begin{bmatrix} C_0 & 0 & 0 \\ T_1^c & C_0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ T_M^c & 0 & C_0 \end{bmatrix} \tag{8.24}$$

both have a lower-triangular-block structure. Although the system size is enlarged by parametrically adding decaps in this fashion, the ports of the augmented system are still the input ports of I/O currents. The size of augmented system can be still reduced by the model order reduction. In contrast, the impedance-based approach [29, 41] needs to increase the port number dramatically to add those decaps.

8.2.6.3 I/O Current Correlation and Spectral Clustering

Due to the large number of input ports, the macromodel by model reduction applied by [29] is still ineffective. Because the input current vectors show redundancy, the time/space-variant input I/O currents are not mutually independent. If the various inputs are correlated, then they are expressible as a function of a smaller number of independent variables by principal component analysis (PCA) using eigen-decomposition (ED).

This becomes the motivation to apply the singular value decomposition (SVD) [49–51]-based terminal reduction as SVD is equivalent to ED when the matrix to be decomposed is symmetric positive definite. These approaches study the correlation or similarity of inputs based on the moment, and compress the system transfer

function by a low rank approximation. Port compaction in fact studies the similarity of the system since it is based on the singular value (pole) analysis of the system transfer function. In contrast, the real correlation of inputs is dependent on the input signals. Therefore, finding the representative ports or ignoring some "insignificant" ports by the system similarity may lead to simulation errors, because there could be one significant output response caused by one significant signal that is applied at one port ignored from the system pole analysis. In this chapter, we propose to directly study the similarity or correlation of I/O currents. As a result, the large number of I/Os are clustered into K groups, each with one principal I/O current as input.

Given a typical set of I/O input vectors applied in a sufficiently long period, the sampled transient current $I(t_k, n_i)$ $(k = 1, \ldots, T,\ i = 1, \ldots, P)$ at time-instant t_k for each I/O n_i can be described by a random process as follows

$$S_{n_1} = \{I(t_1, n_1), \ldots, I(t_T, n_1)\}, \quad S_{n_2} = \{I(t_1, n_2), \ldots, I(t_T, n_2)\} \ldots\ S_{n_\mathrm{p}} = \left\{I(t_1, n_\mathrm{p}), \ldots, I(t_T, n_\mathrm{p})\right\}.$$

A current spatial-correlation matrix is defined by

$$C(i,j) = \frac{\mathrm{cov}(i,j)}{\sigma_i \cdot \sigma_j}, \tag{8.25}$$

where $\mathrm{cov}(i, j)$ is the covariance between nodes n_i and n_j, and σ_i, σ_j are standard variations of nodes n_i and n_j. This can be precomputed and the correlation coefficients $C(i,j)$ can be built into a table.

After extracting the correlation for input currents, we can build a correlation graph by assigning the weight of edge between I/Os n_i and n_j by the correlation value $C(i,j)$. A fast clustering method based on spectral analysis [52] can be applied to efficiently handle a large-scale correlation graph to find K clusters $A_1, \ldots, A_k$ using K-means method, where the I/Os in one cluster all show a similar current waveform. In addition, the number of I/O current sources can be reduced by PCA

$$J_x = VJ = VBI(s) \ \in R^{1\times K}.$$

It is equivalent to reduce the port matrix

$$B_x = VB \ \in R^{N\times K}.$$

As such, there is only one principal port selected to represent each cluster.

The overall clustering algorithm is outlined in Fig. 8.21. Usually, 1,000 sources can be approximated by around 10 sources if the inputs are strongly correlated. In addition, with the use of spectral analysis, the results by PCA and K-means is equivalent [52]. Therefore, there is only one principal port for each cluster.

Spectral Clustering Algorithm
1 **Input**: Cluster number K, correlation matrix $\mathcal{C} \in R^{N \times N}$, and I/O port matrix $\mathcal{B} \in R^{N \times p}$
2 Compute normalized Laplacian: $\mathcal{L} = \mathcal{D}^{-1/2}\mathcal{C}\mathcal{D}^{1/2}$, where $\mathcal{D} = diag(\mathcal{C})$;
3 Compute the first K eigenvectors $v_1, \ldots, v_K$ of L;
4 Let $V = [v_1, \ldots, v_K] \in R^{N \times K}$;
5 Let $y_i \in R^K$ $(i = 1, \ldots, N)$ be the vector of i-th row of V;
6 Cluster y_i $(i = 1, \ldots, N)$ by K-means into $C_1, \ldots, C_K$;
7 Transform $\mathcal{B} \in R^{N \times p}$ by PCA: $\mathcal{B}_x = V\mathcal{B} \in R^{N \times K}$;
8 **Output**: Clusters $\mathcal{A}_1, \ldots, \mathcal{A}_K$ with with $\mathcal{A}_i = \{j \vert y_j \in C_i\}$, and a new I/O port matrix $\mathcal{B}_x$

Fig. 8.21 Algorithm for spectral analysis of input current sources with PCA and K-means

8.2.6.4 Localized Integrity Analysis

Network Decomposition

Because the I/O currents are distributed non-uniformly in space, it has different impact to voltage bounces along different orientations. Therefore, it is possible that the one level of ring can be non-uniformly allocated with different types of decaps. Consequently, it better to decompose the I/O cells, the RLC-network for power supply, and the M templates into K blocks (see Fig. 8.20). A localized analysis can be then preformed to decide how many decaps for one block of I/Os.

The decomposition needs to partition the network based on physical properties such as couplings and latency. The TBS method in [43] leverages the property of latency, which is more suitable for a timing simulator. But, for the verification of power integrity, it is more meaningful to study the partition based on I/O inputs. Moreover, the partition in TBS [43] is to tear nodal voltage variables v_{n} for conductance and capacitance matrices, which is not suitable for inductance/susceptance partition because inductance/susceptance is described by the branch current/vector-potential. This can be solved as follows.

The flat VNA network (G_0, C_0, B_X) in (8.18) is first mapped into a circuit graph, where three different weights (2,1,0) are assigned for the resistor, capacitor, and self-susceptor (branch L^{-1}). A fast multilevel min-cut partition h-Metis [53] is applied to tear those interconnection branches with specified ports $A_1, \ldots, A_K$ obtained from the spectral clustering. As a result, the network is decomposed into two-levels with the torn resistors, capacitors, and self-susceptors in an interconnection block, and all remaining blocks are connected with the interconnection block by incident matrices as shown below

$$G_{ap} \to G_{ap} = \begin{bmatrix} G_1 & \cdots & 0 & X_{1,0} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & G_K & X_{K,0} \\ -X_{1,0}^T & \cdots & -X_{K,0}^T & Z_r \end{bmatrix}$$

$$C_{ap} \to C_{ap} = \begin{bmatrix} C_1 & \cdots & 0 & X_{1,0} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & C_K & X_{K,0} \\ -X_{1,0}^T & \cdots & -X_{K,0}^T & Z_i \end{bmatrix}$$

$$B_x \to B = \begin{bmatrix} B_1 & & & \\ \vdots & \ddots & & \\ & & B_K & \\ & & & 0 \end{bmatrix} \tag{8.26}$$

with

$$G_i = \begin{bmatrix} G_i & 0 & \cdots & 0 \\ T_{1,i}^g & G_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ T_{M,i}^g & 0 & \cdots & G_i \end{bmatrix}, \quad C_i = \begin{bmatrix} C_i & 0 & \cdots & 0 \\ T_{1,i}^c & C_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ T_{M,i}^c & 0 & \cdots & C_i \end{bmatrix} \tag{8.27}$$

where G_0 and C_0 are partitioned into K blocks G_j and C_j ($j = 1, \ldots, K$). Accordingly, those parameterized templates T_i are also partitioned into T_{ij} ($i = 1, \ldots, M \quad j = 1, \ldots, K$). Note that a block matrix structure is implemented to avoid building the large sized matrix.

Because the couplings are relocated into one interconnection block $Z_{r,i}$, each partitioned block in diagonal can be analyzed or reduced individually but with the same accuracy. However, the system poles are not determined only by those blocks in diagonal. To achieve a high-order accuracy but with only a low-order reduction, the TBS reduction in [43] is extended to consider inductance and is presented in the next section.

Triangular Block-Structured Reduction

After tearing the VNA network into a two-level form, we further transform it into a localized triangular block form with the use of replication [43]. Basically, as shown by (8.28), a replica block of G_{ap} is first stacked diagonally to construct a size-doubled G_{tb}, and then those lower-triangular blocks are moved to the upper triangular parts of G_{tb}. The resulting triangularized system is

$$G_{tb}=\left[\begin{array}{cccc|cccc} G_1^x & \cdots & 0 & X_{1,0} & G_1^y & \ldots & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & G_K^x & X_{K,0} & 0 & 0 & G_K^y & 0 \\ 0 & \ldots & 0 & Z_r & -X_{1,0}^T & 0 & -X_{K,0}^T & 0 \\ \hline & & 0 & & & & G_{ap} & \end{array}\right], \tag{8.28}$$

where

$$G_i^x = \mathrm{diag}\left[\underbrace{G_i,\ldots,G_i}_{M}\right],\quad G_i^y = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ T_{1,i}^g & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ T_{M,i}^g & 0 & \cdots & 0 \end{bmatrix}. \tag{8.29}$$

C_{tb} can be transformed in a similar fashion. The tranquilized system has a localized pole distribution, where poles are determined only by those blocks in the diagonal. In addition, the factorization cost only comes from those block in diagonal. However, due to the replica block, the overall factorization cost of the triangulated system is still the same as the original. To reduce the overall computational cost, we further apply a block-structured projection to reduce the system size.

As the network is decomposed and further triangularized, each block (G_i, C_i, B_i) can be reduced independently [42, 43] by finding a qth projection matrix Q_i ($R^{n_{bi}\times q}$) ($1 \le i \le K$) to contain the moment space of the diagonal block

$$\{R_i, A_iR_i, \ldots, A^q-1_iR_i\},$$

where $A_i = G_i^{-1}C_i$ and $R_i = G_i^{-1}B_i$, and $(n_b)_i$ is the size of original block. Accordingly, a block-diagonal projection matrix

$$Q = \mathrm{diag}\left[\underbrace{Q_1,\ldots Q_1}_{M},\ldots,\underbrace{Q_K,\ldots Q_K}_{M},Q_0,Q_{ap}\right] \tag{8.30}$$

is constructed to reduce the original matrix G_{tb}, C_{tb}, and B_{tb}, respectively.

$$\tilde{G}_{tb} = Q^{\mathrm{T}}G_{tb}Q,\quad \tilde{C}_{tb} = Q^{\mathrm{T}}C_{tb}Q,\quad \tilde{B}_{tb} = Q^{\mathrm{T}}B_{tb}. \tag{8.31}$$

In addition, note that Q_0 is an identity matrix to project those interconnection branches, and Q_{ap} is either obtained by directly applying a lower-order PRIMA to (G_{ap}, C_{ap}, B_X), or it can be accurately approximated by $[Q_1, Q_2, \ldots, Q_K, Q_0]^{\mathrm{T}}$ [43].

Moreover, one important observation is that, since only one principal port at each block is selected, a SIMO-reduction can be easily applied to achieve qth order moment matching for each block, and the reduced macromodel for each block can be used for any input signals.

As a result, a localized integrity responses and sensitivities in time-domain:

$$\begin{aligned} \left(\tilde{G}_{tb} + \frac{1}{h}\tilde{C}_{tb}\right)\tilde{x}_{tb}(t) &= \frac{1}{h}\tilde{C}_{tb}\tilde{x}_{tb}(t-h) + \tilde{B}_{tb}I(t) \\ \tilde{y}_{tb}(t) &= \tilde{B}_{tb}^{\mathrm{T}}\tilde{x}_{tb}(t). \end{aligned} \tag{8.32}$$

The kth block power integrity at one principle I/O perturbed by ith template is

$$\tilde{y}_{tb}(t) = \tilde{y}_{tb}^{(0)}(t) + \tilde{y}_{tb}^{(1)}(t). \tag{8.33}$$

Note that although it is a localized solution, the couplings between different blocks are still taken into account due to the two-level network decomposition and the triangularization. In the next section, we present the decap allocation algorithm using the block integrity including nominal responses and sensitivities.

8.2.6.5 Algorithm and Experimental Results

Sensitivity-Based Optimization

The problem in Sect. 8.2.6.1 can be efficiently solved by the sensitivity-based optimization. The key is to calculate sensitivities from the structured and parameterized macromodel in Sect. 8.2.6.2. Then, the decap is allocated for each block according to the sensitivity of I/O power integrity with respect to templates. The partitioned template $T_{i,j}$ is recursively added according to the order of the gain. As a result, a minimum number of decaps are added to reduce the voltage violations in problem formulation (8.17). Such a greedy flow is able to solve large-scale designs efficiently and effectively.

The overall optimization is outlined in Fig. 8.22. The nominal value and sensitivity are computed one time from the structured and parameterized macromodel from (8.32). Afterwards, the decap is added into each block independently. In kth block, the template vector T is ordered according to the magnitude of sensitivities

$$\left\{\delta y_{i_1,k}, \delta y_{i_2,k}, \ldots, \delta y_{i_M,k}\right\}$$

and is added according to this order until the integrity constraint of kth block is satisfied. The algorithm then iterates to the next block until all the power integrities of all blocks are satisfied. Because each input-template is legalized initially to exclude those illegal positions, the output template vector T can be directly used for the detailed placement of decaps.

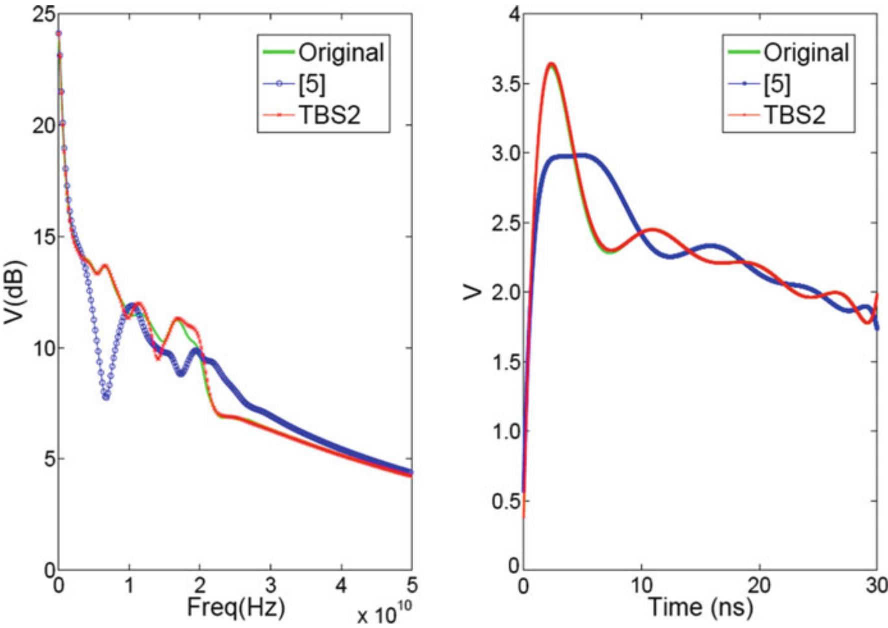


Fig. 8.22 Algorithm for sensitivity-based decap allocation

Experimental Results

The proposed macromodeling and allocation algorithm has been implemented in C and Matlab. We call our macromodeling method TBS2, and our optimization multi-ring-based allocation (MRA). Experiments are run on a Linux workstation with 2 GB RAM. A typical FPGA package model is assumed with the specific application inputs. Four packages P/G planes are assumed with the same size of 1 cm × 1 cm. The V_{dd} is assumed to be 2.5 V, and the targeted noise is 10 % of V_{dd}, i.e., 0.25 V. The worst-case I/O current sources are modeled as triangle-waveform with rising time of 0.1 ns, width of 1 ns and period of 150 ns. The current sources are randomly distributed in a square of 0.2 cm × 0.2 cm located in the center of a 1 cm × 1 cm package plane. The 30 % of remaining area are reserved for legal positions. We keep the same four decap types as summarized in Table 8.1. The total number of decaps is bounded by 80, and the total number of rings is five, each decomposed into four levels (0–3). We increase the circuit complexity by increasing the number of discretized tiles, and need more levels for legal positions when the tile number becomes larger. We allocate decaps by MRA and SA methods to satisfy the power integrity at I/Os under constraints of either the noise amplitude (NA) or the noise integral (NI).

We first compare our method with the macromodeling method in [29] as follows. The packages planes are discretized into 4,096 tiles, described by a RLC-mesh with 4,096 resistors, 6,144 capacitors, and 64,000 susceptors. There

MULTIPLE RING-BASED ALLOCATION ALGORITHM
1 **Input**: Integrity vector **Vc**
2 Compute initial $y^{(0)}$ and $y^{(1)}$ using (33);
3 Reorder $\mathcal{T}_k = \{\mathcal{T}_{i_1,k}, \mathcal{T}_{i_2,k}, \ldots, \mathcal{T}_{i_M,k}\}$ $(k = 1, \ldots, K)$;
4 **Do** allocation with max $\mathcal{T}_k$ for block k
5 Delete max $\mathcal{T}_k$ from $\mathcal{T}_k$ and $M = M - 1$;
6 Compute $y_k = y_k^{(0)} + y_k^{(1)}$;
7 **Until** y_k satisfies the block integrities $\mathbf{Vc}_k$
8 **Output**: Allocated template-vector **T** for detailed decap placement

Fig. 8.23 Waveform accuracy comparisons between the original, the method in [16], and TBS2 in (**a**) time-domain and (**b**) frequency-domain for 4th principal port. The original and our method are visually identical

are 420 I/O current sources as inputs. The sequences of I/O currents are generated by simulating the specified application of input vectors for millions of cycles. One spatial-correlation matrix C is extracted from the sequences. Then, the spectral clustering finds eight principal ports by PCA and clusters the ports into eight groups. Accordingly, the network is partitioned into eight blocks by h-Metis. Figure 8.23 compares the frequency- and time-domain responses at the fourth principal port. Due to the I/O port reduction and a localized reduction and analysis, our method is 21× faster (765 s vs. 35.2 s) to build and 25× faster (51 min vs. 2 min) to simulate compared to [29]. Moreover, because the TBS reduction can achieve a higher accuracy with the use of triangularization, the waveform by TBS2 is visually identical to the original. But the reduced waveform by [29] has non-negligible error. A detailed analysis shows that [29] has about 3.06× larger waveform error in the time-domain.

We also compare the runtime and the cost of allocated decaps between SA and MRA. During this comparison, both methods use the noise amplitude as the constraint. As shown in Table 8.7, due to the systematical allocation with use of sensitivity, MRA reduces the allocation time by 97× on average compared to SA. In addition, SA can only handle circuits up to about 10,000 nodes. To obtain a result in a reasonable time, SA usually cannot find the minimal solution. For a circuit with 10,680 nodes, MRA finds a solution with cost about 216 in 13 min, but SA finds a solution with cost about 233 (+9 %) in 1 day.

In addition, Fig. 8.24 shows the voltage-bounce map (at 80 ns) across the top plane. The initial noise amplitude is around 1.0 V, and its voltage-bounce profile is shown in Fig. 8.24a. In contrast, the decap allocation by MRA results in a smaller voltage bounce that closely approaches the targeted bounce (0.25 V) as shown in Fig. 8.24b.

Table 8.7 Results of decap allocations by SA and our MRA method

ckt (#node + #I/O)	#level	#legal-pos	#partition	SA-NA		MRA-NA		MRA-NI	
				opt	$cost	opt	$cost	opt	$cost
280 + 40	0,1	20	4	192.2 s	16	5.2 s	10	5.4 s	10
1,160 + 160	0,1	80	4	2 h	55	62.3 s	50	64.2 s	40
4,720 + 640	0,1	320	4	7 h	102	277.1 s	96	280.2 s	80
10,680 + 1,440	0,1,2	720	8	1 day	233	783.7 s	216	773.5 s	200
19,521 + 3,645	0,1,2	1,701	8	NA	NA	932.4 s	277	972.2 s	265
55,216 + 10,880	0,1,2,3	5,440	16	NA	NA	51 min	340	54 min	312

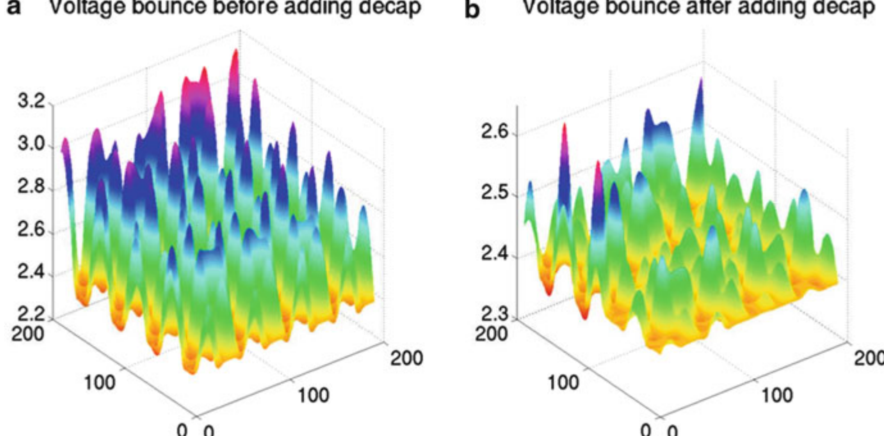


Fig. 8.24 Voltage bounce at P/G-plane (**a**) before decap allocation and (**b**) after decap allocation

We further compare the runtime and the cost of allocated decaps by noise amplitude (NA) and noise integral (NI), both using MRA for allocation. As shown in Table 8.7, compared to the optimization with NA, the optimization with NI reduces the cost of allocated decaps by up to 7 % within a similar allocation time. This is because the constraint by the noise amplitude ignores the accumulated effect of the transient noise waveform. In contrast, the constraint by noise integral can consider the noise pulse width, and can accurately predict the decap allocation using the transient noise waveform. As a result, NI reduces the cost by up to 16 % compared to the SA using NA [41].

In the next section, we discuss a little bit more about three-dimensional packaging, the recent advancements in ICPS.

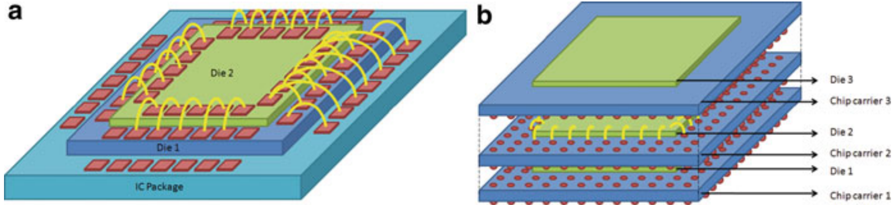


Fig. 8.25 3D SIP stacking with wire bonding (**a**), 3D SIP package-on-package stacking (**b**)

8.3 TSV-Based 3D Stacking: The Good, the Bad, and the Powerful

8.3.1 Introduction About 3D IC Stacking Techniques

Recent years, modern IC design starts to steer into the track of 3D realm with the help of 3D IC stacking techniques. Potentially superior performance can be obtained from 3D IC stacking in terms of smaller footprint size, higher clock speed, and heterogeneous integration, which in turn reduce the chip cost. Also, numerous new design freedoms are given by the 3D IC stacking. 3D IC stacking techniques can be mainly categorized as 3D SIP (System-in-Package) and 3D IC with TSVs (through-silicon via) [54, 55].

The 3D SIP packaging service is currently available with many SATS (Semiconductor Assembly & Test Service) providers such as ASE Group, SPIL, and ChipMOS. The 3D SIP technology uses two stacking strategies: wire bonding and PoP stacking [55]. The former mainly targets at subsystem level integration, e.g., Flash and SRAM (Static Random-Access Memory), Logic and memory, e.g., DRAM (Dynamic Random-Access Memory), non-volatile, etc. As shown in Fig. 8.25a, two dies are stacked vertically in a pyramid shape and they are interconnected through bonding wires on the edges. With this stacking technique, the overall package size and the length of signal and power traces are both largely reduced, thereby improving the power efficiency and signal integrity. However, the interconnections between dies can only be located along the edges, which limits the number of inter-die interconnection. Also the size of the upper tier chip has to be smaller than that of the lower tier one, rending the stacking method not effective to integrate large numbers of tiers. As an “upgraded” version of the wire-bonding method, the PoP stacking technique allows more chip tiers and more interconnections. Figure 8.25b gives a simple demonstration about this technique. Each die is attached to a chip carrier with either bumps or bonding wires first, and then all chip carriers are stacked together like a multi-story building. A great advantage of PoP stacking technique is the heterogeneous integration capability of multiple function modules such as RF, power management, processor, antenna, sensors, and so on, fabricated using different technologies. It is often addressed as “E-CUBE” [56].

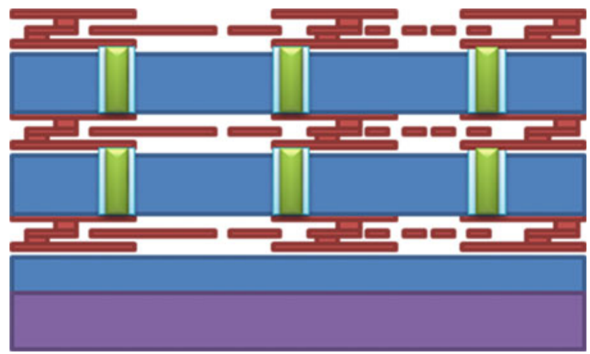


Fig. 8.26 3D IC stacking with TSV

Unlikely the relatively mature 3D SIP packaging technique, TSV-based 3D ICs (Fig. 8.26) are still in the research phase [57, 58]. To the knowledge of authors, there are already some SATS companies that can provide the service, e.g., ALLVIA, austriamicrosystems, AVIZA technology's Versalis fxp system [59], MIT Lincoln Lab, Tezzaron, and Ziptronix. TSV mainly has two advantages in the chip level integration compared with SIP: reduced via size and heterogeneous integration. Compared to the existing 3D SIP packing techniques, TSV technology brings a significant reduction in the via size and pitch size. TSV can be manufactured with diameters down to 1–5 μm with pitch sizes down to 2–10 μm, which is more than ten times smaller than that of the SIP technology [54]. Similar to the "E-CUBE" in 3D SiP, TSVs also allow chips of two or more different technologies to be integrated in the same package, but in a much more compact and thinner way. The thickness of each tier will be less than 20 μm after going through a wafer-thinning process, which is negligible compared to the overall silicon substrate thickness of 700 μm[60]. In addition, with the help of reduced via dimensions and pitch sizes, it offers denser inter-tier connections and better signal integrity, with less area overhead and parasitic capacitance [55].

Generally speaking, there are three kinds of bonding styles for TSV-based 3D ICs: face-to-face, face-to-back, and back-to-back (Fig. 8.27). Face-to-face bonding can be achieved with bumps or copper-to-copper thermal diffusion bonding process [61]. Because of its short and dense interface, face-to-face bonding provides best performance with a reduced cost. However, it can be only used to bond two dies. Face-to-back and back-to-back bonding have to be used for the bonding of more dies, where TSVs are required as well.

In addition, varieties of TSV implementation processes can be categorized as via-first, via-middle, via-last, and via-after by different IC fabrication sequences [62]. Via-first means TSVs are formed on bare Si or SOI (Silicon on insulator) wafer before FEOL process; via-middle means TSVs are formed between FEOL and BEOL processes; via-last means TSVs are formed after BEOL processes and before bonding; via-after means TSVs are formed after bonding process. Typically, via-first has smaller diameter (1–10 μm) than via-last (10–50 μm). All these processed are being pursued, but at the moment via-last and via-middle are the key areas of interest due to the lower implementation cost [62].

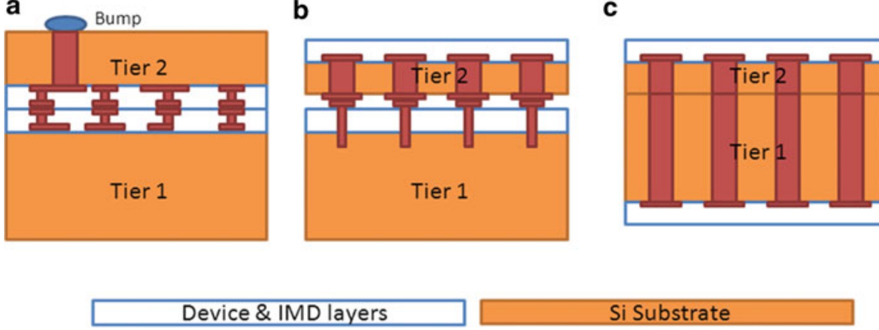


Fig. 8.27 Common bonding styles: face-to-face (**a**), face-to-back (**b**), and back-to-back (**c**)

Fig. 8.28 TSV stacking techniques: wafer-on-wafer stacking (**a**), die-on-wafer stacking (**b**), and die-on-die stacking (**c**)

Finally, as shown in Fig. 8.28, there is a number of TSV stacking techniques that can be used: wafer-on-wafer stacking, die-on-wafer stacking, and die-on-die stacking [54]. Considering the effect of thermal stress, wafer-on-wafer stacking is likely to be used for the bonding of the homogeneous technology, and die-on-wafer stacking and die-on-die stacking are more suitable for the bonding of heterogeneous technology [54]. Wafer-on-wafer stacking has relatively low cost because of the single step operation [63]. But it suffers a cumulative yield lost from bad chips: even though some bad chips are known from pre-bond testing, it still has to be bonded with the whole wafer. Die-on-wafer stacking and die-on-die stacking can be more flexible with the pre-bond test results, e.g., only known good dies (KGD) can be picked out for bonding, and yet they come with a higher cost. We explain this in more detail in Sect. 8.3.2.

With all the 3D IC stacking techniques introduced above, undoubtedly the development of IC technology will be much faster in the 3D realm. And we envision that the application of TSVs plays the fundamentally important role in this revolution. However, on the way of constructing 3D IC world with TSVs, there are many critical unaddressed issues blocking in front. In the second part of this section, we elaborate the key challenges of TSV applications, and briefly review the efforts contributed by many researcher to resolve these issues. In the last part of this section, the solutions of two key challenges are discussed in detail.

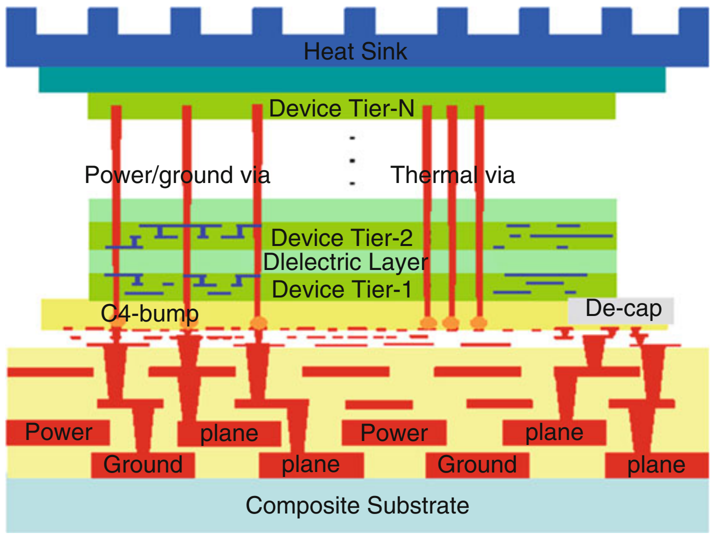


Fig. 8.29 A typical 3D stacking with through-silicon vias

8.3.2 *Challenges*

Many TSV related issues in 3D IC are mentioned in [64], such as KGD (Known Good Die) requirement, test method, design automation, chip cooling, and other fabrication process related issues. Actually, most of the issues are correlated with to each other. And they can be mainly categorized into two aspects: thermal/power issues and testing issues.

8.3.2.1 Thermal and Power Distribution Issues of 3D IC

A common exercise of 3D stacking with TSVs is demonstrated in Fig. 8.29. The figure is shown in a package format to illustrate the power and thermal distribution issues.

The power distribution issues of 3D IC integration with TSVs can be understood from twp perspectives: First, power efficiency and noise compliance. Figure 8.29 shows that the power of 3D IC is supplied from the bottom package through C4-bumpers. Assuming the current consumption of each device tier is identical, the average current passing through Power/Ground via accumulates with the increasing tier number. Therefore a large number of TSVs should be placed to reduce the total resistance loss in TSVs. In real cases, the current consumption of each tier may not be identical. It is therefore more complicated to allocate the vias. Second, in 3D ICs, more currents sinks are connected to the same power/ground trace, which incurs larger on-chip SSN. The off chip decap may not be efficient enough to reduce it, and a number of inductive loop will be generated inside the 3D

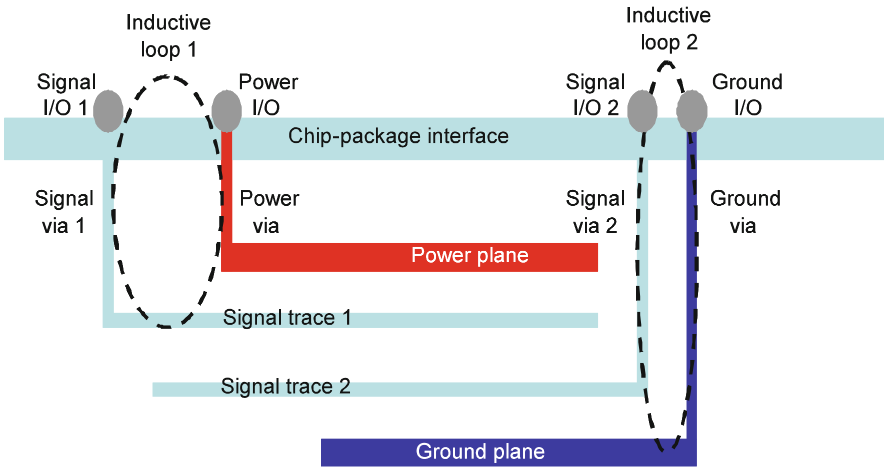


Fig. 8.30 The power delivery by vertical power/ground vias and its impact to inductive current loops

chip (Fig. 8.30) to interfere with each other. Both of these two perspectives are proposed to be resolved by proper allocations of power/ground TSVs [65–67].

Thermal distribution is another grand challenge faced by 3D IC integration. Because of the dielectric layers between each die have much lower thermal conductivity than SI substrate [68], the thermal power generated is easily trapped to generated hot spots [69]. The performance of devices and interconnections will be greatly degraded in high temperature [70–73]. From Fig. 8.29, we know that the thermal resistance gradually decreases with the increasing tier number and it is the highest between bottom Device tier-1 and heat sink. The proper allocation of signal TSVs can lead to a tradeoff between the wirelength and the chip temperature, as have been studied in [74]. Furthermore, in order to alleviate the thermal stress of the bottom tier, (dummy) thermal TSVs are proposed to be placed to improve the effective thermal conductivity of the whole 3D chip [75], which is also shown in Fig. 8.29. However, these thermal TSVs lead to excess chip area overhead and reduce the routing resource of signal and power traces as well.

Modern VLSI design requires dynamic power management inside the chip such as duo-core or four-core processors. Therefore, the actual thermal model is not only spatially variant but also temporally variant. As proposed by [76, 77], the dynamic thermal model can be studied with the inputs of time-varying thermal power which is defined by the running-average of the cycle-accurate power over several thermal time constants (Fig. 8.31). In this way, the output ports temporally and spatially variant temperatures can be estimated by defining an integrity integral according to time and space [78]. Thereby different locations in a chip reach their worst-case temperature at different times. And the reliability of devices can be significantly affected due to the sharp-transition temperature surge to a large value or time-accumulated high temperature.

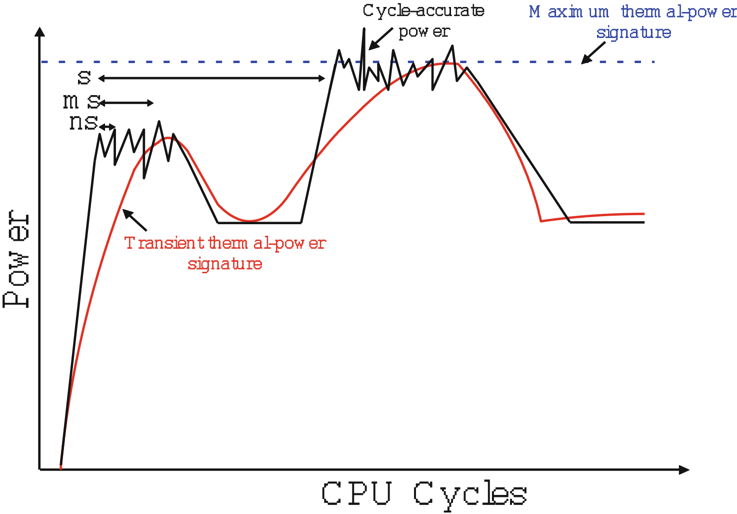


Fig. 8.31 The definitions of the cycle-accurate power, transient thermal power, and maximum thermal power at the different scale of time-constant

To address the issue of dynamic power and thermal integrity, distributed thermal-RC model and electrical-RLC model are proposed to realize computer-aided analysis [79]. They can be described as:

$$Gx(t) + C\frac{dx(t)}{dt} = BI(t), \quad y(t) = L^T x(t) \tag{8.34}$$

or in frequency (s) domain

$$(G + sC)x(s) = BI(s), \quad y(s) = L^T x(s) \tag{8.35}$$

where B is the topology matrix to describe p_i input ports with injected input sources; L is the one to describe p_o output ports for probing thermal or power integrity and adjusting via density.

In order to address vias allocation issue for both power and thermal integrity, of course computer-aided design automation and model-based calculation are required. This is yet a very challenging task which can be resolved in three steps. We give a detailed discussion about it in the third part of this section.

8.3.2.2 Testing Issues of 3D IC

The yield of 3D IC integration drops exponentially with the die number. Taking a 90 % yield die for example, the six-die stacking yield drops to 50 % without considering the yield of stacking up process. Then the cost of each 3D IC is

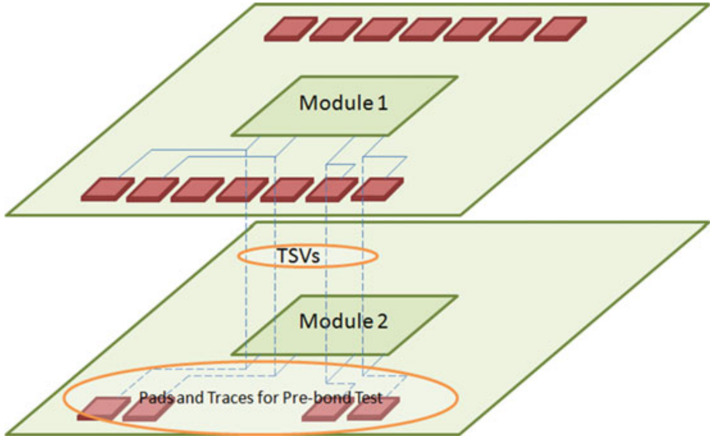


Fig. 8.32 3D IC illustration for testability

accordingly increased drastically. Therefore, in order to have a higher yield after stacking, it is of great importance to use known good dies (KGD) for integration [80–82].

However, it is of non-trivial to find KGD in 3D IC stacking. Pre-bond test must be performed on each tier before stacking them together. Basically there are two essential components for pre-bond testing: the probing interface such as the bonding pads or bumps, and the function integrity which means the chip under test must be able to perform the actual function required during testing.

Conventionally, single tier wafer chip is probed through the bonding pads or bumps to verify the functionality of every individual die. For the single tier 2D IC, all the I/Os can be accessed through the bonding pads or bumpers on the top surface of each die, so the testing can be directly performed. But in 3D ICs, extra pads and traces are required to be placed in inner tiers to maintain testability (Fig. 8.32). The excess pads and traces added inevitably reduce the usable area of inner tier chip, and increase the cost of 3D ICs accordingly. However, area consumed by the inner tier pads will not be totally wasted. As proposed in [81], by properly arranging the location of testing pads, they can be used as on-chip decoupling capacitors.

Function integrity is another challenge in 3D IC pre-bond testing. It is mainly revealed in the aspects of tier partitioned circuitry and synthesized clock trees. As mentioned in [83], the 3D IC partitioning scheme has three levels of granularity: technology level, architectural level, and circuit level. Technology level partitioning is mostly on different tiers, so the pre-bond testing can be addressed with additional traces and testing pads. The key challenges in pre-bond testing are from architectural level and circuit level partitioning. As we can see from Fig. 8.32, if Module 1 and 2 are the architectural level partitioned sub-modules, and they have to be linked together to create the functionality, the question raised is “how to do the pre-bond test for them individually.”

In order to address the pre-bond test issues of incomplete state circuits, scan-chain-based design approach is proposed by [81, 84]. First, modules on each tier are divided into "islands" with registers placed in between. In the normal working condition, data can follow freely between islands, while in the testing mode, these registers can be used as bridges to insert testing patterns into the each "island." LTCs (Layer Test Controllers) can be used to manage all registers on each tier and to connect the testing pads for the pre-bond test, and it can also be used to communicate with globe test controller for the post-bond test. Conventional 2D Bus based TAM (Test Access Mechanism) method is also proposed by [80, 85] to resolved 3D IC pre-bond testing issues. The TAM approach enables the pre-bond testability in a modular basis. Once a module is selected, it is isolated from the surrounding modules and can be tested through the TAM bus.

3D clock tree design is the most fundamental challenge in the pre-bond testability issues. First, there must be a complete 2D clock tree for each tier to perform the pre-bond test. Second, there must be a complete 3D clock tree for the normal operation of 3D IC. Unfortunately, the 2D optimum clock structures are often not identical to the 3D one and the 3D optimum clock structures are often not pre-bond testable [81, 82]. Also, the synthesis of clock signal in 3D IC has several constraints such as skew and latency. It is often required clock signals at different module inputs to have zero skew in the design phase. In order to address the 3D clock distribution issue and pre-bond testability at the same time, pre-bond test clock tree design is introduced by [82]. The major idea of this approach is to add redundant clock trees to the optimized 3D clock trees to realize the pre-bond testability and use TG (Transmission Gate) to disconnect all redundant clock trees to avoid unwanted capacitive loading.

Figure 8.33a illustrates the concept of a 3D clock tree, which utilizes TSVs to distribute the clock signal to different tiers. The 2D redundant clock tree for pre-bond testing is shown in Fig. 8.33b. By using transmission gates in Fig. 8.33c, the 2D redundant tree is connected with the clock tree during the pre-bond testing stage, and is disconnected after die-stacking. Figure 8.33d shows the final 3D clock scheme, which is composed of the 3D clock tree and the 2D redundant tree.

The reliability of TSV is also very important [86]. It is unacceptable to let the whole chip fail because of one broken TSV. In order to increase the reliability of TSVs, various redundancy techniques are proposed. In an 8 GB 3D DDR3 DRAM, Samsung apply a TSV grouping concept including two spare TSVs and four signal TSVs [58]. Another proposed solution to utilize the one spare TSV for each signal TSVs [87], thereby largely increase the reliability. Truly with this method the overall chip yield can be increased, but it requires significantly increased area overhead. Therefore, one can imagine the great challenge of reducing TSVs numbers without reducing reliability. This problem is discussed in detail in the following section.

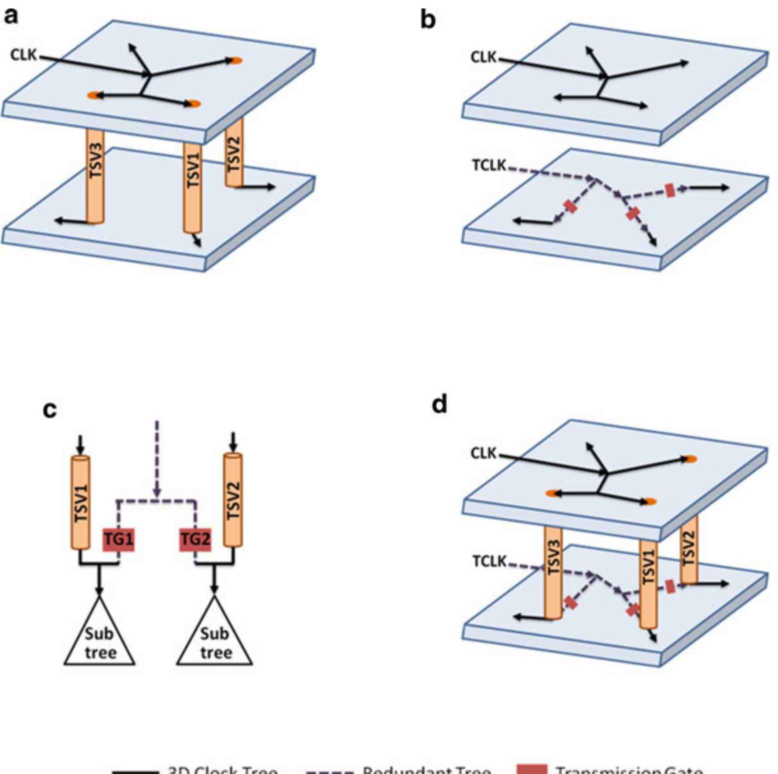


Fig. 8.33 (**a**) The 3D clock tree after die-stacking. (**b**) The 2D clock tree of each die before die-stacking. (**c**) The connection structure of redundant tree. (**d**) A 3D clock scheme

8.3.3 Solutions

The major challenges of TSV application in 3D ICs are elaborated in the last part. We can see huge efforts are spent to resolve these issues by many researchers. Still, there are many unsolved issues blocking in front in terms of power/thermal integrity and pre-bond testing. We would like to focus on two major key issues among them in this section. In the first part, the TSV allocation issue for power and thermal integrity is solved, and in the second part, a novel method to reduce TSV number while maintaining reliability is presented.

8.3.3.1 3D IC Design Automation Considering Dynamic Power and Thermal Integrity

As mentioned in the last section, design automation and model-based calculation are required to resolve the TSV allocation issue for both power and thermal integrity. We first identify the problem formally, and then solve it in three steps: (1) create and

simplify the thermal and power models under consideration, (2) create and simplify the thermal/power I/O ports under consideration, (3) find the minimum TSV density that maintains the both power and power integrity in 3D IC.

TSV Allocation Problem

Generally speaking, metal TSVs are also a good thermal conductor that can be used to dissipate excess heat out of 3D IC. In our approach, TSV density is adjusted to resolve both thermal and power integrity issues with following constraints. The first is the integrity constraint of the temperature gradient (T_t) and voltage bounce (V_t). The second is the resource constraint with provided signal-net congestion, e.g., maximum and minimum TSV density($n_{\min}, n_{\max}$).

Also the thermal-integrity integral [88] is applied to capture the sharp-transition of temperature and the time-accumulated temperature impact as the measure of dynamic thermal integrity at jth ($j = 1, \ldots, p_{\mathrm{o}}$) output port:

$$f_j^T = \int_{t_0}^{t_p} \max\left[y_j(t), Tc\right] \mathrm{d}t = \int_{t_s}^{t_e} \left[y_j(t) - T_{\mathrm{r}}\right] \mathrm{d}t, \tag{8.36}$$

with a pulse width (t_s, t_e) in a sufficient long time-period t_p all in the scale of thermal constant (ms). $y_j(t)$ is the transient temperature waveform at jth output port, and T_{r} is the reference temperature.

The spatial difference of p_{o} output ports at the bottom device tier is considered as well. The overall thermal integrity is defined by a normalized summation:

$$f^T = \frac{\sum_{j=1}^{p_{\mathrm{o}}} f_j^T}{t_p^T \cdot p_{\mathrm{o}}} \tag{8.37}$$

Such a measure of thermal integrity takes into account both the temporal and spatial variation of temperature. Similarly, the power-integrity integral f_j^V is defined at jth power/ground I/O with reference voltages V_{dd} and ground, and integrated at the period t_p^V in the scale of electrical-constant (ns). The overall power integrity f^V is defined similarly to f^T for p_{o} power/ground I/Os with the reference voltage V_{r} (0 for ground vias and V_{dd} for power vias).

Together, the TSV allocation problem can be summarized below:

- Limited Voltage bounce V_t for p_{o} output ports at power/ground I/Os
- Limited Temperature gradient T_t for p_{o} output ports at bottom device tier
- Minimized the total TSV number

It also can be represented in a mathematical format as:

$$\min \sum_{j=1}^{p_o} n_j$$
$$\text{s.t.} f^V \leq V_t, \quad f^T \leq T_t$$
$$\text{and}\, n_{\min} \leq n_j \leq n_{\max} \tag{8.38}$$

Note that n_j is the TSV density at the jth track and V_t and T_t are the targeted voltage bounce and temperature gradient. f^V and f^T are the metrics of power integrity and thermal integrity, respectively.

Algorithm

There are many difficulties if we apply the equations above to resolve the TSV allocation problem, which can be mainly understood from three aspects: (1) there are numerous input and output points to consider, (2) the distributed thermal-RC and electrical-RLC models are too complicated to be solves by limited computing power, (3) as a design optimization approach, sensitivity is out major concern instead of the nominal response. Therefore, the techniques of reducing the states and I/Os and generating sensitivity should be studied for design automation.

Complexity Compression of States

As shown in [89, 90], the dominant state variables and compact macromodels can be obtained via model order reduction. The dominant state variables are related to the block Krylov subspace

$$K(A,R) = \{A, AR, \cdots A^{q-1}R, \ldots\}$$

constructed from moment matrices

$$A = (G + s_0C)^{-1}C, \quad R = (G + s_0C)^{-1}B$$

by expanding the system transfer function

$$H(s) = L^{\mathrm{T}}(G + sC)^{-1}B$$

at one frequency s_0.

By applying the block Arnoldi iteration [90], a small dimensioned projection matrix Q ($N \times q \times p_{\mathrm{i}}$) can be found to contain qth-order block Krylov subspace

$$K(A,R,q) \subseteq Q.$$

Using Q the original system can be reduced by projection

$$G = Q^{\mathrm{T}}GQC = Q^{\mathrm{T}}CQB = Q^{\mathrm{T}}BL = Q^{\mathrm{T}}L.$$

Accordingly, the reduced system transfer function becomes

$$\hat{H}(s) = L^{\mathrm{T}}(G + sC)^{-1}B.$$

As proved [89, 90], the reduced $\hat{H}$ approximates the original system transfer function $H(s)$ by matching the first q block moments expanded at the frequency-point s_0. This procedure can be applied to generate compact macromodels for both thermal-RC and electrical-RLC circuits.

Normally there are large numbers of injecting inputs, and large numbers of output probes are also required to monitor the system integrity. Multiple inputs and multiple outputs (MIMO) in our thermal-RC and electrical RLC models bring challenges for the projection-based model order reduction. The dimension of the reduced MIMO system $\hat{H}$ ($\in R^{p_{\mathrm{o}} \times p_{\mathrm{i}}}$) depends on the input-port number p_{i} and p_{o}. Therefore, to have an effective macromodel, the number of ports should be further compress when p_{i} and p_{o} are both large. In the following, a much smaller number of principal input and output ports are identified by correlation study.

Complexity Compression of I/Os

Generally, there can be thousands of thermal-power sources injected at each active tier or hundreds of switching current sources injected at I/Os. The size of the macromodel increases with the number of ports, and hence the computational cost to solve the macromodel is still high. Since the electrical signals may share the same clock and operate within a similar logic function, their waveforms in the time-domain at certain input ports can show a correlation. Similarly, the thermal power may differ significantly between those regions with and without the clock gating, but can be quite similar inside the region with the same mode since inputs have similar duty-cycles over time. Based on the correlation, we can reduce the redundancy in I/Os by identifying those principal ports.

We call this phenomenon *input similarity*. As the input vector

$$I(t) = [I_1 \quad I_2 \quad \cdots I_{p_{\mathrm{i}}}] \in R^{p_{\mathrm{i}} \times 1}. \tag{8.39}$$

is usually known during the physical design, they can be represented by taking a set of “snapshots” sampled at N time-points

$$\begin{bmatrix} I_1(t_0) & \cdots & I_1(t_N) \\ \vdots & \ddots & \vdots \\ I_{p_{\mathrm{i}}}(t_0) & \cdots & I_{p_{\mathrm{i}}}(t_N) \end{bmatrix} \tag{8.40}$$

in a sufficiently long period $[0, T_p]$. The sampling cycle is in a different time-scale for the thermal power (ms) and switching current (ns). According to the POD analysis [91], the similarity can be mathematically described by a correlation matrix (or Grammian), estimated by a covariance matrix:

$$R = \frac{1}{N}\sum_{\alpha=1}^{N}(I(t_\alpha) - I)(I(t_\alpha) - I)^T \in R^{p_\mathrm{i}\times p_\mathrm{i}}. \tag{8.41}$$

I is a vector of mean values defined by:

$$I = \frac{1}{N}\sum_{\alpha=1}^{N} I(t_\alpha) \tag{8.42}$$

Usually, the input vector $I(t)$ is periodic and the waveform in each period can be approximated by the piecewise-linear model.

An *output similarity* is defined for responses at output ports and measured by a *output correlation matrix*. To extract the output correlation matrix that is independent on the inputs, we assume that p_i inputs in the input vector $I(s)$ are all the unit-impulse source $h(s)$ and define an input-port vector $J(s)$ by

$$J = BI(s), \in R^{1\times N}, \tag{8.43}$$

which has p_i non-zero entries with the unit-value "1". Accordingly, the p_o output responses $y(s)$ are calculated by

$$\begin{aligned} y(s) &= L^\mathrm{T}(G + sC)^{-1}J \\ &= [\, y_1(s) \quad y_2(s) \quad \cdots \quad y_{p_\mathrm{o}}(s)\,] \in R^{p_\mathrm{o}\times 1}. \end{aligned} \tag{8.44}$$

The according output correlation matrix is extracted in the frequency-domain. Similarly, the output signals can be represented by taking a set of "snapshots" sampled at N frequency points

$$\begin{bmatrix} y_1(s_0) & \cdots & y_1(s_N) \\ \vdots & \ddots & \vdots \\ y_{p_\mathrm{o}}(s_0) & \cdots & y_{p_\mathrm{o}}(s_N) \end{bmatrix} \tag{8.45}$$

in a sufficiently wide band $[0, s_{\max}]$. The $s_{\max}$ locates in a low-frequency range for the temperature and in a high-frequency range for the voltage. A covariance matrix is defined in the frequency-domain as follows

$$R = \sum_{\alpha=1}^{N}(y(s_\alpha) - \bar{y})(y(s_\alpha) - \bar{y})^T \in R^{p_\mathrm{o}\times p_\mathrm{o}} \tag{8.46}$$

to estimate the correlation matrix among p_{o} outputs.

$\bar{y}$ is a vector of mean values defined by:

$$\bar{y} = \frac{1}{N}\sum_{\alpha=1}^{N} y(s_\alpha) \tag{8.47}$$

Let $V = [v_1, v_2, \ldots, v_K]$ ($\in R^{N\times K}$) as the first K singular-value vectors of the input correlation matrix R, and $W = [w_1, w_2, \ldots, w_K]$ ($\in R^{N\times K}$) as the first K singular-value vectors of the output correlation matrix R. All singular-value vectors are obtained from the SVD of (V, W). A rank-K matrix P_i can be constructed by $P_i = VV^{\mathrm{T}}$, and a rank-K matrix P_o can be constructed by $P_o = WW^{\mathrm{T}}$. As shown in [90], the correlation matrix (R, R) is essentially the solution that minimizes the least-square between the original states $(I(t), y(s))$ and their rank-K approximations $(P_i \cdot I(t), P_o \cdot y(s))$. As a result, both the input signals $I(t)$ and the output signals $y(s)$ can be approximated by an invariant (or dominant) subspace spanned by the orthonormalized columns of V and W, respectively:

$$I = VI_K, \quad y = Wy_K. \tag{8.48}$$

Based on (8.48), it leads to the following equivalent system equation

$$(G + sC)x_K(s) = B_K I_K(s), \quad y_K(s) = L_K^{\mathrm{T}} x_K(s) \tag{8.49}$$

where

$$L_K^{\mathrm{T}} = W^{\mathrm{T}} L^{\mathrm{T}}, \quad B_K = BV. \tag{8.50}$$

Therefore, both the dimensions of L ($\in R^{N\times p_{\mathrm{o}}}$) and B ($\in R^{N\times p_{\mathrm{i}}}$) are greatly reduced when $K << p_{\mathrm{i}}$ and p_{o}. We call I_K and y_K *principal inputs and outputs* identified by *principal input-port and output-port matrices* B_K and L_K, respectively.

Dynamic Integrity and Sensitivity by Structured and Parameterized Macromodel

Recall that the design parameter in our problem formulation is the TSV density at one track. Blindly allocating the TSV by searching all kinds of combinations would be computationally expensive if not impossible. Therefore, we decide the TSV density based on the changes at outputs, i.e., sensitivities, caused by the change of TSV density.

To calculate sensitivity, let's first parameterize the nominal system (8.35). The added TSV at one track is described by two parameters: n_j the TSV density and X_j the topological matrix that connects the TSV into the nominal system. As such, a parameterized state-space description can be obtained by

$$\left(G + sC + \sum_{j=1}^{p_o} n_j g_j + s\sum_{j=1}^{p_o} n_j c_j\right) x_K(n,s) = B_K I_K(s),$$
$$y_K(n,s) = L_K^{\mathrm{T}} x_K(n,s). \tag{8.51}$$

Similar to [78, 88, 92–95], we expand $x(n,s)$ in the Taylor series with respect to n_j, and introduce a new state variable x_{ap}

$$x_{ap} = \left[x^{(0)}, x_1^{(1)}, \ldots, x_{p_o}^{(1)}\right]^{\mathrm{T}}. \tag{8.52}$$

It contains both the nominal response $x^{(0)}$ and its first-order sensitivities $\left[x_1^{(1)}, \ldots, x_{p_o}^{(1)}\right]$ with respect to p_o parameters $[n_1, \ldots, n_{p_o}]$. The overall response is obtained by

$$x = x^{(0)} + \sum_{j=1}^{p_o} x_j^{(1)}.$$

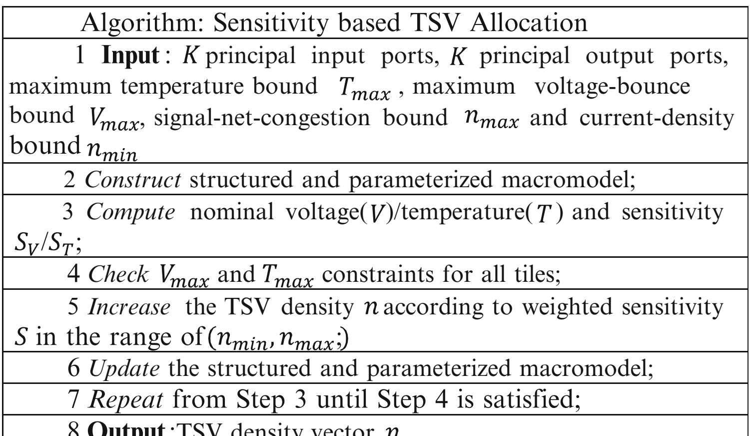
Substituting (8.52) into (8.51), (8.51) can be reformulated into a parameterized system with an augmented dimension by

$$(G_{ap} + sC_{ap})x_{ap} = B_{ap} I_K(s), \quad y_{ap} = L_{ap}^{\mathrm{T}} x_{ap}, \tag{8.53}$$

where G_{ap} and C_{ap} show a lower-triangular-block structure and hence x_{ap} can be solved from block-backward-substitution [77, 88, 92–95].

To further compress the dimension of the state-matrices G_{ap} and C_{ap}, we first construct a lower-dimensioned subspace Q_{ap} from the moment expansion of (8.53), and then transform Q into the block-diagonal form Q_{ap}. After the block-orthonormalization of Q_{ap}, we apply a two-side projection to (8.53) by Q_{ap} and obtain a dimension-reduced system with preserved lower-triangular-block structure. We call the resulted macromodel structured and parameterized macromodel. The accuracy of the macromodel is preserved to match the dominant moments of the original model. More importantly, due to the structure-preservation, both of the nominal response and the sensitivity with regard to the TSV-density change, can be calculated simultaneously. As such, a structured and parameterized macromodel can be easily embed into the optimization flow of TSV allocation problem.

The overall optimization flow to solve the problem formulation (8.38) is outlined in Fig. 8.34. Its inputs are two parts. The first is a principal system by (8.53) with the identified K principal input and output ports. The second is the user provided temperature bound $T_{\max}$, voltage-bounce bound $V_{\max}$, signal-net congestion bound $n_{\max}$, and current-density bound $n_{\min}$. Then, a structured and parameterized macromodel is built once. Both nominal responses and sensitivities at K principal input ports for each perturbed allocation-pattern are solved in one time. If the



Algorithm: Sensitivity based TSV Allocation
1 **Input**: K principal input ports, K principal output ports, maximum temperature bound T_{max}, maximum voltage-bounce bound V_{max}, signal-net-congestion bound n_{max} and current-density bound n_{min}
2 *Construct* structured and parameterized macromodel;
3 *Compute* nominal voltage(V)/temperature(T) and sensitivity S_V/S_T;
4 *Check* V_{max} and T_{max} constraints for all tiles;
5 *Increase* the TSV density n according to weighted sensitivity S in the range of $(n_{min}, n_{max};)$
6 *Update* the structured and parameterized macromodel;
7 *Repeat* from Step 3 until Step 4 is satisfied;
8 **Output**: TSV density vector n

Fig. 8.34 Algorithm for the sensitivity-based TSV allocation with the use of macromodels

integrity constraints are not satisfied for K principal tracks, the TSV density vector n are increased according to the sensitivity. This process repeats until the integrity constraints are satisfied. Details of this Algorithm can be found in [92].

Results

Experimental Setup

Experiments are implemented in C and MATLAB and run on a Sun-Fire-V250 workstation with 2 G RAM. We call the separated allocation of thermal TSVs and power/ground TSVs the *sequential optimization*, and call our allocation of power/ground TSVs for both power and thermal integrity the *simultaneous optimization*. Moreover, the steady-state analysis is employed to calculate a static integrity [96–98]. We use the sequential optimization with the static integrity as the baseline, in comparison to the sequential optimization with the dynamic integrity and the simultaneous optimization with the dynamic integrity proposed in this work. The electrical, thermal constants and dimensions are the same with [92]. The targeted voltage violation V_t is 0.2 V and the targeted temperature T_t is 52 °C. One modest 3D stacking is assumed with 2-device-tier/2-dielectric-layer. Moreover, there are 1-heat-sink and 2-P/G-plane used in the example.

Results Analysis

The steady-state temperature map across the bottom device tier i shown in Fig. 8.35. In this example, it is assumed that all thermal-power sources are located along one side of the device tier. The initial chip temperature at the bottom tier is 150 °C, and

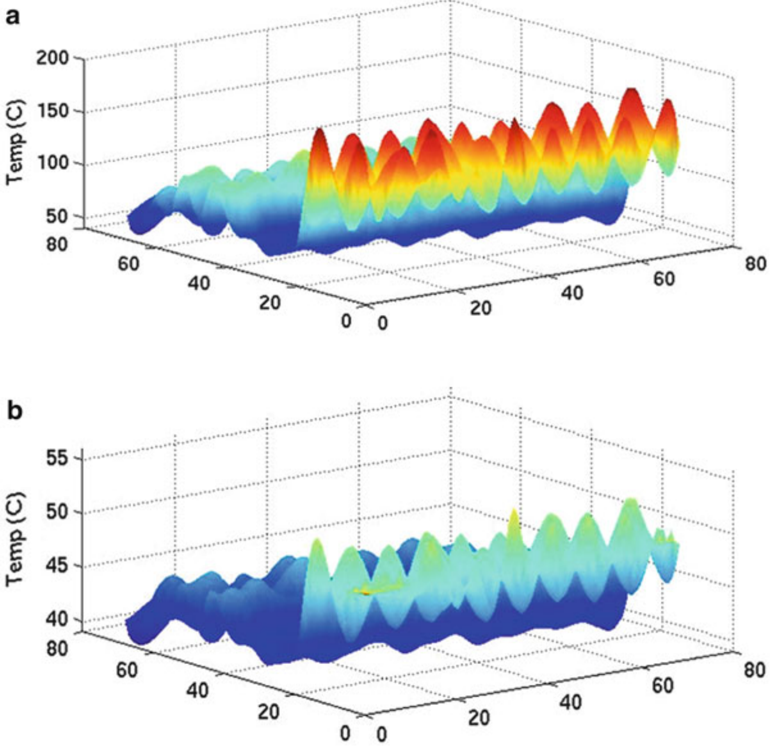


Fig. 8.35 Steady-state temperature maps of bottom device tier (**a**) before allocating TSV, and (**b**) after allocating TSV in different temperature scales

its temperature profile at steady-state is shown in Fig. 8.35a. In contrast, the TSV allocation results in a cooler temperature that closely approaches the targeted temperature as shown in Fig. 8.35b. Clearly, even at steady-state the temperature is still spatially variant. An accurate measure of integrity is therefore needed to consider a time and space averaged integrity at selected probing ports.

The details of the benchmark circuits are summarized in Table 8.8. The runtime and the number of allocated TSVs are compared in Table 8.9. In Table 8.9, columns 2–3 show the runtime and the number of allocated TSV for the baseline, and columns 4–8 show the results for the optimizations using the dynamic integrity. In detail, column 4 shows the runtime of transient analysis using macromodels without the port-compression, and column 5 shows the number of allocated TSVs under the sequential optimization. Column 6 shows the runtime of transient analysis using macromodels with the port-compression, and columns 7–8 show the number of allocated TSVs under the sequential and simultaneous optimizations, respectively.

The use of macromodels reduces the computational cost to solve power and thermal integrity and their sensitivities. Compared to the macromodel without the port-compression, the macromodeling with the port-compression reduces the overall runtime by up to 16× with similar allocation results. Compared to steady-state

Table 8.8 The complexity of the original circuits and the reduced circuits including: the size, number of input ports, and number of output ports

ckt	Total tile#	Reduced size (*T*, *V*)	Input src# (*T*, *V*)	*K*-input (*T*, *V*)	Output track#	*K*-output (*T*, *V*)
ckt1 (2-tier)	1.9 K	(30, 80)	(10, 20)	(10, 20)	4^2	$(4^2, 4^2)$
ckt2 (2-tier)	6 K	(15, 48)	(100, 200)	(5, 8)	4^3	(6, 4)
ckt3 (2-tier)	12 K	(80, 160)	(300, 600)	(10, 16)	4^4	(8, 5)
ckt4 (2-tier)	27 K	(96, 180)	(1 K, 2 K)	(12, 18)	4^4	(10, 8)
ckt5 (2-tier)	52 K	(96, 220)	(1 K, 3 K)	(12, 20)	4^5	(12, 14)

analysis with full-matrix analysis, our macromodel with the port-compression has a 127× smaller runtime. Additionally, the steady-state analysis cannot complete the largest example in a reasonable runtime. The maximum transient-waveform difference introduced by the macromodel is about 7 % when compared to the exact transient waveform.

The sequential thermal/power optimization with the simultaneous thermal/power optimization are further compared. Here both methods allocate TSVs with the use of dynamic integrity. The simultaneous optimization reduces the TSV cost by up to 34 % when compared to the sequential optimization with static integrity, and by up to 22 % when compared to the sequential optimization with dynamic integrity. This demonstrates that the reusing of power/ground TSVs can reduce the TSV cost when compared to allocating the dummy thermal TSVs separately from the power/ground TSVs.

8.3.3.2 Fault-Tolerant 3D Clock Scheme

As discussed in Sect. 8.3.2, the clock signal is essential to the pre-bond testability of 3D IC. And it is also of great importance to increase the reliability of TSVs with reduced TSV number for clock signals. TSV fault-tolerant unit (TFU) is proposed to achieve the same reliability as double TSV method with largely reduced number of TSVs [99].

Clock Design with TFU

As shown in Fig. 8.33, the clock signal is delivered to different tiers through the 3D clock tree which is connected by TSVs. Once a TSV fails, the clock signal cannot be delivered to the sub-tree connected with the TSV, thus resulting in chip malfunction. This problem can be solved by providing a redundant path for

Table 8.9 Comparisons of TSV number and runtime for the sequential optimization with steady-state analysis, the sequential optimization with transient analysis and the simultaneous optimization with transient analysis

	Steady-state (direct)		Transient (MACRO-1)		Transient (MACRO-2)		
ckt	Runtime (s)	Total TSV # by seq-opt	Runtime (s)	Total TSV # by seq-opt	Runtime (s)	Total TSV # by seq-opt	Total TSV # by seq-opt
ckt1 (2-tier)	5.4	178,800	0.63	153,800 (−13 %)	0.63	153,800 (−13 %)	112,800 (−36 %)
ckt2 (2-tier)	29.7	184,900	0.81	159,600 (−13 %)	0.56	159,600 (−13 %)	118,200 (−36 %)
ckt3 (2-tier)	182.2	218,100	18.6	183,800 (−16 %)	4.2	184,200 (−15 %)	136,200 (−38 %)
ckt4 (2-tier)	1,269.2	234,800	165.7	199,000 (−15 %)	10.3	199,600 (−15 %)	145,600 (−38 %)
ckt5 (2-tier)	NA	NA	NA	NA	41.2	208,600 (NA)	154,200 (NA)

Two macromodels are used during the transient analysis. Macromodel-1 does not use the port-compression, and macromodel-2 uses the port-compression

each sub-tree. When a TSV fails, the corresponding redundant path can be used to deliver the clock instead. This idea is conceptually simple, yet challenging in practice: There are many different ways of designing such a redundant path (e.g., double TSVs), each with different area penalties. We would like to choose the one with minimum area overhead.

In the following, we first introduce the TSV fault-tolerant unit (TFU). Then, we discuss the design considerations for a TFU. Finally, we explain how to integrate the proposed TFUs with 3D clock scheme synthesis.

Algorithm

TSV Fault-Tolerant Unit

The main idea of the TFU is to reuse a part of the 2D redundant tree as a redundant path for TSVs. Figure 8.36a shows the design of the proposed TFU, which is composed of two TSVs. In a TFU, each TSV is followed by a multiplexer.

Three transmission gates and two multiplexers are configured to determine the clock signal paths. There are two configurations at the post-bond stage. Firstly, when both TSVs are good, all three transmission gates are turned off, and the two multiplexers select the clock signal from the corresponding TSV. As such, the clock signal is delivered through *TSV1* and *TSV2*. Secondly, when one of the TSVs fails (assuming *TSV2*), *TG1* is turned off, and the other two transmission gates are turned on. In this case, the corresponding multiplexer selects the clock signal which is delivered through *TSV1* and the redundant path. There is one configuration at the pre-bond stage. All three transmission gates are turned on, and the clock signal is delivered through the 2D redundant tree. The resultant fault-tolerant 3D clock scheme with TFUs is illustrated in Fig. 8.36b.

To implement such a TFU, two fundamental problems need to be addressed. First, we need to decide how to pair TSVs, and second, we need to properly balance the clock skew with additional buffers if a redundant path is used in the presence of TSV failure. These two problems are discussed in detail in the next section.

TSV Pairing and Buffer Insertion

We now discuss TSV pairing and buffer insertion in this subsection. To tolerate TSV failure, we insert a TFU for each TSV pair. The two TSVs in a pair act as a "spare" TSV for each other. If the distance between these two TSVs is too long, it will bring about excessive interconnect resistance and capacitance, which in turn will increase the difficulty of clock balancing. Thus, we will only pair two TSVs within a feasible range T to form a TFU. Note that the range T cannot be too small as well. Otherwise we might not be able to find any TSV pairs in the range. In experiments, we find that setting $T = 100\ \mu m$ provides an optimal balance between the number of TSVs that can be paired and the delay overhead. Within the feasible range, the penalty induced by the proposed TFU will not exceed the tolerance bound. Note that it is still possible that a TSV cannot find another TSV in the

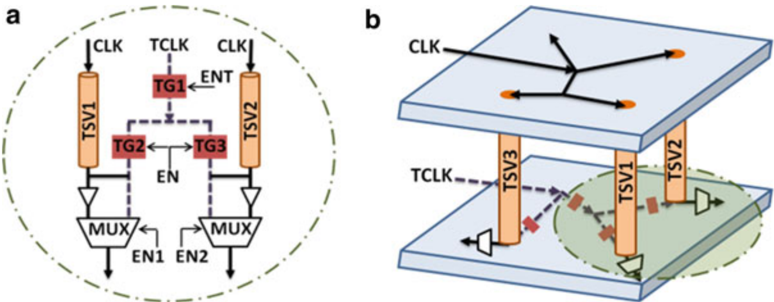


Fig. 8.36 (**a**) The proposed TSV fault-tolerant unit (TFU). (**b**) The fault-tolerant 3D clock scheme using TFUs

feasible range to pair with, and in such case the TSV will be protected by the double TSV technique.

Although the TFU can provide tolerance against TSV failure, the clock latency changes when the clock signal is delivered through a redundant path. As a result, the clock skew may violate the design constraint when the redundant path is used. To tackle this problem, we can insert one delay buffer for each multiplexer (as shown in Fig. 8.36a). When one of the TSV is bad, the buffer is used to balance the additional delay brought by the redundant path. It is clear now that the delay of the buffer should be set to be equal to the delay of the redundant path through *TG2* and *TG3* (the extra time for the clock signal to route). On the other hand, during the pre-bond testing, the paths through the buffers are not selected by the multiplexers, so they have no effect at all.

Note that the buffer insertion may change the upstream load capacitance and thus the delay. In view of this, instead of using a single buffer, we use a buffer chain with proper sizes to match the original upstream load capacitance without inserting the TFU. As such, the clock skew can still be maintained in the entire clock tree.

TFU Integration with Clock Tree Synthesis

We now discuss how to construct a 3D clock tree with TFUs. Since a TFU is constrained by the feasible range T, the TSV positions in the 3D clock tree affect the number of TFUs. As such, it is better to construct TFUs during the course of clock tree synthesis. As a vehicle to demonstrate the efficacy of the proposed TFU structure, we extend the clustering algorithm in [98] to synthesize the 3D clock scheme. However, other bottom-up clock tree synthesis algorithms can be used as well.

For the simplicity of discussion, we illustrate our algorithm using the example in Fig. 8.37, which has two tiers and five sinks. Figure 8.37a shows that a parent node is created by connecting two nodes. Our clock tree topology is generated by iteratively forming a new parent node with the minimum cost in a bottom-up fashion. The cost of connecting two nodes N_a and N_b is based on the distance in 2D geometry and defined as follows:

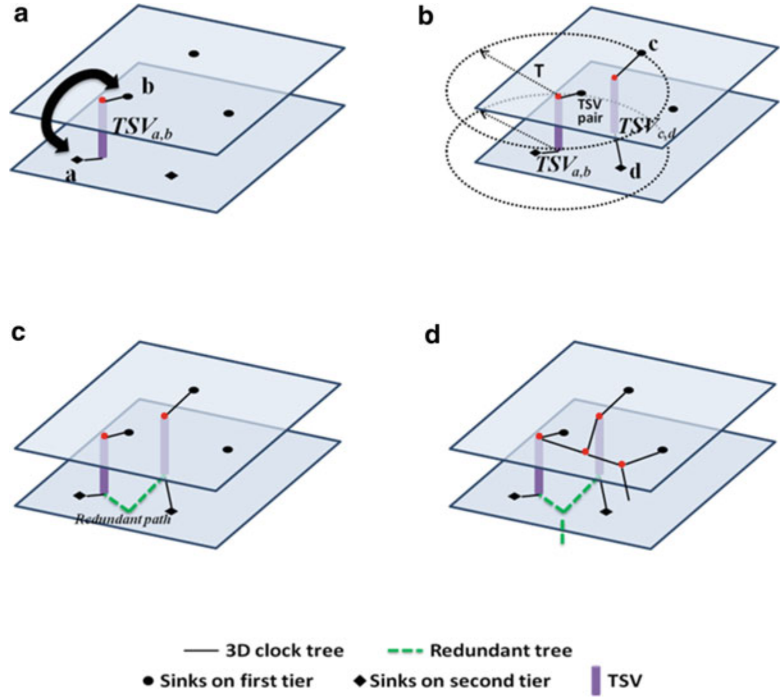


Fig. 8.37 An example for the proposed fault-tolerant clock tree constructing. (**a**) $\mathrm{TSV}_{a,b}$ is generated to connect nodes a and b with the minimum cost. (**b**) Generate a new TSV within the range if possible. (**c**) Determine the redundant path connecting the TSVs. (**d**) The final fault-tolerant 3D clock scheme is constructed

$$\mathrm{Cost}_{N_a,N_b} = \begin{cases} \mathrm{DistanceInGeomerty}(N_a, N_b) & N_a, N_b \text{ in the same tier} \\ \mathrm{DistanceInGeomerty}(N_a, N_b) + \alpha \times \mathrm{TSV}_{\mathrm{length}} & \text{otherwise} \end{cases} \quad (8.54)$$

where the parameter α denotes the weight factor which considers the overhead of using a TSV. When the two nodes with the minimum cost are in different tiers, a TSV is used to connect them. For example in Fig. 8.37a, a TSV $\mathrm{TSV}_{a,b}$ is generated to connect nodes a and b.

When a TSV, $\mathrm{TSV}_{i,j}$ is created, we start to find another TSV to form a TSV pair for TFU. We search for possible *pairing TSVs* within the feasible range T of $\mathrm{TSV}_{i,j}$. The pairing TSV with the minimum distance to $\mathrm{TSV}_{i,j}$ will be inserted for the TFU. For example in Fig. 8.37b, node c and d are the nodes within the feasible range T of $\mathrm{TSV}_{a,b}$, and the pairing TSV, $\mathrm{TSV}_{c,d}$ is generated to form a TSV pair in the TFU. When a TSV pair is created, a redundant path between the TSV pair can be determined in Fig. 8.37c. On the other hand, the double TSV technique will be applied to $\mathrm{TSV}_{i,j}$. when no pairing TSV can be found. The process continues until we build the entire clock tree.

After the 3D clock tree is synthesized, we continue to synthesize the remainder of the 2D redundant tree in each tier according to the algorithm discussed in [100].

Finally, the proposed fault-tolerant 3D clock scheme is constructed in Fig. 8.37d. In addition, we also integrate the slew-aware buffer insertion technique [101, 102] to solve the issue of clock slew rate control. When the downstream capacitance of a node exceeds the maximum capacitance, denoted as C_{max}, a clock buffer is inserted. The overall algorithm is outlined in Algorithm 1. Again, we need to emphasize here that since the TSV paring process is applied during the bottom-up tree construction, various bottom-up clock tree synthesis methods (e.g., [100, 103]) can be integrated with our proposed TFUs.

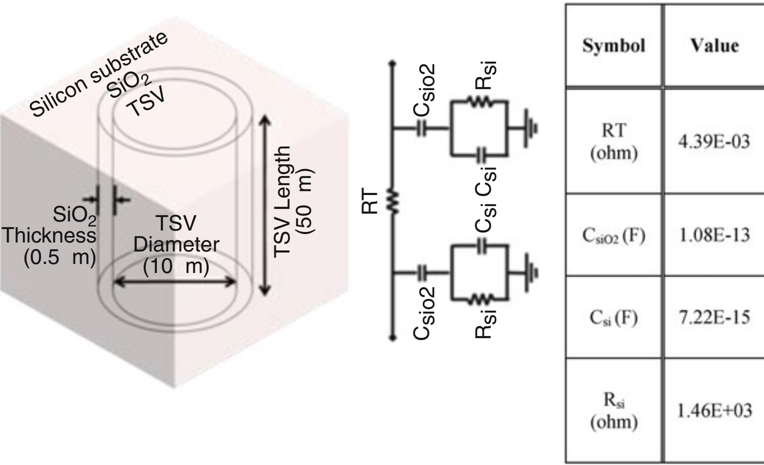
Results

Experimental Setup

We implement our algorithm in C++ and perform experiments on several 3D designs, including one industrial case with two tiers and 55.4 K clock sinks. Other designs are obtained by stacking benchmark circuits from IBM suite or ISCAS89. All designs are synthesized to an industrial 65 nm technology library. Figure 8.38 shows the TSV model used in this work [104]. The wire resistance and capacitance per unit length are $r_w = 0.14\ \Omega$ and $c_w = 0.206$ fF, respectively. To control the slew rate, the maximum load capacitance of each buffer is 200 fF. The feasible range T for our fault-tolerant unit is 100 μm. The clock skew constraint for the industrial design is 150 ps, and for other benchmark circuits is 100 ps. Components such as buffers and multiplexers in TFUs are instantiated from the same library, and the area overhead of a TFU is 11.7 μm^2. In addition, the diameter of a TSV is 10 μm and the failure rate is 1 %. The 3D clock scheme with TFUs is called F3D, while the double TSV technique is called Double3D, and the normal clock tree without these fault-tolerant techniques is called the original design (Orig).

Algorithm 1: Fault-tolerant 3D clock scheme synthesis

Inputs: A set of sinks distributed on N tiers
Outputs: A fault-tolerant 3D clock scheme with TFUs
Initialization: put all sinks to the pool;
while pool is not empty ***do***
  bottom-up tree construction;
  if a $TSV_{i,j}$ is needed ***then***
    search for the possible pairing TSVs within a feasible range T;
    if pairing TSVs are found ***then***
      insert the TFU using the pairing TSV with the minimum distance to $TSV_{i,j}$;
    else
      apply the double TSV technique;
    end if
  end if
  optimize slew and skew by inserting buffers;
  add parent nodes to the pool;
end while
synthesize the remainder of the 2D redundant tree in each tier



Symbol	Value
RT (ohm)	4.39E-03
C_{sio2} (F)	1.08E-13
C_{si} (F)	7.22E-15
R_{si} (ohm)	1.46E+03

Fig. 8.38 A TSV model

Results Analysis

Table 8.10 summarizes the comparison between the original design, Double3D, and F3D. Columns 1 and 2 show the name and the sink number of a circuit, respectively. Columns 3, 4, and 5 show the used TSV numbers of the original design, Double3D, and F3D, respectively. Note that F3D also adopts the double TSV technique for those TSVs which cannot find pairing TSVs within the feasible range. Column 6 shows the number of TSV pairs in F3D. Column 7 shows the area overhead of Double3D. Column 8 shows the area overhead of F3D, which includes the cost of TFUs and the additional TSVs. Column 9 shows the yield rate of the design without any fault-tolerant technique as in (8.55). Columns 10 and 11 show the yield rates of Double3D and F3D as in (8.56), respectively.

$$Y_{\text{original}} = (1 - f_{\text{TSV}})^{\#\text{TSV}} \tag{8.55}$$

$$Y = \left(1 - (f_{\text{TSV}})^2\right)^{\frac{\#\text{TSV}}{2}} \tag{8.56}$$

where f_{TSV} is the failure rate of the TSV and we have assumed that the failure event for each TSV is independent. For a design without TSV redundancy, any faulty TSV results in a bad chip. On the other hand, for Double3D and F3D, the design is faulty when both TSVs in a pair fail. Columns 12 and 13 show the total wire-length of Double3D and F3D, respectively. Columns 14 and 15 show the clock skew of Double3D and F3D, respectively. The clock skew results are based on *HSPICE* simulation. Columns 16 and 17 show the runtime of constructing Double3D and F3D, respectively. On average, Double3D uses 2× TSVs compared with the original design, while our F3D only uses 1.31×. Moreover, the area overhead of

Table 8.10 Comparison between the original, the Double3D, and F3D methods

Circuits		#Sinks	#TSVs			#TPs	Area overhead (μm^2)		Yield rate (%)			Wire-length (mm)		Skew (ps)		Runtime (s)	
			Orig	Double3D	F3D	F3D	Double 3D	F3D	Orig	Double 3D	F3D	Double 3D	F3D	Double 3D	F3D	Double 3D	F3D
2 tiers	Industrial	55,410	871	1,742	1,158	429	87,100	33,728	~0	90.41	94.37	1,302	1,353	98	143	5,691	5,735
	2T_case1	3,199	31	62	40	17	3,100	1,099	68.94	99.63	99.80	30	33	44	72	9.31	9.34
	2T_case2	3,025	31	62	38	15	3,100	876	71.06	99.66	99.81	30	30	38	48	8.27	8.31
	2T_case3	3,025	34	68	42	18	3,400	1,011	67.57	99.61	99.79	31	31	66	93	8.32	8.37
	2T_case4	1,460	5	10	8	3	500	335	93.21	99.93	99.96	7	7	29	43	1.86	1.89
	2T_case5	2,765	8	16	10	5	800	259	90.44	99.90	99.95	12	12	43	57	7.31	7.51
	2T_case6	5,004	15	30	18	7	1,500	382	85.15	99.84	99.91	22	23	51	63	27.71	28.75
3 tiers	3T_case1	4,496	64	128	82	35	6,400	2,210	46.59	99.24	99.59	49	55	35	73	20.45	20.66
	3T_case2	4,753	63	126	74	30	6,300	1,452	51.00	99.33	99.63	49	54	49	83	22.85	22.91
	3T_case3	3,363	15	30	18	7	1,500	382	85.15	99.84	99.91	15	15	28	37	11.72	11.74
	3T_case4	5,866	24	48	28	12	2,400	541	77.00	99.74	99.86	27	27	69	90	40.92	41.22
Average ratio		–	1	2	1.31	–	1	0.36	1	1.34	1.35	1	1.04	–	–	1	1.01

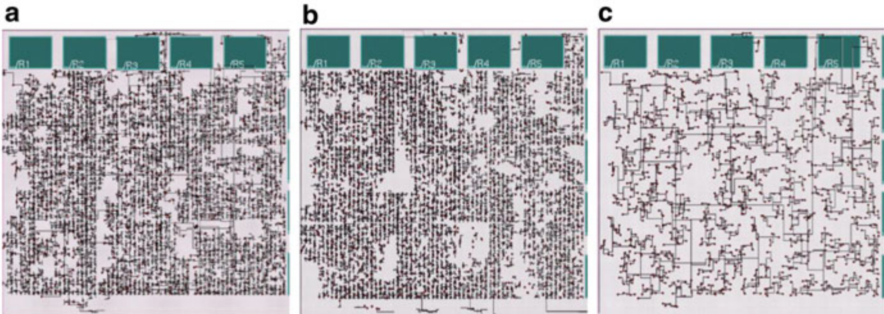


Fig. 8.39 The result of industrial case. (**a**) First tier. (**b**) Second tier : 3D clock tree. (**c**) Second tier : redundant clock tree

F3D is 64 % less than that of Double3D. Finally, the yield rate improvement by Double3D is 34 % and by F3D is 35 %.

Particularly, for the industrial case, the TSV number of the 3D clock tree without any fault-tolerant technique is 871. Considering Double3D and F3D, the TSV number becomes 1,742 and 1,158, respectively. The area overheads of Double3D and F3D are 87,100 (μm^2) and 33,728 (μm^2). The area overhead of F3D is 61 % less than that of Double3D. Without any fault protection, the yield rate of the industrial design will be almost zero. The yield rate can then be improved to 90.41 % by Double3D and 94.37 % by F3D. Note that F3D can provide relative higher yield rate, since it used fewer TSVs than Double3D. The wire-lengths of Double3D and F3D are 1,302 (mm) and 1,353 (mm) respectively. Compared to Double3D, F3D induces 4 % wire-length overhead. Furthermore, all resultant 3D clock schemes satisfy the clock skew constraints. Finally, both F3D and Double3D exhibit similar runtime.

To conclude this section, Fig. 8.39 shows the 3D clock scheme of the industrial circuit after applying the proposed fault-tolerant technique. The clock source is located in the first tier, and the 55,400 sinks are distributed on two tiers. The red spots in the figure denote the TSV positions, and the sinks are not shown for clarity. Figure 8.39a, b show the 3D clock trees in the first tier and in the second tier, respectively. Figure 8.39c shows the redundant clock tree in the second tier.

8.4 Summary

While there exists a continuous effort in industry to integrate more functionalities into the same area, the prohibitive scaling cost at 45 nm and beyond makes it difficult to continue the trend. Towards this, ICPS solution has emerged as a cost-effective and flexible solution that offers an alternative path beyond the device scaling. In this chapter, we start with an overview of various design considerations and design space exploration for ICPS, and then zoom into two specific topics of

importance: decoupling capacitor insertion for noise suppression, and three-dimensional integrated systems. With the limited space, however, it is not possible to touch every aspect of them. We hope that our effort will provide at least some insightful thoughts to motivate and inspire future works and promote the ICPS technologies.

Acknowledgments Some works in this chapter are supported in part by the University of Missouri Research Board (UMRB). The authors would also like to express their great gratitude to the editors for their hard work, and to the anonymous reviewers for their constructive comments, which led to considerable improvements in the technical content.

References

1. Zhang G, Roosmalen A (2009) More than Moore—creating high value micro/nanoelectronics systems. Springer, New York
2. Palmer R, Poulton J, Dally WJ, Eyles J, Fuller AM, Greer T, Horowitz M, Kellam M, Quan F, Zarkeshvari F (2007) A 14 mW 6.25 Gb/s transceiver in 90 nm CMOS for serial chip-to-chip communications. IEEE international solid-state circuits conference, 2007 (ISSCC 2007), Digest of technical papers, San Francisco, CA, 11–15 Feb 2007, pp 440–614
3. Dally WJ, Poulton J (1998) Digital systems engineering. Cambridge University Press, Cambridge
4. Lee M-JE et al (2000) Low-power area-efficient high-speed I/O circuit techniques. IEEE J Solid-State Circ 35:1591–1599
5. Hatamkhani H et al (2006) Power-centric design of high-speed I/Os. In: DAC, pp 867–872
6. Balamurugan G et al (2008) A scalable 5–15 Gbps, 14–75 mW low-power I/O transceiver in 65 nm CMOS. IEEE J Solid-State Circ 43:1010–1019
7. Horowitz M et al (1998) High-speed electrical signaling: overview and limitations. IEEE Micro 18:12–24
8. Dabral S, Maloney T (1998) Basic ESD and I/O design. Wiley, New York
9. Lai M-F, Chen H-M (2008) An implementation of performance-driven block and I/O placement for chip-package codesign. 9th International symposium on quality electronic design, 2008 (ISQED 2008), San Jose, CA, 17–19 Mar 2008, pp 604–607
10. Mak WK (2004) I/O placement for FPGAs with multiple I/O standards. IEEE Trans Comput Aided Des Integrated Circ Syst 23:315–320
11. Caldwell A, Kahng AB, Mantik S, Markov IL (1998) Implications of area-array I/O for row-based placement methodology. Proceedings 1998 IEEE symposium on IC/package design integration, 1998, Santa Cruz, CA, 2–3 Feb 1998, pp 93–98
12. Kar J, Shukla R, Bhattacharyya BK (1999) Optimizing C4 bump placements for a peripheral I/O design. Proceedings of the 49th electronic components and technology conference, 1999, San Diego, CA, pp 250–254
13. Hsieh H-Y, Wang T-C (2005) Simple yet effective algorithms for block and I/O buffer placement in flip-chip design. IEEE international symposium on circuits and systems, 2005 (ISCAS 2005), vol 2, Kobe, Japan, 23–26 May 2005, pp 1879–1882
14. Yasar G, Chiu C, Proctor RA, Libous JP (2001) I/O cell placement and electrical checking methodology for ASICs with peripheral I/Os. International symposium on quality electronic design, 2001, San Jose, CA, pp 71–75
15. Lai M-F, Chen H-M (2008) An implementation of performance-driven block and I/O placement for chip-package codesign. 9th International symposium on quality electronic design, 2008 (ISQED 2008), San Jose, CA, 17–19 Mar 2008, pp 604–607

16. Chrzanowska-Jeske M, Her S-K (1992) Improved I/O pad assignment for sea-of-gates placement algorithm. Proceedings of the 35th midwest symposium on circuits and systems, 1992, vol 2, Washington, DC, 9–12 Aug 1992, pp 1396–1399
17. Kozhaya JN, Nassif SR, Najm FN (2001) I/O buffer placement methodology for ASICs. The 8th IEEE international conference on electronics, circuits and systems, 2001(ICECS 2001), vol 1, pp 245–248
18. Malta Buffet PH, Natonio J, Proctor RA, Sun, Yu H, Yasar G (2000) Methodology for I/O cell placement and checking in ASIC designs using area-array power grid. Proceedings of the IEEE 2000 custom integrated circuits conference (CICC) 2000, Orlando, FL, pp 125–128
19. Mechaik MM (2000) Effects of package stackups on microprocessor performance. ISQED 2000:475–481
20. Darnauer J et al (1999) Electrical evaluation of flip-chip package alternatives for next generation microprocessors. IEEE Trans Adv Packag 22:407–415
21. Libous JP (1998) Characterization of flip-chip CMOS ASIC simultaneous switching noise on multilayer organic and ceramic BGA/CGA packages. IEEE 7th topical meeting on electrical performance of electronic packaging, 1998, West Point, NY, 26–28 Oct 1998, pp 191–194
22. Audet J, O'Connor DP, Grinberg M, Libous JP (2004) Effect of organic package core via pitch reduction on power distribution performance. Proceedings of the 54th electronic components and technology conference, 2004, vol 2, Las Vegas, NV, 1–4 June 2004, pp 1449–1453
23. O'Connor DP, Hamel H, Spring C, Audet J (2003) Electrical modeling and characterization of packaging solutions utilizing lead-free second level interconnects. Proceedings of the 53rd electronic components and technology conference, 2003, New Orleans, LA, 27–30 May 2003, pp 1270–1276
24. Libous JP, O'Connor DP (1997) Measurement, modeling, and simulation of flip-chip CMOS ASIC simultaneous switching noise on a multilayer ceramic BGA. IEEE transactions on components, packaging, and manufacturing technology, part B: advanced packaging, vol 20, no 3, Aug 1997, pp 266–271
25. Garben B, Huber A, Kaller D, Klink E, Grivet-Talocia S (2002) Novel organic chip packaging technology and impacts on high speed interfaces. Electrical performance of electronic packaging, 2002, Monterey, CA, 21–23 Oct 2002, pp 231–234
26. Goto Y, Hosomi E, Harvey PM, Kawasaki K, Noma H, Mori H, Miura M, Takiguchi I, Audet J, Mandrekar R, Nishio T (2006) Electrical design optimization and characterization in cell broadband engine package. Proceedings of the 56th electronic components and technology conference, 2006, San Diego, CA, p 9
27. Harvey P, Zhou Y, Yamada G, Questad D, Lafontant G, Mandrekar R, Suminaga S, Yamaji Y, Noma H, Nishio T, Mori H, Tamura T, Yazawa K, Takiguchi, Ohde T, White R, Malhotra A, Audet J, Wakil J, Sauter W, Hosomi E (2007) Chip/package design and technology trade-offs in the 65 nm cell broadband engine. Proceedings of the 57th electronic components and technology conference, 2007 (ECTC 2007), Reno, NV, 29 May–1 June 2007, pp 27–34
28. Suryakumar M, Cui W, Parmar P, Carlson C, Fishbein B, Sheth U, Morgan J (2004) Power delivery validation methodology and analysis for network processors. Proceedings of the 54th electronic components and technology conference, 2004, Las Vegas, NV, vol 1, 1–4 June 2004, pp 589–592
29. Zheng H, Krauter B, Pileggi L (2003) On-package decoupling optimization with package macromodels. Proceedings of the IEEE 2003 custom integrated circuits conference, 2003, San Jose, CA, 21–24 Sept 2003, pp 723–726
30. Zhao S et al (2002) Power supply noise aware floorplanning and decoupling capacitance placement. In: ASPDAC, pp 489–494
31. Su H et al (2003) Optimal decoupling capacitor sizing and placement for standard-cell layout designs. IEEE Trans Comput Aided Des Integrated Circ Syst 22:428–436
32. Pant MD et al (2002) On-chip decoupling capacitor optimization using architectural level prediction. IEEE Trans Very Large Scale Integr Syst 10:319–326

33. Chen HH, Neely JS, Wang MF, Co G (2003) On-chip decoupling capacitor optimization for noise and leakage reduction. Proceedings of the 16th symposium on integrated circuits and systems design, 2003 (SBCCI 2003), San Paulo, Brazil, 8–11 Sept 2003, pp 251–255
34. Chen HH, Schuster SE (1995) On-chip decoupling capacitor optimization for high-performance VLSI design. International symposium on VLSI technology, systems, and applications, 1995. Proceedings of technical papers, Taipei, Taiwan, 31 May–2 June 1995, pp 99–103
35. Fu J, Luo Z, Hong X, Cai Y, Tan SX-D, Pan Z (2004) A fast decoupling capacitor budgeting algorithm for robust on-chip power delivery. Proceedings of the Asia and South Pacific design automation conference, 2004 (ASP-DAC 2004), Yokohama, Japan, 27–30 Jan 2004, pp 505–510
36. Chen Y, Chen Z, Fang J (1996) Optimum placement of decoupling capacitors on packages and printed circuit boards under the guidance of electromagnetic field simulation. Proceedings of the 46th electronic components and technology conference, 1996, Orlando, FL, 28–31 May 1996, pp 756–760
37. Yang X, Chen Q-L, Chen C-t (2002) The optimal value selection of decoupling capacitors based on FDFD combined with optimization. Electrical performance of electronic packaging, 2002, Monterey, CA, 21–23 Oct 2002, pp 191–194
38. Kamo A, Watanabe T, Asai H (2000) An optimization method for placement of decoupling capacitors on printed circuit board. IEEE conference on electrical performance of electronic packaging, 2000, Scottsdale, AZ, pp 73–76
39. Hattori I, Kamo A, Watanabe T, Asai H (2002) A searching method for optimal locations of decoupling capacitors based on electromagnetic field analysis by FDTD method. Electrical performance of electronic packaging, Monterey, CA, 21–23 Oct 2002, pp 159–162
40. Zhao J, Mandhana OP (2004) A fast evaluation of power delivery system input impedance of printed circuit boards with decoupling capacitors. IEEE 13th topical meeting on electrical performance of electronic packaging, Portland, OR, 25–27 Oct 2004, pp 111–114
41. Chen J, He L (2006) Noise driven in-package decoupling capacitor optimization for power integrity. Proceedings of the 2006 international symposium on physical design (ISPD '06), San Jose, CA, pp 94–101
42. Yu H, Shi Y, He L, Smart D (2006) A fast block structure preserving model order reduction for inverse inductance circuits. IEEE/ACM international conference on computer-aided design, 2006 (ICCAD '06), San Jose, CA, 5–9 Nov 2006, pp 7–12
43. Yu H, Shi Y, He L (2006) Fast analysis of structured power grid by triangularization based structure preserving model order reduction. 43rd ACM/IEEE design automation conference, 2006, San Francisco, CA, pp 205–210
44. http://www-device.eecs.berkeley.edu/ptm/
45. http://www.eda.org/pub/ibis/
46. Ruehli AE (1974) Equivalent circuit models for three-dimensional multiconductor systems. IEEE Trans Microw Theor Tech 22:216–221
47. Devgan A, Ji H, Dai W (2000) How to efficiently capture on-chip inductance effects: introducing a new circuit element K. IEEE/ACM international conference on computer aided design, 2000 (ICCAD-2000), San Jose, CA, pp 150–155
48. Li X, Li P, Pileggi LT (2005) Parameterized interconnect order reduction with explicit-and-implicit multiparameter moment matching for inter/intra-die variations. IEEE/ACM international conference on computer-aided design, 2005 (ICCAD-2005), San Jose, CA, 6–10 Nov 2005, pp 806–812
49. Feldmann P, Liu F (2004) Sparse and efficient reduced order modeling of linear subcircuits with large number of terminals. IEEE/ACM international conference on computer aided design, 2004 (ICCAD-2004), San Jose, CA, 7–11 Nov 2004, pp 88–92
50. Li P, Shi W (2006) Model order reduction of linear networks with massive ports via frequency dependent port packing. 43rd ACM/IEEE design automation conference, San Francisco, CA, 24–28 July 2006, pp 267–272

51. Liu P, Tan SX-D, Li H, Qi Z, Kong J, McGaughy B, He L (2005) An efficient method for terminal reduction of interconnect circuits considering delay variations. IEEE/ACM international conference on computer-aided design, 2005 (ICCAD-2005), San Jose, 6–10 Nov 2005, pp 821–826
52. Ding C (2005) Spectral clustering, principal component analysis and matrix factorizations for learning. In: Conference on machine learning (Tutorial)
53. Karypis G et al (1999) Multilevel hypergraph partitioning: application in VLSI domain. IEEE Trans Very Large Scale Integr Syst 7:69–79
54. Franzon PD, Rhett Davis W, Thorolffson T (2010) Creating 3D specific systems: architecture, design and CAD. Proceedings of the conference on design, automation and test in Europe (DATE ’10), European design and automation association, 3001 Leuven, Belgium, 8–12 Mar 2010, pp 1684–1688
55. Beyne E, Swinnen B (2007) 3D system integration technologies. IEEE international conference on integrated circuit design and technology, 2007 (ICICDT ’07), Austin, TX, 30 May–1 June 2007, pp 1–3
56. Peter R, Armin K (2008) Through-silicon via technologies for extreme miniaturized 3D integrated wireless sensor systems (e-CUBES). International interconnect technology conference, 2008 (IITC 2008), Burlingame, CA, 1–4 June 2008, pp 7–9
57. Rao VS, Ho SW, Vincent L, Yu LH, Ebin L, Nagarajan R, Chong CT, Zhang X, Damaruganath P (2009) TSV interposer fabrication for 3D IC packaging. 11th Electronics packaging technology conference, 2009 (EPTC ’09), Singapore, 9–11 Dec 2009, pp 431–437
58. Kang U, Chung H-J, Heo S, Ahn S-H, Lee H, Cha S-H, Ahn J, Kwon D, Kim JH, Lee J-W, Joo H-S, Kim W-S, Kim H-K, Lee E-M, Kim S-R, Ma K-H, Jang D-H, Kim N-S, Choi M-S, Oh S-J, Lee J-B, Jung T-K, Yoo J-H, Kim C (2009) 8 Gb 3D DDR3 DRAM using through-silicon-via technology. IEEE international solid-state circuits conference, 2009 (ISSCC 2009), Digest of technical papers, San Francisco, CA, 8–12 Feb 2009, pp 130–131, 131a
59. AVIZATechnology (2011) 3D-IC system. http://www.aviza.com/products/ic.html
60. Iker F, Tezcan DS, Teixeira RC, Soussan P, De Moor P, Beyne E, Baert K (2007) 3D embedding and interconnection of ultra thin (<20 μm) silicon dies. 9th Electronics packaging technology conference, 2007 (EPTC 2007), Singapore, 10–12 Dec 2007, pp 222–226
61. Tezzaron (2011) http://www.tezzaron.com/technology/FaStack.htm
62. Johnson SC (2009) Via first, middle, last, or after? 3D Packaging 1–6.
63. Maeda N, Kitada H, Fujimoto K, Suzuki K, Nakamura T, Kawai A, Arai K, Ohba T (2010) Wafer-onwafer (WOW) stacking with damascene-contact TSV for 3D integration. International symposium on VLSI technology systems and applications (VLSI-TSA), Hsin-Chu, Taiwan, 26–28 Apr 2010, pp 158–159
64. Lau JH (2010) TSV manufacturing yield and hidden costs for 3D IC integration. Proceedings of the 60th electronic components and technology conference (ECTC), Las Vegas, NV, 1–4 June 2010, pp 1031–1042
65. Zhao J, Zhang J, Fang J (1998) Effects of power/ground via distribution on the power/ground performance of C4/BGA packages. IEEE 7th topical meeting on electrical performance of electronic packaging, West Point, NY, 26–28 Oct 1998, pp 177–180
66. Hong Y-S, Lee H, Choi J-H, Yoo M-H, Kong J-T (2005) Analysis for complex power distribution networks considering densely populated vias. Sixth international symposium on quality of electronic design, 2005 (ISQED 2005), San Jose, CA, 21–23 Mar 2005, pp 208–212
67. Ryu C, Chung D, Lee J, Lee K, Oh T, Kim J (2005) High frequency electrical circuit model of chip-to-chip vertical via interconnection for 3-D chip stacking package. IEEE 14th topical meeting on electrical performance of electronic packaging, Austin, TX, 24–26 Oct 2005, pp 151–154
68. Beyne E (2004) 3D interconnection and packaging: impending reality or still a dream? IEEE international solid-state circuits conference, 2004 (ISSCC 2004), Digest of technical papers, vol 1, San Francisco, CA, 15–19 Feb 2004, pp 138–139

69. Banerjee K et al (2001) 3D ICs: a novel chip design for improving deep submicron interconnect performance and systems-on-chip integration. In: Proc. IEEE, vol 6, pp 602–633
70. Teng CC et al (1997) iTEM: a temperature-dependent electromigration reliability diagnosis tool. IEEE Trans Comput Aided Des Integrated Circ Syst 16(8):882–893
71. Wang TY, Chen CCP (2003) Thermal-ADI: a linear-time chip-level dynamic thermal simulation algorithm based on alternating-direction-implicit (ADI) method. IEEE Trans Very Large Scale Integr Syst 11(4):691–700
72. Li P et al (2004) Efficient full-chip thermal modeling and analysis. In: ICCAD, 319–326
73. Zhan Y, Sapatnekar SS (2007) High efficiency green function-based thermal simulation algorithms. IEEE Trans Comput Aided Des Integr Circ Syst 6:1661–1675
74. Cong J et al (2011) Thermal-aware cell and through-silicon-via co-placement for 3D ICs. Design Automation Conference (DAC), 2011 48th ACM/EDAC/IEEE, 5–9 June 2011, pp 670–675
75. Chiang TY et al (2001) Compact modeling and SPICE-based simulation for electrothermal analysis of multilevel ULSI interconnects. In: ICCAD, pp 165–172
76. Liao W et al (2005) Temperature and supply voltage aware performance and power modeling at microarchitecture level. IEEE Trans Comput Aided Des Integr Circ Syst 24(7):1042–1053
77. Tiwari V, Singh D, Rajgopal S, Mehta G, Patel R, Baez F (1998) Reducing power in high-performance microprocessors. Proceedings of the design automation conference, 1998, San Francisco, CA, 19 June 1998, pp 732–737
78. Yu H, Shi Y, He L, Karnik T (2006) Thermal via allocation for 3D ICs considering temporally and spatially variant thermal power. Proceedings of the 2006 international symposium on low power electronics and design, 2006 (ISLPED'06), Tegernsee, 4–6 Oct 2006, pp 156–161
79. Jiang L, Huang L, Xu Q (2009) Test architecture design and optimization for three-dimensional SoCs. Design, automation and test in Europe conference and exhibition, 2009 (DATE '09), Nice, 20–24 Apr 2009, pp 220–225
80. Zhao X, Lewis DL, Lee H-HS, Lim SK (2009) Pre-bond testable low-power clock tree design for 3D stacked ICs. IEEE/ACM international conference on computer-aided design, 2009 (ICCAD 2009), Digest of technical papers, San Jose, CA, 2–5 Nov 2009, pp 184–190
81. Lewis DL, Lee HHS (2007) A scanisland based design enabling prebond testability in die-stacked microprocessors. In: ITC, pp 1–8
82. Zhao X, Lewis DL, Lee H-HS, Lim SK (2009) Pre-bond testable low-power clock tree design for 3D stacked ICs. IEEE/ACM international conference on computer-aided design, 2009 (ICCAD 2009), Digest of technical papers, San Jose, CA, 2–5 Nov 2009, pp 184–190
83. Lewis DL, Lee H-HS (2009) Testing circuit-partitioned 3D IC designs. IEEE computer society annual symposium on VLSI, 2009 (ISVLSI '09), Tampa, FL, 13–15 May 2009, pp 139–144
84. Wu X, Falkenstern P, Xie Y (2007) Scan chain design for three-dimensional integrated circuits (3D ICs). 25th International conference on computer design, 2007 (ICCD 2007), Lake Tahoe, CA, 7–10 Oct 2007, pp 208–214
85. Wu X, Chen Y, Chakrabarty K, Xie Y (2008) Test-access mechanism optimization for core-based three-dimensional SOCs. IEEE international conference on computer design, 2008 (ICCD 2008), Lake Tahoe, CA, 12–15 Oct 2008, pp 212–218
86. Loi I, Mitra S, Lee TH, Fujita S, Benini L (2008) A low-overhead fault tolerance scheme for TSV-based 3D network on chip links. IEEE/ACM international conference on computer-aided design, 2008 (ICCAD 2008), San Jose, CA, 10–13 Nov 2008, pp 598–602
87. Laisne M et al (2010) System and methods utilizing redundancy in semiconductor chip interconnects. United States Patent Application 20100060310
88. Yu H, Ho J, He L (2006) Simultaneous power and thermal integrity driven via stapling in 3D ICs. IEEE/ACM international conference on computer-aided design, 2006 (ICCAD '06), San Jose, CA, 5–9 Nov 2006, pp 802–808

89. Grimme EJ (1997) Krylov projection methods for model reduction. PhD Thesis. University of Illinois at Urbana-Champaign
90. Odabasioglu A et al (1998) PRIMA: passive reduced-order interconnect macromodeling algorithm. IEEE Trans Comput Aided Des Integr Circ Syst 17(8):645–654
91. Astrid P et al (2008) Missing point estimation in models described by proper orthogonal decomposition. IEEE Trans Automat Contr 53(10):2237–2251
92. Yu H et al (2009) Allocating power ground vias in 3D ICs for simultaneous power and thermal integrity. ACM Trans Des Autom Electron Syst
93. Yu H et al (2008) Thermal via allocation for 3-D ICs considering spatially and temporally variant thermal power. IEEE Trans Very Large Scale Integr Syst 16(12):1609–1619
94. Yu H, Chu C, He L (2007) Off-chip decoupling capacitor allocation for chip package co-design. 44th ACM/IEEE design automation conference, 2007 (DAC '07), San Diego, CA, 4–8 June 2007, pp 618–621
95. Yu H, Shi Y, He L (2006) Fast analysis of structured power grid by triangularization based structure preserving model order reduction. 43rd ACM/IEEE design automation conference, San Francisco, CA, 24–28 July 2006, pp 205–210
96. Yu H, Hu Y, Liu C, He L (2007) Minimal skew clock embedding considering time variant temperature gradient. Proceedings of the 2007 international symposium on physical design (ISPD '07), Austin, TX, pp 173–180
97. Goplen B, Sapatnekar S (2005) Thermal via placement in 3D ICs. Proceedings of the 2005 international symposium on physical design (ISPD '05), San Francisco, CA, pp 167–174
98. Cong J, Wei J, Zhang Y (2004) A thermal-driven floorplanning algorithm for 3D ICs. IEEE/ACM international conference on computer aided design, 2004 (ICCAD-2004), San Jose, CA, 7–11 Nov 2004, pp 306–313
99. Lung C-L et al (2011) Fault-tolerant 3D clock network. Design Automation Conference (DAC), 48th ACM/EDAC/IEEE, 5–9 Jun 2011, pp 645–651
100. Edahiro M (1994) An efficient zero-skew routing algorithm. 31st Conference on design automation, 1994, San Diego, CA, 6–10 June 1994, pp 375–380
101. Albrecht C et al (2003) On the skew-bounded minimum-buffer routing tree problem. IEEE Trans Comput Aided Des Integr Circ Syst 22:937–945
102. Shiyan H et al (2007) Fast algorithms for slew-constrained minimum cost buffering. IEEE Trans Comput Aided Des Integr Circ Syst 26:2009–2022
103. Cong J et al (1993) Matching-based methods for high-performance clock routing. IEEE Trans Comput Aided Des Integr Circ Syst 12:1157–1169
104. Wang H et al (2011) The effects of substrate doping density on the electrical performance of through-silicon-via. In: Proc. of Asia-Pacific EMC Symposium, 2011

Chapter 9
Thermal Management of Flip Chip Packages

Richard C. Chu, Robert E. Simons, Madhusudan Iyengar, and Lian-Tuu Yeh

Abstract Generally speaking, the electrical energy that is supplied to electronic devices is ultimately transformed into and dissipated as heat. This generation of heat is accompanied by a temperature rise at the heat source followed by the transport of heat to regions of lower temperature within and outside the electronics module or package. Within the package transport of heat occurs via a process of thermal conduction in the solid material making up the package. As the heat reaches the external surfaces of the package it is usually transferred to a cooling fluid (e.g., air) via a thermal convection process. In the case of lower power components thermal radiation may also play a role in transferring heat to the surrounding environment. The temperatures within the electronics package will continue to rise until the rate of heat removal from the package is equal to the rate of heat generation. It is worthwhile to note that, even if purposeful active measures were not taken to cool the package, the laws of nature or physics would prevail and limit the temperature rise. However, in most instances, the resulting temperatures would be too high. As shown in Fig. 9.1, based upon the results of a study conducted under a US Air Force Avionics Integrity Program, temperature was identified as a causal factor in 55 % of electronic failures [1]. It might be noted that in most commercial

R.C. Chu (✉)
IBM Corporation, Hopewell Junction, NY USA
e-mail: churc@us.ibm.com

R.E. Simons
IBM Corporation, Poughkeepsie, NY USA
e-mail: resimons@us.ibm.com

M. Iyengar
Thermal Engineer, Storage Hardware Systems, Facebook, Bldg. 17-3,
1601 Willow Road, Menlo Park, CA 94025, USA
e-mail: madhusudan@fb.com

L.-T. Yeh
Thermal Consultant, Dallas, TX USA
e-mail: jjyeh2@aol.com

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_9, © Springer Science+Business Media New York 2013

applications, electronic packages are not subjected to nearly as severe an environment in terms of vibration, dust or humidity as military avionics, so the percentage of failures caused by temperature are likely to occupy a larger "piece of the pie." In addition to the effect of temperature on electronic device reliability, it can also play an important role on CMOS circuit performance. Consequently, it is necessary to provide satisfactory cooling for electronic packages by design and not by accident.

9.1 Introduction

Generally speaking, the electrical energy that is supplied to electronic devices is ultimately transformed into and dissipated as heat. This generation of heat is accompanied by a temperature rise at the heat source followed by the transport of heat to regions of lower temperature within and outside the electronics module or package. Within the package transport of heat occurs via a process of thermal conduction in the solid material making up the package. As the heat reaches the external surfaces of the package it is usually transferred to a cooling fluid (e.g., air) via a thermal convection process. In the case of lower power components thermal radiation may also play a role in transferring heat to the surrounding environment. The temperatures within the electronics package will continue to rise until the rate of heat removal from the package is equal to the rate of heat generation. It is worthwhile to note that, even if purposeful active measures were not taken to cool the package, the laws of nature or physics would prevail and limit the temperature rise. However, in most instances, the resulting temperatures would be too high. As shown in Fig. 9.1, based upon the results of a study conducted under a US Air Force Avionics Integrity Program, temperature was identified as a causal factor in 55 % of electronic failures [1]. It might be noted that in most commercial applications, electronic packages are not subjected to nearly as severe an environment in terms of vibration, dust or humidity as military avionics, so the percentage of failures caused by temperature are likely to occupy a larger "piece of the pie." In addition to the effect of temperature on electronic device reliability, it can also play an important role on CMOS circuit performance. Consequently, it is necessary to provide satisfactory cooling for electronic packages by design and not by accident.

Fig. 9.1 Major causes of electronic failures in U.S.A.F. Avionics Integrity Program Study [1]

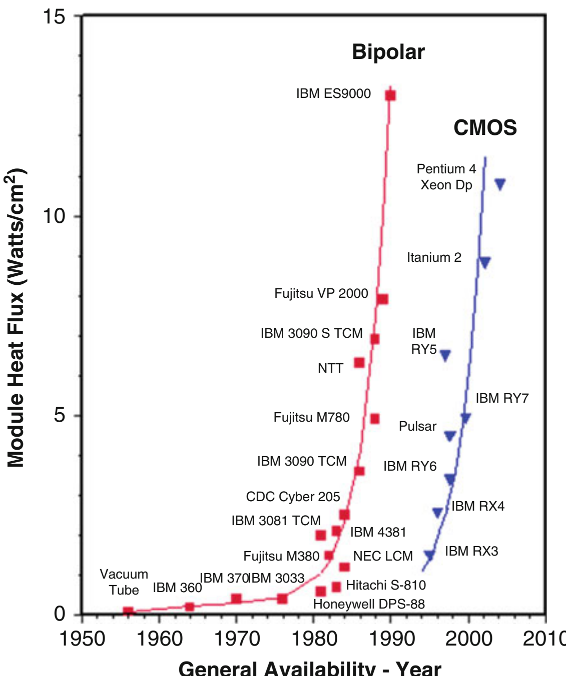


Fig. 9.2 Increasing module heat flux trend [2]

There has been a continuing trend for many years towards increased circuit packaging density. Increased packaging density has been accompanied by increased power dissipation per circuit to provide reductions in circuit delay (i.e., increased speed). This has led to increased heat flux levels at both the chip and module package level. An example of the historical trend in module level heat flux, for both the earlier Bipolar circuit technologies and the now widely used CMOS technologies, is shown in Fig. 9.2. There was a brief respite with the transition from Bipolar to CMOS circuit technologies in the 1990s. However, the trend towards increased heat flux reasserted itself with the demand for greater and greater performance from CMOS circuit packages and continues today.

9.2 Heat Transfer Fundamentals

There are three fundamental mechanisms by which heat is transported from a heat generating device to an external environment. These mechanisms are thermal conduction, convection and radiation and are covered in depth in several heat transfer textbooks [3–5]. In this discussion, we briefly introduce them, along with the recommendation to the interested reader to find more details in the references. The basic heat transfer mechanisms of conduction and convection are illustrated in Fig. 9.3 along with their defining equations.

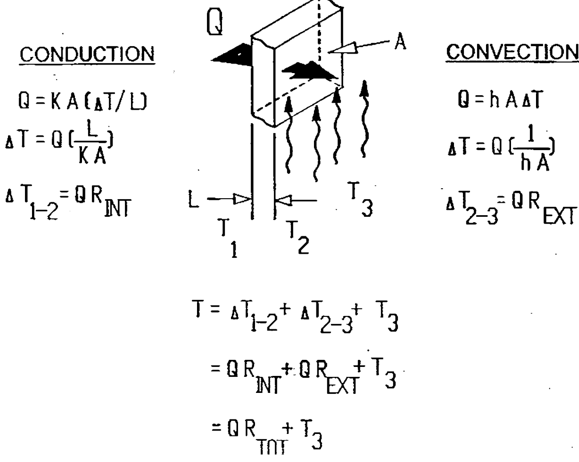


Fig. 9.3 Basic heat transfer mechanisms and defining equations

Thermal conduction may be described as the process by which thermal energy is transferred from a region of higher temperature to one of lower temperature by a progressive exchange of energy between the molecules of a substance. The rate of heat flow is given by the Fourier equation (a simple form of which is shown on the left side in Fig. 9.3), where Q is the heat flow rate, k is the thermal conductivity of the substance, A is the cross-sectional area, ΔT is the temperature difference across the substance, and L is the thickness or distance the heat travels through the substance. As shown the Fourier equation may be rearranged to solve for ΔT across the substance. When this is done, the $L/(KA)$ term is often expressed as an internal thermal resistance, usually in units of °C per watt. Once the heat reaches an external surface it may be removed by convection or radiation or a combination of the two.

Convection may be described as the process by which heat is transferred from a surface to a gas or liquid moving over the surface by the combined actions of thermal conduction, energy storage and mixing motion. The fluid motion may be caused by density variations in the fluid due to temperature. This type of convective heat transfer is termed natural or free convection. Alternatively, coolant may be induced to flow over a heated surfaced by some mechanical means such as a fan or blower. Convective heat transfer in this situation is termed forced convection. The quantity of heat transferred by either process is given by the Newton rate equation (shown on the right side in Fig. 9.3), where Q is the heat flow rate, h is the heat transfer coefficient, A is the area of the surface transferring heat, and ΔT is the temperature difference from the surface to the cooling fluid. As was done for the Fourier equation, the Newton rate equation may be rearranged to solve for ΔT from the surface to the cooling fluid. When this is done, the $1/(hA)$ term is often expressed as an external thermal resistance, usually in units of °C per watt. As illustrated at the bottom of Fig. 9.3, conductive and convective thermal resistance may in the simplest form, be combined to determine the total temperature difference from the heat source to the cooling medium and thereby determine the source temperature for give heat load and sink (i.e., cooling medium) temperature. This elemental concept has been employed extensively in creating complex thermal network models of electronic packages.

Forced convection heat transfer usually provides a substantially greater heat transfer rate for a given temperature difference than does natural convection. As discussed above, natural convection heat transfer is driven by the temperature difference between the heated surface and the adjacent cooling air. Consequently the heat transfer coefficient is proportional to the temperature difference between the heated surface and the cooling air to some power, typically 0.25. Typically, the surface heat flux, *Q/A*, that can be accommodated with forced convection heat transfer is a factor of 5 or more greater than with natural convection heat transfer. Forced convection heat transfer is greatly influenced by the forced flow over the surface effectively moving the heated coolant away at a much greater rate than is possible with natural convection heat transfer. Forced convection may be divided into two flow regimes: laminar and turbulent. Laminar flow may also be termed streamline flow and is characterized by the movement of fluid in parallel layers with no disruption between the layers. Turbulent flow is much more complex, with chaotic mixing and eddy currents moving in irregular patterns constantly fluctuating in time, while the overall flow is in one direction. This type of flow produces higher heat transfer coefficients than laminar flow. The heat transfer coefficient in turbulent flow, as well as developing laminar flow, will be proportional to the average velocity in the flow channel to some power, typically 0.5–0.8.

Thermal radiation is the process by which heat is transferred by means of electromagnetic waves, without any intervening medium playing an essential role in the process. The rate of heat transfer between a warm body at temperature T_1 and a cooler body at temperature T_2 is given by,

$$Q = \sigma A_1 F_\varepsilon F_s ({T_1}^4 - {T_2}^4)$$

where σ = Stefan–Boltzman constant (5.67×10^{-8} W/m^2 K^4); A_1 = radiating area of surface 1 (m^2); F_ε = emissivity factor taking into account the emissivity and reflectivity of both surfaces; F_s = shape factor taking into account the relative positions of the two surfaces, and the effect of shape and size on the fraction of energy received by the two surfaces.

As may be seen in the above equation, the driving potential for the net transfer of heat between the two surfaces is the difference between the source and sink temperatures raised to the fourth power. It should be noted that unlike in the Newton rate equation, in the radiation transfer equation absolute temperatures (i.e., degrees Kelvin) must be used. It can also be seen that the heat transfer rate is affected by both the surface emissivity factor, F_ε, and the shape or view factor, F_s. An electronic module or printed circuit card surface which "sees or has a good view of" other surfaces near room temperature can transfer as much heat by thermal radiation as by natural convection. On the other hand, a module or card surrounded by modules and cards that are nearly equally hot will transfer very little or no heat by radiation. Similarly, the relative cooling contribution from thermal radiation is nearly negligible for module and card surfaces cooled by forced convection. Detailed discussions of radiation heat transfer may be found in the literature [6, 7].

9.3 Electro-thermal Analog Model

As discussed later in this chapter detailed thermal analyses are generally performed for electronic packages using sophisticated and complex thermal conduction and computational fluid dynamic (CFD) computer codes. However, relatively simple “back of the envelope” calculation models based upon electro-thermal analogs can often provide initial insight to identify those parts of a package structure representing significant impediments to heat flow and allow the thermal engineer to assess possible package thermal improvements. Figure 9.4 provides a simple illustration of the electro-thermal analog concept.

As is well known from Ohm’s law, voltage drop, V, across an electrical resistor is equal to the current flow, I, multiplied by the electrical resistance, R_{Ω}. The heat flow, Q, across a thermal resistor is analogous to electrical current flow. Just as voltage difference is the driving potential for electric current flow, temperature difference, ΔT, across thermal resistance, R_{th}, is the driving potential for heat flow. The relationship between temperature difference, heat flow and thermal resistance has the same form as its electrical counterpart. By applying this analogy, the thermal paths in any electronic package may be represented by a network of thermal resistors and the chip and package temperatures may be predicted using the same techniques electrical engineers use to analyze electrical resistor networks.

A simple example of the application of the electro-thermal analog approach to determine the temperature of a flip chip mounted on a ceramic pin grid is illustrated in Fig. 9.5. The thermal resistance circuit is shown both superimposed on a cross-sectional view of the module package and alone at the bottom of the figure for clarity. This simple network captures all the major heat flow paths from the chip heat source

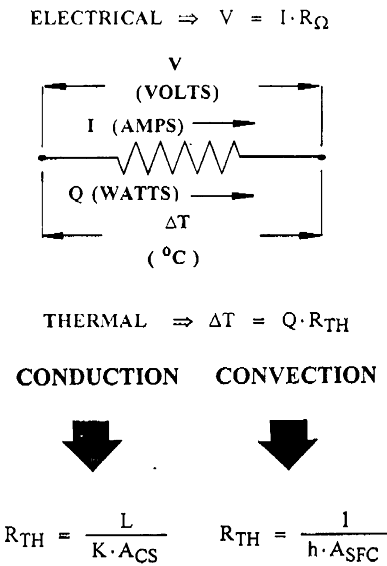


Fig. 9.4 Basic electro-thermal analog concept

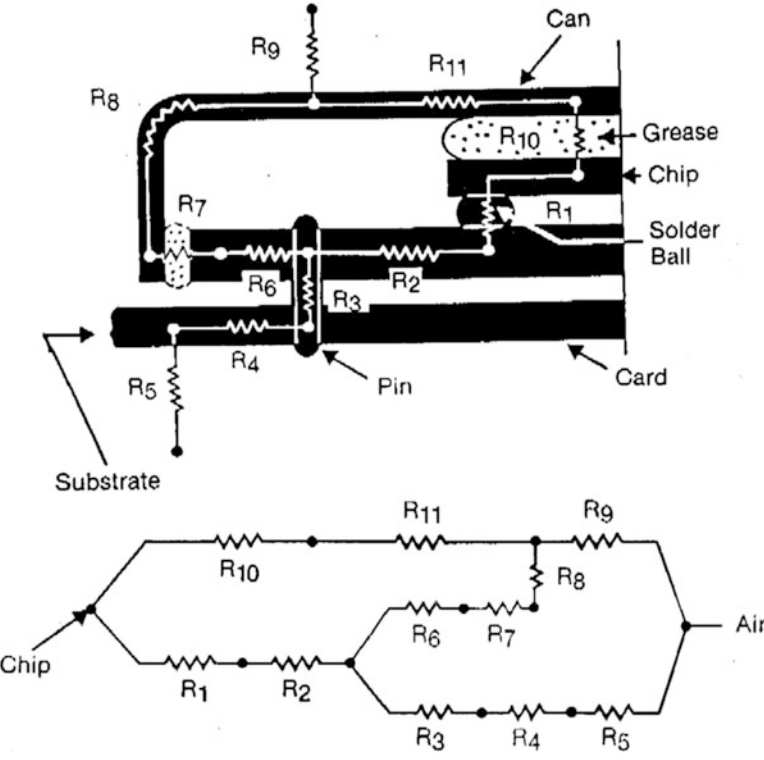


Fig. 9.5 Cross-sectional air-cooled flip chip module with corresponding electro-thermal analog

out to the cooling air. These paths include the thermal resistances across the solder pads (R1), across the substrate (R2 and R6), through the pins (R3), across the card (R4), across the epoxy seal (R7), in the cap (R8 and R11), and across the thermal grease layer from the back of the chip to the inner surface of the cap (R10). The magnitude of each of the thermal conduction resistances is calculated using the Fourier thermal conduction equation or a modified form of the equation. It should be noted that the thermal resistance across the pin connecting the substrate to the card actually represents the resistance across the entire pin array treated as a lumped resistance situated one-half the distance from the chip edge to the outer edge of the substrate. The resistances coming off the cap (R9) and the card (R5) are convective thermal resistances. These are obtained by estimating the convective heat transfer coefficient, h, and the effective area, A, and calculating the reciprocal of their product, $1/(hA)$. If there were no thermal grease in the module package, the thermal resistance between the chip and the inside of the cap would be so great that this path could be ignored. All that would be necessary then to obtain the overall thermal resistance from chip to cooling air, would be to combine the remaining resistances as a series–parallel network.

With the thermal grease path present, to reduce the network to a single overall thermal resistance is more complex and it is necessary to apply Kirchoff's Laws as would be required to analyze a topologically similar electrical resistance network.

As in an electrical network, in a thermal network the heat flow into a node must equal the flow out of the node. A detailed example of the application of this approach and the solution of such a network using Kirchoff's Laws to set up a set of simultaneous linear equations solvable by matrix inversion is given in [8].

9.4 Thermal Management Objectives

As noted in the introduction, circuit packaging density and the associated heat fluxes at the chip and module level have exhibited an increasing trend since the introduction of CMOS circuit technologies. In fact, the same was true for the Bipolar circuit technologies that preceded the introduction of CMOS. The task of supporting increased heat fluxes while maintaining relatively low device and package temperatures has been one of the challenges facing today's thermal engineers. In accomplishing this task a variety of thermal design objectives must be fulfilled which encompass more than simply removing heat from a flip chip package. These diverse thermal design objectives include the following:

- Maintain the temperatures of all components within specified functional limits.
- Ensure that all device/package temperatures satisfy reliability objectives.
- Ensure that the temperatures of any exposed surfaces are within acceptable safety limits.
- Provide a cooling system design that meets overall system availability and reliability objectives.
- Provide a cooling system design that is consistent with the customer heat rejection capability.
- Provide a cooling system that meets the overall system cost objectives.

As illustrated in Fig. 9.6, there are three options or thermal management strategies that might be considered to control or limit device or chip temperature [9]. One strategy would be to reduce the cooling sink temperature, which in most cases would mean reducing the cooling air temperature. While this can and has been done in some instances, in most applications it would be more costly, inefficient or impractical. Alternatively, it could be suggested that the engineers responsible for the electrical design reduce the power dissipation. While this would certainly be a worthwhile objective, it is usually viewed as performance limiting and therefore unacceptable. The third option is that which has been traditionally pursued and is to improve the relevant cooling technology. This option may be subdivided into three categories. The most commonly employed approaches are packaging related. Changes and enhancements either directly affect or require changes to the integrated circuit package. The changes made here serve to either reduce the internal or external thermal resistance of the package. Some examples of enhancements to reduce internal thermal resistance (R_{int}) include the following: (1) adding thermal grease to bridge an air gap within the package, (2) providing a thermal spreader within a package to improve thermal conduction transport of heat to exterior surfaces of the package, (3) changing

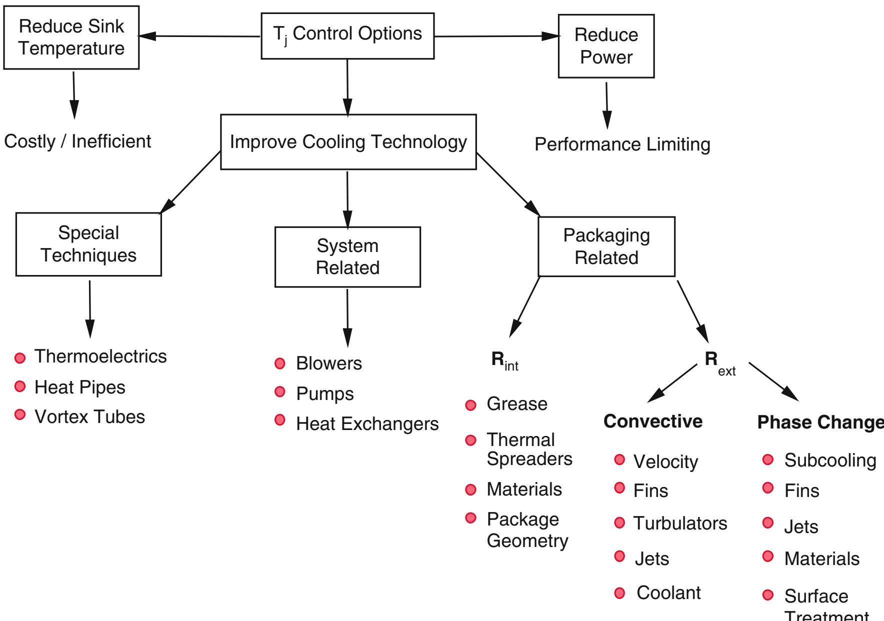


Fig. 9.6 Thermal management strategies and options

the substrate material to one with a higher thermal conductivity, and (4) changing package geometry to provide improved thermal paths. To reduce external thermal resistance (R_{ext}) either single phase convective or phase change cooling methods may be employed. The most widely used approach is natural or forced convection air-cooling. Examples of enhancements that have been applied to improve air-cooling include the following: (1) increasing air velocity to increase the heat transfer coefficient, (2) adding extended surfaces (e.g., fins) to increase the effective heat transfer surface area, (3) using impinging air jets to obtain higher heat transfer coefficients, and (4) applying turbulence promoters to increase heat transfer. The remaining approaches to improve cooling are to apply special cooling techniques or improve system level cooling. The application of special techniques incorporates such enhancement devices as thermoelectric coolers, heat pipes and vapor chambers, or vortex tubes. Heat pipes and vapor chambers have been applied in a number of applications with considerable success. Improvements in cooling at the system level often take the form of larger or more powerful fans or blowers to deliver air to air-cooled packages or larger pumps or more effective cold plates and heat exchangers in liquid cooled systems.

9.5 Thermal Management at the Die and Package Level

Figure 9.7 shows a schematic of a typical flip chip package on an organic package [10]. The different components of this package shown in Fig. 9.7 are the silicon chip attached via C4s (controlled collapse chip connections) on an organic laminate

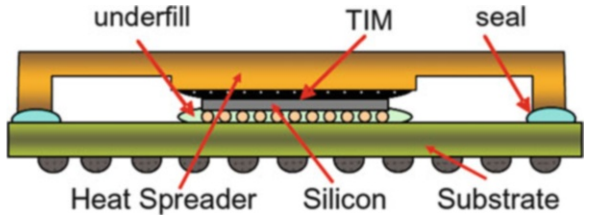


Fig. 9.7 Flip chip package on an organic substrate [10]

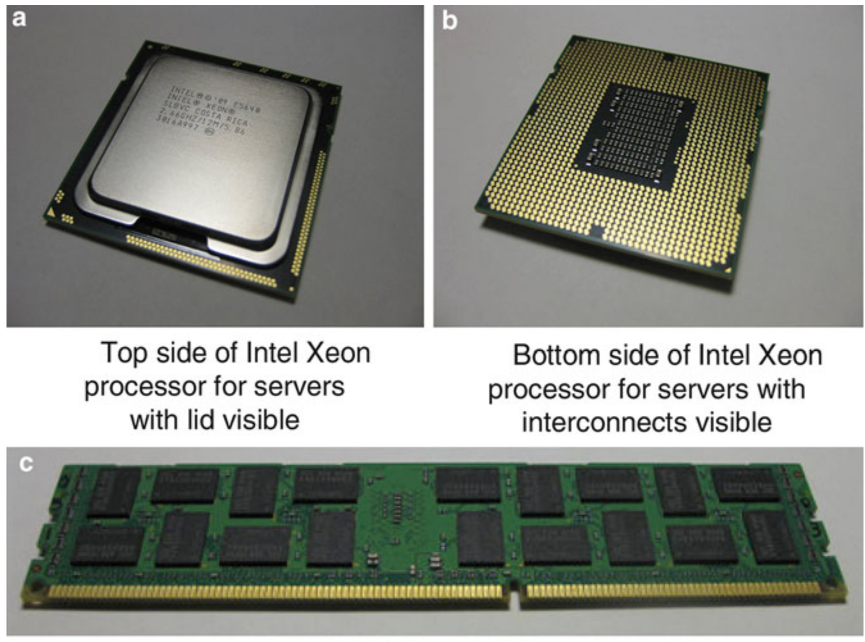


Fig. 9.8 Flip chip packages in commercial server products

substrate with an underfill material filling up the interstitial spaces between the C4s using a cure process. The back side of the chip is attached to a metal lid (heat spreader) using a thermal interface material (TIM), and the lid is attached at its ends to the substrate using a seal material using a second cure process. The chip is first joined to the substrate using the C4s and then the underfill material is inserted into the open spaces. Subsequent to underfill curing, the lid is attached both to the back side of the chip as well as to the substrate.

Figure 9.8 shows photographs of some commercially available flip chip packages that populate today's server products. While Fig. 9.8a, b shows images of Intel microprocessor modules, Fig. 9.8c displays a Samsung Dual-In-Line-Memory card with memory chips.

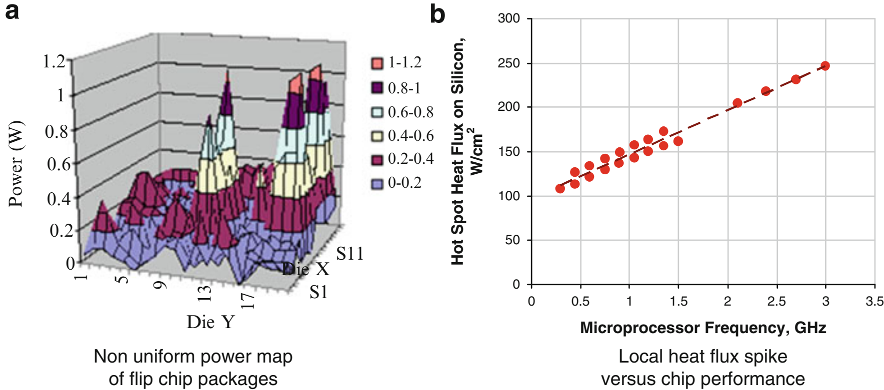


Fig. 9.9 Highly nonuniform power distribution in a flip chip package [11–13], (**a**) power distribution surface plot for 20 × 20 mm chip [11], (**b**) chip performance versus local heat flux [12, 13]

9.6 Hot Spots in Chip Dies

The power distribution on a flip chip package is a direct outcome of the electrical designs to achieve the desired performance and is influenced by the location and density of the various heat producing components. Figure 9.9a shows such a power distribution surface plot for a 20 × 20 mm chip and illustrates the nature of power spikes in some areas of the chip [11]. Figure 9.9b displays a plot of chip performance versus local hot spot heat flux [12] using data provided in [11]. As may be seen from Fig. 9.9b, to achieve higher chip performance using frequency as a metric, higher local heat fluxes have to be accommodated by the thermal design.

Such locally high heat fluxes as displayed in Fig. 9.9 can result in large local temperature increases that are often called hot spots. However, both the actual power maps and the realized temperature contours on the device side of the chips are not trivially determined. Hamann and coworkers [14] have developed a novel methodology using infra red (IR) measurements of temperature contours on the back side of the silicon die. A specially designed liquid flow duct facilitates the forced flow of the dielectric coolant over the back side of the die and the transparent window right over the chip back side allows the IR measurements to be carried out. The back side of the die is coated with a thin black body coating and the chip is thinned to 0.1 mm to reduce the impact of spreading in the chip. Figure 9.10a, b [15] show results of such measurements on an actual functioning microprocessor chip (PowerPC™970MP) for two different conditions, dual core operation and single core operation. As can be seen from Fig. 9.10a, b the hot spot on the chip moves by about 2 mm between the two cases and highlights the dependence of chip hot spots on workload as well as the importance of correct interpretation of thermal on chip sensor data. Figure 9.10c [16] displays a surface plot for a two core processor operation for 13.2 × 11.6 mm die with a frequency of 1.6 GHz and running a typical workload.

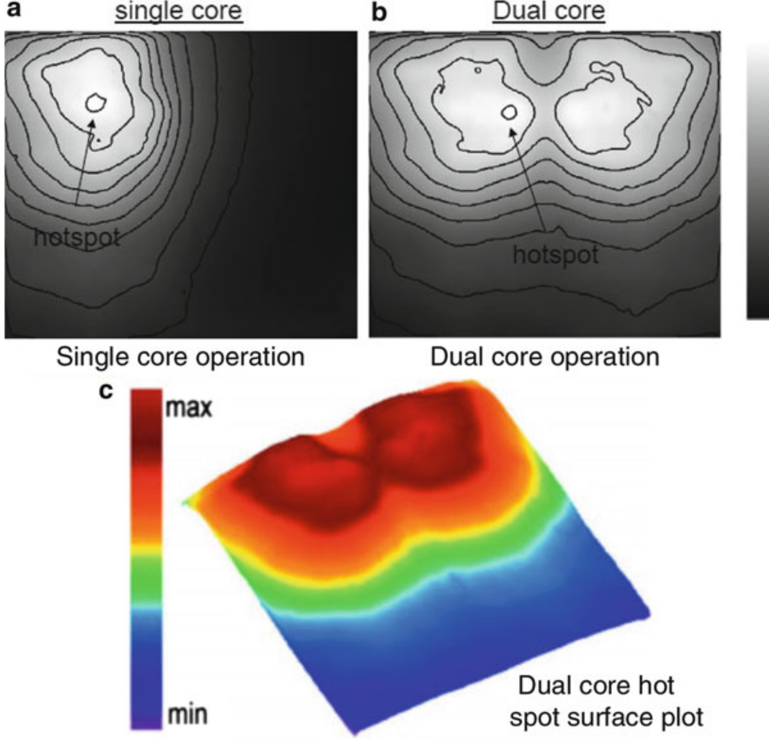


Fig. 9.10 Measured hot spot temperature contours for commercial flip chip microprocessors [15, 16] (**a**) single core operation, (**b**) dual core operation, (**c**) surface contours for dual core typical operation

Figure 9.11 shows transient temperature contours for a 1 cm^2 die from a desktop PC product during initial boot up from room temperature [14]. The average power density was reported to be 100 W/cm^2 with peak heat fluxes of higher than 300 W/cm^2. The temperature contours shown in Fig. 9.11 are for 10 s increments over the boot up process. As can be seen from Fig. 9.11, several different parts of the microprocessor get hot depending on the computational workload being performed by the chip. For the data displayed in Fig. 9.11, the die was also thinned to 0.1 mm to reduce heat spreading effects.

9.7 Analytical Modeling of Chip Hot Spots

Analytical models can be found in the literature to account for hot spots due to power maps. Torresola et al. [17] proposed a density factor to characterize the impact of nonuniform die power maps. Iyengar and Schmidt [13] provided a literature review of multiple discrete heat sources based models and extended some of them to analytically estimate the thermal resistance of a flip chip package

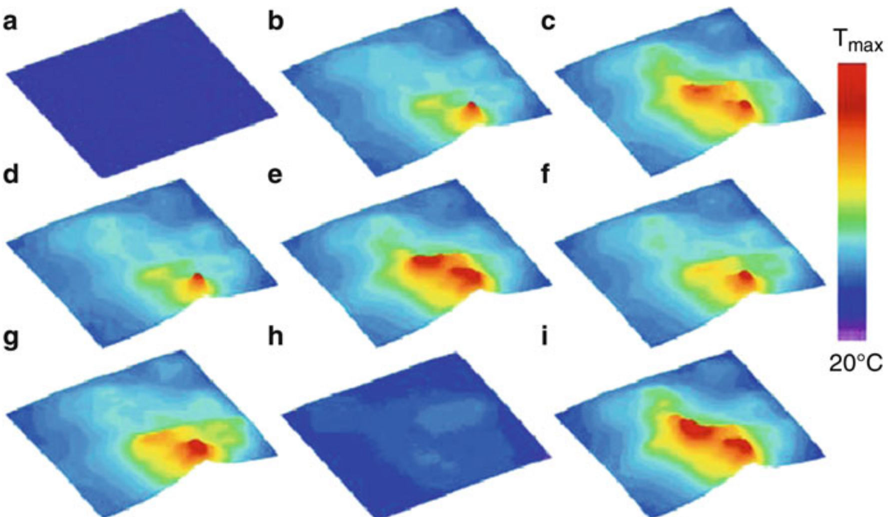


Fig. 9.11 Transient hot spot temperature contours during boot up from room temperature for flip chip die from desktop application [14]

for two configurations, namely, for a lidded module attached to an air-cooled heat sink and a bare die attached to an air-cooled heat sink. The following text revisits some of the work presented in [13].

Culham et al. [18] developed a comprehensive Fourier series based model for three-dimensional (3D) spreading in an electronics package consisting of multiple material layers to which several heat producing components are attached. The general solution for multiple rectangular heat sources located on a rectangular spreader was derived by Muzychka et al. [19], who applied this technique to a heat sink attached to several heat sources. This solution [19] was extended by Muzychka [20] to a simplified model using influence coefficients that enabled easy programming and solution of a multiple heat source—spreader system using state-of-the-art mathematical software tools. Muzychka's [20] method uses the influence coefficient based analytical model for determining the temperature distribution, for local source average or centroidal source values, for a two layer rectangular substrate that has multiple rectangular heat sources attached to it. The relation is presented in the form of a matrix equation (shown in Fig. 9.12), made up of a source temperature matrix, θ, a power input matrix, Q, and a influence coefficient matrix, f, given by [13], where $\theta_1 \ldots \theta_N$ and $Q_1 \ldots Q_N$ are the calculated temperature and input power values for heat sources numbered 1 … N, respectively, and $f_{1,1}$–$f_{N,N}$ are the influence coefficient values for the heat sources one (also shown in Fig. 9.12). The influence coefficient matrix, f, displays reciprocity, for example, the influence of the Nth cell on the 1st cell, $f_{1,N}$, is identical to the influence of the 1st cell on the Nth cell, $f_{N,1}$. The influence coefficient matrix is calculated using an expression that takes the form [13, 20].

$$f = \text{function}\ (t_1, t_2, k_1, k_2, A, B, C, D, h)$$

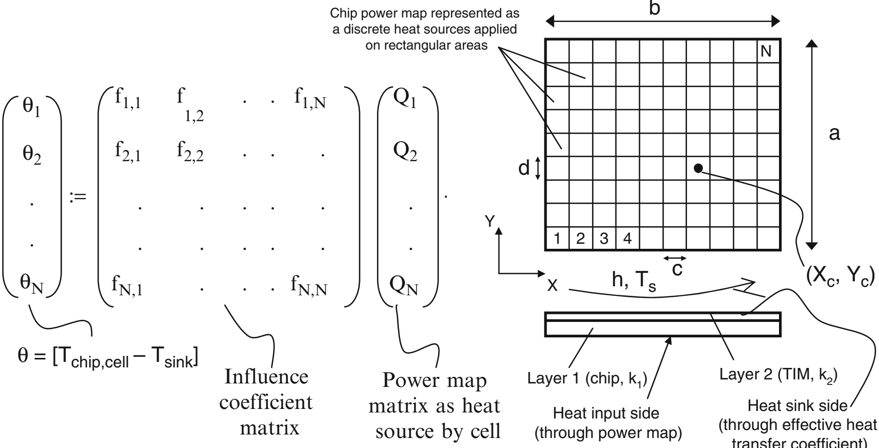
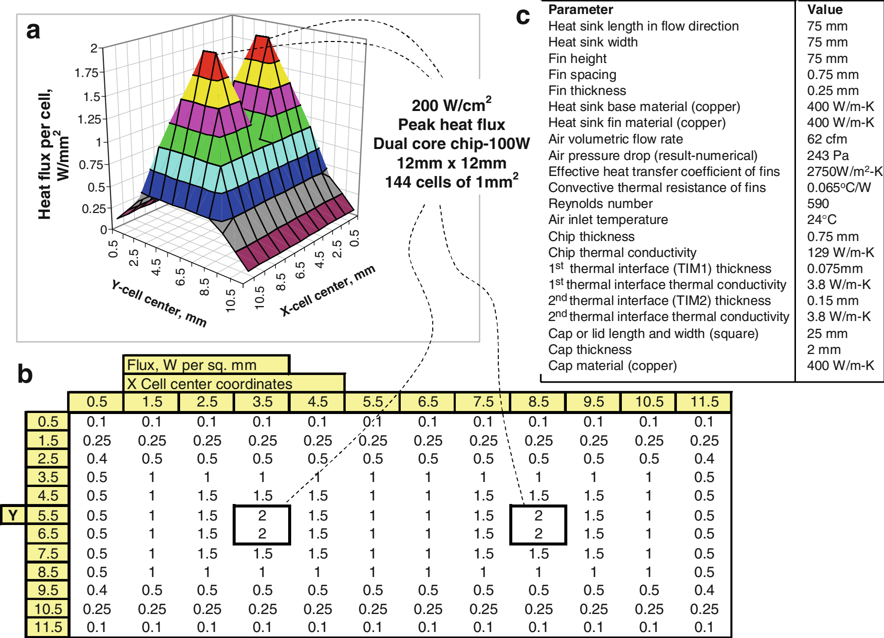


Fig. 9.12 Analytical model formulation for use of chip power map to calculate chip device side temperature contours [13, 20]

where t_1, t_2, k_1, k_2, A, B, C, D, and h, are the first rectangular layer thickness (chip), the second rectangular layer thickness (TIM), the first layer thermal conductivity (chip), the second layer thermal conductivity (TIM), the matrix of rectangular heat source dimensions along the first Cartesian axis, the matrix of rectangular heat source dimensions along the second Cartesian axis, the matrix of rectangular heat source centroidal coordinate location along the first Cartesian axis, the matrix of rectangular heat source centroidal coordinate location along the second Cartesian axis, and the convective boundary condition (uniform heat transfer coefficient) applied at the top of the second layer. The matrices A, B, C, and D, contain values for the sources numbered from 1 to N.

Sikka [21] utilized a known single source series solution [22] and used linear superposition of several individual solutions for each heat source cell of the chip power map, to allow the estimation of the chip junction side temperature profile in a chip-thermal interface-spreader assembly. Sikka neglected the effect of heat spreading from the power map in the chip and the first thermal interface (TIM1) directly over the chip, and superimposed the power map on the bottom surface of the spreader. The 1D conduction in the chip and the first thermal interface were then added to the resulting solution [21]. This disregard of spreading in the chip, and, thermal interface can have a varying impact on the accuracy of junction temperature prediction, depending on the thermal conductivity and thickness of the chip and the first interface (TIM1), respectively, and the value of the effective heat transfer coefficient on the top of the first interface, and the degree and nature of nonuniformity of the input power map. While Sikka [21] superimposed the power map on the heat spreader, the use of Muzychka [20] in Iyengar and Schmidt [4] takes into account the spreading in the chip and the interface (TIM1), but requires an assumption for the uniform effective heat transfer coefficient at the top of the interface.



Flux, W per sq. mm

Y \ X Cell center coordinates	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5
0.5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
1.5	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
2.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.4
3.5	0.5	1	1	1	1	1	1	1	1	1	1	0.5
4.5	0.5	1	1.5	1.5	1.5	1	1	1.5	1.5	1.5	1	0.5
5.5	0.5	1	1.5	2	1.5	1	1	1.5	2	1.5	1	0.5
6.5	0.5	1	1.5	2	1.5	1	1	1.5	2	1.5	1	0.5
7.5	0.5	1	1.5	1.5	1.5	1	1	1.5	1.5	1.5	1	0.5
8.5	0.5	1	1	1	1	1	1	1	1	1	1	0.5
9.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.4
10.5	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
11.5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Parameter	Value
Heat sink length in flow direction	75 mm
Heat sink width	75 mm
Fin height	75 mm
Fin spacing	0.75 mm
Fin thickness	0.25 mm
Heat sink base material (copper)	400 W/m-K
Heat sink fin material (copper)	400 W/m-K
Air volumetric flow rate	62 cfm
Air pressure drop (result-numerical)	243 Pa
Effective heat transfer coefficient of fins	2750W/m^2-K
Convective thermal resistance of fins	0.065°C/W
Reynolds number	590
Air inlet temperature	24°C
Chip thickness	0.75 mm
Chip thermal conductivity	129 W/m-K
1st thermal interface (TIM1) thickness	0.075mm
1st thermal interface thermal conductivity	3.8 W/m-K
2nd thermal interface (TIM2) thickness	0.15 mm
2nd thermal interface thermal conductivity	3.8 W/m-K
Cap or lid length and width (square)	25 mm
Cap thickness	2 mm
Cap material (copper)	400 W/m-K

Fig. 9.13 Power map studied by [13] using analytical and numerical modeling

The use of a uniform effective heat transfer coefficient at the top of the TIM1 results in neglecting the spreading effect of the hot spot heat source in the spreader.

Muzychka [20] applied this analytical model to solve for temperature distributions in printed circuit boards housing several components. The technique developed in Muzychka [20] was used by Iyengar and Schmidt [13] and extended to air-cooled flip chip packages that were either directly attached to an air-cooled heat sink or were attached to the heat sink via a second thermal interface material that was applied between the lid and the heat sink. Figure 9.13a, b illustrate and detail the power map analytically modeled by Iyengar and Schmidt [13] for a dual core 100 W silicon chip that is 12 $\times$ 12 mm with a hot spot heat flux of 200 W/cm^2. Using the information from Fig. 9.13b, such a chip would have a performance of 2 GHz. Figure 9.13c lists some of the inputs to the model. The two configurations modeled [13] are shown in Fig. 9.14a, b which represent an air-cooled heat sink attached to a lidded module (as also shown in Fig. 9.7) and a direct attach air-cooled heat sink that is attached to the back side of the chip with a thermal interface material. In addition to the power map modeling scheme discussed in the preceding text and depicted in Fig. 9.13, several other simplifications and modeling methods were employed to obtain the results shown in Fig. 9.15. For example, to facilitate the use of the power map model, the heat sink fin based air-cooling had to be represented as an effective heat transfer coefficient on the top of the heat sink base. Also, in addition to the multiple heat source power map model discussed in this

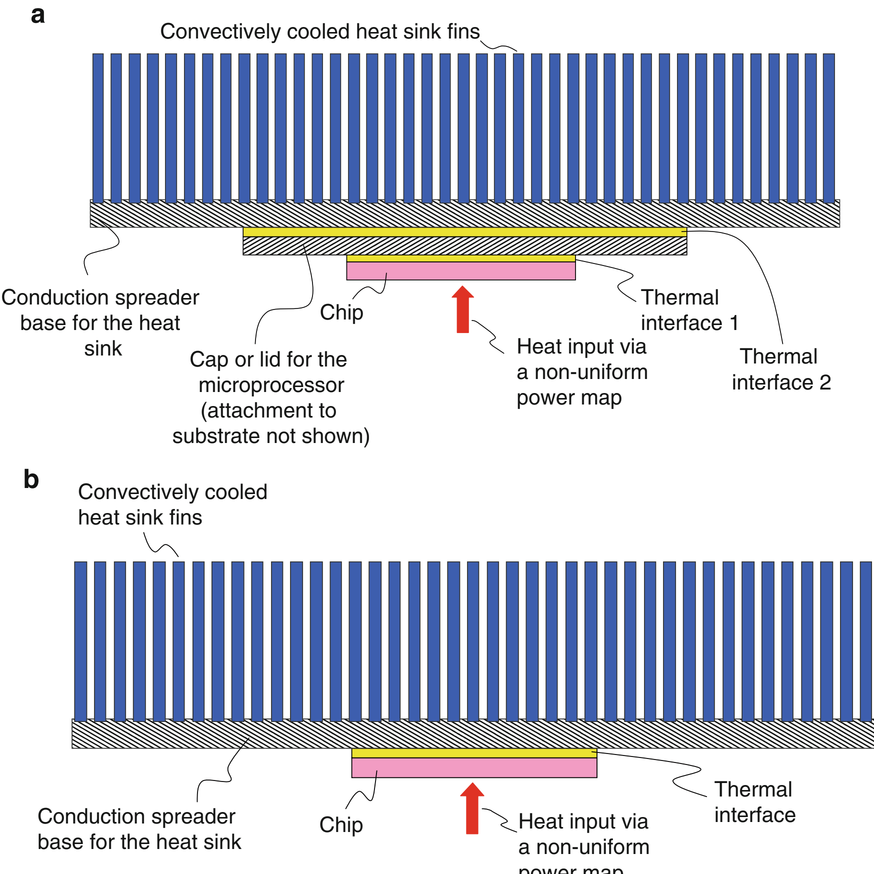
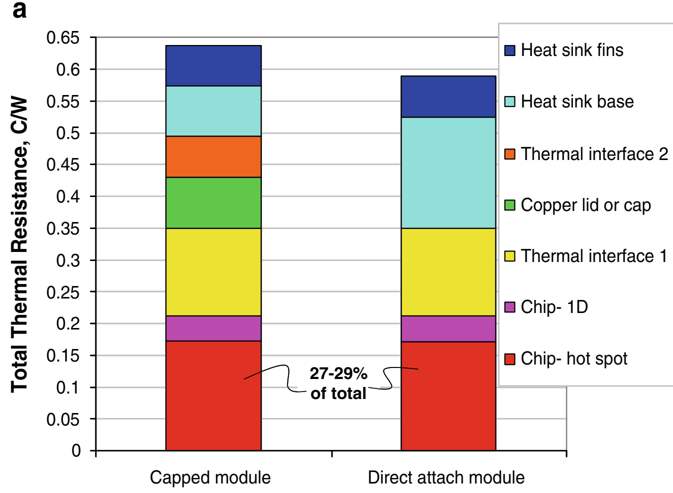


Fig. 9.14 Flip chip cooling configurations studied by [13]; (**a**) lidded module with air-cooled heat sink and (**b**) bare die with direct attach air-cooled heat sink

section, single heat source spreading models were also used which are further described in the next section when describing heat spreaders. Air-cooled heat sinks are also discussed in detail later in this chapter.

Figure 9.15 summarizes the breakup of the total junction to ambient thermal resistance for capped and direct attach modules that are depicted in Fig. 9.14a, b, respectively. Except for the chip 1D resistance, every other conduction resistance was calculated by dividing the average temperature difference between the top and bottom material surfaces by the total heat input (100 W). The hot spot thermal resistance is defined as the temperature difference between the hot spot chip side temperature and the average chip device side temperature, divided by the total chip power. The convective resistance posed by the heat sink fins is calculated numerically using the temperature difference between the ambient reference and the average of the top of the heat sink base. As seen in Fig. 9.15, the absence of the spreader and the

a



b

		Cell average temperature, C											
		X Cell center coordinates, mm											
		0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5
	0.5	53.0	53.3	53.8	54.2	54.5	54.7	54.7	54.5	54.2	53.8	53.3	53.0
	1.5	56.5	57.1	57.8	58.5	58.8	59.0	59.0	58.8	58.5	57.8	57.1	56.5
	2.5	62.0	63.6	65.1	66.1	66.5	66.6	66.6	66.5	66.1	65.1	63.6	62.0
	3.5	67.8	71.3	74.3	75.9	76.3	75.9	75.9	76.3	75.9	74.3	71.3	67.8
	4.5	72.1	76.8	82.2	84.8	84.7	82.6	82.6	84.7	84.8	82.2	76.8	72.1
Y Center Coordinates mm	5.5	74.3	79.6	86.1	90.2	88.9	86.0	86.0	88.9	90.2	86.1	79.6	74.3
	6.5	74.3	79.6	86.1	90.2	88.9	86.0	86.0	88.9	90.2	86.1	79.6	74.3
	7.5	72.1	76.8	82.2	84.8	84.7	82.6	82.6	84.7	84.8	82.2	76.8	72.1
	8.5	67.8	71.3	74.3	75.9	76.3	75.9	75.9	76.3	75.9	74.3	71.3	67.8
	9.5	62.0	63.6	65.1	66.1	66.5	66.6	66.6	66.5	66.1	65.1	63.6	62.0
	10.5	56.5	57.1	57.8	58.5	58.8	59.0	59.0	58.8	58.5	57.8	57.1	56.5
	11.5	53.0	53.3	53.8	54.2	54.5	54.7	54.7	54.5	54.2	53.8	53.3	53.0

c

		Cell average temperature, C											
		X Cell center coordinates, mm											
		0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5
	0.5	49.6	49.9	50.3	50.7	51.0	51.2	51.2	51.0	50.7	50.3	49.9	49.6
	1.5	52.9	53.5	54.2	54.8	55.2	55.3	55.3	55.2	54.8	54.2	53.5	52.9
	2.5	58.3	59.8	61.3	62.2	62.6	62.7	62.7	62.6	62.2	61.3	59.8	58.3
	3.5	63.9	67.3	70.2	71.8	72.1	71.7	71.7	72.1	71.8	70.2	67.3	63.9
	4.5	67.9	72.6	77.9	80.5	80.3	78.2	78.2	80.3	80.5	77.9	72.6	67.9
Y Center Coordinates mm	5.5	70.0	75.2	81.6	85.7	84.3	81.4	81.4	84.3	85.7	81.6	75.2	70.0
	6.5	70.0	75.2	81.6	85.7	84.3	81.4	81.4	84.3	85.7	81.6	75.2	70.0
	7.5	67.9	72.6	77.9	80.5	80.3	78.2	78.2	80.3	80.5	77.9	72.6	67.9
	8.5	63.9	67.3	70.2	71.8	72.1	71.7	71.7	72.1	71.8	70.2	67.3	63.9
	9.5	58.3	59.8	61.3	62.2	62.6	62.7	62.7	62.6	62.2	61.3	59.8	58.3
	10.5	52.9	53.5	54.2	54.8	55.2	55.3	55.3	55.2	54.8	54.2	53.5	52.9
	11.5	49.6	49.9	50.3	50.7	51.0	51.2	51.2	51.0	50.7	50.3	49.9	49.6

Fig. 9.15 Results obtained for configurations described in Figs. 9.13 and 9.14 [13]

second thermal interface (TIM2) results in a larger area ratio for spreading of heat from the top of the first thermal interface to the top of the heat sink base in the case of the direct attach design. Thus, the spreading resistance of the heat sink base is a much larger value of 0.18 °C/W for the direct attach case versus the 0.08 °C/W for the capped module, a difference of 0.1 °C/W. However, the nonexistence of the spreader and second thermal interface (TIM2) adds 0.14 °C/W to capped module thermal resistance. It is important to note the significantly large contribution of the hot spot thermal resistance, making up about 27–29 % of the total, for both types of designs.

9.8 Chip Thinning

Chip thinning is the relatively recent process of reducing the chip thickness in the depth direction with the focus on reducing the conduction thermal resistance for heat transfer through the die. The impact of chip thinning can also significantly affect the heat spreading from locally high power densities within die as well impact chip warpage and thermo-mechanical characteristics of the flip chip package.

9.9 Thermal Interface Materials

Thermal interface materials (TIMs) are used to bond the lid or heat spreader with the back side of the chip package as well as between the lid and the heat sink, the former being denoted as TIM1 and the latter as TIM2. Figures 9.7 and 9.14a, b, each show the location of these TIM1 and TIM2 materials. The goal of TIM1 is to mechanically couple the back side of the chip to the lid or spreader while also providing for satisfactory heat conduction for the heat dissipation to occur. Thus, attributes such as compression and shear strength and thermal conductivity are important metrics for TIM1. Another important attribute of the TIM1 material is its coefficient of thermal expansion (CTE), because it acts as an interlayer between a silicon chip and a metal (typical) lid which each have significantly different CTEs. Since the flip chip package assembly process involves several cure steps and temperature cycling the CTE is an important property. In some cases the TIM1 may mechanically and thermally couple the back side of the chip directly to the bottom of a heat sink as depicted in Fig. 9.14b. Such configurations are known as direct attach. The TIM2 on the other hand often serves to thermally couple an air-cooled heat sink or a liquid cooled cold plate to the lid. In most applications the TIM2 will be a detachable interface since the heat sink may need to be removed or reattached in the field or during system level assembly. The TIM1 is often a permanent joint between the chip and the lid.

Both Yeh and Chu [5] and Prasher et al. [23] provide similar equations for TIM thermal resistance. The equation presented by Prasher et al. [23] for TIM thermal resistance, R_{TIM}, is,

$$R_{\mathrm{TIM}} = \mathrm{BLT}/k_{\mathrm{TIM}} + R_{\mathrm{c1}} + R_{\mathrm{c2}}$$

Table 9.1 Thermal properties of typical TIMs

Thermal interface material	Typical BLT, mm	Typical k_{TIM}, W/m K	Thermal resistivity, C mm^2/W
Mineral oil	0.1	0.2	500
Grease	0.05	3.8	13
Gap filler putty	0.5	4	125
PCM pads	0.25	4	63
Silver filled epoxy	0.025	1.4	18
Low M.P. metal	0.2	30[a]	7

[a]Corrected thermal conductivity taking contact resistance into account

where BLT, k_{TIM}, R_{c1}, and R_{c2} are the bond line thickness of the TIM layer, the thermal conductivity of the TIM material, the first thermal contact resistance between the TIM and the first bonding surface and the second thermal contact resistance between the TIM layer and the second bonding surface, respectively. In the case of particle laded TIMs both the BLT and the k_{TIM} are a function of the particle volume fraction and size [23]. A detailed discussion of particle laden TIM materials can be found in Prasher et al. [23] relating the thermal performance of particle laden TIMs to the pressure applied to achieve a minimum BLT.

In most modeling exercises, the two contact resistances are included in a corrected thermal conductivity, and the equation for TIM thermal resistance, R_{TIM}, used is,

$$R_{TIM} = BLT_{estmated}/k_{TIM,corrected}$$

where $BLT_{estimated}$ and $k_{TIM,corrected}$ are the estimated TIM bond line thickness and the empirically corrected TIM thermal conductivity arrived at by reducing the bulk material value using experimentally determined values for total TIM thermal resistance in lab tests for specific known application geometries and configurations.

Table 9.1 provides some information about commonly used TIM materials such as mineral oils, greases, thermal putty, phase change material (PCM) pads, particle filled epoxy, and gels. Mineral oils are usually silicone or hydrocarbon oils and greases are such oils filled with higher thermal conductivity particles to some fill volume.

Table 9.2 provides some details about advanced TIM materials such as metals that lend themselves to TIM applications. Metal TIMs have the advantage of a significantly higher thermal conductivity.

While discussing thermal interface materials (TIMs) in flip chip packages, it is very important to note the effect of chip warpage on the first level TIM [10]. As the underfill material cures and the package reaches room temperature after the underfill process, the chip package undergoes warpage due to the mismatch of thermal expansion between the organic substrate and the silicon chip [10]. As may be seen in Fig. 9.16, the BLT can vary significantly depending on the location over the back side of the chip.

Table 9.2 Thermal properties of metal TIMs [24]

Solder	Liquidus Temperature C	Thermal conductivity (W/m C)
InSnGa(21.5, 16, 62.5)	10.7	35
InSnBi(51, 16.5:32.5)	60	
InBi(66:34)	72	40
InSn(52:48)	118	34
InAg (97:3)	143	73
In (100)	157	86
SnPb(63:37)	183	50
SnAg(96.5:3.5)	221	33
Sn	223	73
AnSn(80:20)	280	57

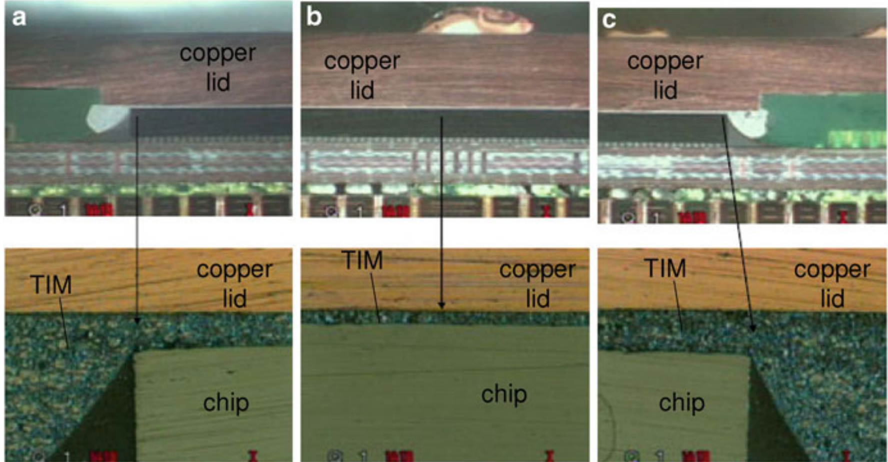


Fig. 9.16 Section view photograph of flip chip package illustrating impact of warpage on TIM [10]—(**a**) at one end of the chip–lid interface, (**b**) in the *middle*, (**c**) at the other end of the chip lid interface

9.10 System Level Thermal Management

The cooling design of electronic packages at the board and rack level can be considered to be "system level thermal management." The primary heat transfer phenomena at the chip and module level is most commonly heat conduction through several different materials including silicon, solder ball arrays, plastic encapsulating structures, thermal interface materials, and metal spreaders. At the system level there are several additional processes that can be more dominant, such as natural and forced convection using coolant transport over enhanced surfaces, boiling and condensation in spreaders, and radiation to the ambient environment.

Figure 9.17 depicts two examples of server boards of recently released server products that are each made up of different components including printed circuit

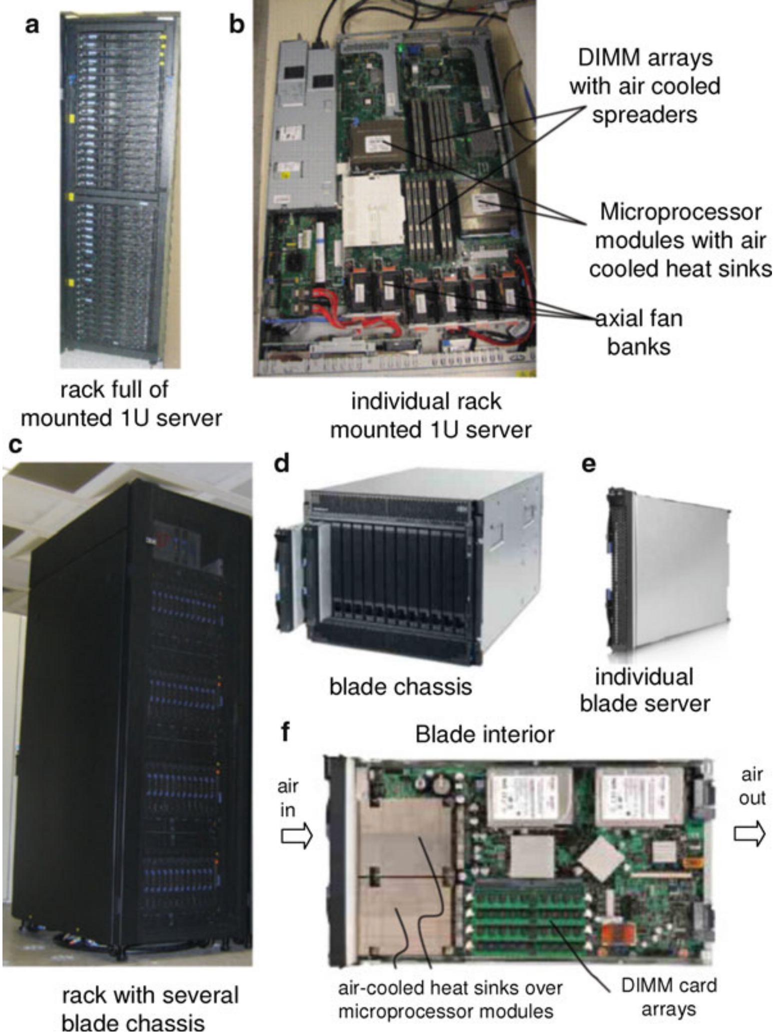


Fig. 9.17 Different air-cooled server node and rack designs—(**a**) rack mounted 1U server, (**b**) rack full of 1U servers, (**c**) blade chassis of servers mounted into rack, (**d**) blade server in a chassis, (**e**) blade server internals

board, several on board components including chip packages and capacitors, microprocessor modules, memory chips on cards, storage devices including disk drives and I/O devices. Since the focus of this book is on flip chip packages, the current discussion is devoted to the cooling of the microprocessor modules and the Dual-In-Line-Memory (DIMM) card arrays which both comprise such packages.

In the rack mounted server system depicted in Fig. 9.17b, there are fans included in the server node assembly which suck air through the front of the server and blow

it over the server components such as the air-cooled microprocessor heat sinks and the air-cooled DIMM spreaders. Figure 9.17a shows a rack full of such servers which may comprise as many as 45 1.75″ tall servers mounted on rails provided inside the rack. The blade servers illustrated in Fig. 9.17c–f do not contain fans, instead powerful blowers are provided at the chassis level to suck air through the blades thus providing air-cooling to the microprocessor heat sinks and the DIMM arrays.

9.11 Conduction Heat Spreaders

The heat generated within flip chip packages is dissipated in a relatively small volume. To accommodate the heat transfer rates achievable with common cooling methods such as air-cooling, it has to be rejected over a much larger surface area to ensure that the temperature rise between the electronic device and the ambient coolant is not excessive. This transfer of heat from a small heat source surface to a much larger heat rejection surface is achieved through heat spreaders. There are two most common types of heat spreaders, conduction spreaders and heat pipes or vapor chambers.

Figure 9.18 depicts heat spreading using a conduction heat transfer process. The heat input surface could be the central area at the bottom of a microprocessor heat sink such as illustrated in Fig. 9.18a, b or may be the DIMM spreader surface directly in contact with the memory chips through a thermal interface material as seen in Fig. 9.18c, d. The conduction spreader is commonly made of metals such as aluminum or copper, but could also be made of more exotic materials such as graphite or diamond.

There are several publications that provide closed form analytical expressions to estimate the spreading resistance of heat spreaders [25–28], Song et al. [25] provide one of the most widely used formulas.

9.12 Heat Pipes

The second popular type of heat spreading structures are heat pipes or vapor chambers, which utilize two-phase heat transfer to provide for highly efficient heat spreading devices. A heat pipe is a cylindrical tube, commonly made of a high thermal conductivity material such as copper which contains a wick structure along the inside surface of the tube as well as a central open region. Figure 9.19 depicts a schematic of a heat pipe and highlights the various structures and regions significant to its thermal operation.

The heat pipe is filled with a liquid such as water that will evaporate at the evaporator region and condense in the condenser region to complete the heat transport process. In a heat pipe such as the one depicted in Fig. 9.19, heat is

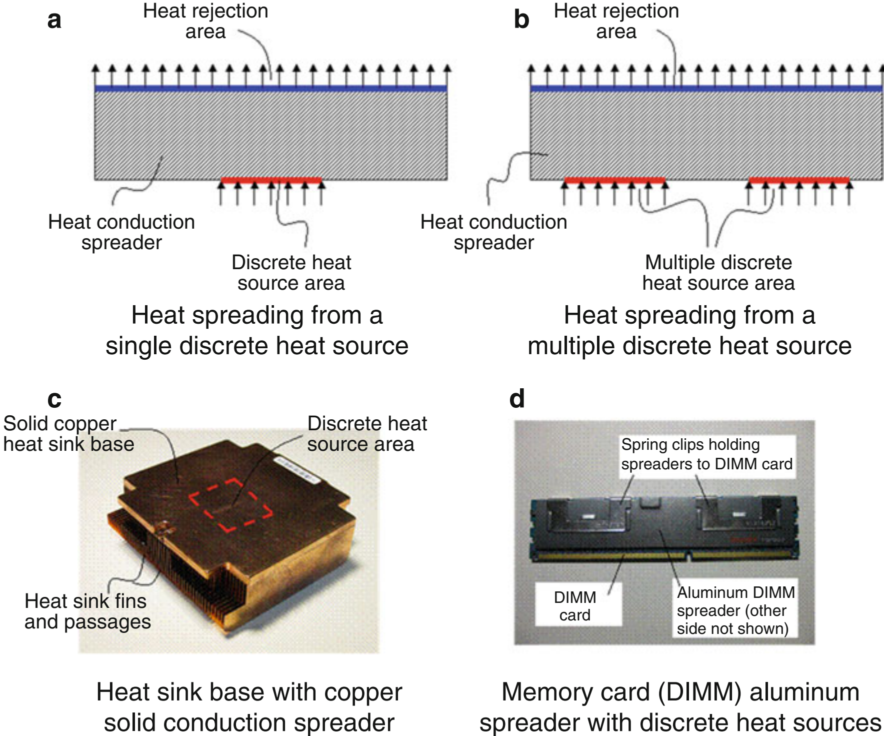


Fig. 9.18 Heat spreading schematic using conduction spreaders—(**a**) heat spreading from a single discrete heat source, (**b**) heat spreading from a multiple discrete heat sources, (**c**) heat sink base with copper solid conduction spreader, (**d**) memory card (DIMM) aluminum spreader with discrete heat sources

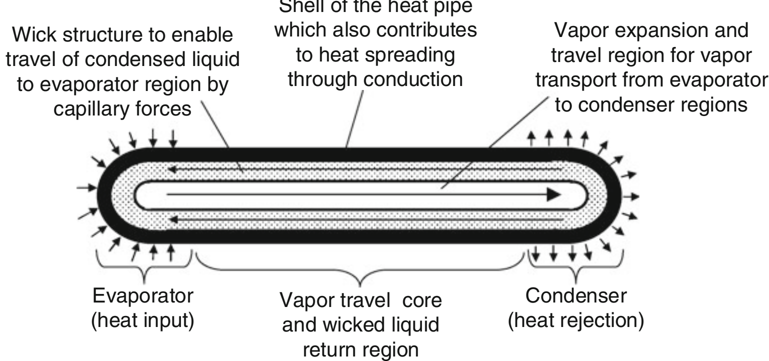


Fig. 9.19 Schematic of heat pipe operation

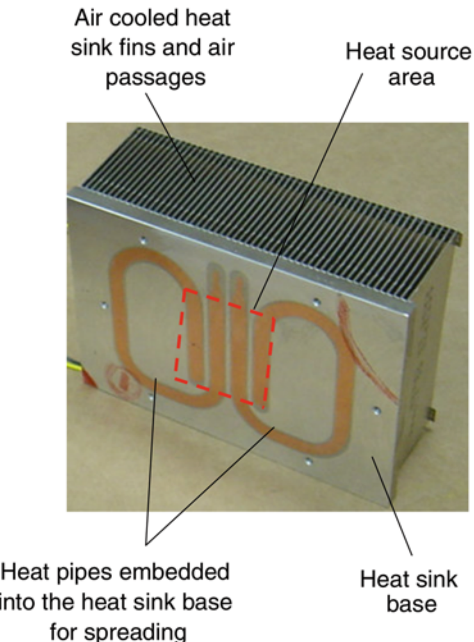


Fig. 9.20 Photograph of an air-cooled heat sink with an embedded heat pipe for spreading heat in the base

conducted into the heat pipe from one end through the outer shell and into the evaporator structure which results in two-phase heat transfer to the resident liquid and subsequent vapor travel through the inner core region to the other end of the heat pipe, i.e., the condenser. At the condenser, the outer surface of the heat pipe shell is cooled using, for example, air-cooling structures which result in the local vapor to condense and form liquid which then travels back through the wick material by means of surface tension driven capillary forces.

Figure 9.20 displays a photograph of the use of heat pipes in the base of an air-cooled heat sink to spread the heat away from a concentrated heat source (the microprocessor module) into a much larger area heat sink base for subsequent rejection into the air stream. The generic heat pipe technology can take different forms and is also available commercially as rectangular slabs with the wick material deposited along the inner cavity. Such devices are called vapor chambers and operate much like the cylindrical heat pipes discussed in this section. Vapor chambers have the advantage that heat spreading structure can be very low profile and yet cover a larger footprint area compared to traditional cylindrical heat pipes. Yet another variant can be flat heat pipes which are similar to the cylindrical variety except that the outer shell tube has been flattened to create a substantially planar heat spreading device.

As may be expected, heat pipes work best when the return liquid flow from the condenser to the evaporator is aided by gravity. While heat pipes can in some configurations operate against gravity, this is not recommended.

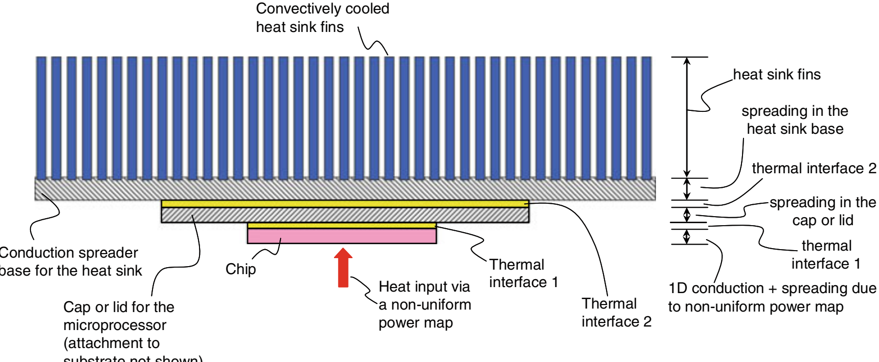


Fig. 9.21 Schematic of an air-cooled heat sink

9.13 Air-Cooled Heat Sinks

Air-cooled heat sinks are commonly used in the electronics cooling industry to reject heat into a flowing air stream that may be driven by fans or blowers [29]. Figure 9.21 shows a schematic of a electronic device that is being cooled by an air-cooled heat sink with a heat spreader as discussed earlier.

Heat spreaders are often integrated with heat sinks to facilitate the spreading of the heat dissipated from a relatively small footprint such as the lid of a microprocessor module and to spread it across a larger area such as the base of heat sinks. The heat then is conducted into the fins which can be rectangular plates attached to the base or can be cylindrical pins protruding from the base. The heated fins transfer the heat to the air flowing across them through convection. Figure 9.21 further depicts the various components of the heat rejection path comprised by the one-dimensional and spreading conduction in the chip, conduction in the first thermal interface between the chip and the lid, conduction through the lid and the second thermal interface between the lid and the heat sink base, conduction (or heat pipe based spreading) in the heat sink base, finally heat conduction and convection through and from the fins, respectively. As may be expected, a higher rate of air flow across the fins can yield higher thermal performance albeit with a higher air pressure drop across the fins which would require greater fan power to deliver the required air flow.

Figure 9.22 illustrates air-cooled heat sinks that have been manufactured using a variety of fabrication techniques and resulting in variants of the generic geometry depicted in Fig. 9.21 [30]. The different techniques illustrated in Fig. 9.22 have a wide range of manufacturing constraints and costs of fabrication which in turn yield a large range of cooling designs spanning a wide range with respect to a cost or performance metric. One of the most economical methods is not depicted in Fig. 9.22. This is extrusions which are used in high volume to make aluminum heat sinks with modest aspect ratios. The higher the aspect ratio the denser (i.e., more fins/unit width) the heat sink design and in most cases the better the thermal performance. In recent years significant effort has been devoted towards advancing manufacturing methods for heat sink fabrication.

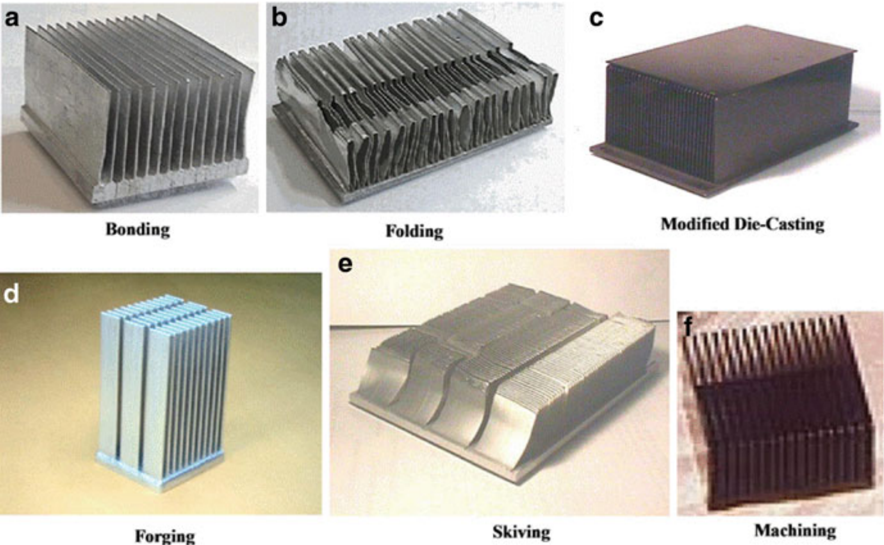


Fig. 9.22 Photographs of air-cooled heat sinks manufactured by a variety of methods

9.14 Liquid Cooled Cold Plates

In several electronics cooling applications, the use of air as a heat rejection coolant is not a viable solution due to the relatively low values for the thermal conductivity and specific heat capacity of air which together greatly influence the capability of the flowing air stream to convectively extract the heat from the air-cooling fins and transport the heat to a remote location. Thus, air-cooling can result in heat transfer coefficients on the fin surface that are too low and caloric heat carrying capacity that is insufficient, resulting in untenable designs with a temperature rise between the air and the base of the heat sink that is extremely high or the air flow rate required to make the design work is prohibitively large.

In contrast, liquids such as water have extremely high values for thermal conductivity and specific heat and are thus excellent coolants with respect to convection heat extraction and transport of the heat from high heat flux sources such as microprocessors used in high end personal computers and servers. Figure 9.23 shows a schematic of a liquid cooled cold plate which topologically is similar to the air-cooled solution depicted in Fig. 9.21. While air-cooling usually utilizes air from the room, liquid cooling required specially designed plumbing to and from the heat generating devices. As seen in Fig. 9.23, the liquid-cooled cold plate usually consists of a rectangular chamber into which coolant is injected and out of which it is extracted. While inside the chamber, the coolant flows across an array of fins that are attached to a base plate which is usually formed on one side of the rectangular chamber. The heat is conducted into the base plate from a module

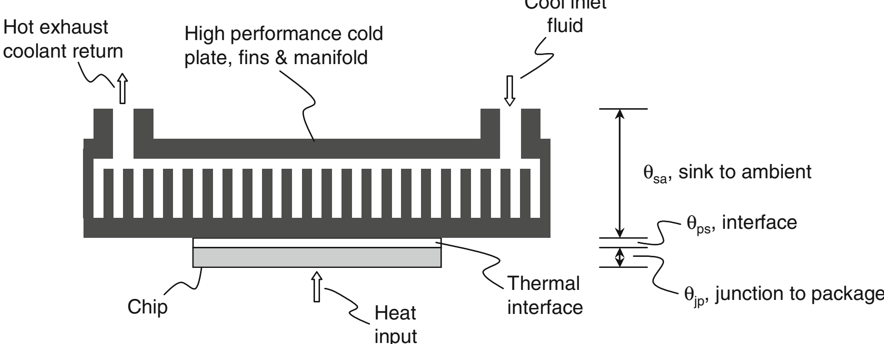


Fig. 9.23 Schematic of a liquid cooled cold plate

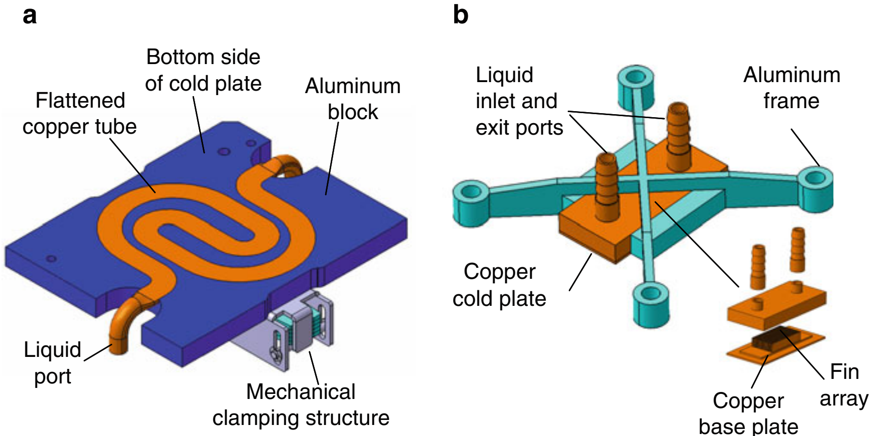


Fig. 9.24 Images of liquid cooled cold plates, (**a**) Copper tube embedded into a aluminum block [31], (**b**) Finned cold plate with liquid flow through the fins [32]

lid or the heat source itself (chip) through a thermal interface material and the fins reject the heat load into flowing liquid through conduction and convection.

Similar to the discussion of air-cooled heat sinks, higher liquid flow rates result in greater cooling capability but at the cost of a higher liquid pressure drop across the cold plates. Owing to the much higher heat transfer coefficients and heat carrying capability from the use of a liquid such as water, the cold plate structures and fins are usually much more compact compared to air-cooled heat sinks.

Figure 9.24 shows images of the two most common types of cold plates, namely, the embedded flattened copper tube based design and the fin array based geometry. In the copper tube based cold plates, the liquid carrying tubes are pressed into grooves in a aluminum block and glued to the groove surfaces yielding moderate

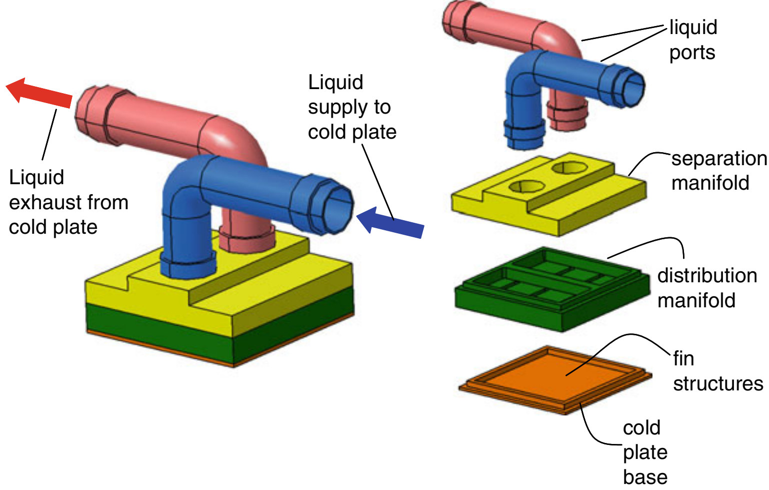


Fig. 9.25 Images of a more complex manifold cold plate with highly parallel liquid flow paths enabled using a manifold layer above the fin array

thermal performance designs which are low cost. In the fin array based designs, the fins are fabricated on a base plate using a variety of processes similar to what are used for making air-cooled heat sinks. These techniques include metal removal (machining), material deformation (forging, skiving) and joining (bonding) processes which each have varying process capabilities with respect to what geometries they can fabricate as well as a range of manufacturing costs.

Figure 9.25 shows a more complex manifold cold plate design in which the liquid distribution is carried out a manifold layer above the fins for a highly parallel liquid distribution into the cold plate to allow for high performance while limiting the pressure drop across the heat sink.

9.15 Internal Hybrid Liquid–Air Cooling Systems

While liquid cooled cold plates can be used exclusively to cool high power components, there are applications in which cold plate based cooling is used to extract the heat from the high power device and an air-to liquid heat exchanger is used to reject the heat into an air stream as shown in Fig. 9.26. In such an application, the liquid cooled cold plate, the pump forcing liquid through it, and the plumbing, all serve as the heat extraction and heat transport assembly, with the air to liquid heat exchanger and its fan playing the role of heat rejection apparatus [33]. Figure 9.27 illustrates such an internal hybrid cooling systems for a high end personal computer application.

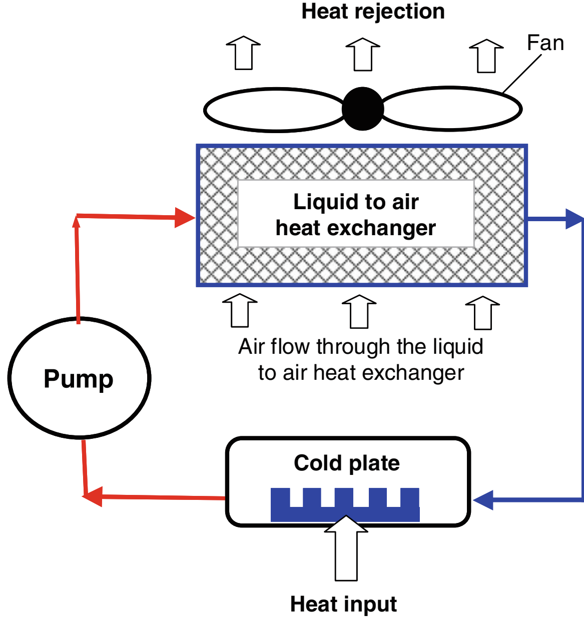


Fig. 9.26 Schematic of an internal hybrid liquid–air cooling system

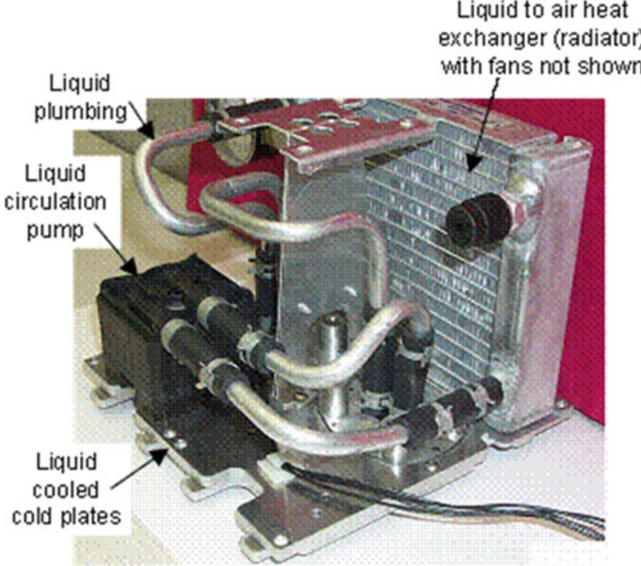


Fig. 9.27 Photograph of an internal hybrid liquid–air cooling for a high end personal computer

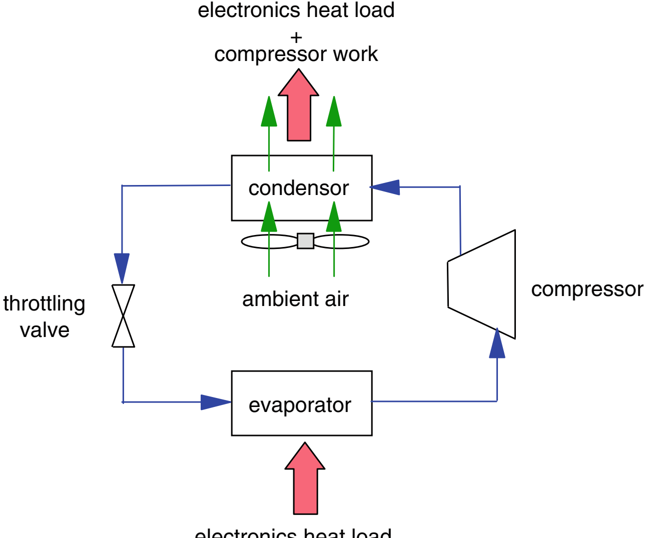


Fig. 9.28 Simple schematic of vapor-compression refrigeration loop for cooling an electronic package

9.16 Refrigeration Cooled Systems

Refrigeration cooled systems as discussed here refer to cooling systems which circulate a liquid refrigerant and utilize a vapor compression cycle to reject the heat load to secondary coolant such as air or water. A principal advantage of such a system is that it offers the ability to provide sub-ambient cooling. That is at a temperature below the temperature of an external coolant such as room air or building water. Within recent years a number of computer manufacturers have taken advantage of refrigeration cooling to achieve computing performance enhancements [34–38] by operating at lower device junction temperatures. As noted in [34], depending on the doping characteristics of the chip, the potential performance improvements range from 1 to 3 % for every 10 °C lower transistor temperature.

A simple schematic of a refrigeration cooling system loop for electronic modules or packages is shown in Fig. 9.28. The loop is somewhat similar to an indirect water cooling loop as discussed earlier in this section, except that the liquid coolant is a pressurized refrigerant which expands to provide a lower temperature, absorbs heat changing to a vapor phase, is recompressed and then condensed into a liquid phase. A typical refrigerant used in such a system is R-134a (CH_2FCF_3), which is commonly used as a refrigerant in automotive and home air-conditioning systems because of its environmental compatibility [36]. It does not contain chlorine and does not contribute to depletion of the ozone layer. Although it does exhibit a global warming potential (GWP), which is an

index describing the relative ability of a greenhouse gas to trap radiant energy compared to an equal mass of carbon dioxide, it is still one of the preferred refrigerants.

As can be seen in Fig. 9.28, the basic components comprising the refrigeration cooling loop are a compressor, condenser, throttling valve or expansion device and an evaporator. The compressor and condenser may be located remote from the electronic packages where space is usually at a premium. High pressure superheated (i.e., at a temperature above saturation temperature) vapor leaves the compressor and passes through a condenser. Within the condenser the vapor refrigerant vapor is condensed and leaves as a low temperature subcooled (i.e., at a temperature below saturation temperature) liquid. The total heat load, which includes the heat dissipated by the electronic package and heating due to compression of the refrigerant vapor, is rejected to cooling air or water via the condenser. Although not shown, a filter may be employed downstream of the condenser to remove any moisture, acids, oil decomposition products, metallic particles or other contaminants in the refrigeration system. To insure satisfactory operation, it is important to keep moisture in the refrigeration system below an allowable limit. Liquid refrigerant leaving the condenser/filter is then passed through an expansion device permitting expansion of the liquid to a low temperature, low quality mixture of vapor and liquid entering the evaporator. Quality is the ratio of the vapor mass flow to the total mass flow (i.e., vapor and liquid). So, the quality of the coolant flow entering the evaporator is nearly zero. Within the evaporator the liquid boils changing phase to vapor producing a relatively high heat transfer coefficient (1,400–1,600 W/m^2 K) to remove heat from the coolant passages [36]. In addition, the high heat of vaporization (215.9 kJ/kg) of R-134a compared to the specific heat of water (4.179 kJ/kg K) means that to transport the same heat load the mass flow rate of refrigerant in the loop can be much less than would be required with single phase water. The coolant exiting the evaporator to return to the compressor is nearly all vapor and has a quality of nearly one. Although, not shown in Fig. 9.28 an accumulator may be placed in the path between the evaporator and the compressor. This is to insure that no liquid enters the compressor as this could damage the compressor. The vapor entering the compressor is then compressed and delivered to the condenser completing the cycle.

Although many people consider the use of refrigeration to cool electronic packages in computers a relatively recent development, for the record it should be noted that the CDC 7600 designed by Seymour Cray and announced in 1964 was perhaps the earliest computer to use refrigeration cooling [39]. Not surprisingly, the CRAY-1 computer announced in 1976 also used refrigeration cooling with Freon [40]. However, it was not until the late 1990s that broader applications of refrigeration cooling began to appear. For example, in 1996 KryoTech teaming with digital equipment corporation demonstrated a refrigeration cooled workstation running at 767 MHz. The KryoTech Super G™ Computer was announced, making KryoTech

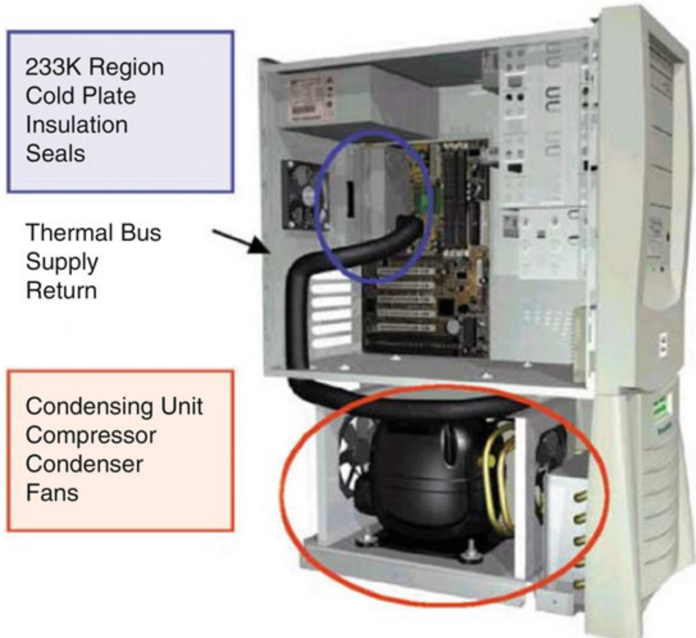


Fig. 9.29 Kryotech's Super G™ Computer with refrigeration cooling system [35]

the first company to commercially offer a 1 GHz PC system. Using a refrigeration cold plate at −40 °C to cool the processor made it possible to operate approximately 33 % faster while remaining within its operating specifications. A picture of the computer with the side cover removed is shown in Fig. 9.29 [35]. In the bottom section of the picture, on the left may be seen the compressor with fans in front and back and on the right side is the condenser. The fans are used to draw air in across the condenser thereby rejecting the heat dissipated by the processor and the heat due to the refrigerant compression process. In the upper left corner of the computer may be seen the insulation surrounding the refrigerant evaporator cold plate. The black hose-like object extending from the compressor region up to the evaporator cold plate is insulation surrounding refrigerant supply and return lines. Thermal insulation is required to prevent condensation of moisture from room air on any surfaces with a temperature that may be below the ambient dew point temperature. A fan may also be seen in the left wall of the box. This is used to provide cooling air circulation for other heat dissipating components within the box that are not cooled by refrigeration.

Refrigeration cooling has also been and continues to be used successfully in a number of IBM high-end servers [34, 36–38]. The first to use refrigeration cooling was the IBM S/390G4 CMOS system which was first shipped in 1997. The decision to use refrigeration cooling instead of other cooling schemes (e.g., high flow

Fig. 9.30 IBM S/390 G4 CMOS system with refrigeration cooling

air-cooling or water cooling) was primarily driven by the system performance improvement that could be obtained through lower temperature operation of processor chips. A brief description of the system follows to illustrate the application and integration of refrigeration cooling to cool a high performance flip chip processor module in a high-end server.

The packaging layout of the G4 system is shown in Fig. 9.30. Below the bulk power compartment is the central electronic complex (CEC) where a multichip module (MCM) housing 12 flip chip mounted processor chips is located. A cross-sectional schematic of the processor module with the evaporator cold plate attached to the top surface is shown in Fig. 9.31. Heat is conducted across a grease filled gap from the back surface of each chip to the hat and then to the evaporator within which it is transferred to R-134a refrigerant. The evaporator mounted on the processor module is fully redundant with two independent refrigerated passages. Although not shown in Fig. 9.31 the evaporator cold plate is surrounded by thermal insulation to prevent moisture condensation. Two modular refrigeration units (MRU's) shown in Fig. 9.30 are located near the middle of the frame and are connected to the evaporator cold plate via thermally insulated hoses. Figure 9.32 illustrates the principal components within an MRU. On the left side of the figure may be seen the air-cooled, finned tube condenser used to reject the heat load picked up by the refrigerant. The condenser is cooled by air flow exiting the I/O expansion cage shown at the bottom of the frame in Fig. 9.30. Only one MRU is operated at a time during normal operation. Refrigerant passing through one passage is adequate to cool the MCM which dissipates a maximum power of 1,050 W. In the event of a refrigeration failure, the failed

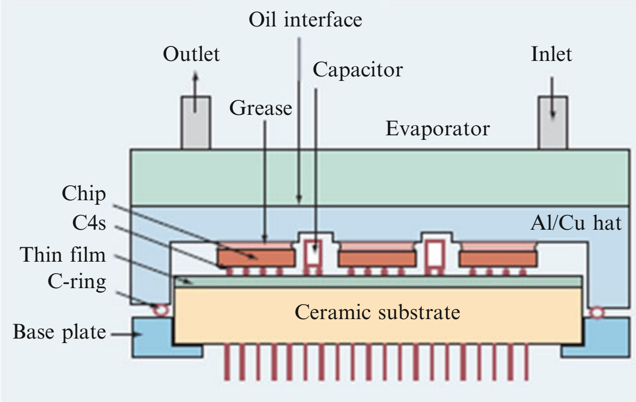


Fig. 9.31 Cross-sectional schematic of CMOS processor module with attached refrigeration evaporator cold plate

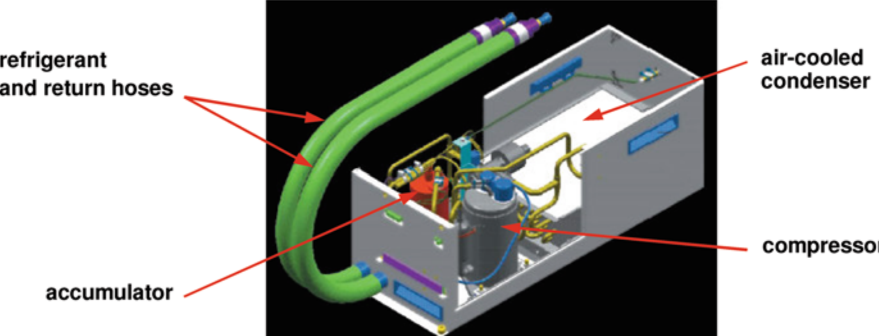


Fig. 9.32 Solid model of IBM modular refrigeration unit (MRU)

MRU maybe disconnected and replaced while uninterrupted cooling is provided by the other MRU. This scheme provides average processor temperatures of 40 °C compared to temperatures of 75 °C in a comparable air-cooled design.

As may be appreciated from the previous discussion and examples, refrigeration cooling may be utilized in some cases to obtain electrical performance improvements that might otherwise be unobtainable. However, as with any design option many factors must be considered before the choice to employ refrigeration cooling is made. Among these are the physical volume, weight and cost of the refrigeration components compared to more conventional options (e.g., air or water cooling). In addition, careful consideration must be given to the refrigerant and cooling component temperatures throughout the cooling loop to take appropriate measures to prevent moisture condensation.

9.17 Emerging Research Areas and Technologies

The continuing demand for increased packaging density and the associated increases in both areal and volumetric heat density has driven research and development of advanced cooling technologies to support the demand. Among these techniques are direct liquid immersion, 3D chip stacks, advanced thermal interfaces, synthetic jets, and application of thermoelectric cooling.

9.18 Direct Immersion Cooling

Although air-cooling continues to be the most widely used method for cooling flip chip packages, it has long been recognized that significantly higher heat fluxes can be accommodated through the use of liquid cooling. The application of liquid cooling for microelectronics may be categorized as either indirect or direct. Indirect liquid cooling has been discussed earlier in this chapter in the form of cold plate cooling utilizing water as a coolant. The focus of this section is direct liquid cooling, which may also be termed direct liquid immersion cooling, because with this form of cooling there are no physical walls separating the flip chips and the surface of the substrate from the liquid coolant. This method of cooling provides the means to remove heat directly from the chip(s) with no intervening thermal conduction resistances, other than that within the silicon between the device heat sources and the chip surfaces contacted by the liquid. Interest in direct liquid immersion as a method for cooling integrated circuit chips may be traced back as early as the 1960s [41–43] and has continued to this day [44, 45].

Direct liquid immersion cooling utilizes convective heat transfer processes which can be classified as natural convection, forced convection or boiling modes of heat transfer. The relative magnitude of surface heat flux that can be accommodated for each of these modes is shown in Fig. 9.33 as a function of wall superheat or surface to liquid temperature difference, for a typical fluorocarbon coolant.

Natural or free convection is the process of transferring heat in which mixing and fluid motion is induced by the variation in coolant density resulting from the heat given off by the heated surfaces. This mode of heat transfer offers the lowest heat flux or cooling capability for a given wall superheat. Nonetheless, the heat transfer rates attainable with direct liquid natural convection easily match or exceed those attainable with forced convection of air.

Higher heat transfer rates may be supported by utilizing a pump to provide forced flow of the liquid over the chip surface. This mode is termed forced convection and as shown in Fig. 9.33, the allowable heat flux for a given wall superheat can be increased by increasing the velocity of the liquid over the chip surface.

Boiling is the most effective mode of direct liquid immersion heat transfer. As the name implies, it is a mode of heat transfer involving liquid to vapor phase

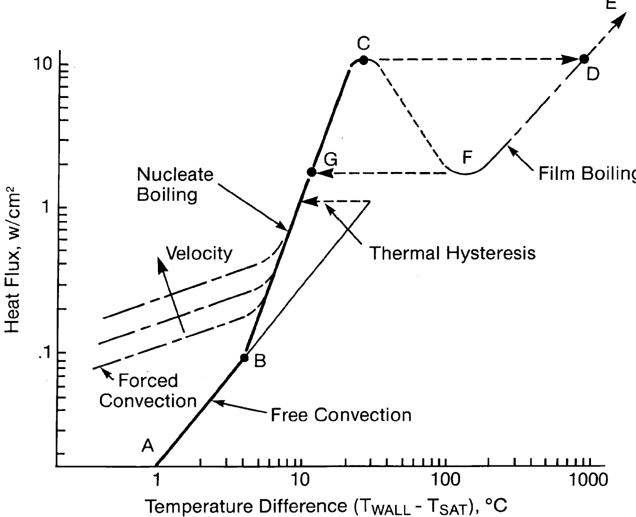


Fig. 9.33 Relative magnitude of chip heat fluxes accommodated with a fluorocarbon coolant for modes of heat transfer

change at heated surface in the form of vapor bubbles. Boiling heat transfer behavior is often characterized in the form of a classical boiling curve, which is usually obtained experimentally for a particular surface and fluid of interest (e.g., silicon and FC-72). For example, Fig. 9.33 describes the cooling path (A-B-C-D-E-F-G) that will be followed in terms of wall superheat as heat flux at the heated surface is increased. If chip power is gradually increased in small steps, cooling occurs first by natural convection (A–B). Eventually a power level is reached at which sufficient superheat is available to initiate the growth of vapor bubbles on the surface and boiling starts (B). As power is increased, more nucleation sites become active and the frequency of bubble departure increases. The region between B and C is termed the nucleate boiling regime. Vigorous agitation of the hot boundary along the heated surface, and gross fluid circulation caused by the motion of the vapor bubbles, provide the ability to accommodate substantial increases in heat flux with minimal increases in surface temperature. Examples of nucleate boiling off chips mounted on a metallized ceramic substrate and a multi-layer ceramic substrate are shown in Fig. 9.34a, b, respectively. As power is increased to point C, the critical heat flux condition is reached. So many bubbles are generated at this point that they begin to form a vapor blanket inhibiting fresh liquid from reaching the surface. Further increases in power will result in a transition to film boiling (D–E). In this regime heat transfer from the surface to the liquid is dependent on thermal conduction through the vapor and it is very poor. In most electronic cooling applications, transition to film boiling will result in failure due to high temperatures. To take advantage of boiling to cool electronic devices, it is desirable to operate in the nucleate boiling regime (B–C).

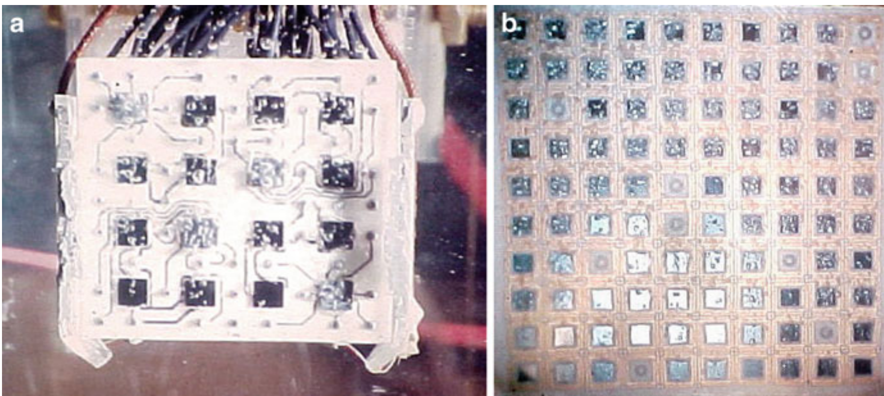


Fig. 9.34 Boiling off chips on a 25 × 25 mm metallized ceramic substrate (**a**) and a 90 × 90 mm multi-layer ceramic substrate (**b**)

A problem sometimes associated with pool boiling of dielectric liquids such as fluorocarbons is that of temperature overshoot (or thermal hysteresis). This behavior is characterized by a delay in the inception of nucleate boiling (i.e., beyond point B), such that the heated surface continues to be cooled by natural convection; with increased surface temperatures unless a sufficient superheat is reached for boiling to occur. This behavior is a result of the good wetting characteristics of the fluorocarbon liquids and the smooth nature of silicon chips. Although much work [46] has been done in this area, it is still a potential problem which must be considered. However, there is usually little or no temperature overshoot associated with flow boiling cooling applications.

When considering the direct liquid immersion cooling option, it must be remembered that the selection of a coolant for direct immersion cooling cannot be made on the basis of heat transfer characteristics alone. Chemical compatibility of the coolant with the chips and other packaging materials exposed to the liquid must be a primary consideration. There may be several coolants which can provide adequate cooling, but only a few will be chemically compatible. Water is an example of a liquid which has very desirable heat transfer characteristics, but which is generally unsuitable for direct immersion cooling on account of its chemical characteristics. Fluorocarbon liquids (e.g., FC-72, FC-86, FC-77, etc.) have generally been considered to be the most suitable liquids for direct immersion cooling, in spite of their poorer thermo-physical properties. However, in recent years because of increasing concerns with global warming there has been increased interest in fluids with lower global warming potential (GWP). Segregated hydrofluoroether and fluoroketones are viewed as safe, sustainable alternatives to the fluorocarbon family of coolants [47].

In spite of prolonged interest in direct immersion liquid cooling to cool high heat flux integrated circuit chips, there have been only a limited number of commercial applications. Two of these applications are discussed here.

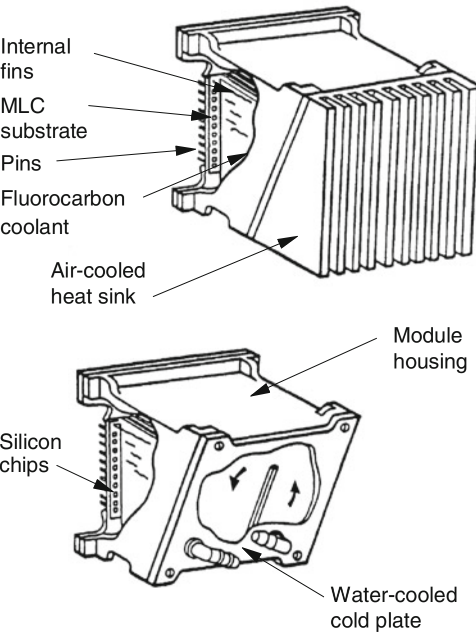


Fig. 9.35 Examples of IBM liquid encapsulated modules (LEMs) with external air or water cooling

The liquid encapsulated module (LEM) developed at IBM in the 1970s provides an example of a package utilizing pool boiling. As shown in Fig. 9.35, a substrate with 100 C-4 mounted circuit chips (100) was mounted within a sealed module cooling assembly containing fluorocarbon coolant (FC-72). Boiling at the exposed chip surfaces provided high heat transfer coefficients (1,700–5,700 W/m^2 K) to meet chip cooling requirements. Internal fins provided a means to condense the vapors and remove heat from the liquid. Either an air-cooled or water cooled cold-plate could be used to remove heat from the back surface of the module. Using this approach, it was possible to cool 4 W chips (4.6 $\times$ 4.6 mm) and module powers up to 300 W. In addition, direct liquid immersion cooling was used within IBM for over 20 years, as a means to cool high powered bipolar chips on multi-chip substrates during electrical testing prior to final module assembly.

A further example of the application of direct liquid immersion cooling is provided by the CRAY-2 supercomputer [48]. In this case a large-scale single-phase (i.e., non-boiling) forced convection fluorocarbon cooling system. As shown schematically on the left side of Fig. 9.36, stacks of electronic module assemblies were cooled by a forced flow of FC-77 in parallel across each module assembly. Each module assembly consisted of eight printed circuit boards on which were mounted an 8 $\times$ 12 array of single chip carriers. A total flow rate of 70 gpm was used to cool 14 stacks containing 24 module assemblies each. The power dissipated by a module assembly was reported to be 600–700 W. Coolant was supplied to the electronics frame by two separate frames containing the required pumps and water-cooled heat exchangers to reject the total system heat load to customer supplied chilled water.

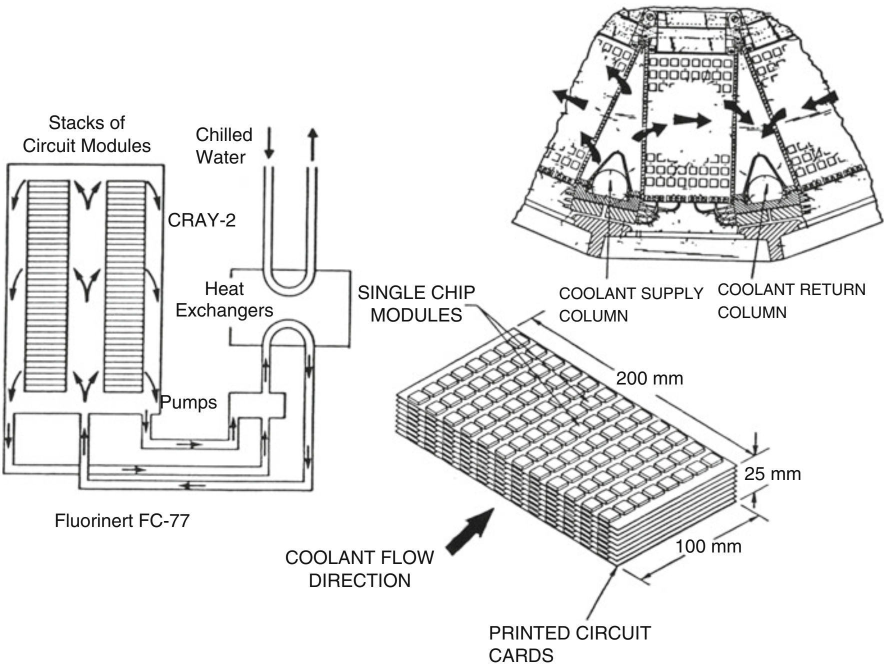


Fig. 9.36 Cray-2 direct liquid immersion cooling system schematic and packaging

9.19 3D Chip Stacks

The flip chip type of package discussed thus far in this chapter is made up of a substantially planar rectangular piece of silicon which has a small thickness in the depth direction. A single chip is attached to a substrate and a lid to form the chip package. However, there have been significant effort research and development efforts in the past decade both in academia and industry pursuing the design and development of 3D chip packages such as the one shown in Fig. 9.37 [49], in which several chips have been stacked together to create a 3D chip stack. A detailed explanation of the prevalent trends and different configurations of 3D chip packages can be found in [49–51] which discuss the many reasons why such technology is an important area of innovation. The term 3D chip package is a generic one which can have various forms of implementation based on the design, manufacturing process and application as has been described in the literature [49–56]. Some of the benefits include reduced power and size [51], reduced chip to chip interconnect distances through vertical integration, interconnection of heterogeneous chips, and much higher I/O density than is possible using today's 2D chips [57].

While there are several good electrical reasons for the migration to 3D chip packages, the thermal challenge of cooling such stacked devices remains one of the key unsolved issues in its implementation. One of the key reasons is the use of

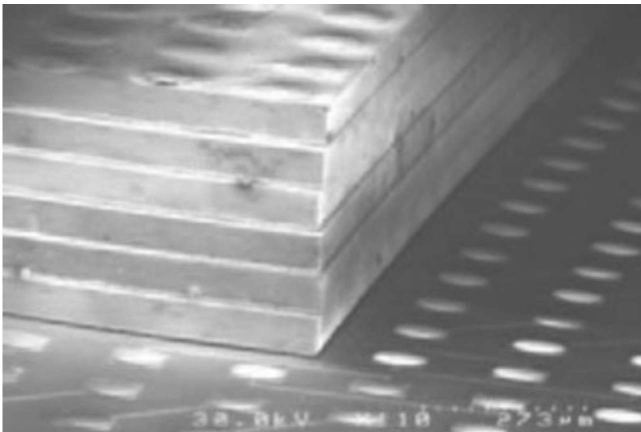


Fig. 9.37 Cross-sectional view of a 3D chip stack package (thinned, stacked silicon test structures on a wafer) [49]

polymer under fills in these 3D chip stacks at the interfaces between the chips. This means that the thermal path for heat rejection from the bottom most chip to the heat sink is poor one, resulting in an enormous challenge for the thermal engineer [57]. Several 3D chip configurations have been studied in the literature including those studied by [52] in which thermal modeling analyses were carried out for several versions of the same 507 W chip stack, with the location of the highest heat dissipation chip being changed. Two different kinds of chips were utilized to build the 3D chip stack with a six core logic or processor chip being the high power one (477 W) and the two memory chips (12 and 18 W) being a low power chip. The stack was made up [52] of two different kinds of memory chips (DRAM and SRAM) and one logic (processor) chip all with a footprint area of 21 mm by 15 mm and the 3D chip stack. In this study [52], the chip was cooled using a liquid cooled cold plate attached to the back side of the chip using a thermal interface material. It was shown that by using the liquid cooled cold plate in this manner, the best configuration was the one where the highest power chip was located at the top of the stack. While this made the best thermal design, it also means that a larger number of vias would be needed to pass through the lower chips to allow electrical connectivity with the processor chip at the top of the stack. Of course this is the least desirable design from an electrical stand point especially because of the longer distance for the signals to travel electrically from the I/O pins as well as due to the number of vias required to pass through the lower chips to communicate with the logic chip. This conflict of design goals between electrical and thermal design of the 3D chip stack highlights the greater need for co-design and co-optimization at the model level as such stacked die technology becomes more pervasive. While the preceding discussion of cooling a 3D chip using a conventional cold plate attached to one side of the top most chip can be considered to be a simple solution, it may not address 3D chip stacks in which there may be multiple high power processor chips stacked on top of each other yielding significantly high total chip stack powers.

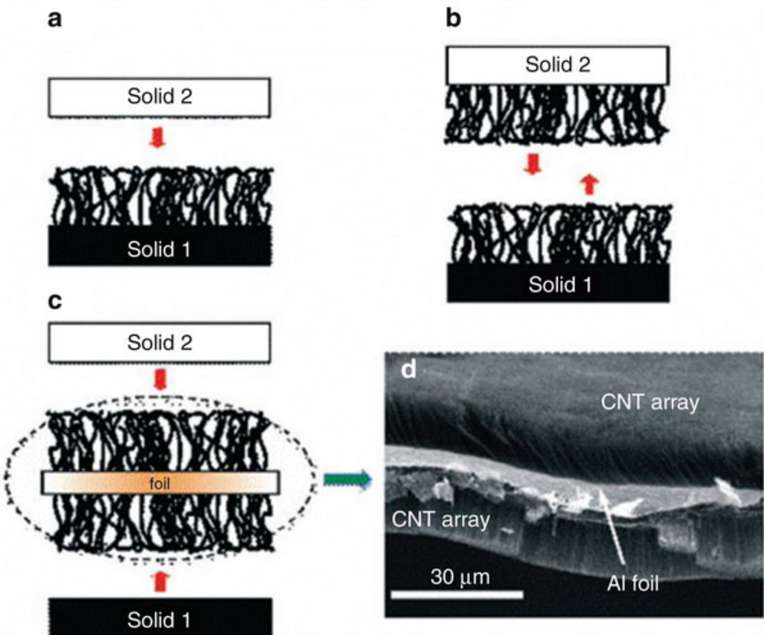


Fig. 9.38 CNT based thermal interface structures [62]

Such a configuration would also result in very large thermal resistances for the heat rejection path for the chips further away from the cooling device if the cooling is only provided via one side of the 3D chip stack.

Cooling and thermal management of 3D chip packages continues to be an active area for research and development. An example of recent studies include the following: (1) characterization of interface and conduction thermal resistances inside the stacked die package (and effects of various interconnects and through vias); and (2) liquid cooling of 3D chips using dielectric [58] and water [59] as the coolant, and thermo-mechanical modeling of 3D chips. A review of recent advances in cooling 3D chip stacks can be found in Venkatadri et al. [60].

9.20 Advanced Thermal Interfaces

Nano thermal interfaces is one the key areas of current research [61] and includes technologies such as double sided carbon nanotube (CNT) foils and metal nano-springs and nano-wires. The goal of these technologies is to further the attributes of existing thermal interfaces with respect to thermal resistance, reworkability, accommodation of lateral shear, and long-term reliability and consistency chip to chip. A description of 3D arrayed CNTs can be found in [62] where one sided and two sided CNT growths on foils have shown promising thermal results as a thermal interface material. Figure 9.38 shows these exploratory structures [62].

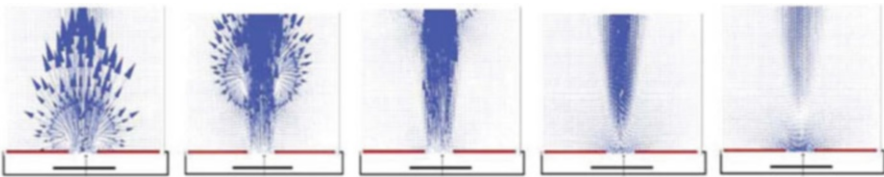


Fig. 9.39 Particle image velocimetry data for a synthetic jet [67]

The entanglement of CNT tubes that are grown on both engaging surfaces can enhance the thermal performance of the interface while allowing for reworkability. One possible application of such a technology would be the application of CNTs to the bottom of a heat sink as well as the top of the lid, both of which are metal and lend themselves to the growth of CNTs which require a high temperature (>700 °C). Another application of this technology can be CNT growth on both sides of a 22 μm aluminum foil to form a stand-alone interface material that can be inserted between the two surfaces being engaged in thermal contact [62].

Nano-springs have been described in [63] in which C shaped and S-shaped Graphene springs are used to construct a compliant thermal interface which could also enhance thermal performance. Graphene is a sheet of carbon atoms that is bound together to form a thin film and CNTs can be formed by rolling such Graphene films [63]. Other types of nano thermal interface include foam like nano-structures that are filled with metal particle filled polymer materials (e.g., epoxies) [64]. Such nano-structured foam-like interfaces also have the potential to provide a compliant thermal interface which also performs well from a thermal perspective.

9.21 Synthetic Jets and Actuated Flow Devices

While many of the robust cooling techniques described in this chapter are focused on higher power applications, the advent of hand held and mobile computing and communication devices have led to a proliferation of lower power devices which also need cooling. While devices such as cell phones are typically cooled via natural convection heat transfer, synthetic jets [65–67] are one type of advanced technology that focus on improving the cooling capability using very low power fans and compact air moving structures. Synthetic jets usually employ a piezoelectric disk or cantilever which vibrates upon the application of oscillating voltage and this mechanical movement in conjunction with a specially designed casing can create both air disturbances and even air movement to a small degree. Such air disturbances and movement can provide significant enhancement to the natural convection heat transfer capability. Figure 9.39 shows particle image velocimetry data of a synthetic jet. Other techniques of actuation include electromagnetic, electrostatic, and combustion driven pistons.

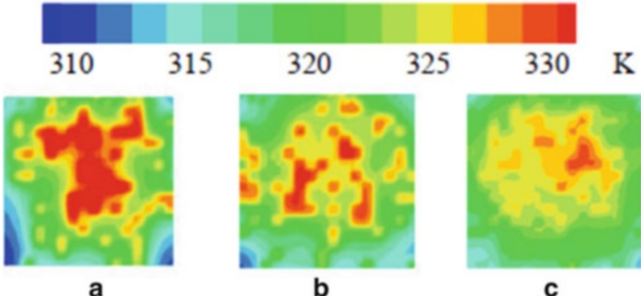


Fig. 9.40 Numerical results for chip temperature contours with and without multiplexing—(**a**) no multiplexing, (**b**) with multiplexing (**c**) with multiplexing and random core migration policy [70]

In addition to synthetic jets, cantilever based air movers [68] and ionic wind driven devices [69] are also being investigated for their product application potential for low flow and fan power applications.

9.22 Multi-Core Focused Chip Cooling

Another technique reported in the literature to address the cooling challenge is the control of multi-core microprocessors using a technique called multiplexing [70] or core hopping. Applying this technique, different cores of a microprocessor are fired up sequentially using a predetermined algorithm, so as allow some time for core cool down so that the time average temperature of each core is less than if a particular core was always running at peak performance. Such methods also serve to create a more uniform temperature over the surface of the multi core chip. Figure 9.40 shows numerical results for chip device side temperature contours without multiplexing and using two different algorithms (two different time constants for core workload movement) for multiplexing [70].

9.23 Thermoelectric Enhanced Cooling

Thermoelectric coolers provide the potential to enhance the cooling of integrated circuit chip packages to reduce chip operating temperatures at a given power dissipation level or to allow higher power dissipation levels at the same temperature. A thermoelectric cooler is a solid-state heat pump which transfers heat from one side of the device to the other side electronically against the temperature gradient (from cold to hot), with consumption of electrical energy. Thermoelectric coolers offer the advantages of being compact, quiet, and free of moving parts. The degree of cooling which they produce may be controlled by regulating the direct current supplied.

Unfortunately, compared to vapor-compression refrigeration, they are limited in the heat flux that they can accommodate. They also exhibit a lower coefficient of performance (COP), which simply is the ratio of the heat pumped to the energy required to perform the pumping. These two limitations have generally confined thermoelectrics to niche applications characterized by relatively low heat flux.

Although the principle of thermoelectricity dates back to the discovery of the Peltier effect in 1834 [71], there was little practical application of the phenomenon until the middle 1950s. Before that, the poor thermoelectric properties of known materials made them unsuitable for use in a practical refrigerating device. From the mid-1950s to the present, the major thermoelectric material design approach was that introduced by Ioffe, leading to semiconducting compounds such as Bi_2Te_3, which is currently used in commercial thermoelectric coolers [72]. These materials made possible the development of practical thermoelectric devices for attaining temperatures below ambient without the use of vapor-compression refrigeration.

In recent years there has been increased interest in the application of thermoelectrics to cooling electronic packages [73–77]. The usefulness of thermoelectric materials for solid-state cooling is often characterized by the dimensionless product, ZT, of the thermoelectric figure of merit Z and temperature T (in K). The value of the thermoelectric figure of merit of a material is given by

$$Z = \frac{\alpha^2}{\rho \cdot k}$$

where α is the Seebeck Coefficient, ρ is the electrical resistivity, and k is the thermal conductivity. Typically, the maximum value of ZT for the materials used in commercial thermoelectric coolers today is around one. However, effort is being focused on improving the performance of thermoelectric cooling devices through the development of new bulk materials and thin film microcoolers [78].

Thermoelectrically enhanced cooling may be applied at either the module or chip level. An example of the reduction in module case temperature that could be achieved applying a thermoelectric cooler to a 126 $\times$ 126 mm MCM containing thirty 15 $\times$ 15 mm integrated circuit chips is shown in Fig. 9.41 [74]. In this study, commercial thermoelectric cooling modules were assumed to be packaged between the top surface of the MCM and either a water-cooled cold plate or an air-cooled heat sink. The reduction in case temperature that could be achieved is shown as a function of total MCM power dissipation. The chip temperatures within the MCM would be reduced by an equal amount.

Comparisons of allowable module power dissipation with and without thermoelectric augmentation are shown in Fig. 9.42 as a function of maximum allowable chip temperature for the example MCM for both air-cooling and water-cooling conditions. It can be seen that for the air-cooled case, at a chip temperature of 47 °C, the allowable module power dissipation is the same with or without augmentation. At higher chip temperatures the allowable module power dissipation is higher without thermoelectric augmentation. A similar situation occurs for the water-cooled case at a chip temperature of 31 °C. In both cases the power dissipation at which this occurs at is about

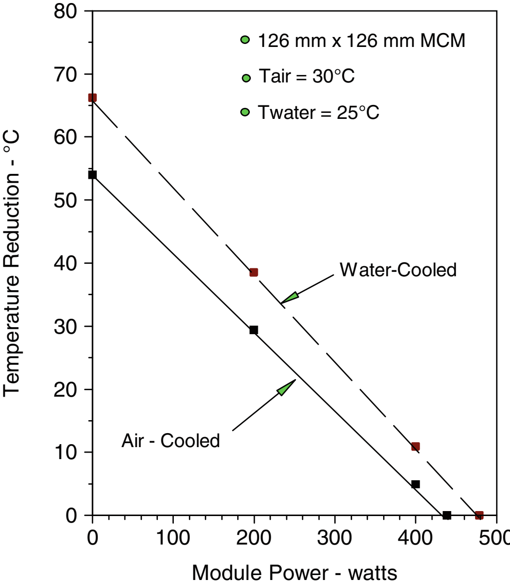


Fig. 9.41 Module temperature reduction with thermoelectric cooler augmentation for air or water-cooled 126 × 126 mm MCM package

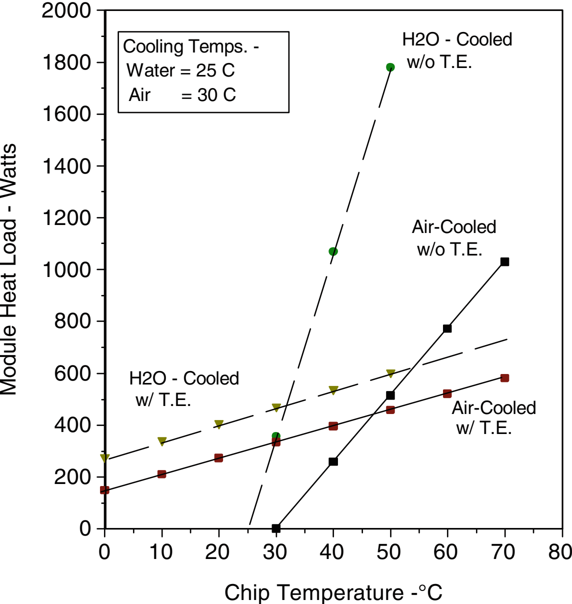


Fig. 9.42 Comparison of allowable module heat load with and without thermoelectric cooling enhancement

450–500 W. Clearly, specific cases for which the application of thermoelectrics are being considered must be carefully assessed to determine if the use of thermoelectrics will be beneficial.

More recently, interest has focused on the application of thermoelectric cooling at the chip level to address chip hot spots [79–85]. For example, one of the approaches [84]

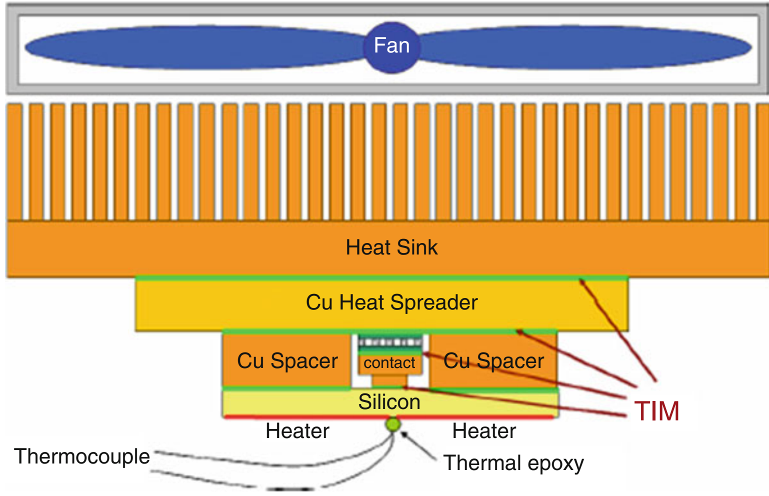


Fig. 9.43 Schematic of test vehicle for mini-contact enhanced TEC for chip hot spot cooling [84]

reported utilizes a miniaturized thermoelectric with a mini-contact directly over the hot spot as shown in Fig. 9.43. It was reported that numerical simulation of a chip with 400×400 μm hot spot with a heat flux of 1,250 W/cm^2, the peak hot spot temperature could be reduced by as much as 17 °C using a 20 μm thick thermoelectric cooler combined with a near perfect contact well below 10^{-5} K m^2/W.

9.24 Measurements and Modeling

Thermal measurements play an essential role in the development and characterization of any electronic package, including flip chip packages. Equally important is the ability to model such packages to predict their thermal response in an operational environment and to trade-off different options for improving their thermal performance. The following section addresses these two areas in terms of package thermal measurements, temperature measurement devices and methods, thermal measurement standards, compact thermal models, and finite element/computational fluid dynamic modeling.

9.25 Package Thermal Measurements

Thermal measurements have longed played an important role in the evolution and continued development of electronic components and systems. Accurate measurements of temperature and power are required to characterize the thermal resistance of integrated circuit chip packages from the chip heat source to the

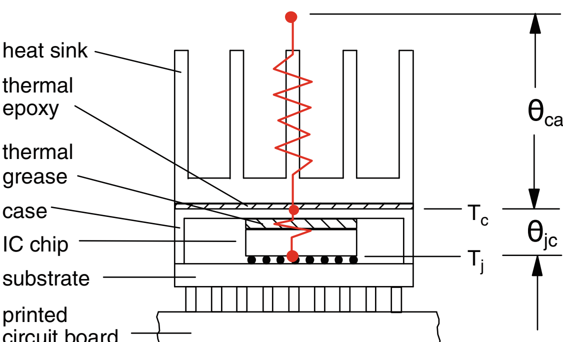


Fig. 9.44 Example of basic thermal measurement to characterize thermal performance of a flip chip module package

cooling sink. This can also include the need for accurate measurement of coolant flow rates and temperature. Temperature measurements may also be helpful in determining individual thermal resistances in one or more of the heat flow paths within an integrated circuit chip package. Temperature measurement on prototype packages has been a long standing approach to verify the adequacy of cooling designs. Thermal measurements can also play an important role in assessing and verifying the quality of incoming vendor packages.

A simple example of a basic thermal measurement to characterize a flip chip module package is illustrated in Fig. 9.44. In this case, it is desired to characterize the thermal resistance from the chip to cooling air, which is represented as Θ_{ja} or R_{ja} (as is often found in the literature) and is given by

$$\Theta_{ja} = (T_j - T_a)/P_c$$

where T_{j} is chip or junction temperature, T_{a} is the ambient air-cooling temperature and P_{c} is the chip power. Thus, if temperature is expressed in °C and power in watts, the measured thermal resistance will have units of °C/W. Quite often it is of interest to determine how much of the thermal resistance occurs within the package from chip to case (termed internal thermal resistance) and how much occurs external to the package from case to air (termed external thermal resistance). In Fig. 9.44, the internal thermal resistance, Θ_{jc}, is comprised of the thermal conduction resistances across the chip, a thermal grease layer, and package lid and is given by

$$\Theta_{\mathrm{jc}} = \frac{T_{\mathrm{j}} - T_{\mathrm{c}}}{P_{\mathrm{c}}}$$

where T_{c} is the case temperature. The external thermal resistance, Θ_{ca}, is comprised of the thermal conduction resistance across a layer of thermal epoxy, the conduction resistance across the heat sink base and then a combined conduction and convective resistance from the heat sink to ambient air.

$$\Theta_{\mathrm{ca}} = \frac{T_{\mathrm{c}} - T_{\mathrm{a}}}{P_{\mathrm{c}}}$$

The ambient air temperature is usually the temperature of the air approaching the package. By breaking down the overall thermal resistance in terms of Θ_{jc} and Θ_{ca}, it is often possible to determine how much of the allowable temperature rise budget is being consumed within and outside the package, and where effort is best directed to reduce thermal resistance to receive the maximum benefit.

9.26 Temperature Measurement Device and Method

A variety of sensors and techniques exist for measuring temperature. The various sensing devices include thermocouples, thermistors, resistance thermometers also known as resistance temperature detectors or RTDs [86–88]. Typically, the most commonly used devices for temperature measurements external to the flip chip or other types of electronic package assemblies are thermocouples. Thermocouples consist of two dissimilar metal conductors (wires) and produce a voltage proportional to a temperature difference between either end of the pair of conductors. Thermocouples offer the advantages of being simple, relatively inexpensive, rugged and flexible, yet reasonably accurate over a wide range of temperatures. However, it is generally intrusive and not practical to attempt to attach temperature sensors such as thermocouples to a flip chip within a module package. The method usually used to measure so-called chip or junction temperature within a flip chip package using temperature sensitive electrical parameters (TSEP) on the chip itself [89, 90]. For example, the temperature sensitive electrical device may be an isolated diode on an actual product chip or on a special test chip that will be used to mimic the thermal performance of an actual product chip. Prior to thermal testing, the temperature sensitive parameter is calibrated in a constant temperature bath or oven [9, 89, 91]. When the sensor is a diode, a very low current is passed through the emitter base junction acting as a diode, and the forward voltage drop is measured as a function of the oven or bath temperature. It is important that measurement current is large enough to produce a voltage reading not influenced by surface leakage effects and small enough not to result in significant self heating. The same measurement current is then applied to the sensing device in the application environment and the voltage drop is read. By entering the calibration curve or table, the corresponding temperature may be determined. Further details relating to this temperature measurement technique may be found in the EIA/JEDEC Standard JESD51-1 [92].

9.27 Temperature Measurement Standards

Depending under what conditions a given module package is tested, the thermal resistances (Θ_{ja} or R_{ja}) could vary greatly from one test to another. For example, referring to Fig. 9.44 imagine Test 1 wherein a forced convection air flow is applied

on the heat sink side of the test board with natural convection air on the opposite side and also imagine that the board is only slightly larger than the module plan form. Under these conditions the thermal resistance path depicted in Fig. 9.44 would be a quite reasonable one with most of the heat going out via the heat sink. Now, imagine Test 2 wherein a forced convection flow is applied on both sides of the board and the test board area is a good deal larger than the module plan form area. Under these conditions the thermal resistance path depicted in Fig. 9.44 would not be so reasonable, with a fairly large fraction of the heat flowing through the pins and out to cooling air not only from the board area directly beneath the test module but also via the surrounding board area. Comparing the results of these tests, the thermal resistances measured in Test 2 could be considerably lower than in Test 1. In addition, the so-called internal thermal resistance (Θ_{jc}) measured in Test 2 would appear smaller than that observed in Test 1, even though no change was made to the module package itself. If these two tests were conducted by two different vendors offering a similar module package, without understanding the test conditions, based upon the reported thermal resistance results it might appear to an engineer choosing between the two vendors that one had a better thermal design than the other. To avoid such situations it is important that the test conditions for which thermal results are reported are clearly understood.

Thermal standards are essential to clearly defining and understanding test conditions for which thermal performance data is reported. The JEDEC (Joint Electron Devices Engineering Council) JC15.1 Subcommittee has been involved with the generation of thermal measurement and modeling standards for microelectronic packages since 1990 [93]. Although these standards are too broad to discuss in detail here, it may be noted that the published, proposed and suggested standards cover the areas of thermal measurement, thermal environment, component mounting, device construction, thermal modeling and measurement application. The various thermal standards are available to download via the JEDEC website (http://www.jedec.org).

9.28 Compact Thermal Models

By the late 1980s thermal researchers began to focus and report on the variability of measured package resistances depending on test cooling conditions. For example, in 1988, Andrews [94] reported a detailed study of the problem in the context of the variability of overall thermal resistance (R_{ja}) in extrapolating from test conditions to application conditions. A model for translating R_{ja} from evaluation conditions to thermal performance under actual application conditions was described. The following year, a paper by Bar-Cohen [95] considered this problem and proposed a new approach providing “a universal, though modified θ_{jc} for use with a particular package in every possible packaging configuration.” Then in the early 1990s, a European group made up primarily of end-user companies formed the DELPHI (Development of Libraries of Physical models for an Integrated design

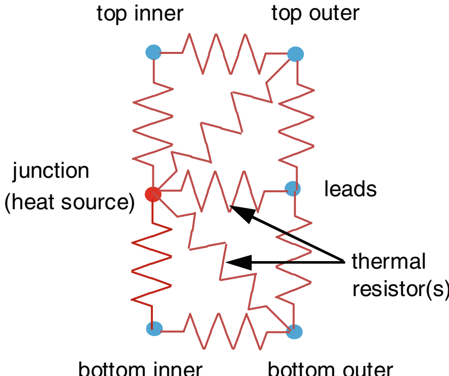


Fig. 9.45 A typical DELPHI compact thermal model topology (adapted from [97])

Environment) consortium [96]. This project ran from 1993 to 1996 with the objective of developing a methodology that could be adopted by component manufacturers and would enable them to produce validated thermal models to pass on to the users of their components.

A major outcome of the DELPHI effort was the widespread recognition of the utility of the compact thermal model (CTM) and its adoption throughout the industry. The CTM differs from the classical detailed thermal model (DTM). The DTM attempts to represent the actual physical geometry and major heat flow paths in a package to the extent that it is feasible. To achieve the required degree of accuracy, a DTM may be quite complex which limits its inclusion in models representing a board or system with many components. Its construction also requires a detailed knowledge of the internal construction of a component such as a flip chip package, which most vendors consider proprietary and are unwilling to disclose. However, as noted by Shidore [97], a properly constructed DTM is by definition boundary condition independent (BCI) and will accurately predict the temperature a various points within the package and regardless of the cooling environment in which it is placed. In contrast to the DTM, a CTM is a behavioral model intended to accurately predict the temperature of a package only at a few key points. It does not attempt to capture the topology and material properties of the component it represents. Typically a CTM will be composed of many fewer resistor elements than a DTM. It may therefore represent a much lesser computational challenge to incorporate CTMs of many such packages in a large board or system model.

The key challenge in creating a CTM is to create one that is boundary condition independent over a wide range of system application environments. A typical DELPHI compact model is shown in Fig. 9.45.

To generate such a model, the resistor network and the values of each element in it, are derived via a step-by-step process of simulation and statistical optimization minimizing the errors in junction temperature over a wide range of environments. For example, the range of environments may include high and low thermal conductivity printed circuit board, bare package and package with heat sink, and natural and forced convection cooling conditions. The end result is a CTM, which is a

major improvement of the simple two resistor model, providing prediction accuracy of junction temperature and the major heat fluxes within 10 % [97].

As noted earlier, component vendors are usually unwilling to reveal details of the internal make-up and construction of their packages. Unlike a DTM, a CTM does not reveal anything about a vendor package other than its projected thermal performance in the system users' application environment. Further details on the development and application of compact thermal models may be found in the JEDEC DELPHI Compact Thermal Model Guideline document [98].

9.29 FEM/CFD Modeling

In the last two decades there has been a tremendous growth in the use of commercial numerical solvers that can be found as generic solvers or programs that specialize in solving electronics cooling problems. The pervasive use of these tools as a indispensible part of any engineering research or development process is due to both the scientific progress in the development of numerical methods as well as improvements in computer computational performance and the increased access of these resources for packaging and thermal engineers. These solvers solve for the various heat transfer physics prevalent in electronics cooling applications such as conduction through materials and convection and radiations from surfaces. These programs use mathematical techniques available in literature such as fine element modeling (FEM), finite difference modeling (FDM), and finite volume modeling (FVM) just to name a few [99, 100].

A typical chip electronics cooling model may be for a module that includes a chip power map on the transistor side of the chip, a silicon die, a first thermal interface material between the back side of the die and the lid, the metal lid, and a heat transfer boundary condition at the top of the lid. Figure 9.46a shows a schematic of such a model. Another common model may be that for an air-cooled heat sink shown in Fig. 9.46b, with a base and fins with air flow between the fins and a heat flux boundary. Figure 9.46a, b could be combined for a more complete solution with the addition if a second thermal interface material between the top of the lid and the bottom of the heat sink.

Typical inputs into these models include a description of the geometry, thermo-physical parameters for the various materials and fluids in the domain, boundary conditions for the problem such as a heat flux into the domain and ingress of coolant into the domain at a particular temperature and egress of coolant from the domain. The numerical program aides the engineers in building the model by providing software utilities to help build the geometry and the assignment of properties and boundary conditions. Comprehensive libraries for solid and liquid materials commonly used in electronics applications (copper, aluminum, air, water, FR4, grease, etc.) are provided with typical properties listed from literature. Temperature variant properties can often be used.

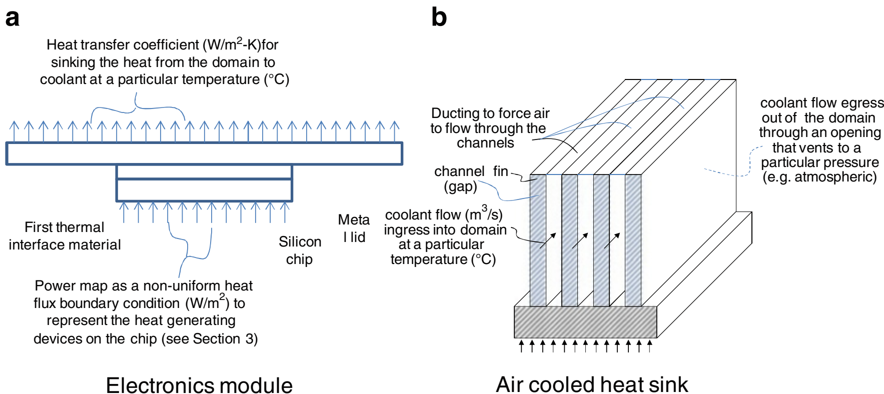


Fig. 9.46 Schematic of typical module level models in electronics cooling, (**a**) electronics module with power map, (**b**) air-cooled heat sink

The model discretizes the problem domains into small geometric pieces (e.g., nodes or volumes) which is called a mesh, constructs a large number of simultaneous equations where the variables represent parameters for each geometric entity (e.g., temperature or fluid velocity), and then solves these equations to yield values for these parameters. The solution of the model equations occur using iterative methods and the solutions results are reported after each iteration. The users either uses default recommendations (from software manufacturer) for solution residuals or input their own values. Solution residuals characterize the extent of numerical error arising from the solution of the equations. A large error means the solution data is not meaningful in describing the physics of the problem and vice versa. The mesh size determines the number of the equations that need to be solved and thus the requisite computation power and solution time. Unsurprisingly, effort has been devoted to allow for finer mesh in regions of greater interest to the engineer or in physical parts of the domain where there might exist large gradients in solved quantities (e.g., temperature or fluid velocity). Also, while 3D structures are very commonly used to represent physical entities, shell elements or two-dimensional (2D) elements are also mixed with 3D ones to reduce the solution time. Two-dimensional elements are commonly used where the gradients in the third dimension can be ignored, e.g., a thermal interface material or a heat sink can be modeled as 2D plane elements.

As may be expected, there is significant software code that is under the covers that addresses sophisticated mesh generation as well as the solutions of the equations by leveraging substantial literature that has been done over several decades in these areas. With the advent of supercomputer clusters as well as work stations or laptops with multi-core microprocessors, parallel computing of these models is an emerging field and impacts how different parts of the domain are solved separately (on different cores or microprocessors) and periodically the

boundaries of these solutions are coupled so that the final result is consistent across the full geometry.

Although there has been an exponential increase in the computational capability available to thermal engineers with stand alone work stations or even a mobile device such as a laptop, engineers will often significantly simplify the problem before building a "thermo fluid" version of it for analyses. There are several practical reasons for this approach. For example, while various mechanical design details are indeed important for a structural analyses (flanges, chamfers, fillets, nuts and bolts, through holes), such details are often not required to determine the thermal behavior of the system and would only serve to increase the size of the model and thus the computational time for solution. Such detail could also complicate the model and lead to unsatisfactory convergence and thus an inaccurate solution. However, having said this, there is significant interest in creating synergy between various engineering disciplines related to the design of electronics packages and systems so as to allow the sharing of CAD designs and rapid interaction and streamlines technical information dissemination between different engineering teams as the design progresses from the concept stage to the final part/system. One of the key aspects of a software system that allows such cross-physics sharing of design data would be the appropriate filtering of design details so as to customize the use of the information appropriate for each engineering discipline (thermal, mechanical, acoustics, thermo-mechanical, EMC) without excessive involvement of an expert.

References

1. Reynell M (1990) Thermal analysis using computational fluid dynamics. Electron Packag Prod
2. Chu RC, Simons RE, Ellsworth MJ, Schmidt RR, Cozzolino V (2004) Review of cooling technologies for computer products. IEEE Trans Device Mater Reliab 4(4):568–585
3. Incropera FP, DeWitt DP (2006) Introduction to heat transfer. Wiley, Hoboken, NJ
4. Holman JP (2009) Heat transfer. McGraw-Hill, New York
5. Yeh L-T, Chu RC (2002) Thermal management of microelectronic equipment: heat transfer theory, analysis methods and design practices. ASME, New York
6. Howell JR, Siegal R, Menguc MP (2010) Thermal radiation heat transfer, 5th edn. CRC Press, New York
7. Modest MF (2003) Radiative heat transfer, 2nd edn. Academic, New York
8. Simons RE Using a matrix inverse method to solve a thermal resistance network. http://www.electronics-cooling.com/2009/05/using-a-matrix-inverse-method-to-solve-a-thermal-resistance-network/, Accessed Feb 2013
9. Simons RE, Antonetti VW, Nakayama W, Oktay S (1997) Heat transfer in electronic packages. In: Tummala RR, Rymazewski EJ, Klopfenstein AG (eds) Microelectronics packaging handbook—technology drivers (Part I). Chapman & Hall, New York, NY
10. Wei X, Marston K, Sikka K (2008) Thermal modeling for warpage effects in organic packages. In: Proceedings of the intersociety conference on thermal phenomena (ITherm), Lake Buena Vista, FL

11. Watwe A, Vishwanath R (2003) Thermal implications of non-uniform die power and CPU performance. In: Proceedings of ASME InterPack conference, Maui, Hawaii, USA
12. Banerjee K, Mahajan R. Advanced cooling strategies—assembly technology development. Intel corporation. http://www.intel.com, Accessed Dec 2005
13. Iyengar M, Schmidt R (2006) Analytical modeling for prediction of hot spot chip junction temperature for electronics cooling applications. In: Proceedings of the InterSociety conference on thermal phenomena (ITherm), San Diego, CA
14. Hamann H, Lacey J, Weger A, Wakil J (2006) Spatially-resolved imaging of microprocessor power (SIMP): hotspots in microprocessors. In: Proceedings of the InterSociety conference on thermal phenomena (ITherm), San Diego, CA
15. Hamann H, Weger A, Lacey J, Cohen E, Atheton C Power distribution measurements of the dual core PowerPCTM970MP microprocessor. In: Proceedings of 2006 Solid-State Circuits Conference (ISSCC 2006), San Francisco, CA
16. Etessan-Yazdani K, Hamann H, Asheghi M (2007) Spatial frequency domain analysis of heat transfer in microelectronic chips with applications to temperature aware computing. In: Proceedings of the ASME InterPack conference, Vancouver, Canada
17. Torresola J, Chiu C, Chrysler G, Grannes D, Mahajan R, Prasher R, Watwe A (2005) Density factor approach to representing impact of die power maps on thermal management. IEEE Trans Adv Packag 28(4):1521–3323
18. Culham JR, Yovanovich YY, Lemczyk TF (2000) Thermal characterization of electronic packages using a three dimensional Fourier series solution. Trans ASME J Electron Packag 122:233–239
19. Muzychka YS, Culham JR, Yovanovich YY (2003) Thermal spreading resistance of eccentric heat sources on rectangular flux channels. Trans ASME J Electron Packag 125:178–185
20. Muzychka YS (2004) Influence coefficient method for calculating discrete heat source temperature on finite convectively cooled substrates. In: Proceedings of the ITHERM conference, vol 1, pp 394–402
21. Sikka K (2004) An analytical temperature prediction method for a chip power map. In: Proceedings of the SEMITHERM conference, San Jose, CA
22. Kadambi V, Abuaf N (1985) An analysis of the thermal response of power chip packages. IEEE Trans Electron Devices ED-32(6):1024–1033
23. Prasher R, Chang J, Sauciuc I, Narasimhan S, Chau D, Chrysler G, Myers A, Prstic S, Hu C (2005) Nano- and micro technology-based next-generation package-level cooling solutions. Intel Technol J 9(4):285–296
24. Furman BK, Lauro PA, Shih DY, Van Kessel T, Martin Y, Colgan EG, Zou W, Iruvanti S, Wakil J, Schmidt R, Iyengar MK (2005) Metal TIM's for high power cooling applications. International microelectronics and packaging society—IMAPS, advanced thermal workshop, Palo Alto
25. Song S, Lee S, Au V (1994) Closed form equation for thermal constriction/spreading resistances with variable resistance boundary condition. In: Proceedings of the 1994 IEPS conference, pp 111–121, Atlanta, GA
26. Kennedy DP (1960) Spreading resistance in cylindrical semiconductor devices. J Appl Phys 31:1490–1497
27. Lee S, Song S, Au V, Moran KP (1995) Constriction/spreading resistance model for electronic packaging. In: Proceedings of the 4th ASME/JSME thermal engineering joint conference, vol 4. pp 199–206, Maui, Hawaii
28. Fischer TS, Zell FA, Sikka KK, Torrance KE (1996) Efficient heat transfer approximation for the chip-on-substrate problem. J Electron Packag 118:271–279
29. Kraus AD, Bar-Cohen A (1995) Design and analysis of heat sinks. Wiley, New York
30. Iyengar M, Bar-Cohen A (2001) Design for manufacturability of SISE parallel plate forced convection heat sinks. IEEE Trans Compon Packaging Technol 24(2):150–158
31. Ellsworth J et al (2011) An overview of the IBM power 775 supercomputer. In: Proceedings of the ASME InterPack conference, Portland, OR

32. Iyengar M, David M, Kamath V, Kochparambil B, Graybill D, Schultz M, Gaynes M, Simons R, Schmidt R, Chainer T (2012) Server liquid cooling with chiller-less data center design to enable significant energy savings. In:Proceedings of 28th IEEE SEMI-THERM Symposium, San Jose, CA
33. Iyengar M, Garimella S (2006) Analytical modeling for prediction of hot spot chip junction temperature for electronics cooling applications. In: Proceedings of the InterSociety conference on thermal phenomena (ITherm)
34. Schmidt RR (2000) Low temperature electronic cooling. 6(3):18–24
35. Peeples JW (2001) Vapor compression cooling for high performance applications. Electron Cooling 7(3):16–24
36. Schmidt RR, Notohardjono BD (2002) High-end server low-temperature cooling. IBM J Res Dev 46(6):739
37. Goth GF, Kearney DJ, Meyer U, Porter D (2004) Hybrid cooling with cycle steering in the IBM eServer z990. IBM J Res Dev 48(3/4):409
38. Torok JG, Bosco FE, Brodsky WL, Furey EF, Goth GF, Kearney DJ, Loparco JJ, Peets MT, Pizzolato KL, Porter DW, Ruehle G, White WH (2009) Packaging design of the IBM system z10 enterprise class platform central electronic complex. IBM J Res Dev 53(1):9
39. Bar-Cohen A, Wang P (2011) On-chip thermal management of nanoelectronic hot spots. In: InterPack tutorial, ASME InterPack conference, Portland, OR
40. Kolodzey JS (1981) CRAY-1 computer technology. IEEE Trans CHMT-4(2):181–186
41. Armstrong RJ (1967) Cooling components with boiling halocarbons. IEEE Trans PMP-3 (4):135–142
42. Cochran DL (1968) Boiling heat transfer in electronics. Electron Packag Prod 8(7):CL3–CL7
43. Preston SB, Shillabeer RN (1970) Direct liquid cooling of microelectronics. In: InterNEPCON proceedings:10–31
44. Simons RE (1996) Direct liquid immersion cooling for high density microelectronics. Electron Cooling 2(2):24–29
45. Bar-Cohen A, Arik M, Ohadi M (2006) Direct liquid cooling of high heat flux micro and nano electronic components. Proc IEEE 94(8):1549–1570
46. Bergles AE, Bar-Cohen A (1994) Immersion cooling of digital computers. In: Kakac S, Yuncu H, Hijikata K (eds) Cooling of electronic systems. Kluwer Academic Publishers, Boston, MA
47. Tuma P, Hesselroth D, Brodbeck T (2009) Next-generation dielectric heat transfer fluids for cooling military electronics. http://www.mil-embedded.com/articles/id/?4039, Accessed Feb 2013
48. Danielson RD, Krajewski N, Brost J (1986) Cooling a superfast computer. Electron Packag Prod 26(7):44–45
49. Knickerbocker JU, Andry PS, Dang B, Horton RR, Interrante MJ, Patel CS, Polastre RJ, Sakuma K, Sirdeshmukh R, Sprogis EJ, Sri-Jayantha SM, Stephens AM, Topol AW, Tsang CK, Webb BC, Wright SL (2008) Three-dimensional silicon integration. IBM J Res Dev 52 (6):571–581
50. Rhett W, Wilson J, Mick S, Xu J, Hua H, Mineo C, Sule A, Steer M, Franzon P (2005) Demystifying 3D ICs: the pros and cons of going vertical. IEEE Des Test Comput 22(6):498–510
51. Hummler K (2011) Stacked IC packaging with TSV true 3D without glasses. In: Presentation at the electronics packaging symposium in Binghamton, New York
52. Sri-Jayantha SM, McVicker G, Bernstein K, Knickerbocker JU (2008) Thermomechanical modeling of 3D electronic packages. IBM J Res Dev 52(6):539–540
53. Emma PG, Kursun E (2008) Is 3D chip technology the next growth engine for performance improvement? IBM J Res Dev 52(6):541–552
54. Dang B, Wright SL, Andry PS, Sprogis EJ, Tsang CK, Interrante MJ, Webb BC, Polastre RJ, Horton RR, Patel CS, Sharma A, Zheng J, Sakuma K, Knickerbocker JU (2008) 3D chip stacking with C4 technology. IBM J Res Dev 52(6):571–581

55. Koester SJ, Young AM, Yu RR, Purushothaman S, Chen K-N, La Tulipe DC Jr, Rana N, Shi L, Wordeman MR, Sprogis EJ (2008) Wafer-level 3D integration technology. IBM J Res Dev 52(6):583–597
56. Sakuma K, Andry PS, Tsang CK, Wright SL, Dang B, Patel CS, Webb BC, Maria J, Sprogis EJ, Kang SK, Polastre RJ, Horton RR, Knickerbocker JU (2008) 3D chip-stacking technology with through-silicon vias and low-volume lead free interconnections. IBM J Res Dev 52(6):611–622
57. Oprins H, Cherman V (2012) Modeling and experimental characterization of hot spot dissipation in 3D chip stacks. Electron Cooling:18–23
58. Bar-Cohen A, Geisler K, Rahim E (2008) Pool and flow boiling in narrow gaps—application to 3D chip stacks. In: 5th European thermal-sciences conference, Netherlands
59. Brunschwiler T, Paredes S, Drechsler U, Michel B, Cesar W, Leblebici Y, Wunderle B, Reichl H (2011) Heat-removal performance scaling of interlayer cooled chip stacks. In: Proceedings of the thermal and thermomechanical phenomena in electronic systems (12th ITherm), San Diego, CA
60. Venkatadri V, Sammakia B, Srihari K, Santos D (2011) A review of recent advances in thermal management in three dimensional chip stacks in electronic systems. ASME J Electro Packag 133(4)
61. Avram Bar-Cohen Thermal management technologies. DARPA presentation. http://www.darpa.mil/WorkArea/DownloadAsset.aspx?id=2147484801, Accessed Feb 2013
62. Cola B (2010) Carbon nano tubes as high performance thermal interface materials. Electron Cooling 16(1):24–29
63. Tilak V, Nagarkar K, Tsukalakos L, Wetzel T (2010) Thermal management system with graphene based thermal interface material. US Patent Application US 2010/0128439A1, 2010
64. Pashayi K, Fard HR, Lai F, Iruvanti S, Plawsky J, Borca-Tasciuc T (2012) High thermal conductivity epoxy-silver composites based on self-constructed nanostructured metallic networks. J Appl Phys 111, 104310 (2012)
65. Smith B, Glezer A (1998) The formation and evolution of synthetic jets. Phys Fluids 10(9):2281–2297
66. Li R, Sharma R, Arik M (2011) Energy conversion efficiency of synthetic jets. In: Proceedings of the ASME 2011 Pacific Rim technical conference & exposition on packaging and integration of electronic and photonic systems (InterPACK), paper number IPACK2011-52034, Portland, USA
67. Mahalingam R, Heffington S, Jones L, Williams R (2007) Synthetic jets for forced air cooling of electronics. Electron Cooling 13(2):12–18
68. Acikalin T, Garimella SV, Raman A, Petroski J (2007) Characterization and optimization of the thermal performance of miniature piezoelectric fans. Int J Heat Fluid Flow 28(4):806–820
69. Schiltz DJ, Garimella SV, Fisher TS (2004) Microscale ion-driven air flow over a flat plate. In: ASME 2004 heat transfer/fluids engineering summer conference, paper no. HT-FED2004-56470, pp 463–468, Charlotte, NC
70. Gupta MP, Cho M, Mukhopadhya S, Kumar S (2012) Thermal investigation into power multiplexing for homogenous many-core processors. ASME J Heat Transf 134:061401
71. Godfrey S (1996) An Introduction to thermoelectric coolers. Electron Cooling 2(3):30–33
72. Nolas GS, Slack GA, Cohn JL, Scujman SB (1998) The next generation of thermoelectric materials. In: Proceedings of the 17th international conference on thermoelectrics, Nagoya, Japan
73. Simons RE (2000) Application of thermoelectric coolers for module cooling enhancement. Electron Cooling 6(2):18–24
74. Simons RE, Ellsworth MJ, Chu RC (2005) An assessment of module cooling enhancement with thermoelectric coolers. J Heat Trans 127(1):76–84
75. Johnson DA, Bierschenk J (2005) Latest developments in thermoelectrically enhanced heat sinks. Electron Cooling 11(3):24–32

76. Lasance CJM, Simons RE (2005) Advances in high-performance cooling for electronics. Electron Cooling 11(4):22–39
77. DeBock HP, Icoz T (2007) Evaluation on use of thermoelectric devices for electronics cooling. In: Proceedings of IPACK 2007, Vancouver, Canada
78. Habbe B, Nurnus J (2011) Thin film thermoelectrics today and tomorrow. Electron Cooling:24–31
79. Shakouri A, Zhang Y, Fukutani K (2006) Solid-state microrefrigerator on a chip. Electron Cooling 12(3)
80. Ramanathan S, Chrysler GM (2006) Solid-state refrigeration for cooling microprocessors. IEEE Trans Compon Packag Technol 29(1):179–183
81. Snyder GJ, Soto M, Alley R, Koester D, Conner R (2006) Hot spot cooling using embedded thermoelectric coolers. In: Proceedings 22nd IEEE semi-therm symposium, Dallas, TX
82. Lee KH, Kim OJ (2007) Simulation of the cooling system using thermoelectric micro-coolers for hot spot mitigation. In: Proceedings of 2007 international conference on thermoelectrics, Jeju, Korea
83. Yang B, Wang P, Bar-Cohen A (2007) Mini-contact enhanced thermoelectric cooling of hot spots in high power devices. IEEE Trans Compon Packaging Technol 30(3):432–438
84. Wang P, Bar-Cohen A, Yang B (2007) Enhanced thermoelectric cooler for on-chip hot spot cooling. In: Proceedings of IPACK 2007, Vancouver, Canada
85. Alley R, Soto M, Kwark L, Crocco P, Koester D (2008) Modeling and validation of on-die cooling of dual-core CPU using embedded thermoelectric devices. In: Proceedings 24th IEEE SEMI-THERM symposium, San Jose, CA
86. Azar K (1997) Thermal measurements in electronics cooling. CRC, Boca Raton, FL
87. Michalski L, Eckersdorf K (2001) Temperature measurement. Wiley, New York, NY
88. Childs PRN (2001) Practical temperature measurement. Butterworth-Heinemann, Woburn, MA
89. Sofia JW (1997) Electrical temperature measurement using semiconductors. Electron Cooling 3(1):22–25
90. Rencz M (1996) The increasing importance of thermal test dies, Electron Cooling 2(2)
91. Claassen A, Shaukatullah H (1997) Comparison of diodes and resistors for measuring chip temperature during thermal characterization of electronic packages using thermal test chips. In: Proceedings of the 13th SEMI-THERM symposium, Austin, TX
92. EIA/JESD51-1 (1995) Integrated circuits thermal measurement method—electrical test method (single semiconductor device). http://www.jedec.org/standards-documents/results/jesd-51-5, Accessed Feb 2013
93. Guenin B (2002) Thermal standards for the 21st Century. In: Proceeding of 18th IEEE SEMI-THERM symposium, San Jose, CA
94. Andrews JA (1988) Package thermal resistance model dependency on equipment design. In: Proceedings of 4th IEEE SEMI-THERM symposium, San Diego, CA
95. Bar-Cohen A (1989) θ_{jc} characterization of chip packages—justification, limitations, and future. In: Proceedings of the 5th IEEE SEMI-THERM symposium, San Diego, CA
96. Rosten HI, Parry JD, Lasance CJM, Vinke H, Temmerman W, Nelemans W, Assouad Y, Gautier T, Slattery O, Cahill C, O'Flattery M, Lacaze C, Zelianoy P (1997) Final report to SEMITHERM XIII on the European-funded project DELPHI—the development of libraries and physical models for an integrated design environment. In: Proceedings of the 13th IEEE SEMI-THERM symposium, Austin, TX
97. Shidore S (2007) Compact thermal modeling in electronics design. Electron Cooling 13(2)
98. JEDEC JESD 15-4 (2008) DELPHI thermal l compact model guideline. http://www.jedec.org/standards-documents/docs/jesd-15-4, Accessed Feb 2013
99. Mincowycz WJ, Sparrow EM, Schneider GE, Murthy JY (2006) Handbook of numerical heat transfer, 2nd edn. Wiley, New York
100. Patankar SV (1980) Numerical heat transfer and fluid flow. Hemisphere Publishing Corporation, New York

Chapter 10
Thermo-mechanical Reliability in Flip-Chip Packages

Li Li and Hongtao Ma

Abstract Reliability of the flip-chip packages is highly dependent on the properties of the constituent components and the interfaces formed among them. The relative mechanical compliances and thermal mismatch between the silicon chip, the underfill material and the package substrate (organic or inorganic based) are particularly important to the design and performance the package. Strong, thermo-mechanical, Chip–Packaging Interaction (CPI) can cause chip cracking, solder bump cracking, package substrate trace cracking, delamination of Interlayer Dielectrics (ILD) of the silicon chip, delamination of underfill encapsulation, and problems associated with the board-level interconnection when the package is assembled to a printed circuit board (PCB). These problems became more severe as we migrate to lead-free packaging materials for the leading silicon technology nodes such as 32 and 28 nm nodes where low-k and extreme low-k ILD materials are widely used. In addition to the thermo-mechanical stresses, moisture absorption in packaging materials especially at the critical interfaces, electrical current, manufacturing defects can also be the drivers for some the failure modes.

In this Chapter attention has been focused on reliability of flip-chip packages especially those with Cu/low-k chips. Combined experimental and modeling methods were used to investigate the thermo-mechanical behavior and failure mechanisms controlling the package reliability. Thermal deformation in flip-chip package assembly was first studied for minimizing the chip-substrate thermo-mechanical coupling. Thermo-mechanical response of the package was measured and analyzed using high-resolution moiré interferometry, analytical and numerical modeling techniques. Four-point bending test was also used to characterize interfacial fracture energy for the critical interface between die passivation and underfill material. The experiments and modeling were correlated with the JEDEC standard component-level reliability testing results. The combined experimental and numerical analysis provided a systemic approach for reliability assessment, package

L. Li (✉) • H. Ma
Cisco Systems, Inc., 170 West Tasman Drive, San Jose, CA 95134, USA
e-mail: lili2@cisco.com; hongtma@cisco.com

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_10, © Springer Science+Business Media New York 2013

materials selection. It also demonstrated that significantly improved reliability of the flip-chip PBGA packages can be achieved by controlling thermo-mechanical coupling of the silicon die and the package, and by enhancing various important interfaces within the package.

10.1 Overview of Thermo-mechanical Reliability Issues in Flip-Chip Packages

Reliability of the flip-chip packages is highly dependent on the properties of the constituent components and the interfaces formed among them. The relative mechanical compliances and thermal mismatch between the silicon chip, the underfill material and the package substrate (organic or inorganic based) are particularly important to the design and performance of the package [1]. As shown in Fig. 10.1, strong, thermo-mechanical, Chip–Packaging Interaction (CPI) can cause chip cracking, solder bump cracking, package substrate trace cracking, delamination of Interlayer Dielectrics (ILD) of the silicon chip, delamination of underfill encapsulation and also problems associated with the board-level interconnection when the package is assembled to a printed circuit board (PCB). These problems became more severe as we migrate to lead-free packaging materials for the leading silicon technology nodes such as 32 and 28 nm nodes where low-k and extreme low-k ILD materials are widely used. In addition to the thermo-mechanical stresses ("*T*" and "*F*" as shown in Fig. 10.1), moisture absorption in packaging materials especially at the critical interfaces ("*RH*"), electrical current ("*j*"), and manufacturing defects can also be the drivers for most of the failure modes shown in Fig. 10.1.

Driven by continued device integration, Cu interconnect and low-k dielectrics (Cu/low-k) became the desired choice staring from the 90 nm technology node. Benefits of Cu/low-k include the following:

- Reduced RC delay
- Reduced power dissipation
- Improved performance.

However, because of the methods employed to lower the dielectric constant of ILD, challenges do exist for the low-k ILD materials due to the nature of their porous structures. These challenges include the following:

- Difficulties in back-end process integration because of the intrinsically lower modulus, hardness, poor adhesion and tensile residual stresses of low-k ILD films
- Needs to optimize packaging processes including wafer sawing process
- Maintaining product reliability due to strong chip–package interaction (CPI).

Interfacial delamination in low-k ILD has been widely observed in the semiconductor industry. For instance, Chen et al. [2] reported low-k ILD delamination and

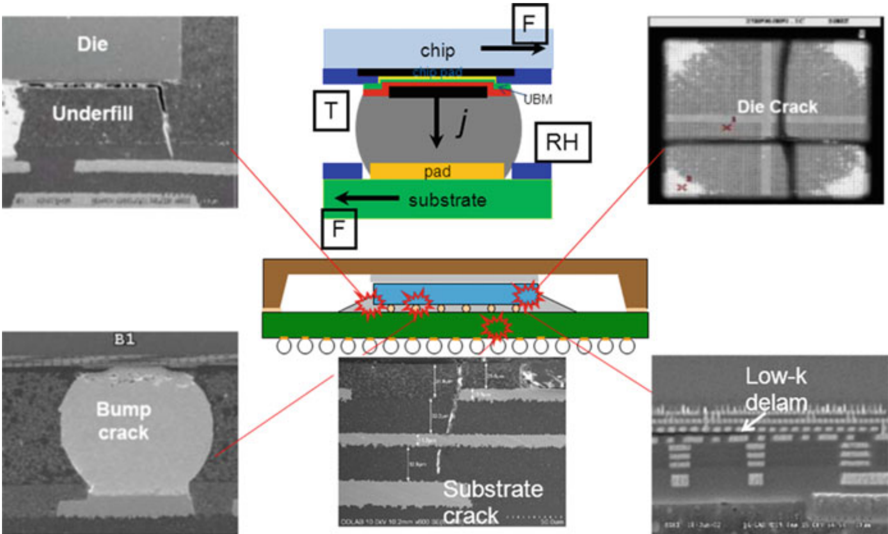


Fig. 10.1 CPI can cause chip cracking, solder bump cracking, package substrate trace cracking, delamination of ILD layer in the silicon chip, and delamination of underfill encapsulation for the flip-chip assembly

solder bump cracking during −55 to 125 °C temperature cycling test. They concluded that the failure is due to improper selection of underfill materials. Tsao et al. [3] also found low-k ILD layer delamination near the low-k ILD and silicon substrate interface. The ILD delamination occurred at the die corner where higher thermally induced stress is expected in the ILD to silicon substrate interface.

In flip-chip packages, the underfill material effectively redistributes the thermo-mechanical load on solder bumps providing a practical solution to the solder bump fatigue problem. To prevent solder bumps from fatigue cracking, the requirements for underfill are high T_g, high modulus and with a Coefficient of Thermal Expansion (CTE) closely matched to that of the solder material. While with a low-T_g underfill material we can effectively reduce die and package stresses at lower temperatures, we need to check if solder bump reliability is still intact. The recent study showed that the underfill material has a profound impact not only on the solder bump reliability but also on the chip–packaging interaction [4].

In this chapter attention has been focused on reliability of flip-chip packages especially those with Cu/low-k chips. Combined experimental and modeling methods were used to investigate the thermo-mechanical behavior and failure mechanisms controlling the flip-chip package reliability. Thermal deformation in flip-chip package assembly was first studied for minimizing the chip-substrate thermo-mechanical coupling. Thermo-mechanical response of the package was measured and analyzed using high resolution moiré interferometry, analytical and numerical modeling techniques. Four-point bending test was also used to characterize interfacial fracture energy for the critical interface between die passivation

and underfill material. The experiments and modeling were correlated with the JEDEC standard component-level reliability testing results. The combined experimental and numerical analysis provided a systemic approach for the reliability assessment and the packaging materials selection. It also demonstrated that significantly improved reliability of the flip-chip PBGA packages can be achieved by controlling thermo-mechanical coupling of the silicon die and the package, and by enhancing various important interfaces within the package. Modeling of solder bump reliability is also included with a brief discussion on the constitutive equations for solder and reliability analysis based both empirical and numerical modeling techniques.

10.2 Thermal Deformation in Flip-Chip Assembly

In flip-chip plastic ball grid array (FC-PBGA) packages, silicon die is attached on a laminate substrate by solder joints. The underfill material is filled in the gap between the die and the substrate to protect the solder joints for better reliability. After the underfill process, the die and the substrate are rigidly bonded together and no interface delamination and separation should be present. A silicon die, with a coefficient of thermal expansion (CTE) of 2.6 ppm/°C, and a laminate substrate, with a CTE from 15 to 25 ppm/°C, are connected by underfill material. As a result of CTE mismatch, significant thermal stress occurs in the die and the substrate during thermal cycles. In component level reliability testing, this thermal stress is the major cause of many failure modes including die cracking, underfill delamination and low-k ILD layer delamination. Lately in the development of FC-PBGA packages, reducing die stress and improving reliability has been a very serious issue.

10.2.1 Consistent Composite Plate Model

The flip-chip package can be treated as a multiplayer composite system. The consistent plate model treats the chip, the underfill and the carrier substrates, i.e., the chip/carrier module, as plies of an integrated laminated plate. The only assumptions made are those normal to thin and classical laminate plate theory [5].

The term consistent plate model comes from the assumption of consistent deformations in the chip and substrate. The strain in the system will therefore depend only on the strain on the reference plane somewhere in the middle of the plate, ε^0, and the curvature of the laminated plate, κ,

$$\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{Bmatrix} = \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \varepsilon_z^0 \end{Bmatrix} + z \begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_z \end{Bmatrix} \tag{10.1}$$

The constitutive relation for any ply of a laminated plate is

$$\sigma = \mathbf{Q}\,(\varepsilon - \Lambda), \tag{10.2}$$

where the stress vector σ and induced strain vector Λ are

$$\sigma = \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \sigma_{xy} \end{Bmatrix}, \quad \Lambda = \begin{Bmatrix} \Lambda_x \\ \Lambda_y \\ \Lambda_{xy} \end{Bmatrix} = \begin{Bmatrix} \alpha_x \Delta T \\ \alpha_y \Delta T \\ \alpha_{xy} \Delta T \end{Bmatrix} \tag{10.3}$$

Here α is the coefficient of thermal expansion and ΔT is the temperature change.

The matrix **Q** is the transformed reduced stiffness of the lamina and is given by Ref. [6] as following

$$\mathbf{Q} = \begin{Bmatrix} \overline{Q}_{11} & \overline{Q}_{12} & \overline{Q}_{16} \\ \overline{Q}_{12} & \overline{Q}_{22} & \overline{Q}_{16} \\ \overline{Q}_{16} & \overline{Q}_{16} & \overline{Q}_{66} \end{Bmatrix} \tag{10.4}$$

The load-deformation relationship for the consistent plate is given by

$$\begin{bmatrix} \mathbf{N} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} \begin{Bmatrix} \varepsilon^0 \\ \kappa \end{Bmatrix} - \begin{bmatrix} \mathbf{N}^{\Lambda} \\ \mathbf{M}^{\Lambda} \end{bmatrix}, \tag{10.5}$$

where the conventional mechanical stress resultants, the mechanical forces and moments, are

$$\mathbf{N} = \int_t \sigma \, \mathrm{d}z \tag{10.6}$$

$$\mathbf{M} = \int_t \sigma \, z \mathrm{d}z \tag{10.7}$$

The matrices **A**, **B**, and **D** are the usual extensional stiffness, bending-stretching coupling stiffness, and bending stiffness of the plate [6]. The integrations in the above equations are carried out through the composite plate thickness. The equivalent thermal forces and moments are

$$\mathbf{N}^{\Lambda} = \int_t \mathbf{Q}\Lambda \mathrm{d}z \tag{10.8}$$

$$\mathbf{M}^{\Lambda} = \int_t \mathbf{Q}\Lambda \, z \mathrm{d}z \tag{10.9}$$

The total potential energy stored in the plate is given by

$$U = \frac{1}{2}\iint\limits_{\Omega} \begin{Bmatrix} \varepsilon^0 \\ \kappa \end{Bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} \begin{Bmatrix} \varepsilon^0 \\ \kappa \end{Bmatrix} \mathrm{d}\Omega - \iint\limits_{\Omega} \begin{bmatrix} \mathbf{N}^{\Lambda} \\ \mathbf{M}^{\Lambda} \end{bmatrix}^{\mathrm{T}} \begin{Bmatrix} \varepsilon^0 \\ \kappa \end{Bmatrix} \mathrm{d}\Omega \tag{10.10}$$

This strain energy equation together with a Ritz approximate solution method can be used to solve for the approximate strains and curvatures in the multiplayer system.

10.2.2 *Free Thermal Deformation*

When there is no external mechanical load, (10.5) can be reduced to

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} \begin{Bmatrix} \varepsilon^0 \\ \kappa \end{Bmatrix} = \begin{bmatrix} \mathbf{N}^{\Lambda} \\ \mathbf{M}^{\Lambda} \end{bmatrix} \tag{10.11}$$

Equation (10.11) can be solved directly for ε^0 and κ.

As shown in Fig. 10.2, there are three special cases that are of particular interests for the electronics packaging analysis, namely, the beam bending, the cylindrical bending, and the axisymmetrical bending. When material anisotropy for each layer is small, (10.4) can be written as

$$\mathbf{Q} = \begin{Bmatrix} \frac{E_i}{1-v_i^2} & \frac{v_i E_i}{1-v_i^2} & 0 \\ \frac{v_i E_i}{1-v_i^2} & \frac{E_i}{1-v_i^2} & 0 \\ 0 & 0 & \frac{E_i}{2(1+v_i)} \end{Bmatrix} \tag{10.12}$$

Here E_i and v_i are the Young's modulus and Poisson's ratio for the ith layer lamina. Note (10.12) is given for the axisymmetrical case. For the cylindrical and beam bending cases, we can simply replace the term $\frac{E_i}{1-v_i}$ with E_i.

The deformation of the composite plate can be further simplified such that

for beam bending, $\sigma_y << \sigma_x$;

for cylindrical bending, $\begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix} = \begin{Bmatrix} \kappa_x \\ 0 \\ 0 \end{Bmatrix}, \quad \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \varepsilon_{xy}^0 \end{Bmatrix} = \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_x^0 \\ 0 \end{Bmatrix}$

for axisymmetrical bending, $\begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix} = \begin{Bmatrix} \kappa_x \\ \kappa_x \\ 0 \end{Bmatrix}, \quad \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \varepsilon_{xy}^0 \end{Bmatrix} = \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_x^0 \\ 0 \end{Bmatrix}$

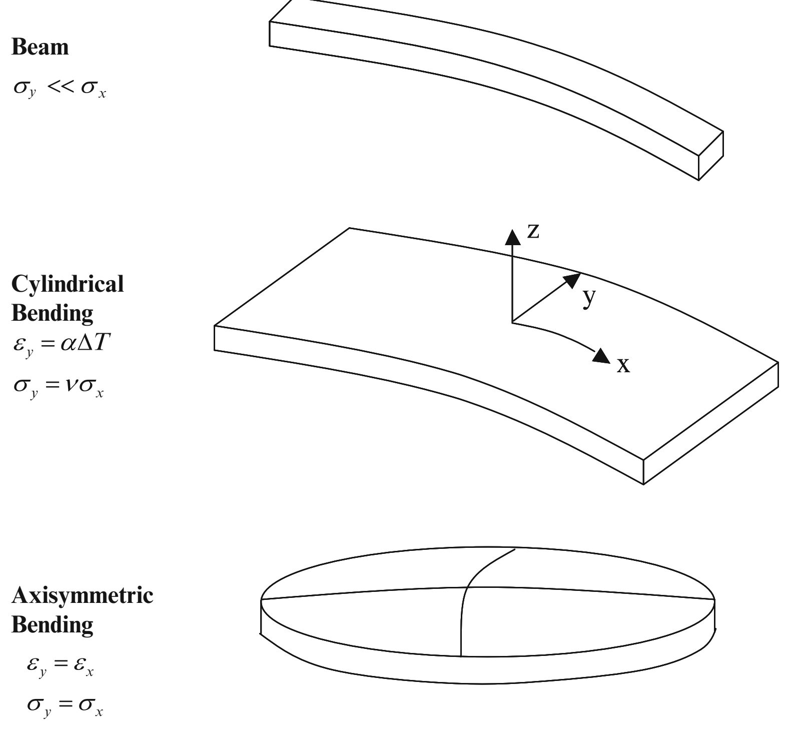


Fig. 10.2 Three possible deformation modes for multilayered structures

The force and the moment equations for the above three cases become

$$\begin{bmatrix} A_x & B_x \\ B_x & D_x \end{bmatrix} \begin{Bmatrix} \varepsilon_x^0 \\ \kappa_x \end{Bmatrix} = \begin{bmatrix} N_x^\Lambda \\ M_x^\Lambda \end{bmatrix} \tag{10.13}$$

Here coefficients $A_x, B_x, D_x, N_x^\Lambda$, and M_x^Λ are given by the following equations [5]

$$\begin{cases} A_x = \int\limits_t \frac{E_i}{1-\nu_i} \mathrm{d}z \\ B_x = \int\limits_t \frac{E_i}{1-\nu_i} z\mathrm{d}z \\ D_x = \int\limits_t \frac{E_i}{1-\nu_i} z^2\mathrm{d}z \\ N_x^\Lambda = \int\limits_t \frac{E_i\alpha_i\Delta T}{1-\nu_i} z\mathrm{d}z \\ M_x^\Lambda = \int\limits_t \frac{E_i\alpha_i\Delta T}{1-\nu_i} z\mathrm{d}z \end{cases} \tag{10.14}$$

Again, (10.14) is given for the axisymmetrical case. For cylindrical and beam bending cases, we simply replace the term $\frac{E_i}{1-\nu_i}$ with E_i in the above equation.

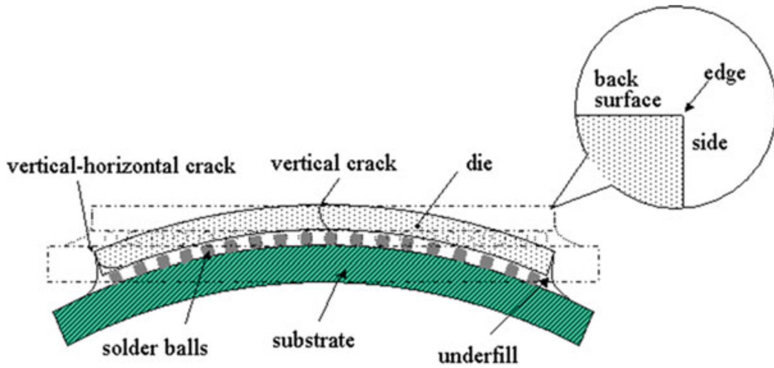


Fig. 10.3 A schematic diagram of a flip-chip package and die cracking failure modes due to die bending. The *dashed lines* represent the package at underfill curing temperature, which is stress free. The *solid lines* represent the package at room temperature, at which the package bends due to the CTE mismatch between the die and the substrate

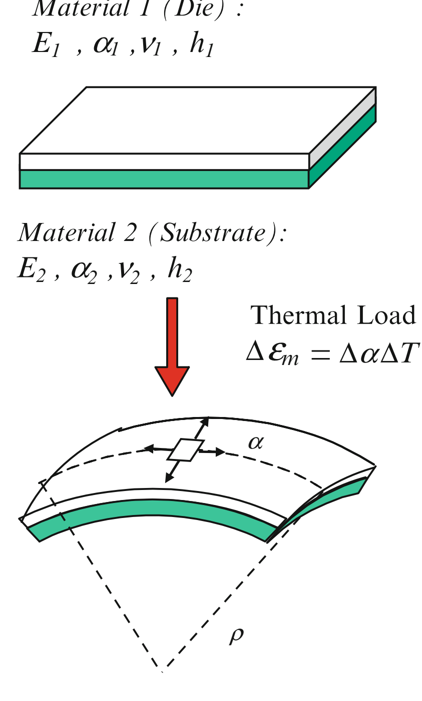
10.2.3 Die Stress Estimation with Bi-material Plate Model

As an example, we consider a flip-chip package shown in Fig. 10.3. Using the consistent composite plate model, the flip-chip package can be treated as a system of two plates bonded together if only the die bending curvature and the stress on top of the die is interested. The underfill and solder joins can be treated as part of the substrate in this scenario. The Young's modulus, Poisson's ratio, CTE, and thickness of upper plate (plate 1 or die) are E_1, ν_1, α_1, and h_1, respectively (Fig. 10.4). The corresponding quantities for the lower plate (plate 2 or substrate) are E_2, ν_2, α_2, and h_2, respectively. In the flip-chip package, the two plates were bonded together at a certain temperature T_0. The bi-material plate (BMP) system will bend into a surface which is part of a sphere (axisymmetrical bending) when temperature is reduced to T (the temperature change is ΔT as shown in Fig. 10.4). Solving (10.13) for the BMP system, we have

$$\begin{cases} \varepsilon_x^0 = \frac{D_x N_x^\Lambda - B_x M_x^\Lambda}{A_x D_x - B_x{}^2} \\ \kappa_x = \frac{A_x M_x^\Lambda - B_x N_x^\Lambda}{A_x D_x - B_x{}^2} \end{cases} \tag{10.15}$$

Please note that ε_x^0 is dependent on the location of x–y plane while κ_x is consistent through the plate thickness. Substituting (10.14) into (10.15) and after some mathematical manipulations we have

Fig. 10.4 Bi-material plate model for a flip-chip module



$$\begin{cases} \varepsilon_x^0 = \frac{hm(4+3h+h^3m)(\alpha_2+\alpha_1)\Delta T}{(1+hm(4+6h+4h^2+h^3m))} \\ \kappa_x = \frac{6\varepsilon_m hm(1+h)}{h_1(1+hm(4+6h+4h^2+h^3m))} \end{cases}, \tag{10.16}$$

where $h = h_2/h_1$ is the thickness ratio of the lower plate to the upper plate, $m = M_2/M_1$ is the ratio of the biaxial modulus of the lower plate ($M_2 = E_2/(1 - \nu_2)$) respect to the upper plate ($M_1 = E_1/(1 - \nu_1)$), and $\varepsilon_m = (T - T_0)(\alpha_1 - \alpha_2)$ is the thermal mismatch strain between the two plates.

In practice, it is convenient to express the curvature in a dimensionless quantity or, so-called, characteristic curvature:

$$\bar{\kappa} = \frac{h_1}{6\varepsilon_m}\kappa_x = \frac{hm(1+h)}{1 + hm(4 + 6h + 4h^2 + h^3m)} \tag{10.17}$$

The characteristic curvature is a function of only the thickness ratio h and the biaxial modulus ratio m.

From (10.2) and (10.16), the stress on the top surface of plate 1, the die, can be expressed as

$$\sigma_{\text{top}} = \frac{\varepsilon_m M_1 hm(2 + 3h - h^3m)}{1 + hm(4 + 6h + 4h^2 + h^3m)} \tag{10.18}$$

The stress is uniform at any point on the top of the die which is a direct result of neglecting the edge effect. The dimensionless stress or characteristic stress is defined as

$$\bar{\sigma} = \frac{\sigma_{\mathrm{top}}}{\varepsilon_m M_1} = \frac{hm(2 + 3h - h^3 m)}{1 + hm(4 + 6h + 4h^2 + h^3 m)} \tag{10.19}$$

It is instructive to express the stress on top of the die in terms of the curvature or characteristic curvature.

$$\sigma = \kappa_x \frac{h_1 M_1 (2 + 3h - h^3 m)}{6(1 + h)} \tag{10.20}$$

$$\bar{\sigma} = \bar{\kappa}_x \frac{2 + 3h - h^3 m}{1 + h} \tag{10.21}$$

According to (10.20) and (10.21), the stress contains the curvature κ_x. Therefore, the curvature κ_x can provide the direct information of the die and package stresses.

10.2.4 Minimizing Chip–Packaging Interaction

The chip–packaging interaction (CPI) is a result of CTE mismatch between the silicon chip and the packaging materials. As shown in the previous thermal deformation analysis, CPI is high dependent on the material properties of the packaging materials.

In order to minimize chip–packaging interaction and its impact on reliability of Cu/low-k interconnects, a thorough understanding on the thermo-mechanical characteristics of FC-PBGA is essential. To investigate the materials effect on CPI, fully assembled FC-PBGA packages with various underfill materials were evaluated following a carefully designed analysis and screening flow. Two FC-PBGA packages are chosen for the thermal deformation verification. These two packages are identical. They are built with a build-up laminate substrate, copper stiffener, and heat spreader. The package size is 40 $\times$ 40 mm. The difference between the two packages is the underfill material used. Package A is made with an underfill (A) with a high Glass Transition Temperature (T_{g}) and Package B is made of a low-T_{g} underfill (B). A schematic of the flip-chip package is shown in Fig. 10.4. Material properties of the two underfills that are important to package thermal deformation are listed in Table 10.1.

The moiré interferometry technique is used to measure thermal deformations of the FC-PBGA packages. Moiré interferometry is an optical method with high sensitivity and spatial resolution for measuring in-plane displacements and strains. During the experiment, a cross section of the package was made along the die edge (as shown in Fig. 10.5). A schematic of the package cross section is given in Fig. 10.6.

Table 10.1 Material properties for the underfill materials evaluated

Underfill	A	B
Filler content (%)	50	55
T_g (°C)	150	70
CTE under T_g (ppm/°C)	45	32
CTE above T_g (ppm/°C)	141	120
Modulus under T_g (GPa)	4	8
Modulus above T_g (GPa)	0.1	0.04

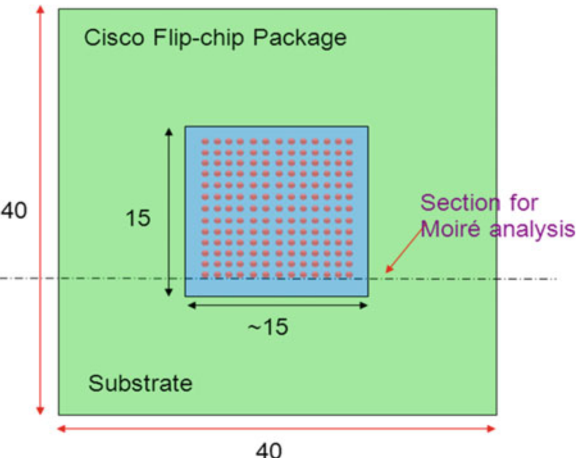


Fig. 10.5 Schematic of the FC-PBGA analyzed with moiré interferometry

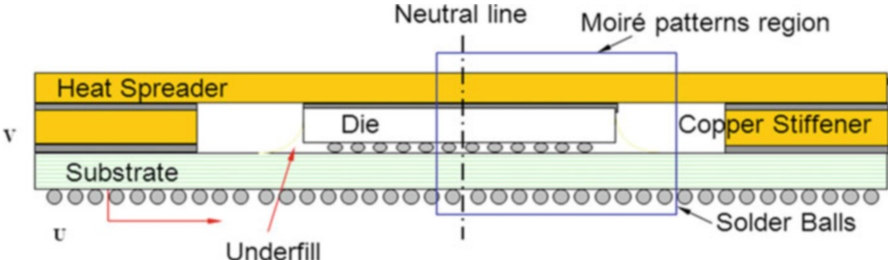


Fig. 10.6 Cross section on which moiré gratings are attached to flip-chip assembly

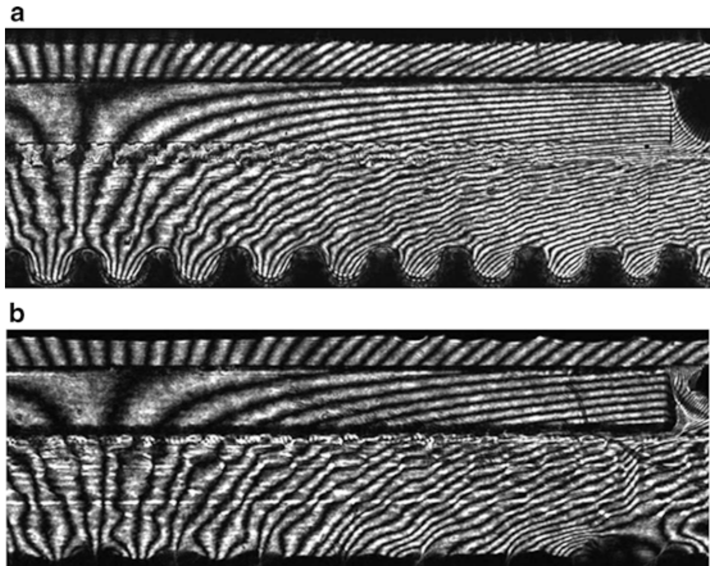


Fig. 10.7 Horizontal displacement contour maps: (**a**) Package A, (**b**) Package B

A very thin layer of epoxy adhesive (less than 5 μm) was used to replicate a 1,200 lines/mm grating at an elevated temperature of 122 °C (or 80 °C) on the cross section of the specimen. The deformation at this temperature was considered as a reference (zero deformation state). The specimen was then cooled down to room temperature (22 °C) where moiré interferometry was conducted. The moiré fringes recorded represent displacement contours induced by the temperature difference between 22 and 122 °C or by a thermal loading of (−) 100 °C.

Moiré fringe patterns obtained for the FC-PBGA packages are shown in Figs. 10.7 and 10.8. Package thermal deformation can be quantitatively determined by analyzing the spatial distribution of the moiré fringe pattern where the displacement sensitivity is 417 nm per fringe. A comparison in horizontal displacement, U, for the two packages is given in Fig. 10.7. The bending displacement, V, for the two FC-PBGA packages is shown in Fig. 10.8.

From the moiré analysis we can see the package with the high-T_g underfill (A) had a larger displacement in both vertical and horizontal directions. This increase in bending and stretching displacements indicated strong chip and packaging coupling as comparing with the package made with the low-T_g underfill (B). This effect is further illustrated in Fig. 10.9 where the chip bending displacement, V, under different thermal loadings was plotted for the two packages.

Therefore, for the package with high-T_g underfill, higher die stresses are expected at temperatures below the T_g of the underfill. On the other hand, low-T_g underfill can effectively decouple the die from the package substrate when temperature is near and above the (relatively low) T_g of the underfill. This in turn will lower the "stress-free" temperature and will induce less thermal stress for the

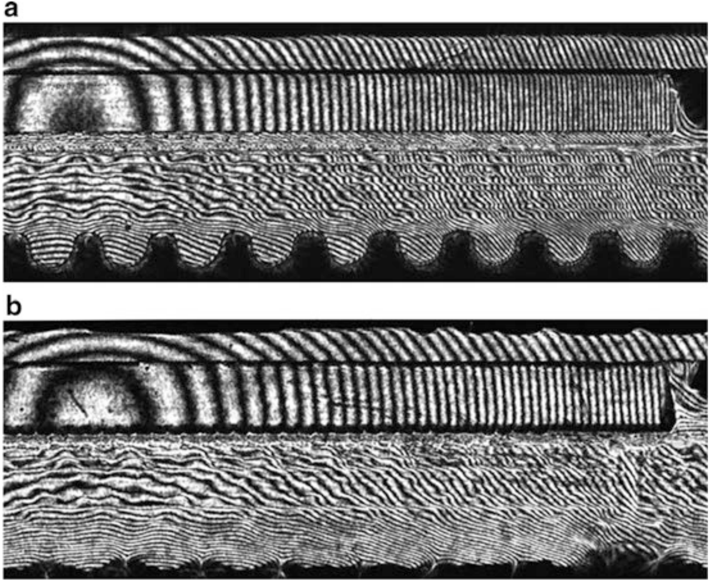


Fig. 10.8 Bending displacement contour maps: (**a**) Package A, (**b**) Package B

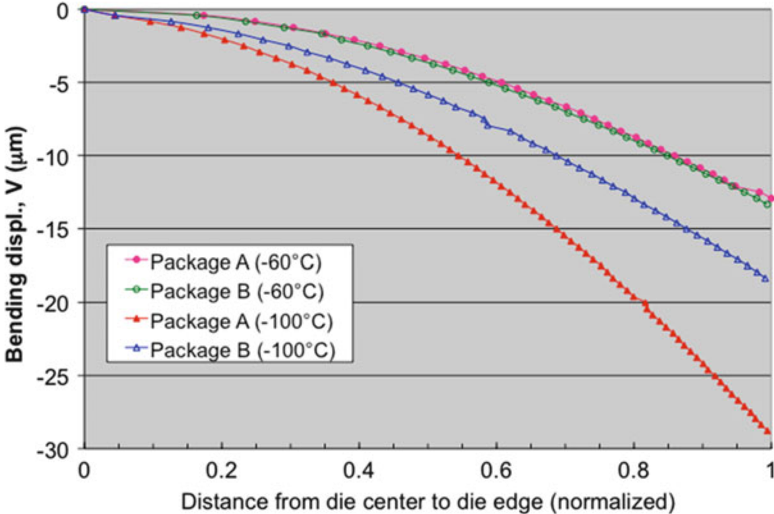


Fig. 10.9 Comparison of bending deformation in Package A and Package B

FC-PBGA package (B). The lowered thermal stresses in the flip-chip assembly are very effective in minimizing reliability risks associated with die cracking, substrate cracking, underfill delamination, and low-k ILD layer delamination.

Another observation made by comparing the moiré fringe patterns for the two packages is relative large deformation occurred in the underfill region for the package with low-T_g underfill (B). This is reflected in the moiré fringe patterns where very dense fringes showed at the underfill interface for Package B. This observation raised some concerns on the effectiveness of low-T_g underfill for solder bump protection.

10.2.5 Summary of Thermal Deformation in Flip-Chip Assembly

When the flip-chip packaging is subjected to a thermal loading of ΔT, thermal deformation is introduced in the package as a result of CTE mismatch between the chip and the package. Chip–packaging interaction is a critical reliability issue for flip-chip packages with Cu/low-k chips. Combined experimental and modeling methods were used to investigate the thermo-mechanical behavior and failure mechanisms controlling package reliability. Materials effect of new generation of underfill materials was studied for minimizing the chip-substrate thermo-mechanical coupling. The combined experimental and numerical analysis provided a systemic approach for the reliability assessment and the packaging materials selection.

A wide range of reliability experiments were performed to evaluate the reliability of the FC-PBGA package. These tests include JEDEC Moisture Level 4 pre-conditioning, Thermal Cycling (Condition B, −55 °C/125 °C) and Highly Accelerated Temperature and Humidity Stress Test (HAST). The testing results correlated well with both the numerical and experimental evaluation data. The testing results showed that Package B performed consistently better than Package A in all the reliability tests.

10.3 Solder Bump Reliability in Flip-Chip Assembly

Due to thermal expansion mismatch between the chip and the package, thermal stresses are introduced in the cooling steps of the chip joining and underfill processes. Time-dependent thermo-mechanical stresses are further introduced in the package and package-board assembly, especially in the solder joints, when the assembly is subjected to power cycling, thermal cycling, or thermal shock. The time-dependent stress/strain has a great impact on the long-term reliability of the solder joint.

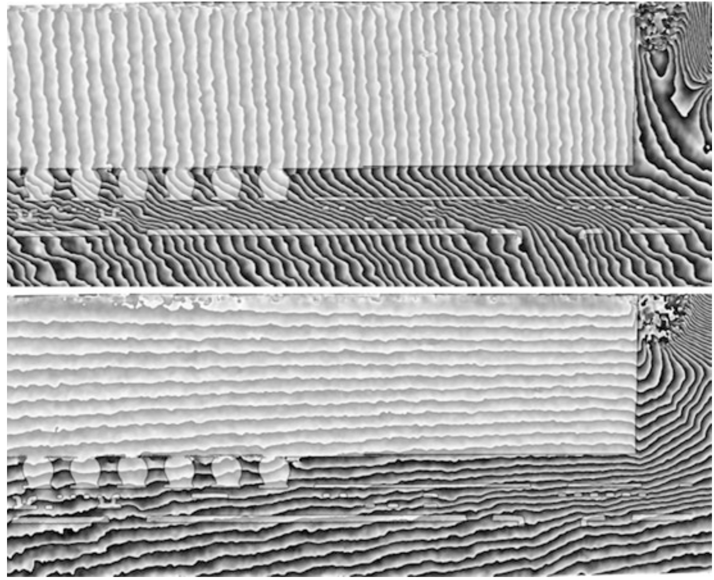


Fig. 10.10 Phase maps for Package A. The thermal loading is −60 °C, (*top*) U field, and (*bottom*) V field

10.3.1 Thermal Strain Measurement in Solder Bumps

In the flip-chip package, the underfill material effectively redistributes the thermo-mechanical load on solder bumps providing a practical solution to the solder bump fatigue problem. To prevent solder bumps from fatigue cracking, the requirements for underfill are high T_g, high modulus and with a Coefficient of Thermal Expansion (CTE) closely matched to that of the solder material. While with a low-T_g underfill material we can effectively reduce die and package stresses at lower temperatures, we need to determine if solder bump reliability is still intact.

High resolution moiré interferometry was developed to measure localized thermal strain distributions in the critical solder joint during temperature excursion [4, 7]. High resolution moiré interferometry is achieved using the phase-shifting method. Instead of using the amplitude of the moiré fringes, the phase of the optical field was recorded and analyzed to extract the displacement information. As an example, the phase maps for Package A are shown in Fig. 10.10, for a thermal loading of −60 °C. The corresponding displacement maps are shown in Fig. 10.11. We can see the displacement sensitivity is increased from 417 nm per fringe to 52 nm per fringe by using the phase-shifting technique.

Micro-strain distributions can also be derived from the high resolution moiré patterns. For solder bump and interfacial reliability, interlayer shear strains are of the most interest. The interlayer shear strains are assessed along the two critical interfaces in the FC-PBGA package. As shown in Fig. 10.12, the optical

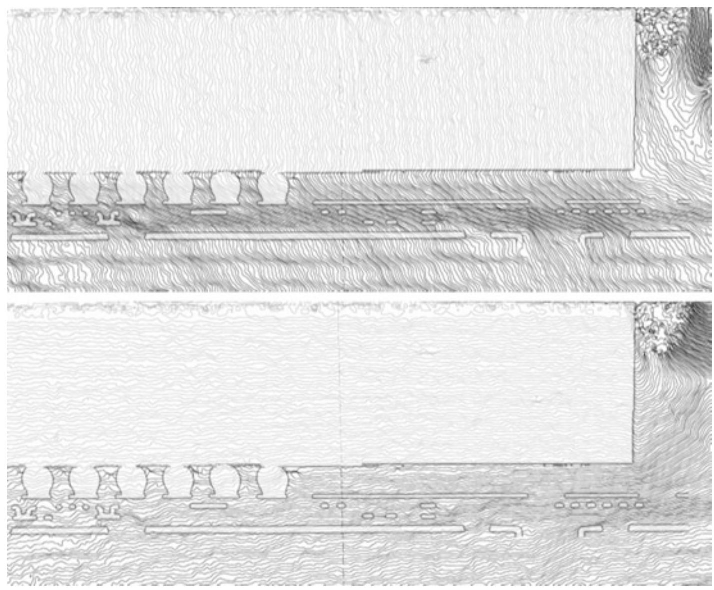


Fig. 10.11 Displacement contour maps showing 52 nm displacement per contour for Package A. Thermal loading is −60 °C, (*top*) U field, and (*bottom*) V field

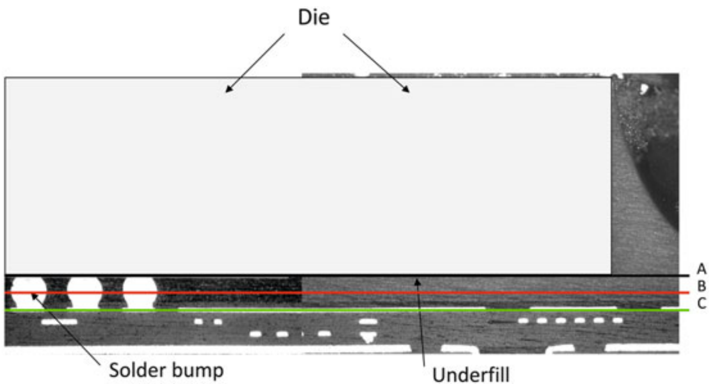


Fig. 10.12 Optical micrograph showing critical bump locations and underfill fillet

micrograph of the package cross section, Line A is along the die to solder bump interface, Line B is along middle of the underfill layer and Line C is for the solder to substrate interface. The shear strain distribution results are shown in Figs. 10.13 and 10.14 for Package A and Package B, respectively.

From the shear strain chart, we can see shear strain increased significantly close to die edge for both packages. Comparing the strain distributions at the die/underfill interface (Line A) and in the underfill (Line B), we can see the shear strain for

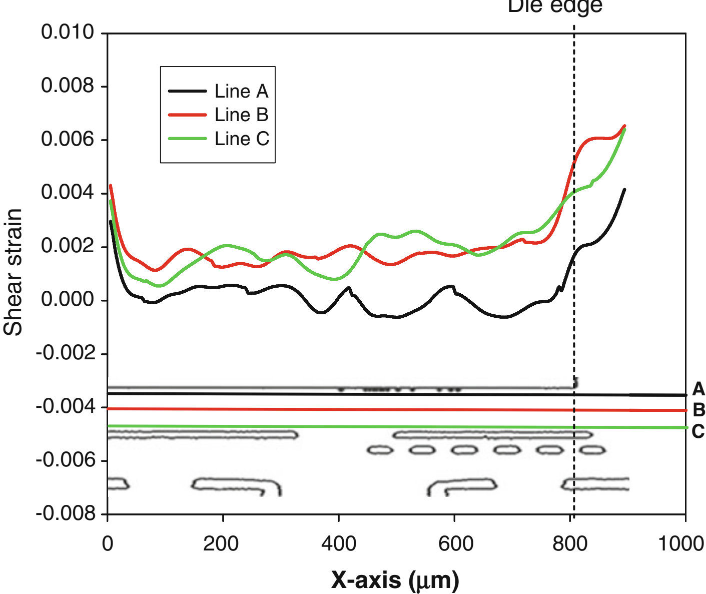


Fig. 10.13 Shear strain distribution in Package A

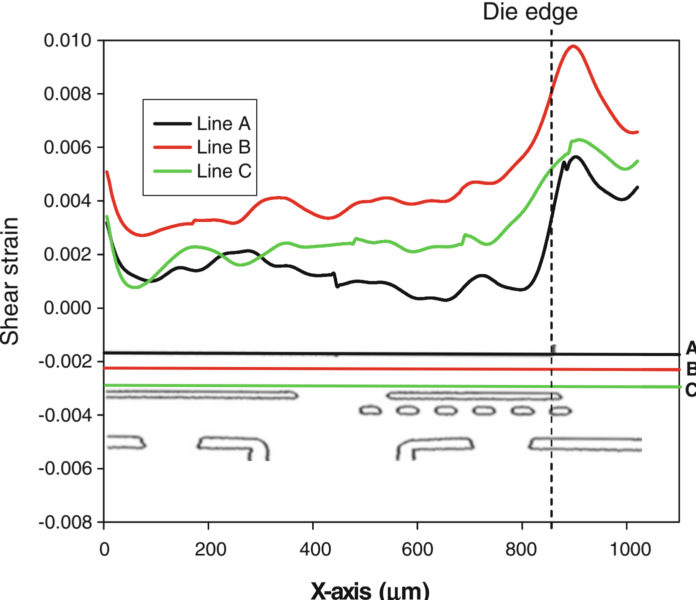


Fig. 10.14 Shear strain distribution in Package B

Package B is indeed larger than that of Package A. This is expected since Underfill B is more compliant than Underfill A at elevated temperatures. Thermal mismatch between the die and the substrate will introduce a higher shear strain for Package B. This result in strain measurement underscores the importance on the need to strike a balance between lowering packaging stress with low-T_g underfill material and controlling mechanical strain and fatigue in solder bumps.

10.3.2 Constitutive Equation for Solder

Accurate constitutive modeling of the solder plays an important role in the simulation of solder bump reliability. The commercial finite element software ANSYS does have viscoplastic elements as a standard option, but they use Anand's constitutive model [8, 9]. The use of these elements is convenient since the user does not have to modify the source code. Anand's model was developed for hot working metals, which unifying plasticity and creep via a set of flow and evolutionary equations:

Flow Equation

$$\frac{\mathrm{d}\varepsilon_p}{\mathrm{d}t} = A[\sin h(\xi\sigma/s)]^{\frac{1}{m}} \exp\left(-\frac{Q}{kT}\right) \tag{10.22}$$

Evolution Equations

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \left\{ h_0(|B|)^a \frac{B}{|B|} \right\} \frac{\mathrm{d}\varepsilon_p}{\mathrm{d}t} \tag{10.23}$$

$$B = 1 - \frac{s}{s^*} \tag{10.24}$$

$$s^* = \hat{s} \left\{ \frac{\frac{\mathrm{d}\varepsilon_p}{\mathrm{d}t}}{A} \exp\left(-\frac{Q}{kT}\right) \right\}^n \tag{10.25}$$

Darveaux et al. [10] were the first to modify the constants in Anand's constitutive relation to account for both time-dependent and time-independent phenomenon. Parameters for near-eutectic 62Sn/36Pb/2Ag are given in Table 10.2 (see Refs. [10, 12]).

In ANSYS, the solder bump can also be modeled in a separate constitutive model taking into account the time-independent plasticity and time-dependent creep deformation modes using the SOLID182 and SOLID185 elements. For solder alloys in the high-homologous-temperature regimes ($>0.5T_m$) primary and secondary (steady-state) creep are the dominant deformation modes and are accompanied by stress relaxation.

Table 10.2 Anand's constants for 62Sn36Pb2Ag solder in ANSYS

ANSYS	Parameter	Value	Definition
C1	S_0 (psi)	1,800	Initial value of deformation resistance
C2	Q/k (1/K)	9,400	Activation energy/Boltzmann's constant
C3	A (1/s)	4.0E6	Pre-exponential factor
C4	ξ	1.5	Multiplier of stress
C5	m	0.303	Stain rate sensitivity of stress
C6	h_0 (psi)	2.0E5	Hardening constant
C7	s^ (psi)	2.0E3	Coefficient for deformation resistance saturation value
C8	n	0.07	Strain rate sensitivity of saturation (deformation resistance) value
C9	a	1.3	Strain rate sensitivity of hardening

Creep generally refers to the time-dependent strain plastic deformations at constant uniaxial stress [13–16]. Creep deformation tends to be rapid when the homologous temperature is above $0.5T_{\mathrm{m}}$. Even though, creep of solders consists of three stages after the initial instantaneous strain when a constant load is applied [13]. However, the secondary, or steady-state creep is the dominate deformation experienced by solder alloys. In this stage, the strain rate is retarded by strain-hardening, which decreases the deformation speed, while the associated recovery and recrystallization (softening) tend to accelerate the creep rate [14]. The steady-state creep rate can be quantitatively estimated, and a series of constitutive models have been proposed. The following two models are the most widely accepted for the characterization of solder alloys by considering the diffusion controlled creep deformation mechanism.

Dorn Power Law [16]:

$$\dot{\varepsilon} = A\sigma^n \exp\left(-\frac{Q}{RT}\right) \tag{10.26}$$

Garofalo Hyperbolic Sine Law [14]:

$$\dot{\varepsilon} = C[\sin h(\alpha\sigma)]^n e^{\left(\frac{-Q}{RT}\right)} \tag{10.27}$$

where R is the universal gas constant, T is the temperature in Kevin, σ is the applied stress, A and C are material dependent constants, n is the stress exponent, and Q is the activity energy. The models show that the steady state creep strain rates are strongly stress and temperature dependent.

The constitutive models can be determined by creep testing at different temperature and stress levels. However, there are large discrepancies in the creep data and materials constants vary over a very large range. The materials constants are important in determining the accuracy of end-of-life prediction for solder joints using finite element analysis. Large discrepancies would degrade the accuracy of these predictions.

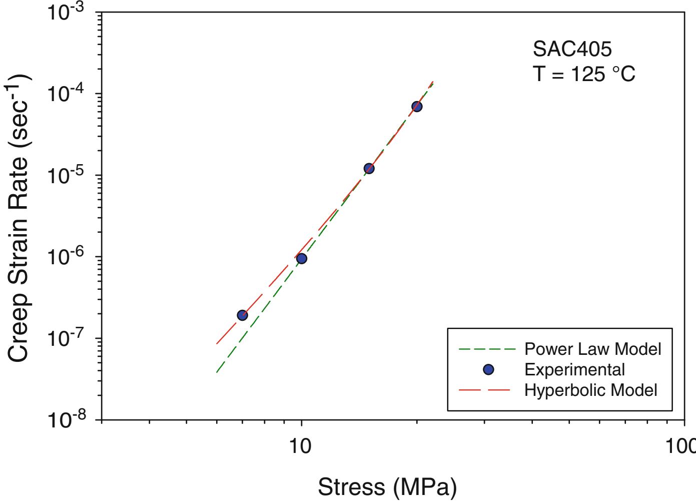


Fig. 10.15 Comparisons of Hyperbolic Sine Model and Power Law Model (SAC405)

The recent results show that room temperature aging effects will dramatically affect the mechanical properties of solder alloys [17, 18]. Any difference in the testing conditions of specimens could seriously affect the accuracy of the data. The data shows that the tensile strength significantly degrades during room temperature aging up to 2 months of aging time by up to 35 %. The creep resistance also degrades at room temperature. The data also shows that the tensile properties of SAC405 (Sn–4.0Ag–0.5Cu) lead-free solders after 10 days of aging at room temperature tend to be relatively stable. In order to minimize the discrepancies in test results due to room temperature aging, Ma et al. performed comprehensive creep test at various stress levels and temperatures on both lead-free solders (SAC405) and SnPb solders after aging the samples at room temperature for 10 days [19]. So when all samples were tested at same period of time after stabilizing at room temperature, the data discrepancies due to room temperature aging can be minimized. Equation 10.28 is the hyperbolic sine law fitting of the SAC405 creep model. Equation 10.29 is the fitting for the power-law model. Figure 10.15 compares the two models at the same testing temperature and stress level for the SAC405 alloy. The results indicate that at higher stress levels above 15 MPa, both models have an almost identical fit to the experimental data. However, at lower stress levels, hyperbolic sine model fits the experimental data better than the power-law model. The obtained constitutive models can be used to predict the end-of-life of the solder joints in numerical techniques such as finite element methods. Overall, Garofalo's hyperbolic model is a better model to fit the current experimental data.

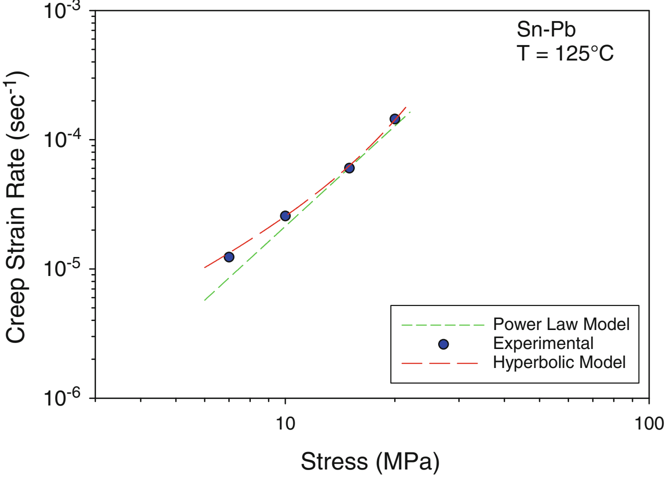


Fig. 10.16 Figure 10.3 Comparisons of Constitutive Creep Models (Sn–Pb)

$$\dot{\varepsilon} = 1.77E + 05[\sin h(5.48E - 02\sigma)]^{4.89} \exp\left(-\frac{76.13}{RT}\right) \tag{10.28}$$

$$\dot{\varepsilon} = 5.09E - 03\sigma^{6.27} \exp\left(-\frac{76.2}{RT}\right) \tag{10.29}$$

Corresponding data for Sn–Pb eutectic solder alloy were also fitted in with multiple variables data fitting for both Garofalo Hyperbolic Sine Law and the Dorn power-law models. The comparison between the two models for Sn–Pb eutectic solder alloy is shown in Fig. 10.16. It indicates similar results to those for lead-free solders. The two models show similar perfect fitting at higher stress levels above 15 MPa, and once again, the hyperbolic sine law model exhibits a better fit at lower stress levels. Overall, the Garofalo Hyperbolic Sine Law provides a better fit for the experimental data for Sn–Pb.

$$\dot{\varepsilon} = 0.908[\sin h(0.105\sigma)]^{1.51} \exp\left(-\frac{35.74}{RT}\right) \tag{10.30}$$

$$\dot{\varepsilon} = 2.87E - 04\sigma^{2.58} \exp\left(-\frac{33.30}{RT}\right) \tag{10.31}$$

Comparing to the materials constants of Sn–Pb (10.30 and 10.31) with that of SAC alloy (10.28 and 10.29), the activation energy of Sn–Pb is significantly smaller than that of the SAC alloy, indicating that the creep resistance of SAC is higher than that of the Sn–Pb. The stress component of Sn–Pb is also significantly smaller than

Table 10.3 Constitutive creep model of SAC and SnPb solders

Hyperbolic model	C	ς (MPa^{-1})	n	Q (kJ/mol)
SAC	1.77×10^5	5.48×10^{-2}	4.89	76.13
Sn–Pb	0.908	0.105	1.51	35.74

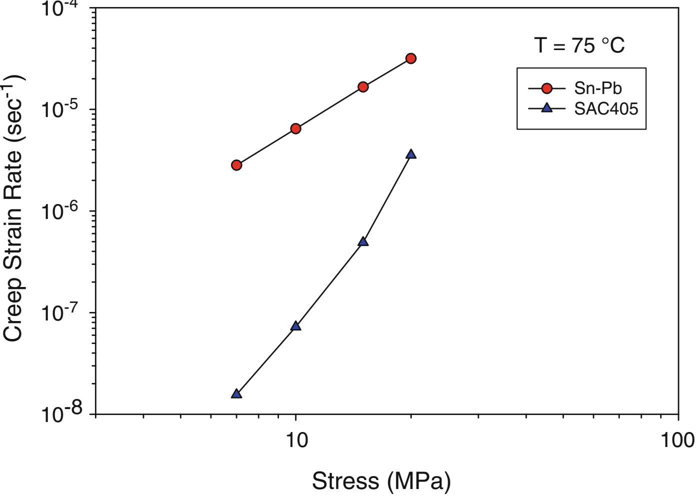


Fig. 10.17 Comparison of Creep Rate of Sn–Pb vs. SAC405

that of SAC, which also indicates lower creep resistance for Sn–Pb than that of SAC alloys. The constants of hyperbolic model for both SAC and Sn–Pb alloy were summarized in Table 10.3.

Comparing Sn–Pb and SAC, SAC has a higher creep resistance than Sn–Pb at the same stress level and testing temperature (Fig. 10.17). The constitutive models also show that the SAC alloy possesses higher activation energy than Sn–Pb, which indicates that the SAC alloy has a higher creep resistance than Sn–Pb. Activation energy represents the height of the energy barrier the atoms have to overcome to diffuse and move to lower energy levels. Dorn, Garofalo, and Weertman independently found that the creep activation energy equals the self-diffusion activation energy when $T_h \geq 0.5T_m$ [13, 14, 20, 21]. The homologous temperatures in the operational temperature range of solder alloys are greater than half of the melting points, indicating that the self-diffusion activation energy is the barrier that atoms must overcome to cause creep deformation in solder alloys, which corresponding to their levels of creep resistance. This higher creep resistance of lead-free solders is contributed by the second phase intermetallic compounds, Ag_3Sn and Cu_6Sn_5 in SAC alloys, which can effectively block dislocation movement and increase the creep resistance of the material.

10.3.3 Reliability Modeling for Solder Joint

Acceleration tests, such as thermal cycling test, are widely used in the industry for solder joints fatigue life prediction. Solder joint fatigue is considered as low cycle fatigue, it is the fatigue life of the solder to the plastic strain under stress. The fatigue of metals can be described in the Coffin–Manson equation as following:

$$N = \left(\frac{A}{\Delta\varepsilon}\right)^{\frac{1}{m}} \tag{10.32}$$

Where N is the number of fatigue cycles to failure, $\Delta\varepsilon$ is the plastic strain range per cycle, n is an empirical material constant, and C is a proportionality factor. Norris–Landzberg model has been developed for solder joint fatigue predication by assuming the plastic strain range is proportional to the range of temperature cycling conditions, it has been used in the past decades for Sn–Pb based solders (10.33) [28]

$$\mathrm{AF} = \frac{N_P}{N_T} = \left(\frac{\Delta T_T}{\Delta T_P}\right)^n \left(\frac{f_P}{f_T}\right)^m \exp\left[Q\left(\frac{1}{T_{\mathrm{Max}P} + 273} - \frac{1}{T_{\mathrm{Max}T} + 273}\right)\right], \tag{10.33}$$

Where N is the number of fatigue cycles, ΔT is the applied temperature difference, T is the absolute temperature in Kelvin, and f is the cyclic frequency. The subscript "P" refers to product operating conditions whereas "T" refers to test conditions.

With the implementation of lead-free solders, there are some noticeable studies for using Norris–Landzberg model for lead-free solder joints fatigue analysis. There are some notable studies in recent years exploring the acceleration factors (AF) for Pb-free solders. Pan et al. first developed acceleration models for Pb-free solders based on standard 0–100 °C testing conditions; however, there is lack of evidence that the model can be used for more extreme conditions [22]. Andersson et al. [23], Vasudevan et al. [24], Xie et al. [25], Salmela et al. [26], also developed similar models with different parameters based on their particular test conditions and package types. Dauksher has also developed generalized models using the data from previous studies [27]. Ma et al. found that the current models fit milder thermal cycling range. However, the current model does not fit for more extreme temperature conditions due to pad cratering failure mode was found at extreme temperature conditions, especially at extreme low temperature. More work needs to be done in solder joint fatigue modeling at extreme temperature conditions [29]. More work is needed to correlate different failure modes at extreme accelerated thermal cycling conditions to appropriated life prediction models.

In addition to the inelastic strain amplitude based Coffin–Manson-type fatigue models, an energy based metric for predicting crack initiation and growth in solder joints was developed and refined over the years, [10–12]. The plastic work accumulated during the last cycle is used for crack growth and solder joint fatigue

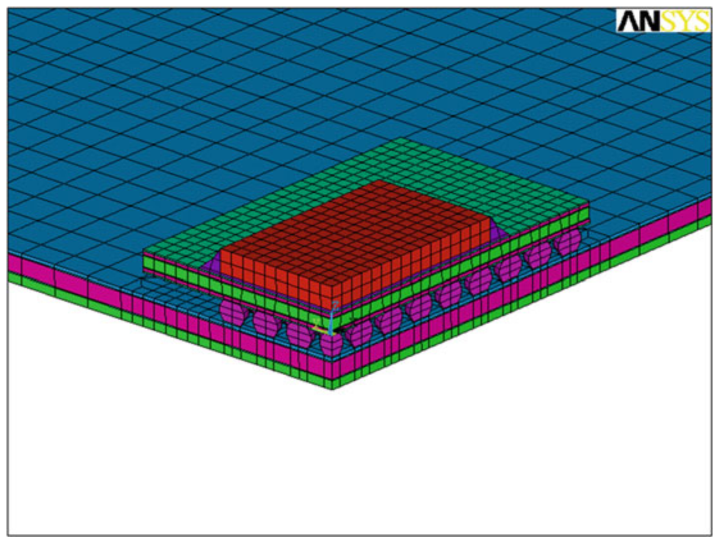


Fig. 10.18 Finite element model of a FC-PBGA assembled on a board

life correlations. The characteristic fatigue life (N_a) for the solder ball with a pad diameter of a can be written as

$$N_a = A(\Delta W_{\text{ave}})^B + C(\Delta W_{\text{ave}})^D, \tag{10.34}$$

where A, B, C, D are constants that depend on material and solder joint design and assembly process and ΔW_{ave} is the volume averaged viscoplastic strain energy density increment per cycle.

In practice, non-linear finite element models have been developed to calculate the plastic work per unit volume (or viscoplastic strain energy density) accumulated per thermal cycle in the solder joints. Several complete thermal cycles are simulated in order to establish a stable stress–strain hysteresis loop. The A, B, C, D constants are determined by correlating the simulated results to the temperature cycling test results. As an example, the finite element modeling of a FC-PBGA assembly under accelerated temperature cycling conditions is given here. ANSYS 6.0 is used for all aspects of the study: preprocessing, solution, and post-processing.

Due to symmetry, only one-quarter of the module-board assembly is modeled. Figure 10.18 shows one quarter of the module-board assembly for the FC-PBGA. The BGA solder material is modeled as a viscoplastic solid. The PBGA substrate along with the printed circuit board is modeled as composite materials with their copper power and ground planes, and their orthotropic, temperature dependent dielectric materials explicitly modeled. The physical and mechanical properties of the silicon chip, underfill material, and board are summarized in Table 10.4.

The cyclic temperature loading imposed on the FC-PBGA/PCB assembly is closely matched to the temperature profile measured in accelerated temperature

Table 10.4 Material properties used in modeling of an FC-PBGA assembly

Material	Modulus at 23 °C (Gpa)	Poisson ratio at 23 °C	CTE (10^{-6}/°C) at 23 °C	T_g (°C)
Si chip	186	0.28	3.2	NA
Underfill	3.4	0.33	24	145
PBGA substrate (power plane)	45.6	0.34	17.6	NA
PBGA substrate (dielectric)	16.4 (*x*, *y*) 2.6 (*z*)	0.48 (*x*, *y*) 0.16 (*z*)	14.8 (*x*, *y*) 67 (*z*)	170
SnPb solder	Viscoplastic, see Ref. [12]	0.29	22	NA
Printed circuit board (power plane)	45.6	0.34	17.6	NA
Printed circuit board (dielectric)	88.6 (*x*, *y*) 2.7 (*z*)	0.48 (*x*, *y*) 0.16(*z*)	22.4 (*x*, *y*) 67 (*z*)	170

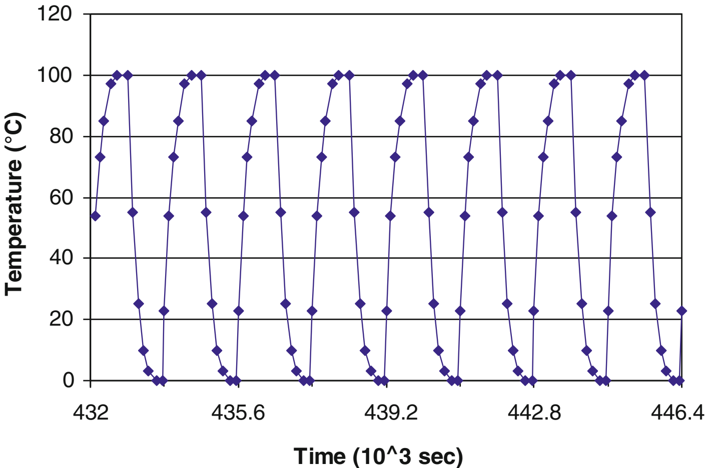


Fig. 10.19 Cyclic temperature loading used in finite element modeling

cycling test and is shown in Fig. 10.19. It can be seen that for each cycle (30 min) the temperature is between 0 and +100 °C, with 10 min ramp, 5 min hold at the two temperature extremes. In addition, pre-cycling, temperature-time history which corresponds to 5 days at 23 °C is also included in the simulation.

Residual stresses in the solder joint after the FC-PBGA module was assembled to the board, cooled down to room temperature but before the temperature cycling test are shown in Fig. 10.20. As expected the residual stresses relaxed very rapidly in the first hour after the assembly process. The stresses shown in Fig. 10.20 are for the solder ball which is under the corner of the chip as indicated in Fig. 10.24. The time-dependent stresses and strains in the same solder ball due to cyclic temperature variation are shown in Figs. 10.21 and 10.22, respectively. It can be seen from Figs. 10.21 and 10.22 the stresses and strains oscillate within certain ranges

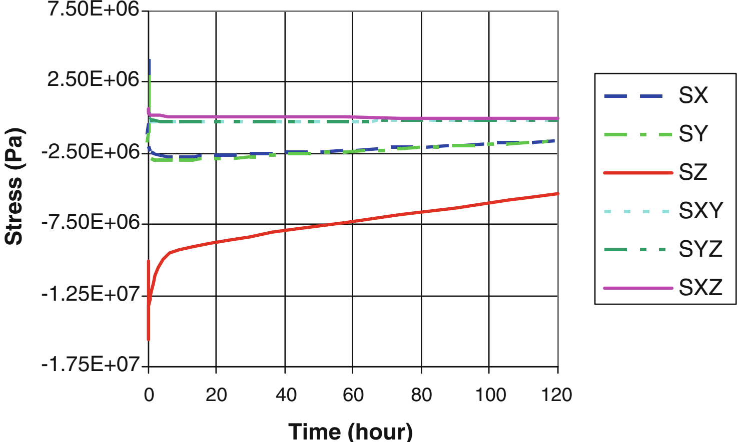


Fig. 10.20 Residual stresses in the solder joint after assembly process

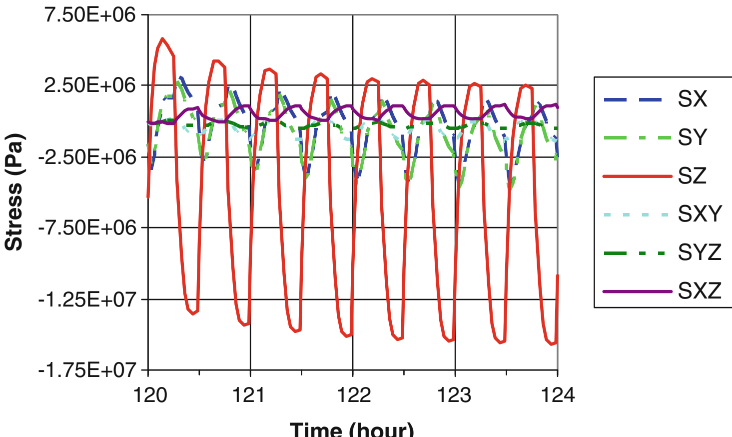


Fig. 10.21 Time-dependent stresses in the solder joint due to cyclic temperatures

in accordance with the cyclic temperature load. Figure 10.23 shows the normal stress (σ_z) and the normal viscoplastic strain hysteresis loops for multiple cycles. For viscoplastic analysis, it is important to study the stress–strain responses for multiple cycles until the hysteresis loops become stabilized. Figure 10.23 indicates that the plastic strain is quite stabilized after the third cycle and the creep response converged after the fourth cycle.

Figure 10.24 shows the viscoplastic strain energy density increment calculated from the eighth to the seventh cycle. It is noticed that the solder ball with the highest accumulation in viscoplastic strain energy density is located just under the corner of the die edge. Experimental results show that the balls under the corner of the die edge usually fail first in accelerated temperature cycling test.

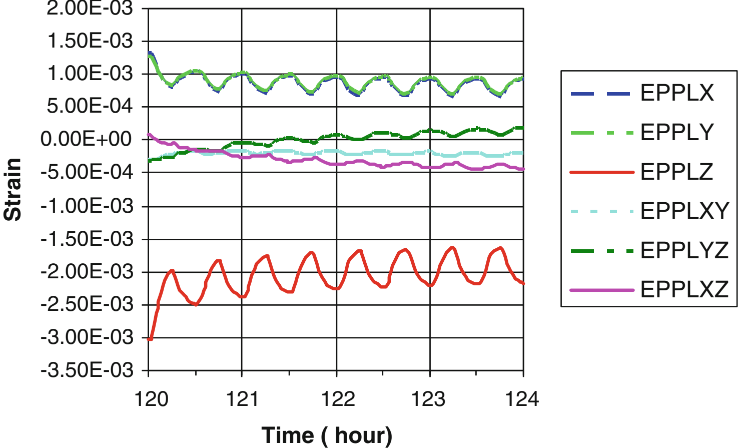


Fig. 10.22 Time-dependent strains in the solder joint due to cyclic temperatures

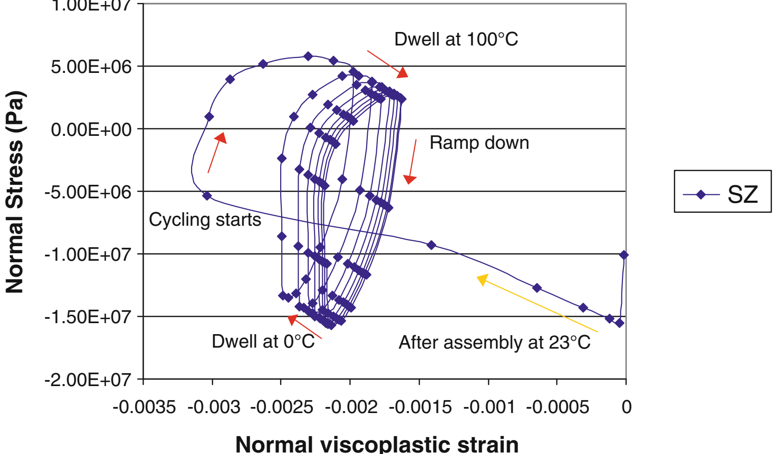


Fig. 10.23 Hysteresis loops of stress and viscoplastic strain

10.3.4 Underfill Adhesion Strength on Solder Bump Reliability

The resistance to initiation and growth of interfacial delamination is a critical issue in determining the overall FC-PBGA package reliability. Using the fracture mechanics approach, interfacial delamination can be analyzed as a crack propagating along the interface between two materials of interest.

The adhesion strength of the die to underfill interface can be evaluated from its interfacial fracture energy which can be determined by measuring the critical energy release rate for the interface. One experimental technique which has been

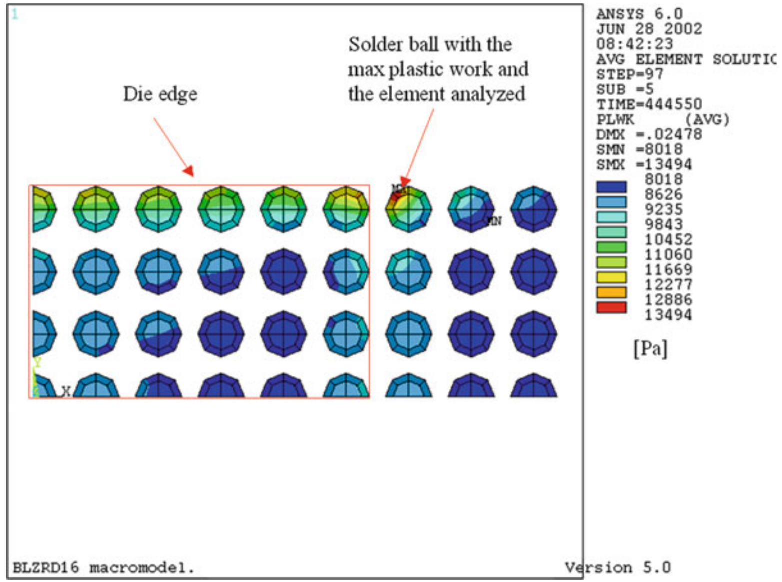
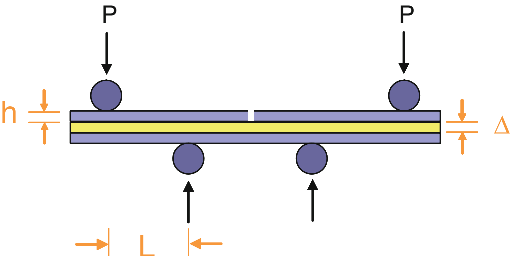


Fig. 10.24 Calculated plastic work (viscoplastic strain energy density) increment

Fig. 10.25 Schematic of 4-point bending test



developed to measure the critical energy release rate is the 4-point bending test [4]. The sample used in the 4-point bending test is made of a layer of underfill material sandwiched by two pieces of silicon wafer as shown schematically in Fig. 10.25. Considering the underfill thickness, we can express the critical energy release rate, G_c, as

$$G_c = \frac{3P_c{}^2L^2(1-v^2)}{2Eb^2h^2}\left\{\frac{1}{h+3\lambda\Delta} - \frac{1}{8h+12\Delta}\right\} \tag{10.35}$$

Here E and v are Young's modules and Poisson's ratio of silicon, b is the sample width, h is the wafer thickness, P is the mechanical load applied, and L is the distance between inner and outer loading points. The modulus ratio between the

Table 10.5 Interfacial adhesion results

Interfaces	Underfill A/die passivation	Underfill B/die passivation
G_c (J/m^2)	~130.0	~130.0

underfill and the silicon material is expressed as λ. The critical energy release rate G_c is determined using the critical load P_c when the crack propagates along the interface.

Samples made of both underfill materials as shown in Table 10.1 are used in the 4-point bending test to measure adhesion at the critical interface between the die passivation and the underfill. Note the 4-point bending test can only measure the critical energy release rate at certain phase angle (mode mixity). The phase angle for the sample geometry illustrated in Fig. 10.25 is 42°. The critical energy release rate G_c for the two underfill materials are summarized in Table 10.5. Both materials showed excellent interface adhesion results.

10.3.5 Summary of Solder Bump Reliability

In the flip-chip package, the underfill material effectively redistributes the thermo-mechanical load on solder bumps providing a practical solution to the solder bump fatigue problem. To prevent solder bumps from fatigue cracking, the requirements for underfill are high T_g, high modulus and with a Coefficient of Thermal Expansion (CTE) closely matched to that of the solder material. The materials effect of the new generation underfill materials on solder bump reliability is evaluated using the high resolution moiré interferometry. The package resistance to underfill delamination was characterized by measuring interfacial adhesion at the critical die-passivation—underfill interface of the FC-PBGA. The results in the solder bump strain measurement and underfill adhesion strength measurement underscore the importance of striking a balance between lowering overall packaging stress with the low-T_g underfill material and controlling the mechanical strain and fatigue in solder bumps.

Time-dependent thermal stress as a result of thermal cycling has a great impact on the long-term reliability and integrity of the flip-chip packages. A nonlinear finite element simulation methodology is introduced and applied to simulate the time-dependent thermal strain or strain energy density in the solder joints of a FC-PBGA assembly. The simulated results can be correlated with temperature cycling test results and can be used to predict solder joint reliability under various use and testing conditions.

References

1. Jimarez M et al (1998) Technical evaluation of a near chip scale size flip chip/plastic ball grid array package. Proceedings of 48th Electronic Components and Technology Conference, Seattle, WA, May 1998, pp 219–225
2. Chen K et al (2006) Effects of underfill materials on the reliability of low-k flip-chip packaging. Microelectron Reliab 46(1):155–163
3. Tsao P-H et al (2004) Underfill characterization for low-k dielectric/Cu interconnect IC. Proceedings of 54th Electronic Components and Technology Conference, Las Vagas, NV, May 2004, pp 767–769
4. Li L et al (2006) Materials effects on reliability of FC-PBGA packages for Cu/low-k chips. Proceedings of 56th Electronic Components and Technology Conference, San Diego, CA, May 2006, pp 1590–1594
5. Ashton JE, Whitney JM (1970) Theory of laminated plates, vol IV, Progress in materials science series. Technomic Publishing, Stamford, CN, p 153
6. Jones R (1975) Mechanics of composite materials. Hemisphere Publishing Corporation, New York
7. Post D, Han B, Ifju P (1994) High sensitivity moire: experimental analysis for mechanics and materials, Chap 2. Springer, New York
8. Anand L (1985) Constitutive equations for hot-working of metals. Int J Plast 1:213–231
9. Brown SB, Kim KH, Anand L (1989) An internal variable constitutive model for hot working of metals. Int J Plast 5:95–130
10. Darveaux R, Banerji K, Mawer A, Dody G (1995) Reliability of plastic ball grid array assembly. In: Lau J (ed) Ball grid array technology. McGraw-Hill, New York
11. Darveaux R (1997) Solder joint fatigue life model. Proceedings of TMS Annual Meeting, Orlando, FL, pp 213–218
12. Darveaux R (2000) Effect of simulation methodology on solder joint crack growth correlation. ECTC, Las Vegas, NV, pp 1048–1063
13. Garofalo F (1966) Fundamentals of creep and creep-rupture in metals. The Macmillan Company, New York
14. Hertzberg RW (1996) Deformation and fracture mechanics of engineering materials, 4th edn. Wiley, New York
15. Evans RW, Wilshire B (1985) Creep of metals and alloys. The Institute of Metals, London
16. Mukherjee AK, Bird JE, Dorn JE (1969) Experimental correlation for high-temperature creep. Trans Am Soc Met 62:155–179
17. Ma H, Suhling JC, Lall P, Bozack M (2006) Reliability of the aging lead-free solder joints. Proceeding of the 56th Electronic Components and Technology Conference (ECTC), San Diego, California, 30 May–2 June 2006, pp 849–864
18. Ma H, Suhling JC, Lall P, Bozack M (2007) The influence of elevated temperature aging on reliability of lead-free solder joints. The Proceeding of The 57th Electronic Components and Technology Conference (ECTC), May 2007, pp 653–668
19. Ma H (2009) Constitutive models of creep for lead-free solders. J Mater Sci 44(14):3841–3851
20. Dorn JE (1957) Creep and recovery. ASM Publication, Metal Park, OH, pp 255–259
21. Weertman J (1968) Dislocation climb theory of steady-state creep, noting necessity of self diffusion mechanism in any high temperature creep theory. ASM Trans Q 61:681–694
22. Pan N et al (2005) An acceleration model for Sn-Ag-Cu solder joint reliability under various thermal cycle conditions. Surface Mount Technology Association International, pp 876–883
23. Andeersson K, Salmela O, Perttula A, Sarkka J, Tammenmaa M (2005) Measurement of acceleration factor for lead-free solder (SnAg3.8CuO.7) in thermal cycling test of BGA components and calibration of lead-free solder joint model for life prediction by finite element analyses, EuraSimE, pp 448–453

24. Vasudevan V, Fan X (2008) An acceleration model for lead-free (SAC) solder joint reliability under thermal cycling. Proceedings of the 58th Electronic Components and Technology Conference, pp 139–145
25. Xie D, Gektin V, Geiger D (2009) Reliability study of high-end Pb-free CBGA solder joint under various thermal cycling test conditions. Proceedings of the 59th Electronic Components and Technology Conference, pp 109–116
26. Salmela O (2007) Acceleration factors for lead-free solder materials. IEEE Trans Compon Packag Technol 30(4):700–707
27. Dauksher W (March 2008) A second-level SAC solder-joint fatigue-life prediction methodology. IEEE Trans Device Mater Reliab 8(1):168–173
28. Norris KC, Landzberg AH (1969) Reliability of controlled collapse interconnections. IBM J Res Dev 13:266–271
29. Ma H, Ahmad M, Liu K-C (2010) Acceleration factor study of lead-free solder joints under wide range thermal cycling conditions. The IEEE 60th Electronic Components and Technology Conference (ECTC), Las Vegas, June 2010, pp 1816–1822

Chapter 11
Interfacial Reactions and Electromigration in Flip-Chip Solder Joints

C.E. Ho, C.R. Kao, and K.N. Tu

Abstract Soldering has been one of the most important assembly and interconnection technologies for electronic products since the beginning of the electronic age. As electronic devices become more complicated and smaller, the number of solder joints on each device increases while the size of these joints decreases—several chemical and physical processes become ever more threatening toward the small joint reliability. These processes include chemical reactions, metal dissolution, diffusion driven by chemical potential gradient, electromigration, Joule heating, thermomigration, and stress migration. There are two threatening issues arise from the combined effect of these process: (1) excessive intermetallic formation and (2) excessive under bump metallurgy (UBM) consumption. In this chapter, we would first discuss the chemical potential gradient-induced dissolution and chemical reaction kinetics between the substrate materials (e.g., Cu and Ni) and the Sn-based solders (e.g., eutectic Sn-Ag and Sn-Ag-Cu). Emerging reliability issues like (1) chemical interaction/cross-interaction between Cu and Ni; (2) effect of alloying with active elements on the reaction; and (3) effect of small solder volume on the solder joints are presented in the section. Next, electromigration issues like effects of electron current stressing on solder and UBM are highlighted. Lastly, mitigation strategy against electromigration is presented.

C.E. Ho
Department of Chemical Engineering & Materials Science, Yuan Ze University, Jhongli, Taoyuan, Taiwan
e-mail: ceho1975@hotmail.com

C.R. Kao (✉)
Department of Materials Science & Engineering, National Taiwan University, Taipei, Taiwan
e-mail: crkao@ntu.edu.tw

K.N. Tu
Department of Materials Science & Engineering, University of California—Los Angeles, Los Angeles, CA, USA
e-mail: kntu@ucla.edu

H.-M. Tong et al. (eds.), *Advanced Flip Chip Packaging*,
DOI 10.1007/978-1-4419-5768-9_11, © Springer Science+Business Media New York 2013

11.1 Introduction

Soldering has been one of the most important assembly and interconnection technologies for electronic products since the beginning of the electronic age. Despite their prevalence, solder joints in electronic products have long been recognized as the weak links responsible for product failures. The numbers of solder joints in modern central or graphical processing units are in the 10^4 range, and for a typical electronic system such as a personal computer the number can be ten times higher. The failure of any one of these solder joints could result in the malfunction of an electronic system, and accordingly the lifespan of the weakest joint determine the reliability of the entire electronic product.

As electronic devices become more complicated and smaller, the number of solder joints on each device increases while the size of these joints decreases. Among the solders joints in an electronic system, the flip-chip solder joints are the smallest and the most technical challenging. At present, 50 μm diameter flip-chip solder joints are under development. To produce solder bumps of such a small size, the traditional electroplating process or the stencil printing process are no longer satisfactory. The electroplating process is capable of producing solder bumps of this size, but offers inadequate solder composition control during electroplating. It will become clear in later parts of this chapter that solder composition control is a critical tool for achieving the required solder joint performance. The stencil printing process does offer a good solder composition control, but printing becomes very difficult when opening size approaches 50 μm. Two new processes offer promising solution for shrinking solder joints to 50 μm and beyond. The first is micro-ball wafer bumping [1]. In this technology, pre-manufactured solder balls are mounted directly onto the Under Bump Metallurgy (UBM) opening of a wafer, just like the mounting of solder balls on Ball Grid Array (BGA) substrates. The second is the IBM C4NP (Controlled Collapse Chip Connection—New Process) [2, 3]. In this technology, molten solder is injected into mold first, forming many micro-solder cylinders or hemispheres in the mold, and then these micro-cylinders or hemispheres are transferred to the matching UBM openings on a wafer, forming the solder bumps. Both processes offer the benefit of exact solder composition control, and the capability to go beyond 50 μm. A schematic drawing for such a solder is shown in Fig. 11.1a.

An alternative approach, the Cu pillar joints, illustrated in Fig. 11.1b, is gaining popularity in the meantime [4–6]. In this process, an electroplated Cu pillar, deposited directly over the UBM, replace bulk of the solder. One advantage of the Cu pillar joint is that the high local current density region is always within Cu, which is much more electromigration-resistant than solder material [7]. The Cu pillar evens out the current distribution before the electrons enter the solder so that electromigration becomes less an issue in this type of joints. This technology also affords a tighter pitch since the molten solder has a less tendency to touch the molten solder from neighboring joints. Nevertheless, since the amount of solder in each joint is very limited, there is a concern that chemical reaction might convert all of the solder into brittle intermetallics [8].

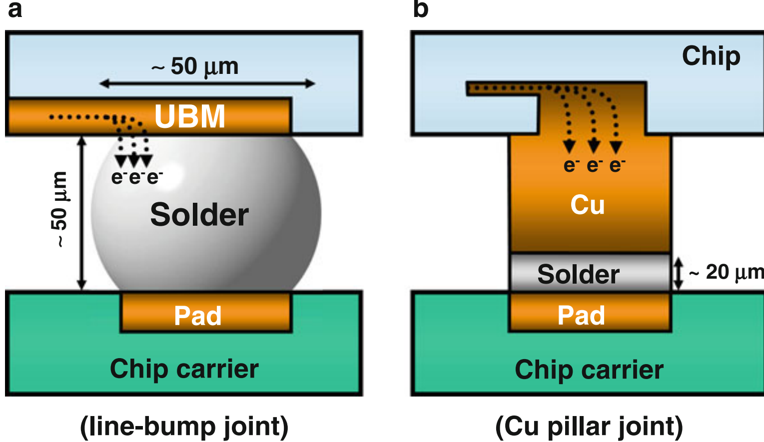


Fig. 11.1 (**a**) Schematic drawing of a flip-chip solder joint. (**b**) Schematic drawing of a Cu pillar bump

As solder joints become smaller, several chemical and physical processes become ever more threatening toward the reliability of flip-chip joints. These processes include the chemical reactions, metal dissolution, diffusion driven by chemical potential gradient, electromigration, Joule heating, thermomigration, and stress migration [7, 9–11]. To make things even more complicated, some of these processed are coupled in that they share common key variables [7, 10, 12]. For example, solder joints stressed under high current density not only generate electromigration, but also Joule heating. The resulting Joule heating can raise the sample temperature, which in turn accelerate metal dissolution and chemical reaction [12]. The metal dissolution and chemical reaction change the microstructure and phases, which in turn disrupt local electron flow, and change the electromigration behaviors. Flip-chip solder joints are truly a unique class of material system in that so many processes become operative at the same time. This is one of the reasons why flip-chip soldering is becoming an ever more challenging technology.

Among these chemical and physical processes, two key processes, chemical reaction and electromigration, are to be discussed in this chapter. The rationale is that these two processes are not only the most important ones but also relatively better understood. The effect of metal dissolution will also be covered mainly because it is closely related to the chemical reaction and electromigration. In the following, chemical reaction is discussed in Sect. 11.2 and electromigration in Sect. 11.3.

11.2 Interfacial Reactions of Lead-Free Solders with Substrates

Soldering by definition involves the chemical reaction(s) between the solder and the two surfaces to be joined together [13], and consequently the importance of understanding the chemical reactions between solders and bonding surfaces cannot

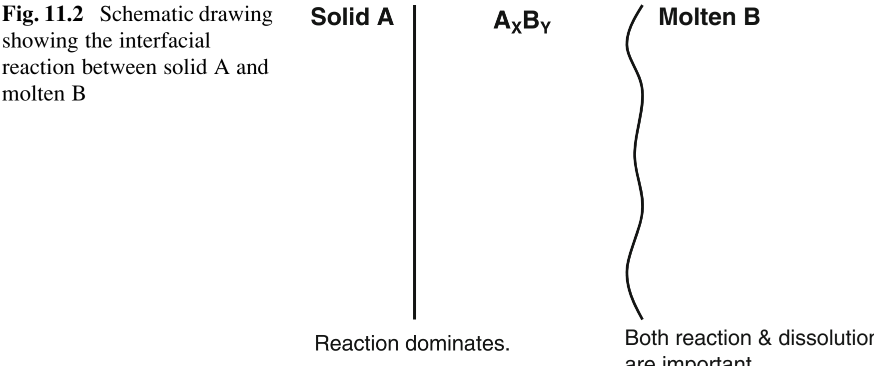


Fig. 11.2 Schematic drawing showing the interfacial reaction between solid A and molten B

be overemphasized. Several review articles for electronic solders have been published in the past few years [9, 10, 14–23], and interfacial reactions were always key elements in these articles.

11.2.1 Dissolution and Its Kinetics During Reflow

Interfacial reaction in flip-chip solder joints can occur during assembly (reflow) and during aging. During reflow, the solder is in the molten state, and the reaction is between a solid (the UBM on the chip side or the surface finish on the substrate side) and a liquid (the molten solder). During aging, the solder is in a solidified state, and the reaction is between a solid and another solid. The key difference between these two cases is that dissolution plays an important role in solid/liquid reactions but in solid/solid reactions dissolution is not as important. As illustrated in Fig. 11.2 showing the reaction between solid A and molten B, both reaction and dissolution at the A_xB_y/B interface are important. At the A/A_xB_y interface, however, dissolution is of lesser an issue. The dissolution effect is particular important if the solid exhibits a considerable solubility in the liquid [24, 25].

Figure 11.3 shows the dissolution rates into molten Sn40Pb (wt%, same below) for several metals that are important for soldering. The data source is from the work of Bader [26]. It should be noted that the data in Fig. 11.3 represent the dissolution of thin metal wires into large solder baths so that the dissolved metal concentrations are always far from reaching saturation at the specified temperature. Because the metal dissolution rate tends to slow down as the metal concentration in solder increases, the data in Fig. 11.3 correspond to the upper bounds at a given temperature. As can be seen in Fig. 11.3, at 250 °C, the typical peak reflow temperature for lead-free soldering, 0.1 ~ 0.2 μm of Cu will be dissolved per second. The dissolution rate of Ni is 30 ~ 40 times slower than that of Cu at the same temperature, illustrating that Ni is a good diffusion barrier materials against soldering reaction. In contrast, the Au dissolution rate is 30 times faster than that of Cu, suggesting that the Au layer in Au/Ni and Au/Pd/Ni surface finishes will disappear from the interface at the very early stage of soldering.

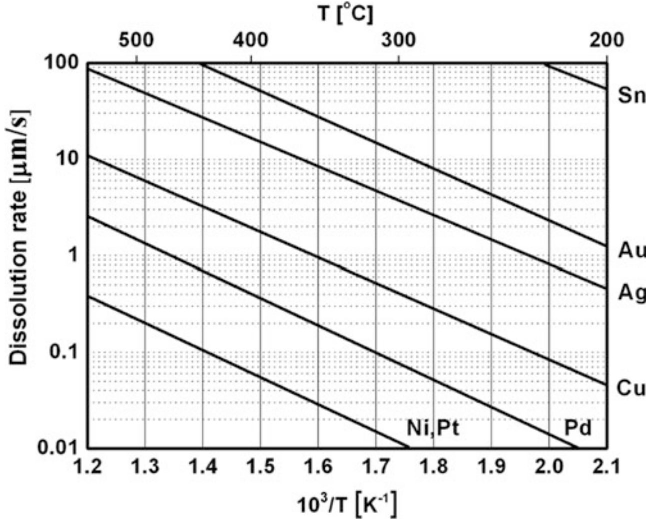


Fig. 11.3 Dissolution rates of various metals into molten Sn40Pb as a function of temperature. Data from reference [26]

In soldering, excessive dissolution of the metal layers of UBMs or surface finishes into molten solders during reflow is one of the key processing concerns. An excessive dissolution of the metals layers might result in failures of devices. Although the data in Fig. 11.3 is a good starting reference point for estimating the consumed thickness, the actual consumed amount is difficult to estimate due to the following complicating factors: the amount of solder used, the contact area between solder and metal, the initial metal concentration in solder, the reflow temperature profile, and the number of reflow. As shown in Fig. 11.4, the consumed Cu thickness after one reflow is plotted versus the initial Cu concentration in Sn3Ag*x*Cu solder [27]. The reflows were carried out on BGA substrates with organic surface preservative (OSP) surface finish over electroplated Cu layer. Each band in Fig. 11.4 represents a different ball diameter and soldering pad diameter combination. The reflow profile used had a peak reflow temperature of 235 ± 2 °C and a 90 s duration during which the solder was molten. The nominal ramp rate and cooling rate were both 1.5 °C/s. As can be seen in Fig. 11.4, a larger solder joint consumed more Cu thickness and exhibited a stronger dependence on the initial Cu concentration. A larger joint size consumed more Cu because it took more Cu to saturate the molten solder. As shown in Fig. 11.5 for the case of 760 μm ball reflowed over 600 μm pad, the first reflow cycle consumed the most Cu, but subsequent reflow cycles only consumed marginally more Cu because the molten solder had almost reached saturation. Typically it takes only 1 ~ 2 reflows to saturate the solder with Cu even for a large BGA joint.

The morphology of the intermetallic after reflow also depends on the joint size. As shown in Fig. 11.6, Cu_6Sn_5 exhibited the classical scallop-type microstructure.

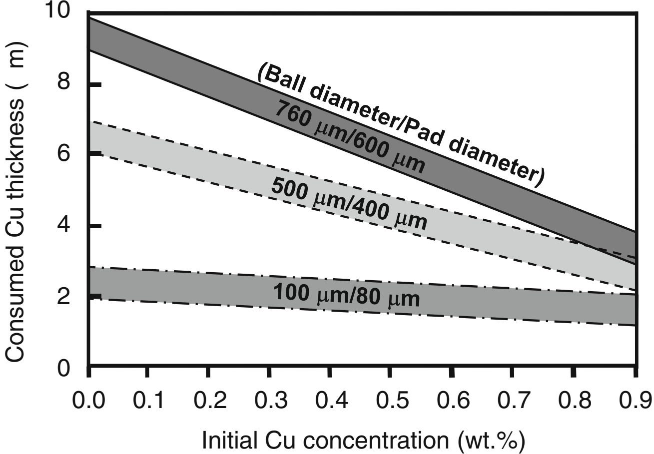


Fig. 11.4 Consumed Cu thickness after one reflow for different initial Cu concentrations and different ball/pad combinations [27]

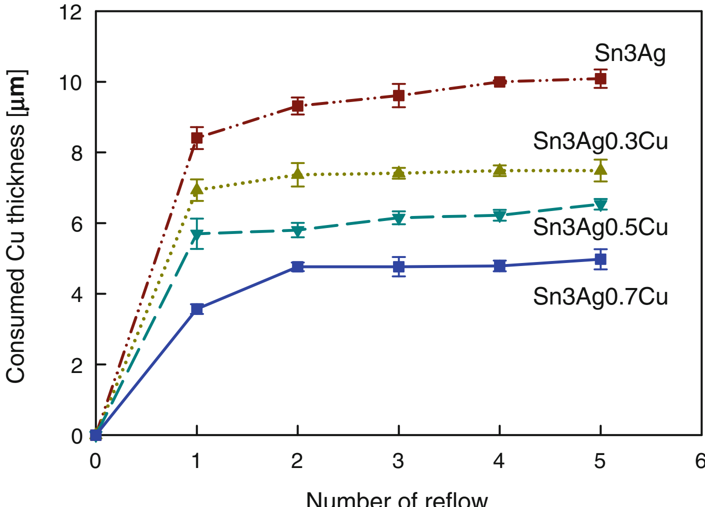


Fig. 11.5 Consumed Cu thickness (accumulated) for different solders plotted versus the number of reflow. The solder ball and pad-opening diameter was fixed at 760 μm and 600 μm, respectively [27]

This scallop-type microstructure is typical from the reaction between Cu and molten Sn-based solder. These scallops would transform into a continuous, layered morphology after solid-state aging [9, 10, 17, 18, 21]. As the diameter of solder balls decreased, the scallop size of Cu_6Sn_5 increased. The average thickness of the intermetallic compound at the interface also increased upon decreasing ball diameter.

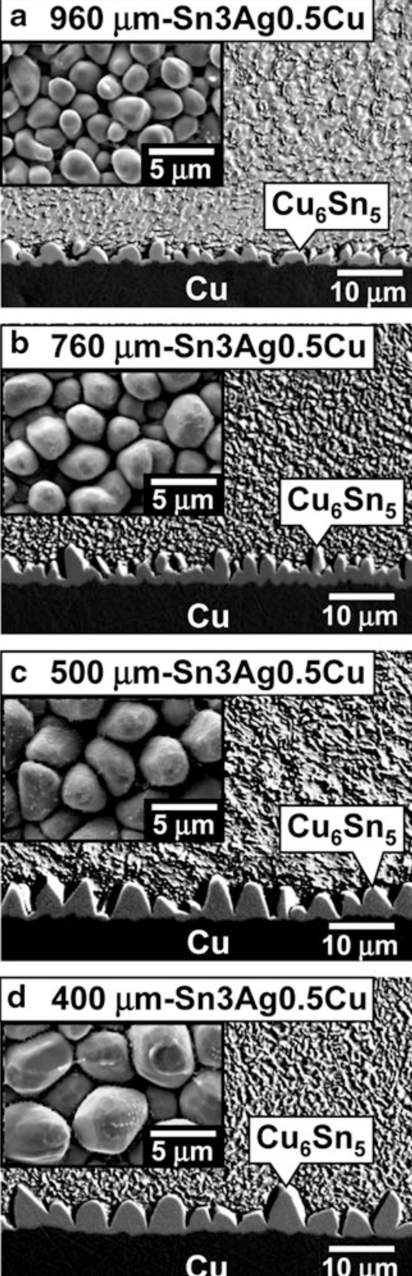


Fig. 11.6 Cross section and top-view (*inset*) micrographs showing the microstructure for Sn3Ag0.5Cu solder balls with different ball sizes. The diameters of solder spheres were (**a**) 960 μm, (**b**) 760 μm, (**c**) 500 μm, and (**d**) 400 μm, respectively [27]

11.2.2 Reaction of Lead-Free Solders with Cu-Based Pads

Copper is one of the most popular choices for the surface layer of UBMs, mainly due to its good wetting property with solders [7, 9, 10, 28]. Copper is also the most common base metal for soldering pad applications. For soldering pad applications, the Cu base metal has to be coated with a surface finish to preserve the wetting property during the storage period before assembly. Popular surface finishes include immersion Sn, immersion Ag, OSP, Ni/Au, and Ni/Pd/Au. In the cases of the first three surface finishes, Sn, Ag, or OSP will be dissolved into the solder (immersion Sn and immersion Ag) or be displaced from the interface (OSP) during soldering, leaving the Cu layer exposed to the solder.

It is well known that at temperatures greater than 50 ~ 60 °C the reactions between Cu and Sn-based solders (pure Sn, eutectic Pb-Sn, and Sn-Ag-Cu, with the exception of Sn-Zn solders [29]) will produce two reaction products, Cu_6Sn_5 and Cu_3Sn, as predicted by the Cu-Sn binary phase diagram shown in Fig. 11.7 [30]. At temperatures lower than 50 ~ 60 °C, only Cu_6Sn_5 is detected at the interface [21]. However, recent TEM (Transmission Electron Microscopy) observation revealed that Cu_3Sn also formed at the early stage, but its thickness was too thin (<100 nm) to be easily resolved. Recently, it was reported that microvoids tend to form within Cu_3Sn when this phase grew [21, 31–35]. As shown in Fig. 11.8a–d, these voids located not only at the Cu/Cu_3Sn interface but also within the Cu_3Sn layer [32]. Without question, the relatively faster diffusion of Cu in Cu_3Sn was the major contributing factor for the formation of these voids. It was indeed experimentally verified that the Cu flux was greater than that of the Sn flux in Cu_3Sn (three times greater at 200 °C) [36].

The microvoids accompanying the Cu_3Sn growth raise serious reliability concerns because excessive void formation increases the potential for brittle interfacial fracture [31, 32, 35, 37]. The voids in Fig. 11.8d had reached such a high population that they almost aggregated into continuous regions. Such a reliability threat is especially serious at high temperature because Cu_3Sn grows relatively fast at temperatures greater than 50 ~ 60 °C. Existing literature results indeed show that the number of drops required in order producing 50 % failures in drop test decreased with the aging time at 125 °C, as shown in Fig. 11.8e [35].

It was recently discovered that minor alloy addition to solders was able to reduce or even completely inhibit the formation of the microvoids. Recent advancements this area are presented in Sect. 11.2.5.

11.2.3 Reaction of Lead-Free Solders with Ni-Based Pads

According to the binary phase diagram, shown in Fig. 11.9 [30], there are three stable intermetallic compounds, Ni_3Sn, Ni_3Sn_2, and Ni_3Sn_4, in the Ni-Sn system. Nevertheless, at temperatures relevant to soldering, only Ni_3Sn_4 was observed

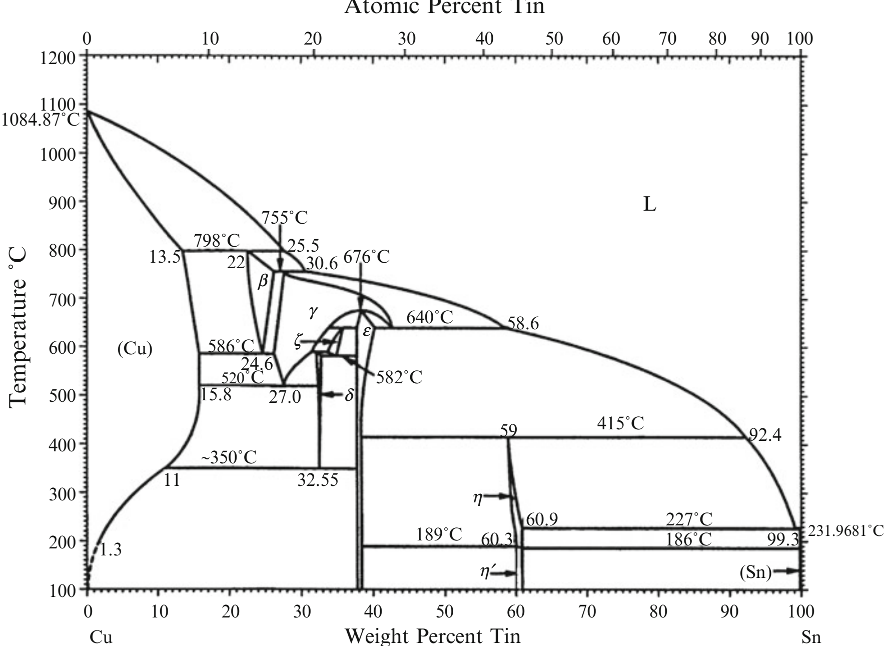


Fig. 11.7 Cu-Sn binary phase diagram [30]

[21, 38, 39]. The other two compounds were observed only at higher temperatures [21, 38, 39]. Although the reaction between Ni and pure Sn only produces Ni_3Sn_4 at temperatures relevant to soldering, the reaction between Ni and Sn alloys with a small amount of Cu is very complicated. Unfortunately, common lead-free solders recommended by various national or international organizations all have Cu as a minor constituent (Table 11.1) [22], and the allowance for the concentration variation for each element is typically ±0.2 wt%, a customary adoption of the practice from the eutectic Pb-Sn solder. As shown below, this ±0.2 wt% difference can change the compound formed at the interface.

11.2.3.1 Uncomplicated Cu Concentration Effect

During the reflow of Sn-Cu or Sn-Ag-Cu solders over those soldering pads that have the Ni, Ni(P), and Ni(V) underlayer, the reaction product(s) at the interface are very sensitive to the Cu concentration in the solders. After the first report [47] of this Cu sensitivity a few years ago, over 40 papers had been published on this subject. As summarized in Table 11.2 [22], the results of these studies as a whole are quite consistent, even though different reflow conditions and different Ni substrates, including Ni, Ni(P), Ni(V), or Au/Ni, had been used. As shown in Table 11.2, when the Cu concentration was low (<0.3 wt%), only $(Ni_{1-y}Cu_y)_3Sn_4$

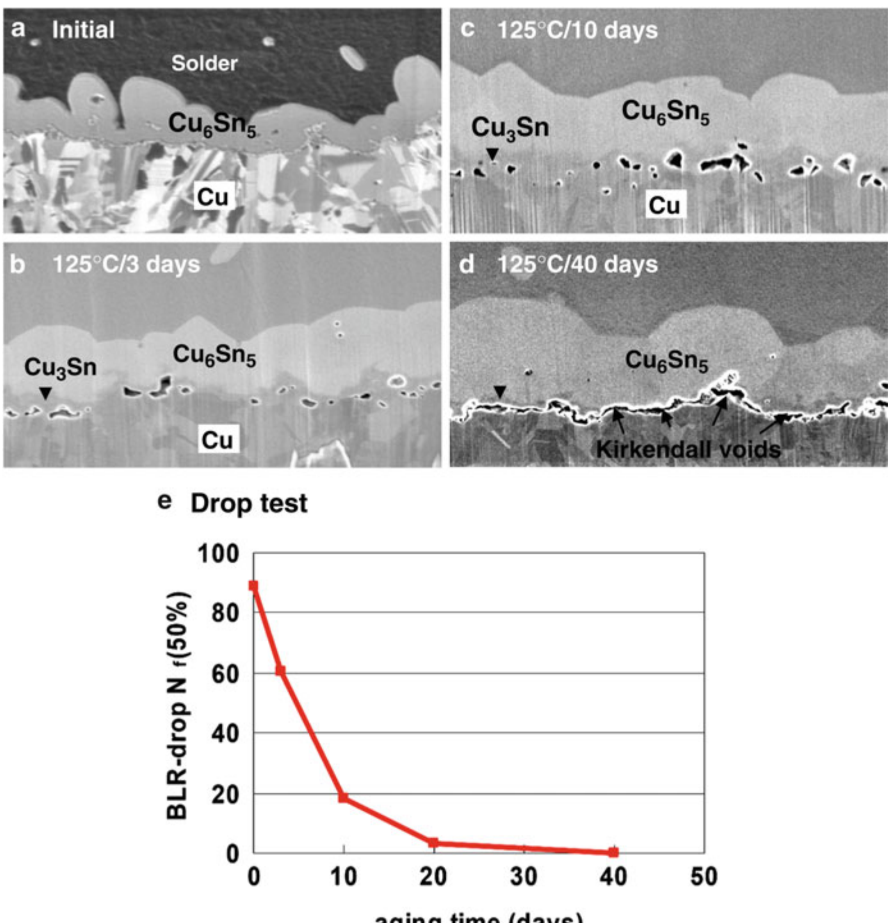


Fig. 11.8 (**a–d**) Focused ion beam (FIB) processed cross sections showing the microstructure of the Sn-Ag-Cu/Cu interfaces after aging at 125 °C for 0∼40 days [32]. (**e**) Cycles required for 50 % failure for board level drop test as a function of aging time [35]

formed at the interface. When the Cu concentration increased to 0.4∼0.5 wt%, both $(Ni_{1-y}Cu_y)_3Sn_4$ and $(Cu_{1-x}Ni_x)_6Sn_5$ formed. When the Cu concentration increased above 0.5 wt%, only $(Cu_{1-x}Ni_x)_6Sn_5$ formed. In a few specific cases [64, 66, 71, 72], where the reported results seemed to differ from the trend shown in Table 11.2 at the first sight, the inconsistency in fact could be attributed to the so-called "solder volume effect," which is to be discussed in the Sect. 11.2.6.

To exclude other factors that may obscure the discussion, the reaction between a large amount of solder and a piece of thick, high purity Ni substrate is examined first. Since the amount of solder was large, one can then assume that the solder composition remained constant during the reaction even though Cu was being

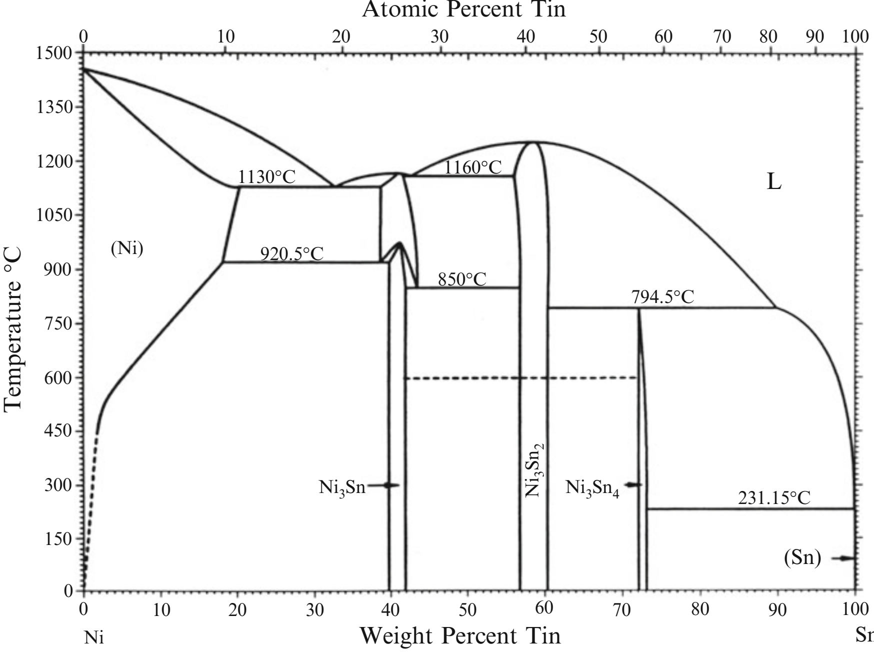


Fig. 11.9 Ni-Sn binary phase diagram [30]

Table 11.1 Lead-free solders recommended by various national or international organizations [22]

Sn-Ag-Cu composition (wt%)	Recommending organizations
Sn–(3.5 ± 0.3)Ag–(0.9 ± 0.2)Cu	NIST (ternary eutectic) [40]
Sn–(3.9 ± 0.2)Ag–(0.6 ± 0.2)Cu	NEMI (N. America) [41]
Sn–(3.4 ~ 4.1)Ag–(0.45 ~ 0.9)Cu	Soldertec-ITRI (UK) [42]
Sn–3.8Ag–0.7Cu	IDEALS (EU) [43]
Sn–3.0Ag–0.5Cu	JEITA (Japan) [44]
Sn–4.0Ag–0.5Cu	–
Sn–2.5Ag–0.8Cu–0.5Sb	AIM, CASTIN alloy [45]
Sn–3.5Ag–0.5Cu–1.0Zn	NCMS [46]

extracting out of the solder to be incorporated into the reaction product(s), i.e., $(Cu,Ni)_6Sn_5$ and $(Ni,Cu)_3Sn_4$. In addition, since the Ni substrate was thick, the Ni substrate was never totally consumed. As shown in Fig. 11.10a, the reaction product at the interface was a continuous layer of $(Ni_{1-y}Cu_y)_3Sn_4$ when the Cu concentration was 0.2 wt%. When the Cu concentration increased to 0.4 wt%, discontinuous $(Cu_{1-x}Ni_x)_6Sn_5$ particles began to form over the $(Ni_{1-y}Cu_y)_3Sn_4$ continuous layer, Fig. 11.10b. When the Cu concentration increased to 0.5 wt%, both $(Cu_{1-x}Ni_x)_6Sn_5$ and $(Ni_{1-y}Cu_y)_3Sn_4$ were continuous, Fig. 11.10c. When the Cu concentration reached 0.6 wt%, only a continuous $(Cu_{1-x}Ni_x)_6Sn_5$ layer was present,

Table 11.2 Summary of the reported reaction products between Sn-based solders and various Ni-bearing soldering pads after reflow

Cu	Ag	Sn	Base metal	Intermetallic(s)	References
(wt%)					
0.0	3.5–3.9	Balance	Ni and Ni(P)	Ni_3Sn_4	[48–52]
			Au/Ni and Au/Ni(P)	Ni_3Sn_4	[37, 53–65]
0.1	0	Balance	Ni	Ni_3Sn_4	[66]
0.2	0–3.9	Balance	Ni and Ni(P)	$(Ni,Cu)_3Sn_4$	[47, 49, 67–69]
0.3	0–3.0	Balance	Ni	$(Ni,Cu)_3Sn_4$	[66, 70]
0.4	0–3.9	Balance	Ni	$(Ni,Cu)_3Sn_4/(Cu,Ni)_6Sn_5$	[49, 67, 69–72]
			Au/Ni(P)	$(Ni,Cu)_3Sn_4/(Cu,Ni)_6Sn_5$	[73]
0.5	1.0–4.0	Balance	Ni and Ni(P)	$(Cu,Ni)_6Sn_5$	[70–72]
			Au/Ni and Au/Ni(P)	$(Cu,Ni)_6Sn_5$	[57, 60, 62, 63, 74–76]
			Ni and Ni(P)	$(Ni,Cu)_3Sn_4/(Cu,Ni)_6Sn_5$	[49, 68, 69]
			Au/Ni and Au/Ni(P)	$(Ni,Cu)_3Sn_4/(Cu,Ni)_6Sn_5$	[54, 64, 74, 77]
0.6	0–3.9	Balance	Ni	$(Cu,Ni)_6Sn_5$	[47, 49, 69–72]
0.7	0–3.8	Balance	Ni and Ni(P)	$(Cu,Ni)_6Sn_5$	[49, 66–68, 78–80]
			Au/Ni and Au/Ni(P)	$(Cu,Ni)_6Sn_5$	[64, 81–88]
0.75	3.5	Balance	Au/Ni and Au/Ni(P)	$(Cu,Ni)_6Sn_5$	[89]
			Ni and Ni(P)	$(Cu,Ni)_6Sn_5$	[54, 58, 61, 89]
0.8	0–3.9	Balance	Au/Ni and Au/Ni(P)	$(Cu,Ni)_6Sn_5$	[49, 69]
			Ni	$(Cu,Ni)_6Sn_5$	[62]
0.9	0	Balance	Ni	$(Cu,Ni)_6Sn_5$	[66]
1.0	3.5–3.9	balance	Ni and Ni(P)	$(Cu,Ni)_6Sn_5$	[49, 67, 68]
			Au/Ni	$(Cu,Ni)_6Sn_5$	[90]
1.5	0	Balance	Ni	$(Cu,Ni)_6Sn_5$	[66]
1.7	4.7	Balance	Ni	$(Cu,Ni)_6Sn_5$	[90]
3.0	0–3.9	Balance	Ni	$(Cu,Ni)_6Sn_5$	[49, 78]
			Au/Ni	$(Cu,Ni)_6Sn_5$	[86]

Only those studies that used reflow conditions similar to the industry practices are included in this table [22]

Fig. 11.10d. The crystal structures of the products had been verified by the X-ray diffraction (XRD) and transmission electron microscopy (TEM) [22, 47, 91]. The discontinuous $(Cu_{1-x}Ni_x)_6Sn_5$ particles in Fig. 11.10b and the continuous $(Cu_{1-x}Ni_x)_6Sn_5$ layer in Fig. 11.10d might have different formation mechanisms. Those $(Cu_{1-x}Ni_x)_6Sn_5$ grains in Fig. 11.10d formed through the direct reaction of Ni and the solder, and accordingly these grains had a preferred growth direction of [0001] [91]. This preferred growth direction had also been observed during the reaction between Ni and the eutectic Pb-Sn solder with 0.5 wt% Cu doping [92]. On the other hand, those discontinuous $(Cu_{1-x}Ni_x)_6Sn_5$ particles in Fig. 11.10b did not show any preferred orientation.

To understand this strong Cu concentration dependency, one needs to have the relevant phase diagram information for the Sn-Ag-Cu-Ni system. Although Ag is an important constituent controlling the solidification microstructure of the solder itself [20, 77, 93], it has been shown that Ag is inert as far as the interfacial reaction

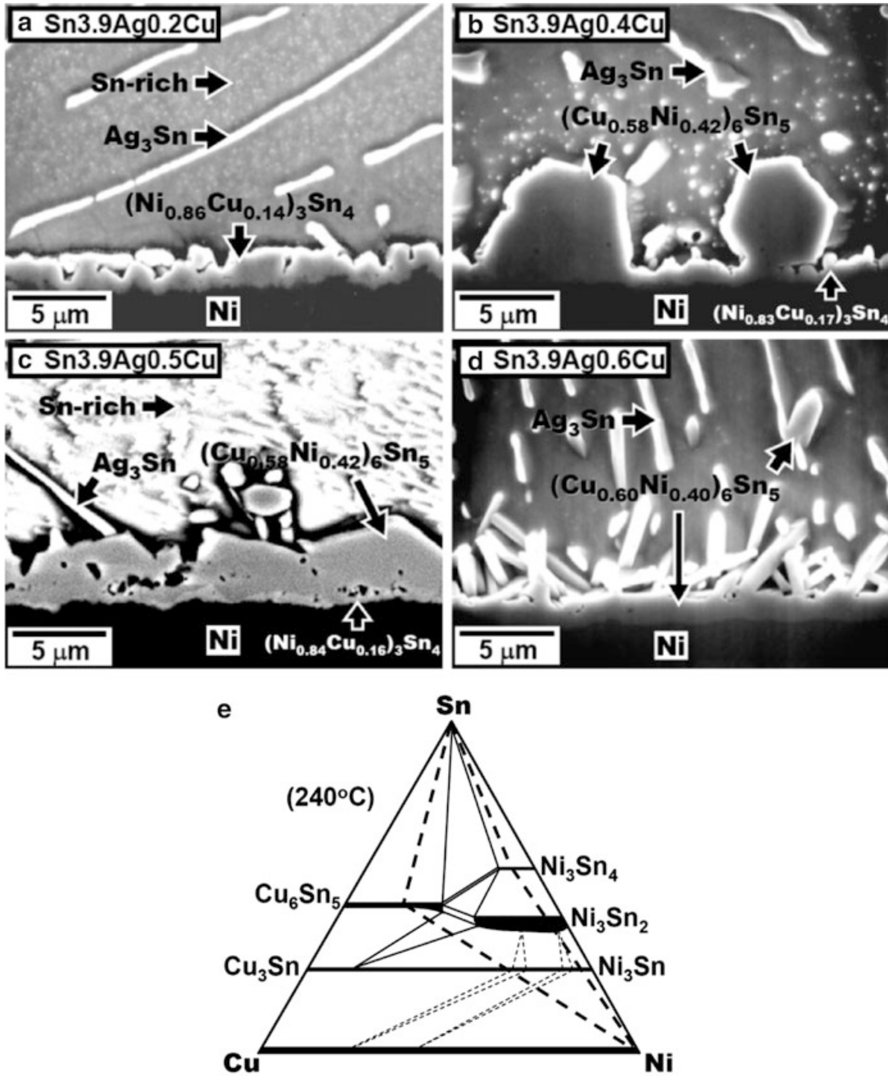


Fig. 11.10 (**a**)–(**d**) Electron micrographs showing the Sn3.9Ag*x*Cu/Ni interfaces after 250 °C soldering for 10 min [22]. (**e**) Cu-Ni-Sn isotherm at 240 °C. This isotherm was replotted according to the diagram by Lin et al. [94]

is concerned [22, 47, 49, 67]. Accordingly, the Sn-Cu-Ni ternary phase diagram is sufficient for the present purpose. The Sn-Cu-Ni isotherm had been measured by three independent groups [21, 94, 95], and the results are reasonably consistent. The 240 °C Sn-Cu-Ni isotherm basing on these two studies is shown in Fig. 11.10e. The Sn-rich corner of this isotherm is shown in Fig. 11.11 [22]. There is some evidence for the existence of a ternary compound (Ni26Cu29Sn45, atomic percent) [96].

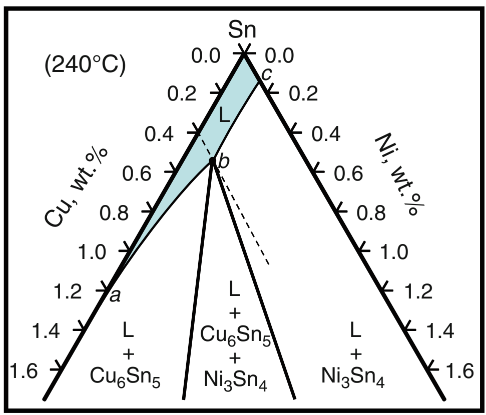


Fig. 11.11 Schematic drawing showing the Sn-corner of Cu-Ni-Sn isotherm at 240 °C [22]

If this compound is indeed stable, then the isotherm in Fig. 11.10e is only a metastable isotherm [21]. Nevertheless, as far as soldering is concerned, the isotherm shown in Fig. 11.10e is still adequate, as results from most soldering reaction experiments were observed to follow the phase relationships shown in Fig. 11.10e.

As shown in Fig. 11.11, the molten Sn phase (L) has a phase boundary a–b–c, which is composed of two segments a–b and b–c. Along a–b, L is in equilibrium with the Cu_6Sn_5 phase, and L is in equilibrium with Ni_3Sn_4 along b–c. The point b represents that L is in equilibrium with both Cu_6Sn_5 and Ni_3Sn_4. As indicated by the dash line passing through b, b has a Cu concentration of about 0.4 wt%. In other words, when the Cu concentration in the solder is 0.4 wt%, this solder is in equilibrium with both Cu_6Sn_5 and Ni_3Sn_4. This explains why both these two phases form as shown in Fig. 11.10b. When the Cu concentration is less than 0.4 wt%, the solder in this range is in equilibrium with only Ni_3Sn_4, and consequently only this phase forms as shown in Fig. 11.10a. When the Cu concentration is higher than 0.4 wt%, L is in equilibrium with only Cu_6Sn_5, and only this phase can be immediately next to L, as shown in Fig. 11.10c, d.

11.2.4 Cross-interaction Between Cu and Ni Across a Solder Joint

In Sects. 11.2.2 and 11.2.3, the reaction between Cu and solder and the reaction between Ni and solder are discussed, respectively. In this section, the interaction between Cu and Ni across a solder joint is considered. The importance of such a cross-interaction had been reported in several publications available in the literature [22, 81, 97–106].

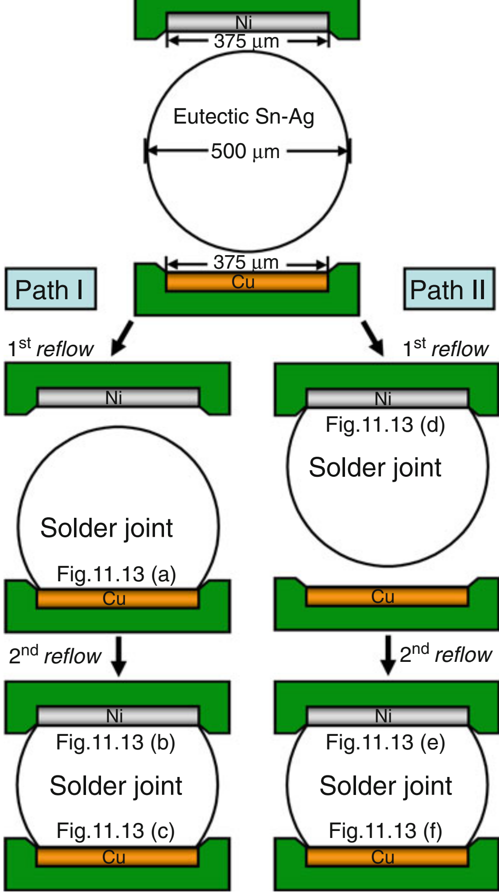


Fig. 11.12 Schematic Illustration showing the two sequences that the solder joints were assembled [22]

11.2.4.1 Cross-interaction During Reaction with Molten Solders (Reflow)

To investigate the cross-interaction during the reflow stage, the solder joints that were assembled by two different sequences are compared. As illustrated in Fig. 11.12, in path I, the solder (Sn3.5Ag) was attached to the Cu substrate in the first reflow, and then the Ni substrate was attached to the other side of the solder joint in the second reflow. In path II, the solder was attached to the Ni substrate first (the first reflow), and in the second reflow Cu substrate was attached to the other side of the solder joint.

Shown in Fig. 11.13 are the solder/Cu and Ni/solder interfaces for path I (left column) and path II (right column) after the first and second reflow. For path I, the solder/Cu interface after the first reflow (Fig. 11.13a) is quite typical: formation of Cu_6Sn_5 scallops. Such a reflow dissolves a large quantity of Cu from the substrate. Our own data shows that as high as 7.5(±0.3) μm of Cu can be dissolved into the Sn3.5Ag, making the Cu concentration in solder become as high as ~1.2 wt%.

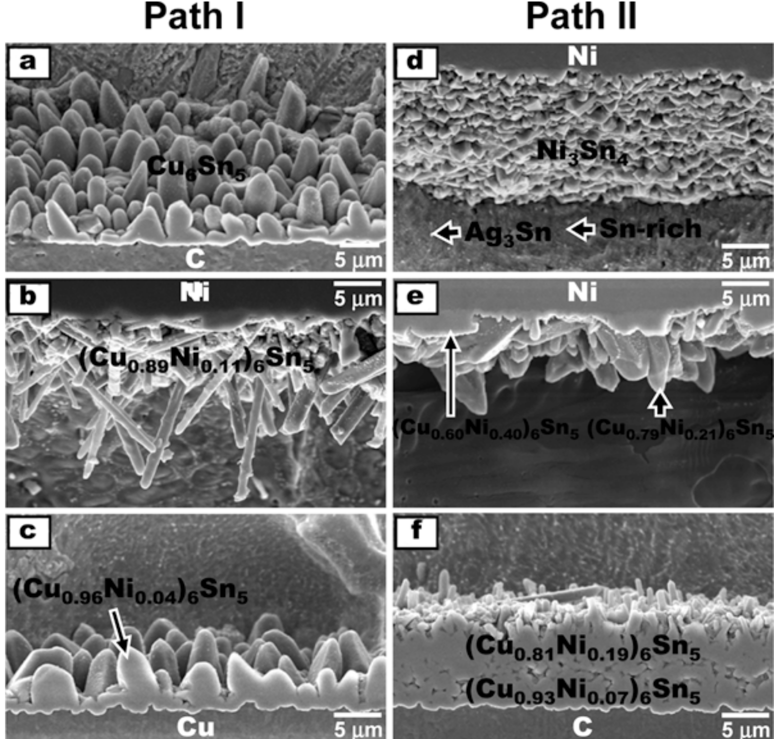


Fig. 11.13 Electron micrographs showing the solder/Cu and Ni/solder interfaces for path I (*left column*) and path II (*right column*) in Fig. 11.12. The micrographs in (**a**) and (**d**) had been tilted by 30° [22]

These dissolved Cu atoms precipitated out as Cu_6Sn_5 inside the solder when the solder solidified. During the second reflow of path I, the dissolved Cu atoms had an effect on the Ni/solder reaction. As shown in Fig. 11.13b, the reaction product under such conditions was $(Cu_{0.89}Ni_{0.11})_6Sn_5$. This microstructure exhibits marked difference with that of Fig. 11.13d, which shows the reaction between fresh Sn3.5Ag and Ni substrate. The second reflow also influenced the original solder/Cu interface. As shown in Fig. 11.13c, a small amount of Ni (~2 at.%) can be detected in Cu_6Sn_5. The only Ni source was from the Ni substrate. This result clearly shows that during the second reflow Ni had the capacity to dissolve at the Ni/solder interface, migrate across the solder joint, and incorporate itself into Cu_6Sn_5. In short, during the assembly of a Cu/solder/Ni solder joint through path I, the two interfaces can cross-interact during the reflow stage.

For path II after the first reflow, a thin layer of Ni_3Sn_4 formed at the Ni/solder interface, as shown in Fig. 11.13d. A small amount of Ni also dissolved into the molten solder. During the second reflow, the dissolved Ni atoms had an effect on the solder/Cu reaction. As shown in Fig. 11.13f, the reaction product at the solder/Cu interface had two distinct regions with their own compositions,

$(Cu_{0.81}Ni_{0.19})_6Sn_5$ and $(Cu_{0.93}Ni_{0.07})_6Sn_5$. Apparently, the reflow condition was sufficient to dissolve enough amount of Ni into solder to produce a clear effect. Quite surprisingly, the dissolved Cu during its fist reflow (i.e., the second reflow in path II) was also high enough to induce the Cu concentration effect. As shown in Fig. 11.13e, the outer intermetallic at the Ni/solder interface had changed from Ni_3Sn_4 to $(Cu_{1-x}Ni_x)_6Sn_5$. This fast transformation from one compound to another in time as short as 90 s does not seem reasonable. A careful cutting by using the Focused Ion Beam (FIB) technique was able to show that this transformation was indeed fast, but was not complete. There was a very thin layer of $(Ni_{1-y}Cu_y)_3Sn_4$ beneath the $(Cu_{1-x}Ni_x)_6Sn_5$ layer [22]. Comparing path I and II, one can conclude that the sequence of assembly has an effect on the intermetallic microstructures.

11.2.4.2 Cross-interaction During Reaction with Solid Solders (Aging)

Figure 11.14 shows the microstructures of the interfaces in Fig. 11.13 after aging at 160 °C for 1,000 h. The Ni/solder interface and the solder/Cu interface that had been soldered only once and consequently had no cross-interaction during aging were shown in the left column for comparison [Fig. 11.14a, b]. In the middle column and the right column, the corresponding interfaces for the path I and path II after aging are shown, respectively. It can be seen that Cu_3Sn in Fig. 11.14d, f are thinner than that in Fig. 11.14b. Moreover, Cu_3Sn in Fig. 11.14f is also thinner than that in Fig. 11.14d. This is reasonable because in path II more Ni was dissolved due to the fact that Ni in path II went through one more reflow than in path I. In addition, no microvoid was observed in Fig. 11.14d, f even by the means of FIB sample preparation. It should also be pointed that the cross-interaction accelerated the consumption rate of the Cu substrate and reduced the consumption rate of the Ni substrate [107]. The Ni consumption rate decreased because the growth of $(Cu_{1-x}Ni_x)_6Sn_5$ consumed a less amount of Ni than did the growth of $(Ni_{1-y}Cu_y)_3Sn_4$ [54, 92, 107, 108].

11.2.5 Effect of Alloying with Other Active Elements

The Sn-Ag-Cu family of solders has obtained a wide acceptance as a replacement for the Pb-Sn eutectic solder in electronic applications. At present, relatively fewer activities are undertaken by researchers worldwide to develop another solder family to replace the Pb-Sn eutectic solder. Instead, the current main research thrust in lead-free solder alloy development is on enhancing or fine-tuning the various properties of Sn-Ag-Cu through adding minor alloying elements. For example, Ni was evaluated for its potential as a minor alloying element to Sn-based lead-free solders [22, 109–118], and so were Ge [109, 111], Fe [118, 119], Co [112, 115, 118, 119], Zn [29, 111, 120, 121], Bi [122], Mn [111], Ti [111], Si [111], and Cr [111]. These previous studies point out minor alloy additions can strengthen the solder performance in many different respects. Among these elements, Fe, Co, Ni, Zn, and

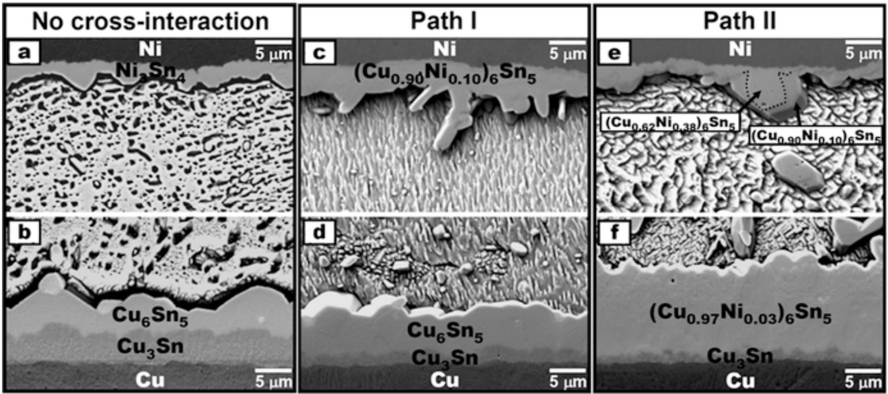


Fig. 11.14 Electron micrographs showing the microstructures of the interfaces in Fig. 11.12 after aging at 160 °C for 1,000 h [22]

Cu are more noteworthy. In the following, the effects of these alloy additions are discussed. As the effects of Fe, Co, and Ni additions are quite similar, these three elements are discussed together.

11.2.5.1 Fe, Co, and Ni Additions

It had been shown that Ni addition to Sn3.5Ag in amounts as minute as 0.01 wt% (100 ppm) was able to substantially retard the Cu_3Sn growth during soldering [97] as well as during the sequential solid-state aging [110, 111, 116–118, 123, 124]. As shown in Fig. 11.15a, b, the Cu_3Sn thickness remained very thin even after aging at 150 °C for 1,000 h [110]. Because the Cu_3Sn growth had been linked to the formation of microvoids, which in turn increased the potential for a brittle interfacial fracture, thinner Cu_3Sn layers might translate into better solder joint strength.

Figure 11.15b also shows that Ni addition increased the amount of Cu_6Sn_5 at the interface compared to that without Ni addition. In addition, the Cu_6Sn_5 phase now possessed a small amount of Ni, and became $(Cu_{1-x}Ni_x)_6Sn_5$ [110]. The explanation for the formation of this thicker $(Cu_{1-x}Ni_x)_6Sn_5$ is presented elsewhere [22] and is omitted here for brevity.

The way Ni was introduced into the reacting system did not seem to matter as far as hindering the Cu_3Sn growth was concerned. It had been reported that the growth of Cu_3Sn was hindered even if Ni-alloyed Cu substrates were used. These substrates included Cu alloyed with 6∼9 wt% Ni [125] and Cu alloyed with 15 at.% Ni [21].

The additions of Fe and Co to lead-free solder produce very similar effects to that of Ni [118] in terms of reducing Cu_3Sn thickness and increasing Cu_6Sn_5 thickness. The minimum amount of Fe or Co required to produce these effects is also very similar to that of Ni. It was experimentally verified that additions of Fe, Co, or Ni in an amount as small as 0.03 wt% were effective in reducing the Cu_3Sn thickness at 160 °C for at least 2,000 h.

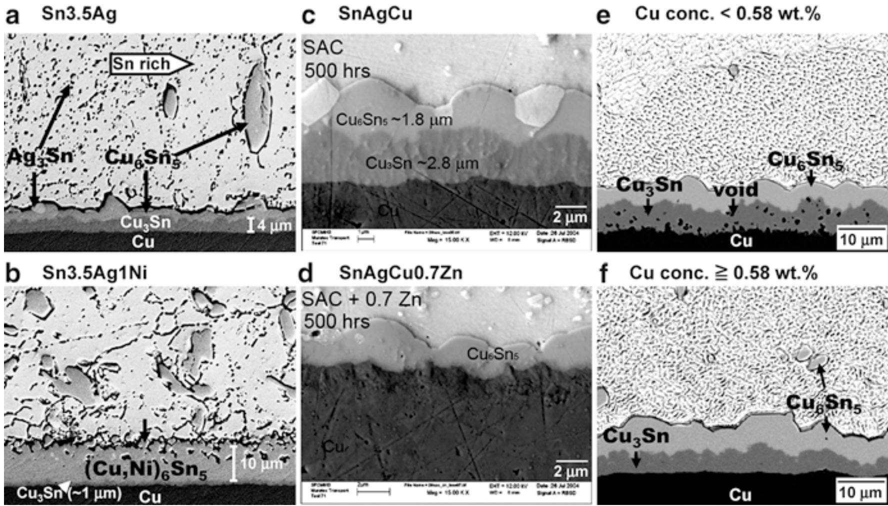


Fig. 11.15 (**a**) Sn3.5Ag/Cu interface after aging at 150 °C for 1,000 h [110]; (**b**) Sn3.5Ag0.1Ni/Cu interface after aging at 150 °C for 1,000 h [110]; (**c**) Sn-Ag-Cu/Cu after aging at 150 °C for 500 h [120]; (**d**) SnAgCu0.7Zn/Cu after aging at 150 °C for 500 h [120]; (**e**) Sn-*x*Cu/Cu interface ($x < 0.58$ wt%) after aging at 150 °C for 1,000 h [126]; (**f**) Sn-*x*Cu/Cu interface ($x \geq 0.58$ wt%) after aging at 150 °C for 1,000 h [126]

11.2.5.2 Zn Addition

While Fe, Co, or Ni additions are able to decrease the amount of microvoids by reducing the thickness of Cu_3Sn. It was recently discovered that minor Zn additions (0.1 and 0.7 wt%) into Sn-Ag-Cu solders could completely suppress the growth of Cu_3Sn, and thus avoided the formation of microvoids [121]. As shown in Fig. 11.15c, without Zn addition, a thick layer of Cu_3Sn can be seen at the interface [120]. However, with just 0.7 wt% Zn, Cu_3Sn disappeared from the interface, as shown in Fig. 11.15d [120].

Adding Zn has the additional benefits of improving the creep resistance of Sn-Ag-Cu [127], and restraining the formation of large Ag_3Sn plates [128]. It was proposed that the suppression of the Cu_3Sn growth was probably associated with the accumulation of Zn at the Cu_3Sn/Cu interface [121].

If the amount of Zn addition is high enough (>2 wt%), Zn could in fact completely inhibit the formation of both Cu_3Sn and Cu_6Sn_5, and produced Cu-Zn compounds instead [29, 129]. Figures 11.16a–c show the reaction products that formed for different Zn concentrations after reaction at 250 °C for 2 min [29]. When the solder composition was Sn/0.5Zn [Fig. 11.16a], a layer of compound with the scallop microstructure existed at the interface. The compound had the Cu_6Sn_5 crystal structure as verified by XRD analysis. The composition of this Cu_6Sn_5 layer was determined by FE-EPMA to be 39.1 at.% Sn, 4.8 at.% Zn, and 56.1 at.% Cu, indicating that a small amount of Zn had been incorporated into the Sn

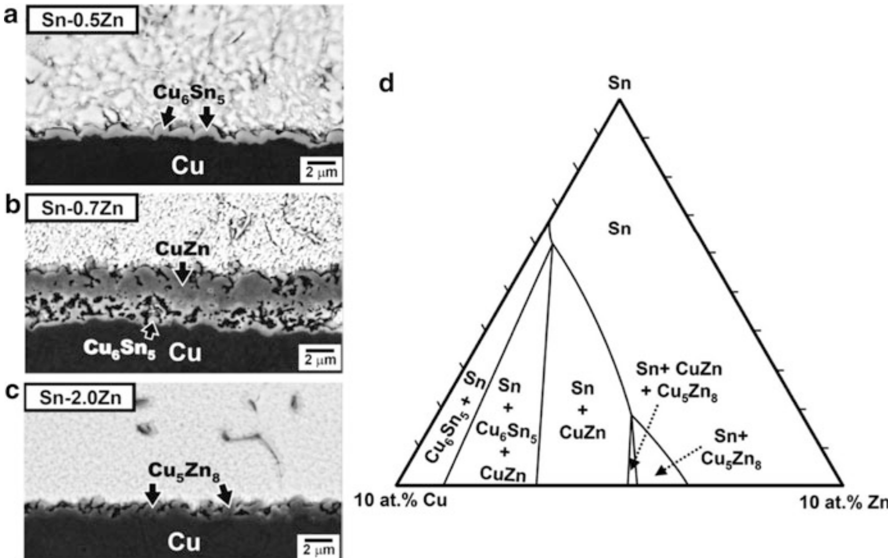


Fig. 11.16 (**a**)–(**c**) Cross-sectional view for the Sn-xZn/Cu interfaces that reacted at 250 °C for 2 min: (**a**) $x = 0.5$, (**b**) $x = 0.7$, and (**c**) $x = 2.0$. The Sn-0.5Zn/Cu reaction produced Cu_6Sn_5. The Sn-2.0Zn/Cu reaction produced Cu_5Zn_8. The Sn-0.7Zn/Cu reaction produced Cu_6Sn_5 and Cu-Zn. A selective Sn etching solution (5 vol % HCl-methanol) had been used to reveal the morphologies of reaction zones [29]. (**d**) Sn-rich corner of the Cu-Sn-Zn ternary isotherm at 250 °C. This isotherm was drawn based on the information determined by Chou et al. [131]

sublattice of Cu_6Sn_5. Consequently, this compound should be more properly referred to as $Cu_6(Sn_{1-x}Zn_x)_5$. It should be noted that accurate composition measurements for this compound layer would have been difficult by using a conventional EPMA because this layer was too thin. The FE-EPMA used in this study has a much smaller interaction volume, and therefore accurate composition measurements could be performed. We also measured the composition using a Sn-0.5Zn/Cu sample that had been reacted for 10 min and had a thicker $Cu_6(Sn_{1-x}Zn_x)_5$ (1.5 μm). The measured $Cu_6(Sn_{1-x}Zn_x)_5$ composition remained the same, suggesting that the thickness was not an issue and the measured composition was accurate. From the fact that only $Cu_6(Sn_{1-x}Zn_x)_5$ formed, one can conclude that the addition of 0.5 wt% Zn to Sn did not change the chemistry of the pure-Sn/Cu reaction as reported in the literature. The effect of Zn addition at this level was that Zn became incorporated into the Sn sublattice of Cu_6Sn_5. However, when the Zn addition was increased to 2 wt%, the chemistry was completely different, as shown in Fig. 11.16c. The reaction product now had the Cu_5Zn_8 crystal structure as verified by X-ray diffraction analysis. Using samples that was reacted for 2 min and 10 min (larger particles size facilitating more accurate FE-EPMA measurements), we determined the composition of Cu_5Zn_8 to be 1.5 at.% Sn, 59.5 at.% Zn, and 39.0 at.% Cu. Apparently, Zn was the most reactive element, and when the Zn concentration was high enough, the Zn-Cu reaction dominated at

the interface. The type of compound formed here, Cu_5Zn_8, was the same as that formed in the reaction between eutectic Sn-Zn (Sn-9Zn) and Cu substrate [130]. Therefore, it can be expected that the reaction product will be Cu_5Zn_8 when the Zn concentration is between 2 and 9 wt%. When the Zn concentration was 0.7 wt%, two compounds were identified at the interface: a Cu_6Sn_5-based compound next to the Cu substrate and a Cu-Zn-based compound next to the solder, as shown in Fig. 11.16b. According to the FE-EPMA measurements, the composition for the Cu_6Sn_5-based compound was 38.3 at.% Sn, 7.1 at.% Zn, and 54.6 at.% Cu, and the composition for the Cu-Zn-based compound was 17.5 at.% Sn, 33.1 at.% Zn, and 49.4 at.% Cu.

This strong Zn effect results can be explained by using the Cu-Sn-Zn phase diagram. Shown in Fig. 11.16d is the Sn-rich corner of the Cu-Sn-Zn 250 °C isotherm according to Chou et al. [131]. Notice that the molten Sn phase can be in two-phase equilibrium with Cu_6Sn_5, CuZn, or Cu_5Zn_8, depending on the Zn concentration. When the Zn concentration in solder was high (2 wt%), the interface represented a tie-line in the Sn + Cu_5Zn_8 two-phase field, and Cu_5Zn_8 formed next to the solder. When the Zn concentration was low (0.5 wt%), the interface represented a tie-line in the Sn + Cu_6Sn_5 two-phase field, and Cu_6Sn_5 forms next to the solder. When the Zn concentration was in-between (0.7 wt%), the Sn + Cu_6Sn_5 + CuZn three-phase field dominated, and both Cu_6Sn_5 and CuZn formed. It should be noted that the CuZn + Sn two-phase field was not encountered in this study, and the CuZn phase did not form alone at the interface. Nevertheless, if a Sn-Zn alloy with Zn concentration between 0.7 and 2 wt% Zn had been used, it was very likely that CuZn would form alone at the interface.

11.2.5.3 Cu Addition

While Cu is a key element in many commercially successful Sn-Ag-Cu lead-free solders, its role in inhibiting the formation of microvoids within the Cu_3Sn layer has not been fully appreciated. As shown in Fig. 11.15e, f, microvoids within the Cu_3Sn layer can be completely inhibited by maintaining the Cu concentration in solders higher than 0.58 wt%. It is believed that the high Cu concentration in solder is able to reduce the Cu flux diffusing through the intermetallics to solder. This reduction in Cu flux reduces the over-saturation of the vacancy within Cu_3Sn, thus eliminating the nucleation and growth of void with this layer [126].

11.2.6 Effect of Small Solder Volume

As the solder joints become smaller, the solder might markedly enrich in the dissolved metal(s), altering the microstructures of the solder joints and/or the chemical interaction between the solder and the metallization pad [155, 156]. Two emerging issues become challenging in the case of small solder volumes.

As the solder joints become smaller, two emerging issues become challenging. The first is due to the gradual consumption of the active element (other than Sn) in solder. The active element becomes incorporated into the reaction product(s) and its effective concentration in solder decrease as more reaction product(s) form. The second issue is the gradual consumption of the main constituent of solder (Sn). When a large portion of Sn is converted into the reaction products, the volume percentage occupied by intermetallic(s) is also large. At this stage, bulk of a solder joint is made up of brittle intermetallics, instead of the much softer solder, and as a result the mechanical properties degrade. This second issue is particular for flip-chip joint smaller than 50 μm.

11.2.6.1 Exhaustion of the Active Element Due to Smaller Solder Volume

During reflow and the subsequent aging, Cu atoms in Sn-Ag-Cu solders become incorporated into the intermetallic(s) [i.e., $(Cu_{1-x}Ni_x)_6Sn_5$ and $(Ni_{1-y}Cu_y)_3Sn_4$] at the interface. As the intermetallic(s) grow, more Cu atoms are being consumed. If the amount of the solder is relatively large so that the supply of Cu is nearly unlimited, the average Cu concentration in the solder could consequently remain constant. The thermodynamic condition at the interface was thus static, and the formation of one compound or another was more or less dictated by the thermodynamics. For a real solder joint in array-array packages, the supply of Cu is actually very limited because the solder volume is quite small in the first place, and secondly the Cu concentration in Sn-Ag-Cu solder is always less than a few atomic percents (see Table 11.1). The Cu concentration can decrease noticeably as the intermetallic (s) grow [72]. As the Cu concentration changes, the type of the equilibrium intermetallic at the interface might also change. Now the condition at the interface becomes dynamic. Now the solder volume must be considered as the volume of the solder determines the total available Cu. As the size of the joints shrinks, the supply of Cu becomes more limited, and the decrease in Cu concentration becomes more critical.

In the reaction between Sn-Ag-Cu solder and Ni substrate, the original amount of Cu in the solder ball before reflow equals the remaining Cu in the solder plus the Cu which is incorporated into the intermetallic(s). Neglecting the Cu atoms in those intermetallic particles located inside the solder, one can obtain the following equation [72]:

$$W_{\mathrm{Cu}} - W_{\mathrm{Cu}}^{0} \cong -40 \frac{d_{\mathrm{pad}}^{2}}{d_{\mathrm{joint}}^{3}} T_{\mathrm{IMC}} \tag{11.1}$$

where W_{Cu}^{0} and W_{Cu} represent the Cu concentration (in wt%) in Sn-Ag-Cu before reflow and the remaining Cu concentration after reflow, respectively. The symbols d_{joint} and d_{pad} represent the diameters (in μm) of the solder ball and the pad's opening of the substrate, respectively. The symbol T_{IMC} represents the thickness (in μm) of intermetallic compound (IMC) at the interface. The compounds

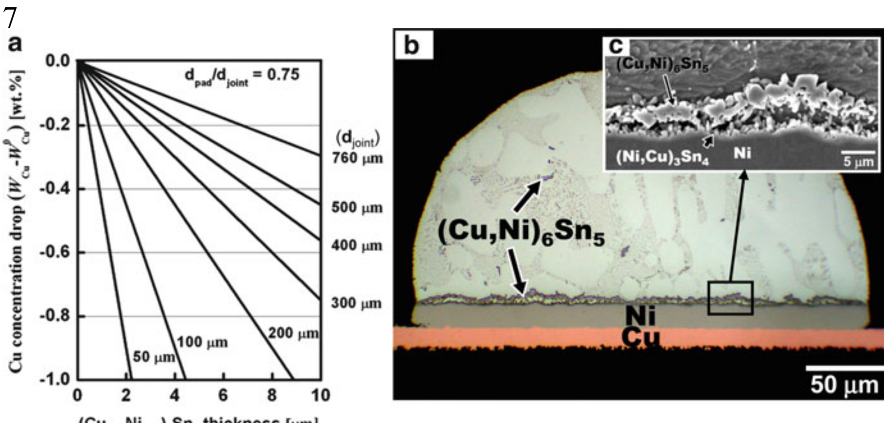


Fig. 11.17 (**a**) Calculated results showing the drop of Cu concentration with different $(Cu_{0.60}Ni_{0.40})_6Sn_5$ thickness in different ball/pad combinations (d_{joint}/d_{pad}). (**b**) Massing spalling of $(Cu,Ni)_6Sn_5$ after Sn3Ag0.6Cu soldering with Ni at 235 °C for 20 min. (**c**) Zoom-in view of (**b**). The solder ball (d_{joint}) and pad-opening diameter (d_{pad}) was fixed at 300 μm and 375 μm, respectively [72]

$(Cu_{1-x}Ni_x)_6Sn_5$ and $(Ni_{1-y}Cu_y)_3Sn_4$ are the only two Cu-bearing intermetallics that can be present at the interface, and their compositions had been determined by EPMA to be around $(Cu_{0.60}Ni_{0.40})_6Sn_5$ and $(Ni_{0.80}Cu_{0.20})_3Sn_4$ for those solders with Cu within 0.4~0.6 wt% [47, 67]. Here, those Cu in $(Ni_{1-y}Cu_y)_3Sn_4$ could be ignored because the thickness of this compound was thin and its Cu concentration was low (~6 wt%). The drop of the Cu concentration according to (11.1) for several different ball/pad combinations ranging from the BGA to flip-chip dimensions is plotted in Fig. 11.17a. In plotting this figure, d_{pad}/d_{joint} is kept constant at 0.75. This figure shows that if 2 μm $(Cu_{1-x}Ni_x)_6Sn_5$, which is the thickness commonly seen in as-reflow solder joints, forms for $d_{joint} = 100$ μm, the Cu concentration will drop by as large as 0.45 wt%. Under such a condition, the residue Cu concentration in solder will be less than 0.3 wt% for most Sn-Ag-Cu solder compositions listed in Table 11.1. If the Cu concentration indeed becomes less than 0.3 wt%, then the Cu_6Sn_5/molten solder interface is no longer thermodynamically stable. There will be a huge driving force for Ni_3Sn_4 [or $(Ni_{1-y}Cu_y)_3Sn_4$] to form. For BGA or flip-chip joints, the design rule usually calls for the condition that $d_{pad} \approx 0.75 d_{joint}$, therefore the drop of the Cu concentration is proportional to the inverse of d_{joint}.

Experimental evidence of the solder volume effect is presented in Fig. 11.17b, c [72]. Here, $d_{pad} = 375$ μm and $d_{joint} = 300$ μm, and the solder used was Sn3Ag0.6Cu. The samples had been reflowed with a typical profile (235 °C peak temperature, 90 s molten solder duration). Using (11.1) and the $(Cu_{1-x}Ni_x)_6Sn_5$ thickness measured from Fig. 11.17c (2.8 μm), one can calculate the Cu concentration drop to be 0.58 wt%. As a result, the remaining Cu concentration of the solders after reflow was nearly zero. Consequently, the Cu concentration was so low that a

new layer of $(Ni_{1-y}Cu_y)_3Sn_4$ had nucleated beneath the $(Cu_{1-x}Ni_x)_6Sn_5$ layer that formed in the early stage of the reaction when the Cu concentration was still high. In Fig. 11.17b, c, the total separation (massive spalling) of the upper $(Cu_{1-x}Ni_x)_6Sn_5$ layer from the interface. That is, the consequence of the shifting equilibrium phase leads to the massive spalling of $(Cu_{1-x}Ni_x)_6Sn_5$. The massive spalling has a higher tendency to occur in smaller joints because these joints will experience a lager Cu concentration drop.

Massive spalling of intermetallic compounds has been reported in the literature not only for the Sn-Ag-Cu soldered on Ni solder/substrate systems [72], but also several other solder/substrate systems including Sn-Zn on Cu [29], high-Pb Pb-Sn on Cu [132, 133], and high-Pb Pb-Sn on Ni [134]. A unified thermodynamic argument has been proposed to explain this rather unusual phenomenon [135]. According to this argument, two necessary conditions must be met. The number one condition is that at least one of the reactive constituents of the solder must be present in a limited amount, and the second condition is that the soldering reaction has to be very sensitive to its concentration. With the growth of intermetallic, more and more atoms of this constituent are extracted out of the solder and incorporated into the intermetallic. As the concentration of this constituent decreases, the original intermetallic at the interface becomes a non-equilibrium phase, and the spalling of the original intermetallic occurs. It should be emphasized that the arguments presented here are purely thermodynamic in nature. Nothing is said about dynamically how massive spalling occurs. In other words, we only identify the driving force for the massive spalling. The step-by-step mechanism is an important and extremely interesting area worthy of more studies. Without question, the interfacial energy must have played an important role in "peeling" the intermetallic layer away from the interface.

The solder volume effect also plays an important role during the solid-state aging. Figure 11.18 shows the solder joints with the same reflow conditions (235 °C peak temperature, 90 s molten solder duration), the same aging conditions (160 °C for 1,000 h), and the same d_{pad} value (375 μm), but with different d_{joint} values and different Cu compositions in solder. When the joint was relatively large (760 μm), the only intermetallic compound at the interface was $(Cu_{1-x}Ni_x)_6Sn_5$ for all the solder compositions, as shown in Fig. 11.18a–c. This is due to the fact that the supply of Cu was relatively abundant for the larger solder joints.

For the solder joints with the medium diameter (500 μm), the supply of Cu became somewhat limited. There was still only the $(Cu_{1-x}Ni_x)_6Sn_5$ layer at the interface if the joint's Cu concentration was high enough (0.5 and 0.7 wt%), as shown in Fig. 11.18d, e. But when the applied Cu concentration became very low (0.3 wt%), as shown in Fig. 11.18f, a layer of $(Ni_{1-y}Cu_y)_3Sn_4$ formed beneath $(Cu_{1-x}Ni_x)_6Sn_5$. Given more time, the remaining $(Cu_{1-x}Ni_x)_6Sn_5$ will also disappear, leaving $(Ni_{1-y}Cu_y)_3Sn_4$ as the only layer. This is because the Cu atoms in $(Cu_{1-x}Ni_x)_6Sn_5$ are extracted out to form more $(Ni_{1-y}Cu_y)_3Sn_4$.

When the solder joints became even smaller (300 μm), the supply of Cu became very limited. For the joint with the highest Cu concentration (0.7 wt%), shown in Fig. 11.18g, the amount of Cu was still enough to sustain a $(Cu_{1-x}Ni_x)_6Sn_5$ layer

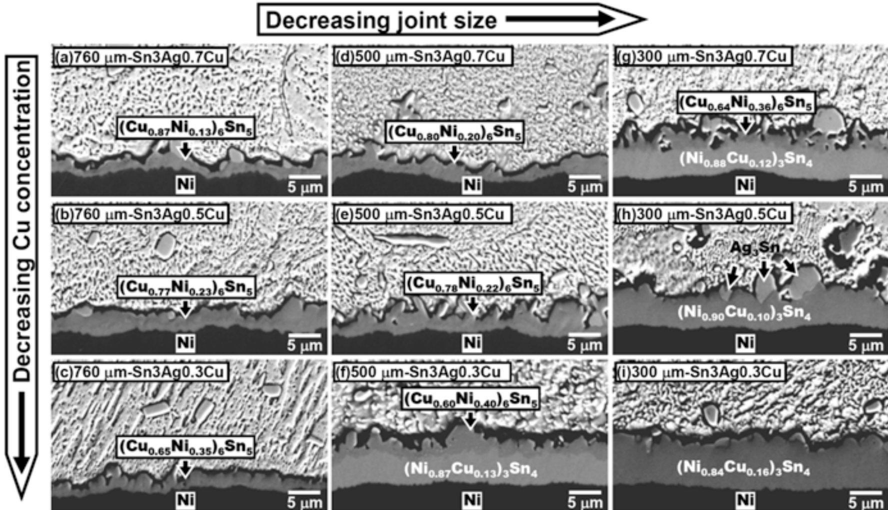


Fig. 11.18 Electron micrographs showing solder joints with the same reflow conditions (235 °C peak temperature, 90 s molten solder duration), the same aging conditions (160 °C for 1,000 h), and the same d_{pad} value (375 μm), but with different d_{joint} values and different Cu compositions [22]

over the $(Ni_{1-y}Cu_y)_3Sn_4$, as shown in Fig. 11.18g. But when the Cu concentration became lower (0.5 and 0.3 wt% Cu), shown in Fig. 11.18h, i, the amount of Cu was not enough to sustain the $(Cu_{1-x}Ni_x)_6Sn_5$ layer, and only a $(Ni_{1-y}Cu_y)_3Sn_4$ layer was present. It should be noted that in these two cases $(Cu_{1-x}Ni_x)_6Sn_5$ did appear over the $(Ni_{1-y}Cu_y)_3Sn_4$ layer once during certain time frame, but $(Cu_{1-x}Ni_x)_6Sn_5$ eventually disappeared and be converted into $(Ni_{1-y}Cu_y)_3Sn_4$. It can be anticipated that the $(Cu_{1-x}Ni_x)_6Sn_5$ layer in Fig. 11.18g will also disappear if this sample is subjected to longer aging.

From our recent study, the formation of both $(Ni_{1-y}Cu_y)_3Sn_4$ and $(Cu_{1-x}Ni_x)_6Sn_5$ layers at the interface after aging has a negative impact on the strength of the joints. The joints tend to fail along the interface of these two compounds. This is because both these two compounds are quite brittle, and the interface between two brittle intermetallics unfortunately tends to be weak, especially for two compounds originate from two different binary systems.

There are two approaches to inhibit the massive spalling during reflow. The first is to use solders with a higher Cu concentration. The second method is to provide an infinite Cu source, such as a thick Cu layer on either side of a solder joint [72].

11.2.6.2 Excessive Intermetallic Formation Due to Small Solder Volume

When the size of solder joints shrinks below a certain level, such as 50 μm, the amount of solder becomes limited. For solder joints of such small sizes, it becomes a real possibility that during high temperature storage test a large portion of Sn be

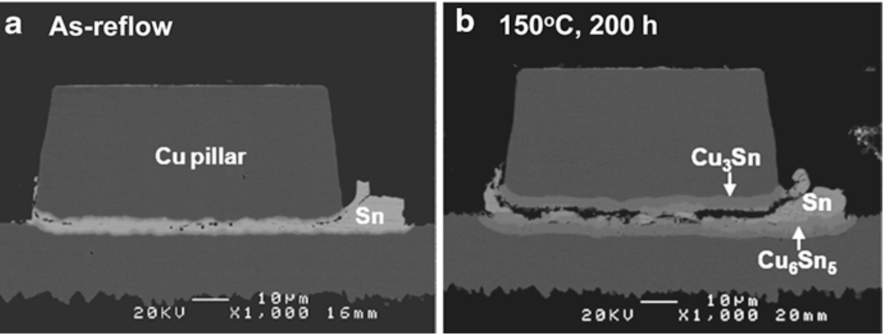


Fig. 11.19 Impact failure for copper pillar solder joint having a high fraction of Cu-Sn Intermetallics. The original thickness of Sn was about 2 μm [8]

converted into intermetallics. At this stage, bulk of a solder joint is made up of brittle intermetallics, instead of solder which is much softer and compliant. As a result, the mechanical properties of the joints degrade. Illustrated in Fig. 11.19 [8] is an example of failure caused by excessive consumption of the solder. As shown in Fig. 11.19b, a crack propagated through the solder joint, where most of the solder had been converted into intermetallics after aging at 150 °C for only 200 h. Typical high temperature storage test requirement is at 150 °C for at least 1,000 h. Clearly, for solder joints smaller than such a size, Cu cannot be left to be indirect contact with Sn-based solder. A diffusion barrier layer, such as Ni, has to be used.

11.3 Electromigration in Flip-Chip Solder Joints

For typical 100 μm flip-chip solder joints in use today, an electric current of about 0.2 A passes through those solder joints that are responsible for power-carrying (instead of signal-carrying). This translates into a current density of about 2×10^3 A/cm^2. While this current density is about two orders of magnitude less than that in Al or Cu interconnect lines, electromigration in the solder joints cannot be ignored [7, 9, 10, 136–139]. This is because of the low melting point and high atomic diffusivity of solder alloys. For the eutectic Sn-Pb solder with a melting point of 183 °C, room temperature is about two thirds of its melting point on the absolute temperature scale. This is also true for Pb-free solders. The second reason is the low "critical product" of solder alloys [7], so that electromigration can occur in solder alloy under a lower current density. The third reason is the line-to-bump geometry, in which a large change of current density exists between the interconnect line and the solder bump, as illustrated in Fig. 11.1a. It leads to a unique current crowding at the line-to-bump contact interface, where the current density is a factor of 10∼20 higher than the average current density in the bump. Effect of the current crowding is one of the most serious reliability issues in flip-chip solder joints from

the point of view of electromigration failure [140]. The fourth reason is Joule heating from the Al or Cu interconnect line contacting the bump. The Joule heating not only will increase the temperature of the solder bump, which in turn will increase the rate of electromigration, but also may produce a small temperature difference across the solder bump to cause thermomigration. A temperature difference of 10 °C across a solder bump 100 μm in diameter will give a temperature gradient of 1,000 °C/cm, which cannot be ignored [141]. Another very unique and important aspect of electromigration behavior in solder joints is that it has two reactive interfaces. The polarity effect on IMC growth exists at the cathode and the anode. Electromigration drives atoms from the cathode to the anode. Therefore, it tends to dissolve or retard the growth of IMC at the cathode but build up or enhance the growth of IMC at the anode [94, 137, 142, 143].

11.3.1 Fundamentals of Electromigration

Electromigration is the result of a combination of thermal and electrical effects on mass transport. If the conducting pathway is kept at a very low temperature there is no atomic mobility of diffusion, even though there is a driving force resulting from current stressing. In other words, electromigration is still a diffusion process in nature. The applied electron current only assists the atom diffusion in the direction of electron flow. There are two main diffusion mechanisms in metals: vacancy and interstitial. Atom diffusion by either mechanism can be assisted by the applied current. Illustrated in Fig. 11.20a1 are a shaded metal atom and a neighboring vacancy in a lattice before they exchange position. When the shaded atom is diffusing halfway toward the vacancy as depicted in Fig. 11.20a2, it is at the activated state, sitting at a saddle point while displacing the four nearest-neighbor atoms. Since the saddle point is not part of the lattice periodicity, the atom at the saddle position is out of its equilibrium position and will make a much larger contribution to the resistance to electrical current than a normal lattice atom. In other words, it experiences a greater electron scattering and hence a greater electron wind force which will push it to an equilibrium position, the vacant site, as illustrated in Fig. 11.20a3. The diffusion of the atom is enhanced in the direction of the electron flow. Similarly, interstitial atom diffusion can also be assisted by the applied current, as illustrated in Fig. 11.20b1–b3.

In typical flip-chip solder joints, the medium through which metal atoms migrate include solders, soldering pads (UBM on the chip side and surface finish on the substrate side), and intermetallics formed between solders and soldering pads. The main constituents for solders are Sn (and maybe also Pb) with minor alloy elements such as Cu, Ag, and Ni. Common soldering pad materials include Cu, Ni, Au, and Pd. Common intermetallic formed at the interfaces include Cu_6Sn_5, Cu_3Sn, and Ni_3Sn_4. All these revenant atoms (Sn, Pb, Cu, Ag, Ni, Au, and Pd) diffuse through the soldering pads and the intermetallics by the vacancy mechanism. The diffusion of Sn or Pb in solders is also through the vacancy mechanism. Nevertheless, all

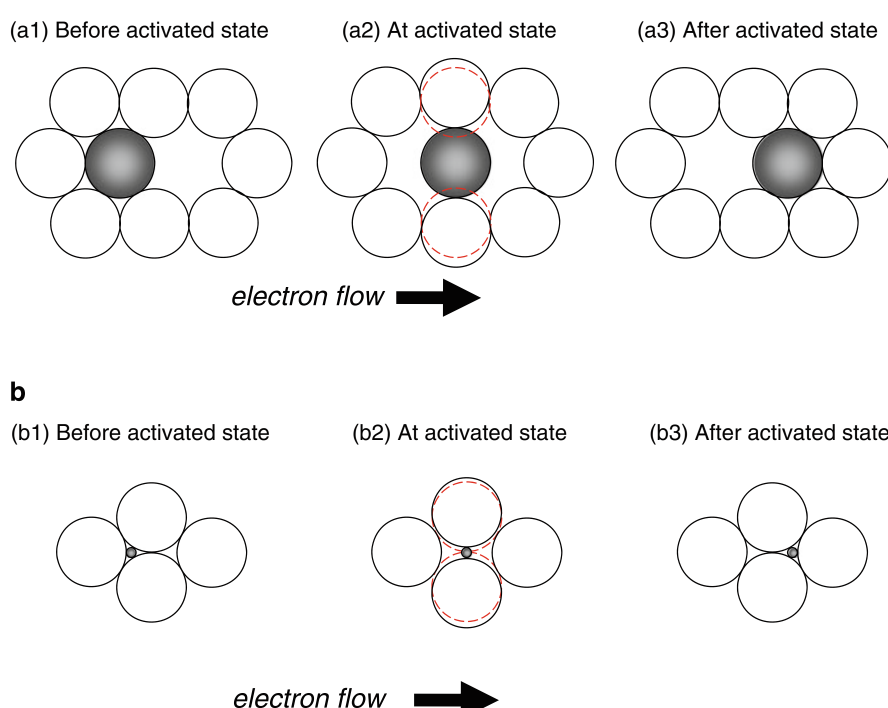


Fig. 11.20 Schematic diagram of electromigration for atoms diffuse by (**a**) vacancy mechanism, and (**b**) interstitial mechanism

other relevant atoms (Cu, Ag, Ni, Au, and Pd) diffuse through the solders by the interstitial mechanism [144]. In fact, these atoms (Cu, Ag, Ni, Au, and Pd) are among those so-called "fast diffuser" in both pure Sn and Pb [144]. As shown in Fig. 11.21, these fast diffusers have diffusion coefficients two to three orders of magnitude faster than those elements that diffuse by the vacancy mechanism in Sn or Pb. In short, many elements used as the soldering pad materials can diffuse very rapidly through solders. This fast brings about important implications that are to be discussed in the following sections.

According to Huntington and Grone [145], the atomic flux of atom A induced by electromigration J_{EM} can be written as

$$J_{\mathrm{EM}} = \frac{cDZ^*}{kTn\mu_{\mathrm{e}}} J_{\mathrm{e}} \tag{11.2}$$

where c is the concentration of A per unit volume, D is the atom diffusivity of A, Z^* is the effective charge number of A, k is the Boltzmann constant, T is the absolute temperature, n is the electron density of the substrate, μ_{e} is the electron mobility in the substrate, and J_{e} is the electron flux.

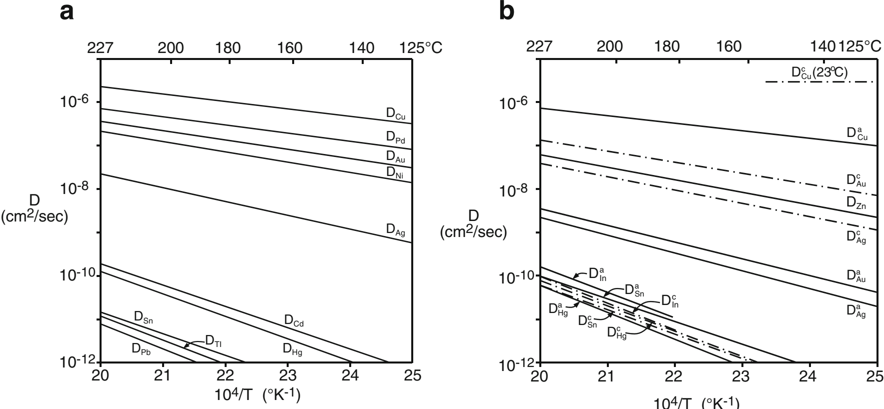


Fig. 11.21 (**a**) Diffusivities of various elements in Pb as a function of inverse temperature. (**b**) Diffusivities of various elements along Sn *a* and *c* axes as a function of inverse temperature. Data are from [144] and replotted here

Consider a Sn-based solder joint with a small amount of fast diffuser element A (such as Cu or Ni) dissolved in the Sn phase under current stressing. As mentioned earlier, Sn diffuses by the vacancy mechanism in solders, and A by the interstitial mechanism in solders. Both Sn and A diffuse toward the anode side as a result of the applied current, and the diffusion of Sn and that of A are not coupled because these two species diffused by different mechanisms. The electromigration flux for Sn and the electromigration flux for A can be written as

$$J_{\mathrm{EM}}^{\mathrm{Sn}} = \frac{c^{\mathrm{Sn}} D^{\mathrm{Sn}} Z^{*\mathrm{Sn}}}{kTn\mu_{\mathrm{e}}} J_{\mathrm{e}} \tag{11.3}$$

$$J_{\mathrm{EM}}^{\mathrm{A}} = \frac{c^{\mathrm{A}} D^{\mathrm{A}} Z^{*\mathrm{A}}}{kTn\mu_{\mathrm{e}}} J_{\mathrm{e}} \tag{11.4}$$

where the superscript Sn and A denote species. The ratio of the Sn flux to that of the A flux at any certain location inside the solder joint can be obtained by dividing (11.3) by (11.4), and is

$$J_{\mathrm{EM}}^{\mathrm{Sn}} / J_{\mathrm{EM}}^{\mathrm{A}} = \frac{c^{\mathrm{Sn}} D^{\mathrm{Sn}} Z^{*\mathrm{Sn}}}{c^{\mathrm{A}} D^{\mathrm{A}} Z^{*\mathrm{A}}} \tag{11.5}$$

The solubility limits of most fast diffusers in pure solid Sn are quite limited, typically much less than 1 at.%. It follows that

$$\frac{c^{\mathrm{Sn}}}{c^{\mathrm{A}}} \tilde{} (x_{\mathrm{A}})^{-1} \tag{11.6}$$

where x_A is the mole fraction of A in solder. Combing (11.5) and (11.6), one obtains

$$J_{EM}^{Sn}/J_{EM}^{A} = \frac{1}{x_A}\frac{D^{Sn}Z^{*Sn}}{D^{A}Z^{*A}} \tag{11.7}$$

The ratio D^{Sn}/D^{A} in (11.7) is a very small number since A is a fast diffuser. For example, if A is Cu, the ratio is $10^{-6} \sim 10^{-7}$ at 125 °C and even smaller at lower temperatures, according to Fig. 11.21.

The order of magnitude number for the ratio Z^{*Sn}/Z^{*A} is one if Sn and A have the same valence number [10]. In addition, the ratio Z^{*Sn}/Z^{*A} does not vary too much with temperature because Z^* is in itself is a weak function of temperature [10].

The ratio in (11.7) plays an important role in determining the key failure mechanisms for flip-chip solder joints. In flip-chip solder joints, two major electromigration failure mechanisms reported in the literature are (1) pancake-type void formation on the cathode side of the solder [146] and (2) UBM metal dissolution [142]. The first mechanism requires the transport of enough Sn atoms to the anode side. The pancake void forms as a result of the accumulation of sufficient incoming vacancies from the anode side. The second mechanism requires the rapid dissolution of the UBM metal. To sustain a rapid and continuing dissolution, the dissolved A atoms have to be sweep away from the cathode side or solder near the cathode side will become saturated with A. If the solder is saturated with A near the cathodes side, the dissolution of A will inevitably stop. These two mechanisms operate in parallel and compete with each other. The eventual mechanism responsible for failure is the one that cause failure first. In the following sections, these two mechanisms are examined in finer detail.

11.3.2 Effect of Current Stressing on Solder and the Resulting Failure Mechanism

At present, the diameter of a solder joint is about 100 μm and it will be reduced to 50 μm and even 25 μm soon. If the diameter of solder bumps and the spacing between them is 50 μm, we can place $100 \times 100 = 10,000$ solder joints on a 1×1 cm chip surface. In the most advanced devices today, there are already over 7,000 solder joints on a chip. When a current of 0.2 A is applied to a bump, the average current density in a 50-μm solder bump is about 10^4 A/cm^2. This current density is about two orders of magnitude smaller than that in Al and Cu interconnects. Nevertheless, electromigration does occur in flip-chip solder joints at such low current density, and it occurs by lattice diffusion. Often the easy electromigration in solder joints is explained by the low melting point of solder or fast atomic diffusion in solder. However, Table 11.3 shows that at the device working temperature of 100 °C, lattice diffusion in solder is not much slower than grain boundary diffusion in Al, nor slower than surface diffusion in Cu. While the

Table 11.3 Self diffusivities of Al, Cu, and Sn-Ag-Cu [7]

	Melting point (K)	373 K/ T_m	Diffusivities at 100 °C (cm^2/s)	Diffusivities at 350 °C (cm^2/s)
Cu	1,356	0.275	Lattice: $D_l = 7 \times 10^{-28}$ Grain boundary: $D_{gb} = 3 \times 10^{-15}$ Surface: $D_s = 10^{-12}$	$D_l = 5 \times 10^{-17}$ $D_{gb} = 1.2 \times 10^{-9}$ $D_s = 10^{-8}$
Al	933	0.4	Lattice: $D_l = 1.5 \times 10^{-19}$ Grain boundary: $D_{gb} = 6 \times 10^{-11}$	$D_l = 10^{-11}$ $D_{gb} = 5 \times 10^{-7}$
Sn-Ag-Cu	~490	0.76	Lattice: $D_l = 2 \times 10^{-9}$ to 2×10^{-10}	Molten state $D_l > 10^{-5}$

total atomic flux in lattice diffusion is much greater than that in grain boundary diffusion or in surface diffusion, so is the larger volume of a void required to cause failure of a solder joint. Therefore, low melting point or fast diffusion is not the key answer. Why electromigration can occur in flip-chip solder joints at such low current densities is explained below; it is due to the low "critical product" of electromigration of solder alloys [7]. Furthermore, the line-to-bump geometry of a flip-chip solder joint leads to a large current crowding at the cathode contact where accelerated electromigration occurs.

11.3.2.1 Low Critical Product of Solder Alloys

The "critical product" is defined is the following equation [10]

$$J_e \Delta x = \frac{Y \Delta \varepsilon \Omega}{Z^* e \rho} \tag{11.8}$$

where Δx is the critical length, Y is Young's modulus, $\Delta\varepsilon = 0.2\ \%$ is the elastic limit, Ω is the atomic volume, e is the charge of an electron, and ρ is the resistivity. To compare the value of "critical product" among Cu, Al, and eutectic Sn-Pb, we recall that eutectic Sn-Pb has a resistivity that is one order of magnitude larger than those of Al and Cu. The Young's modulus of eutectic Sn-Pb (30 GPa) is a factor of 2∼4 smaller than those of Al (69 GPa) and Cu (110 GPa). The effective charge number of eutectic Sn-Pb (Z^* of lattice diffusion) is about one order of magnitude larger than those of Al (Z^* of grain boundary diffusion) and Cu (Z^* of surface diffusion). Therefore, in (11.8), if Δx is kept constant for comparison, the current density needed to cause electromigration damage in eutectic Sn-Pb solder is two orders of magnitude smaller than that needed for Al and Cu interconnects. If Al or Cu interconnect fails in electromigration by a current density of 10^5 to 10^6 A/cm^2, solder joint will fail by 10^3 to 10^4 A/cm^2. This is the major reason why electromigration in flip-chip solder joints can be serious.

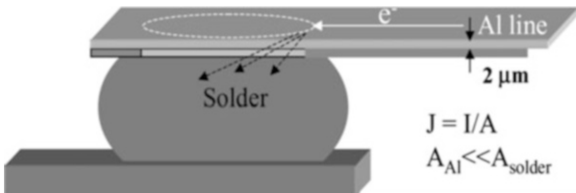


Fig. 11.22 Schematic diagram depicting the geometry of a flip-chip solder bump joining an interconnect line on the chip side (top) and a conducting trace on the board or module side (bottom) [10]

11.3.2.2 Current Crowding in Flip-Chip Solder Joints

Figure 11.22 is a schematic diagram depicting the geometry of a flip-chip solder bump between an interconnect line on the chip side (top) and a conducting trace on the board or module side (bottom). Current crowding occurs at the contact interface between the solder bump and interconnect. The high current density due to current crowding is about one order of magnitude higher than the average current density in the bulk of solder joint. The low threshold of the current density needed to cause electromigration in solder and the high current density induced by current crowding are the key reasons why electromigration in flip-chip solder joints can compete with electromigration in Al and Cu interconnects as the major reliability problem in microelectronic devices.

11.3.2.3 Pancake-Type Void Formation Within Solder on the Cathode Side

Because the cross section of the line on the chip side is at least two orders of magnitude smaller than that of the solder bump, as depicted in Fig. 11.22, there is a very large current density change at the contact between the bump and the line because the same current is passing between them. Since electric current will take the lowest resistance path, electrons will jam at the entrance into the solder bump, resulting in current crowding. The current density in the solder near the entrance point will be about one order of magnitude higher than the average current density in the middle of the bump. It will be 10^5 A/cm^2 near the entrance when the average current density in the middle of the bump is 10^4 A/cm^2. Figure 11.23a is a two-dimensional simulation of current distribution in a solder joint [140]. Figure 11.23b is a display of current density distribution in the joint, where the cross section of the joint is plotted on the x–y plane and the current density is plotted along the z-axis [140]. It is the current crowding or the high current density shown at the upper right corner in Fig. 11.23a, b that leads to electromigration damage in the solder joint, not the average current density in the bulk of the joint. Consequently, electromigration damage in a flip-chip solder joint occurs near the cathode contact on the chip side, i.e., the contact between the interconnect and the bump. The damage begins near the entrance point of the electric current, and propagates across the contact, as illustrated in Fig. 11.24. This mode of failure is sketched schematically in

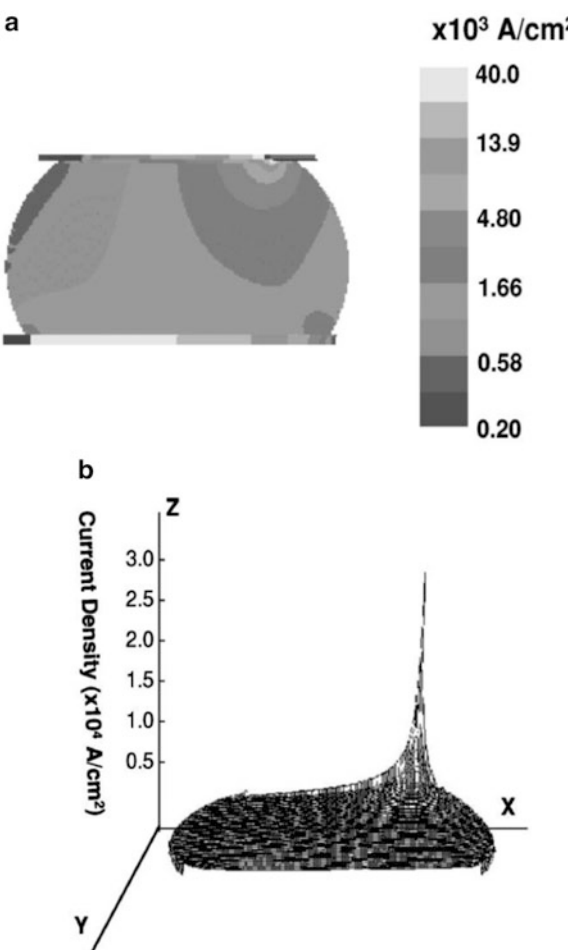


Fig. 11.23 (**a**) Two-dimensional simulation of current distribution in a line-bump solder joint. (**b**) Current density distribution in the joint, where the cross section of the joint is plotted on the *x*–*y* plane and the current density is plotted along the *z*-axis [140]

Fig. 11.25 [140]. The out diffusion of the Sn driven by electromigration creates an incoming vacancy flux. Vacancy flux divergence near the current crowding region leads to vacancy accumulation. As a void nucleates by condensing the supersaturated vacancies, it will displace the current to its peripheral area and leads to a lateral growth of the void. So, the void spreads across the contact.

An incubation time is often needed for the nucleation of the pancake-type void. Figure 11.26 shows the evolutions of the microstructure in the current-up joint (right column) and the current-down joint (left column). In these micrographs, electrons entered the current-up joint from its lower-right corner, exited this joint through its upper-left corner, passed through the Cu line on the chip, entered the current-down joint from its upper-right corner, and exited through its lower-left corner. This specimen failed at the current-down joint after 87 min of accumulated stressing time. During these 87 min, the current-stressing was interrupted four times, at 30, 50, 65, and 75 min, in order to have the specimen examined by

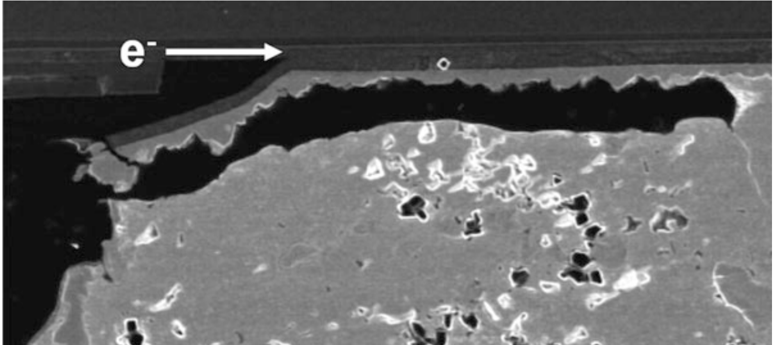


Fig. 11.24 A pancake-type void formation in Sn4Ag0.5Cu solder bump after current stressing at 146 °C for 6 h. The applied current density was 3.67×10^3 A/cm^2 [146]

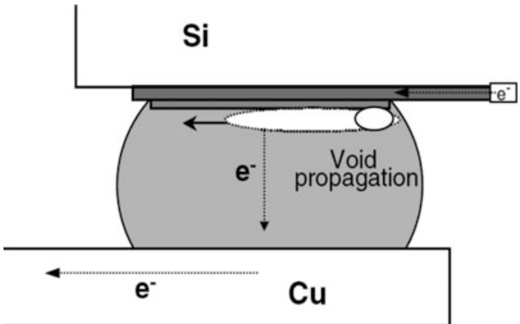


Fig. 11.25 Schematic diagram depicting the cross-section of a solder joint with pancaketype void formation and propagation at the upper interface [10]

SEM. Figure 11.26a, b show the cross-sections of current-down and current-up joints, respectively, before applying current. The bright areas in the joints are Pb-rich regions and dark areas are Sn-rich regions. There were scallop-type Cu_6Sn_5 compounds formed over the disk-shaped Cu UBM. After 30 min of current-stressing, no change in microstructure that could be directly attributed to electromigration was observed, as shown in Fig. 11.26c, d. However, after 50 min, a void was observed near the entrance of electrons into the current-down joint, as shown in Fig. 11.26e. This location of this void was near the current-crowding region. Nothing pronounced occurred at least during the first 30 min, and a rapid change occurred after the observation of the first void at 50 min. Despite the formation of a void in the current-down joint at 50 min, there still was not any clear evidence of change due to electromigration in current-up joint, as shown in Fig. 11.26f. The sequence of events in Fig. 11.26 can be described as below. A void first initiated at the upper-right corner of the current-down joint. This location coincided with the high current crowding region. Once the void formed, the electrons were diverted to the left of the void, causing the void to grow to that

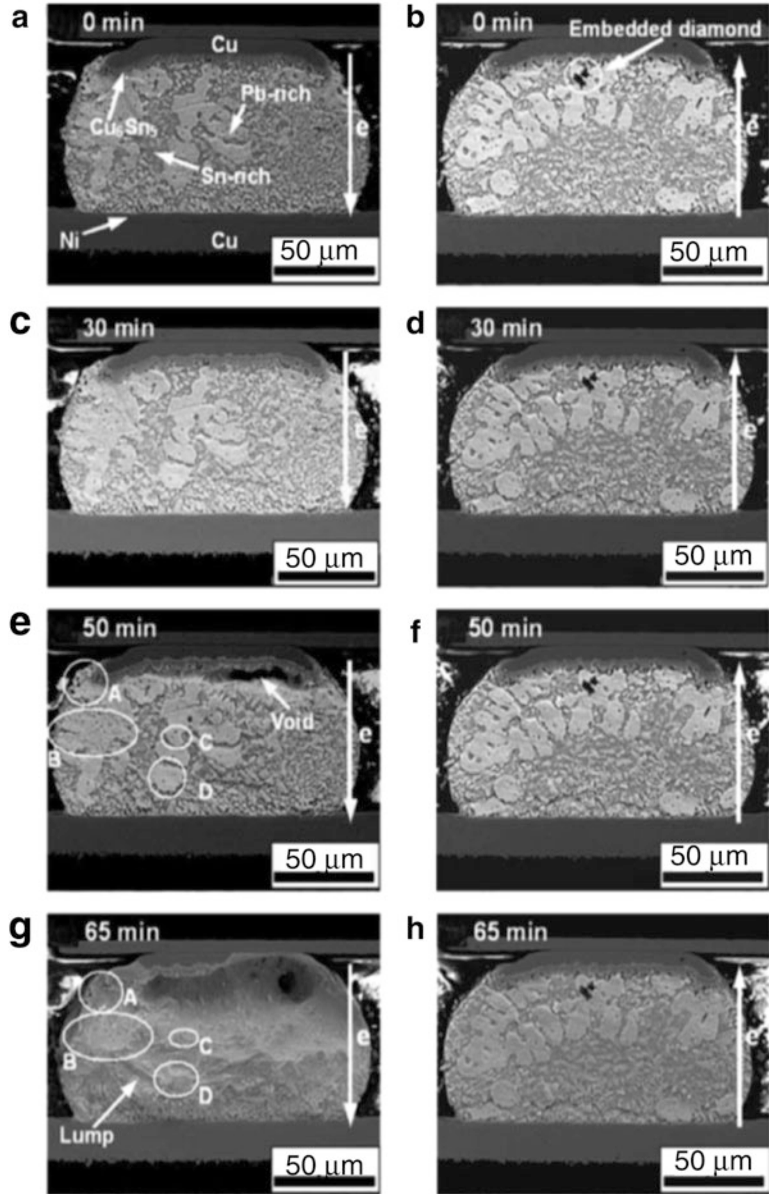


Fig. 11.26 (Continued)

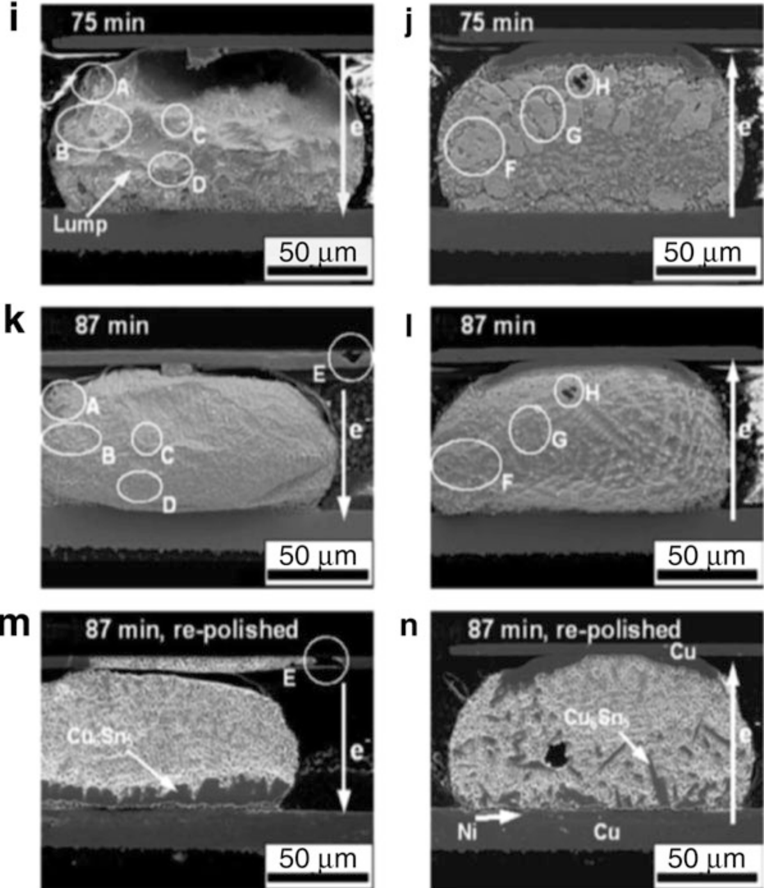


Fig. 11.26 (**a**)–(**n**) SEM images showing the evolution of the microstructure under current-stressing. The *right column* shows the images for the joint with electrons going downward, and the *left column* with electrons going upward. The accumulated stressing time is shown at the *upper-left corner* of each image [12]

direction. Consequently, the void propagated along the UBM/solder interface. The importance of the incubation time before void nucleation cannot be overemphasized. After the void nucleation, the joint failed very quickly. This leads to the suggestion that the incubation time is a good indicator of the overall lifetime of a joint under current-stressing. A theory that can predict the incubation time will then be just as good as a theory than can predict the lifetime of a solder joint for practical applications. Before the void nucleation, the joint underwent very little microstructure change. On the other hand, the joint experienced substantial changes in microstructure before failure. Therefore, it should be much easier to derive a theory for the void nucleation than for the whole failure process.

11.3.2.4 Comparison Between Eutectic Sn-Pb and Sn-Ag-Cu Solders Under Current Stressing

Electromigration in Sn-Ag-Cu is much slower than that in eutectic Sn-Pb, and the MTTF (Mean-Time-To-Failure) of the latter is also shorter than the former [10]. This can be understood by considering the electron wind force and the mechanical force in the following equation

$$J_{\mathrm{EM}} = -\frac{cD}{kT}\frac{\mathrm{d}\sigma\Omega}{\mathrm{d}x} + \frac{cD}{kT}Z^{*}eE \qquad (11.9)$$

where σ is the hydrostatic stress in the metal and E is the electric field. The Sn flux or Pb flux driven by electromigration has the driving force term, $Z^{*}eE$, and the mobility term, D/kT. In driving force, both Z^{*} and resistivity differ for the two solders, yet the difference is small. In mobility, the difference in diffusivity can be very large; the diffusivity in eutectic Sn-Pb may be one order of magnitude faster than that in eutectic Sn-Ag-Cu. This is because the melting temperature of Sn-Ag-Cu (about 217 °C) is higher than that of eutectic Sn-Pb (183 °C). Therefore, at the same stressing temperature, the homologous temperature of the Pb-free is lower than that of eutectic Sn-Pb. Also, the smaller grain size and the eutectic lamellar interfaces in the Sn-Pb solder may enhance the diffusivity. Thus, electromigration in eutectic Sn-Pb will be faster. Furthermore, in (11.9), we note the back stress term. The effect of back stress in resisting electromigration in the Pb-free is larger than that in the Sn-Pb. A distinct difference in electromigration behavior between eutectic Sn-Pb and Sn-Ag-Cu is the squeezing out of IMC at the anode side of the latter [10]. It seems that in eutectic Sn-Pb, the compressive stress at the anode can be relaxed by the bulge of the solder, indicating that lattice sites can be created easily because of more grain and interface boundaries. But in the Sn-Ag-Cu, the Sn matrix is mechanically harder and the surface oxide is protective. The higher compressive stress or back stress was relaxed by squeezing out the hillocks of IMC. If the Pb-free solder bump is confined by underfill which resists surface relief, the buildup of compressive stress at the anode may be even higher.

11.3.3 Effect of Current Stressing on UBM and the Resulting Failure Mechanism

On the condition that $J_{\mathrm{EM}}^{\mathrm{A}}$ in (11.7) is appreciable compared to $J_{\mathrm{EM}}^{\mathrm{Sn}}$ the flip-chip solder joint might fail through the excessive dissolution of the UBM before the nucleation of void within the solder.

11.3.3.1 Dissolution of Cu UBM

Figure 11.27 shows an example of failure caused by excessive Cu UBM dissolution [142]. The solder used in this study was eutectic Sn-Pb, and the nominal diameter of solder bump was 125 μm. In electromigration tests, only a pair of bumps consisting of No. 1 and No. 2 were subjected to current stressing; one with electrons flowing from the substrate side to the die side (No. 1 bump) and the other with the opposite direction (No. 2 bump). As a reference, the third bump (No. 3 bump) has no current passing through it, but it has a very similar thermal history to the neighboring pair. The UBM on the chip side was a layer of 14 μm thick Cu with a dish-shaped rim, and the metallization on the substrate side was a bi-layer of 0.25 μm Au and 3 μm Ni over the Cu conducting trace. The electromigration test was conducted in an oven set at 100 °C. A constant current of 1.27 A was passed through the pair of bumps, producing a nominal current density of 2.5×10^4 A/cm^2 (based on the opening area of the UBM). The chip temperature was monitored by attaching a thermalcouple to the backside of the chip. Due to Joule heating, the backside of the chip reached a steady state temperature of 157 °C after 10 min of testing.

Figure 11.28a shows a cross-sectional SEM image of a failed No. 2 bump. It took 95 min for the bump to fail by showing an abrupt increase in resistance through extensive Cu dissolution in the upper-left corner. Not only the dish-shaped Cu UBM, but also part of the Cu conducting trace has been consumed. The dissolved Cu atoms were driven by electromigration to the anode (substrate) side, and a large amount of Cu_6Sn_5 was formed in the solder bump. For comparison, Fig. 11.28b shows the cross-sectional SEM image of the neighboring No. 3 bump after the test. Without current stressing, no apparent difference in microstructure can be seen in No. 3 bump before and after the test. There was no obvious dissolution of Cu and no formation of Cu_6Sn_5 in the No. 3 solder bump. Figure 11.27 shows the event of microstructure evolution in No. 1 as well as No. 2 bump. The progress of the Cu dissolution in the No. 2 bump as a function of time is clearly visible. The effect of current crowding can be seen very clearly in the No. 2 bump because of the asymmetric dissolution. It started at the upper-left corner near the entrance of electrons into the solder bump and became observable after 15 min of electromigration. It indicated that no incubation was needed for this mechanism. As the time of current stressing increased, more Cu was dissolved into the solder. These dissolved Cu atoms were swept away toward the anode side by the applied current, and formed Cu_6Sn_5 near the anode side. In contrast, electromigration in the No. 1 bump did not produce similar metal dissolution and failure. There are three possible reasons for the difference. The first is that Ni on the substrate side is more resistant to dissolution driven by electromigration than Cu. It is known that the rate of Ni dissolution into molten eutectic Sn-Pb solder is much slower than that of Cu (see Fig. 11.3). The second is that the conducting trace on the substrate side is much thicker than on the chip side. A thicker conducting line can reduce current crowding by spreading out the current distribution. The third is that the chip side had a higher temperature than the substrate side.

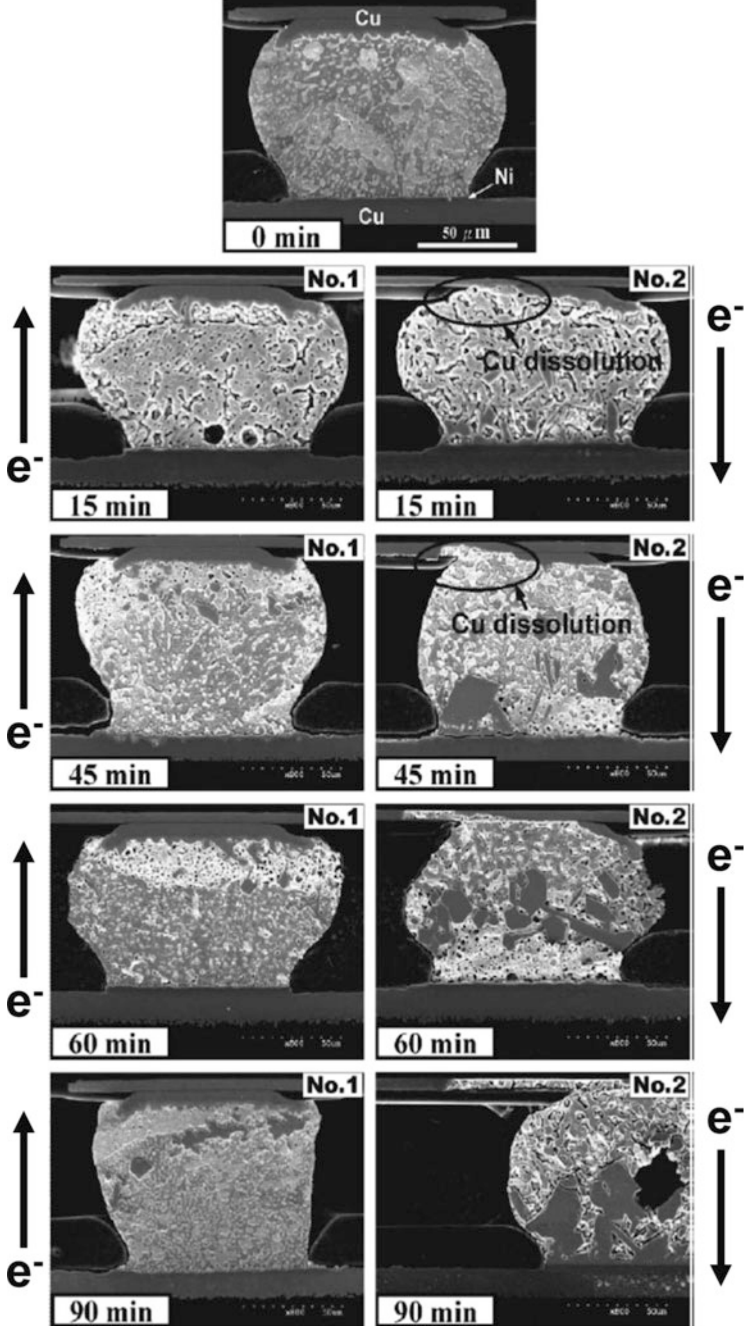


Fig. 11.27 Microstructure evolution in the No. 1 bump (the *left column*) as well as the No. 2 bump (the *right column*) [142]

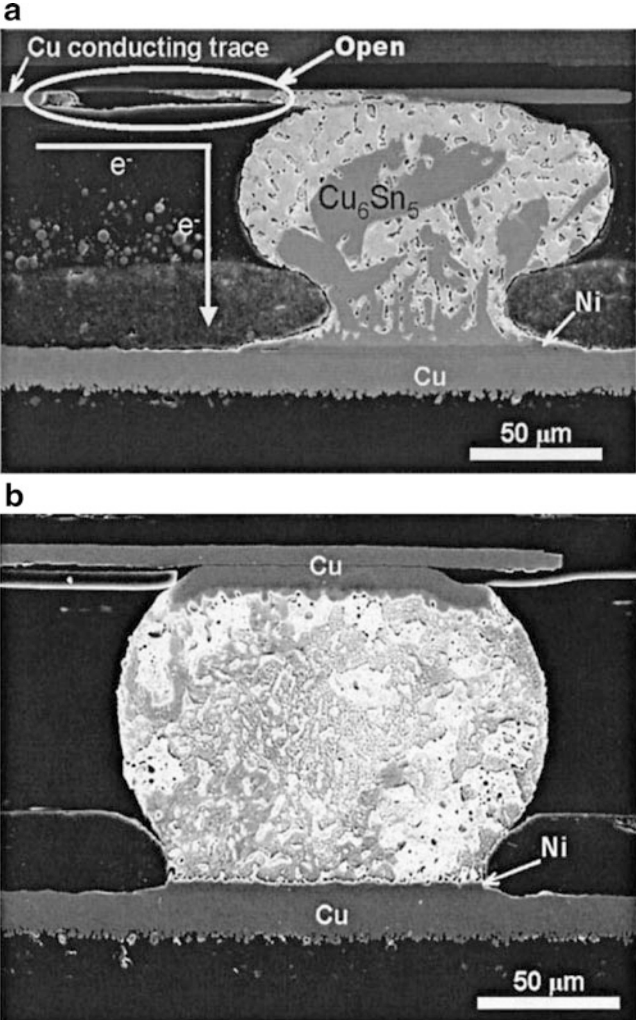


Fig. 11.28 (**a**) A secondary electron micrograph of the No. 2 bump after 95 min current stressing at a nominal current density of 2.5×10^4 A/cm^2. This flip chip failed through an opening at the Cu conducting trace. Extensive Cu dissolution can be seen. (**b**) A secondary electron micrograph of the No. 3 bump from the same flip chip after 95 min current stressing through No. 1 and No. 2 bumps. A small amount of Cu_6Sn_5 formed over the Cu layer [142]

11.3.3.2 Dissolution of Ni UBM

Dissolution can also cause failures in flip-chip solder joints with Ni UBMs. Figure 11.29a–c shows microstructures of the joints that had been stressed at 150 °C for (a) 150, (b) 300, and (c) 400 h [147]. The electrons entered the joints from the upper-right corner. The solder joints had a configuration illustrated

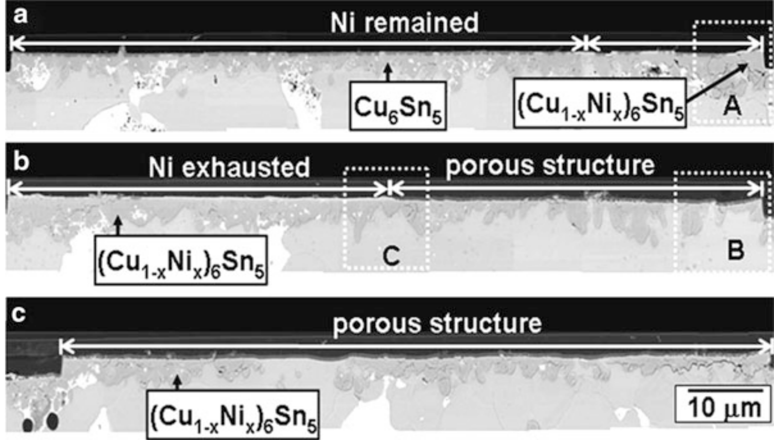


Fig. 11.29 Electron micrographs showing the Ni(V) UBM after current stressing for (**a**) 150 h; (**b**) 300 h; and (**c**) 400 h [147]

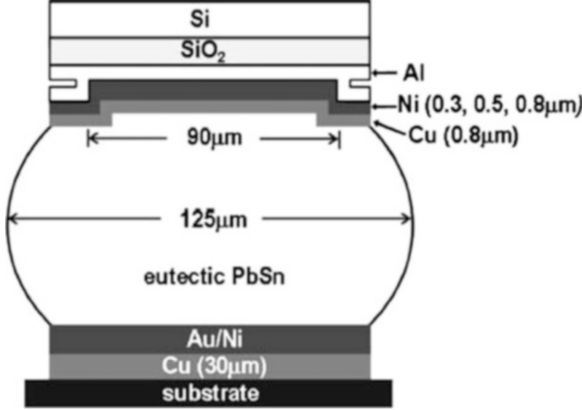


Fig. 11.30 Schematic drawing showing the solder joint configuration used in Fig. 11.29 [147]

schematically in Fig. 11.30. The chips had the Al interconnect lines, which were 60 μm wide and 1 μm thick. The UBM on the chips had the Cu/Ni/Al structure. The Cu (0.8 μm) and the Ni (0.3 μm) layers were sputter-deposited. On the substrate side, the Au/Ni surface finish was used. The solder was 63Sn37Pb eutectic solder, and the solder joints had a nominal diameter of 125 μm. The 0.8 μm Cu layer in the Cu/Ni/Al UBM had been completely consumed and was converted into Cu_6Sn_5 during the assembly of the joint. During the electromigration test, the assembled chips and substrates were kept in an oven set at 150 °C. A constant 0.32 A current was passed through the solder joints, producing a nominal current density of 5×10^3 A/cm^2 (basing on the opening area of the contact window). After 150 h of current stressing, as shown in Fig. 11.29a, Cu_6Sn_5 near the upper-right corner of the joint had become $(Cu_{1-x}Ni_x)_6Sn_5$, but elsewhere on the chip side no Ni can be

detected inside Cu_6Sn_5. After 300 h of current stressing, as shown in Fig. 11.29b, all Cu_6Sn_5 had become $(Cu_{1-x}Ni_x)_6Sn_5$. After 400 h of current stressing, as shown in Fig. 11.29c, no void occurred in the solder joint even though the MTTF (mean-time-to-failure) value (384 h) suggested that it was about to fail. The joint failed at 420 h. Figure 11.29a showed that only the part of the Ni UBM near the entrance of the electrons had been consumed. Meanwhile, the Ni UBM at other location was still intact. No Ni could be detected in those Cu_6Sn_5 beneath the intact Ni UBM. When the stressing time reached 300 h, all the Ni UBM had been completely consumed and Ni could be detected in all $(Cu_{1-x}Ni_x)_6Sn_5$, as shown in Fig. 11.29b. The simultaneous disappearance of the Ni layer and the appearance of Ni in $(Cu_{1-x}Ni_x)_6Sn_5$ suggests that the dissolution of Ni into Cu_6Sn_5 was the main consumption mechanism for the Ni layer. At this stage, the so-called porous structure [148] was limited to the right side of the UBM, but after 400 h of current stressing, as shown in Fig. 11.29c, the porous structure had extended across the entire UBM. Shown in Fig. 11.31a–c are the zoom-in micrographs for the A, B, and C regions in Fig. 11.29, respectively. It is clearly visible that the Ni UBM consumption initiated at the upper-right corner, as shown in Fig. 11.31a. No void formed at the interface. Figure 11.31b shows that the Ni UBM had been completed replaced by the porous structure. The leading edge of the porous structure is shown in Fig. 11.31c. Summarizing the features in Fig. 11.31a–c, one notes that the Ni consumption and the formation of the porous structure was a sequence of events that initiated at the right side and extended to the left under current stressing.

In the pancake-type void mechanism, a void forms in the solder at the location with the highest local current density, and the electrons are diverted around the void, making the solder next to the edge of the void become the highest current density region. This causes the void to grow toward the highest current density region. This process repeats itself, and the void gradually propagates until an open circuit is created. However, in this Ni UBM consumption mechanism, the role of void was replaced by the porous structure, which presumably was non-conductive. Once the porous structure formed, the electrons were diverted to nearby region, and caused the porous structure to extend toward that region. The joints failed when the porous structure extended all the way across the joint. In the pancake-type mechanism, one key factor determining the MTTF is the incubation time for the void nucleation. Once the void nucleates, it propagation can cause the joint failure very quickly. In the Ni UBM consumption mechanism, the key factor determining the MTTF was the time required to consume a certain thickness of the Ni UBM. As can be seen in Fig. 11.32, the thicker the Ni UBM, the longer was the MTTF. Figure 11.32 shows the Weibull plots of the samples with different Ni UBM thicknesses. The MTTF of the solder joints were determined to be 384, 736, and 1,269 h for 0.3, 0.5, and 0.8 μm Ni, respectively. The MTTF values reported here are based on 22 samples for each Ni thickness. The MTTF results clearly show that a thicker Ni UBM had a longer life time.

The pancake-type void and the Ni UBM consumption mechanism are competing mechanisms. If a void nucleates before the Ni UBM is consumed, then the void

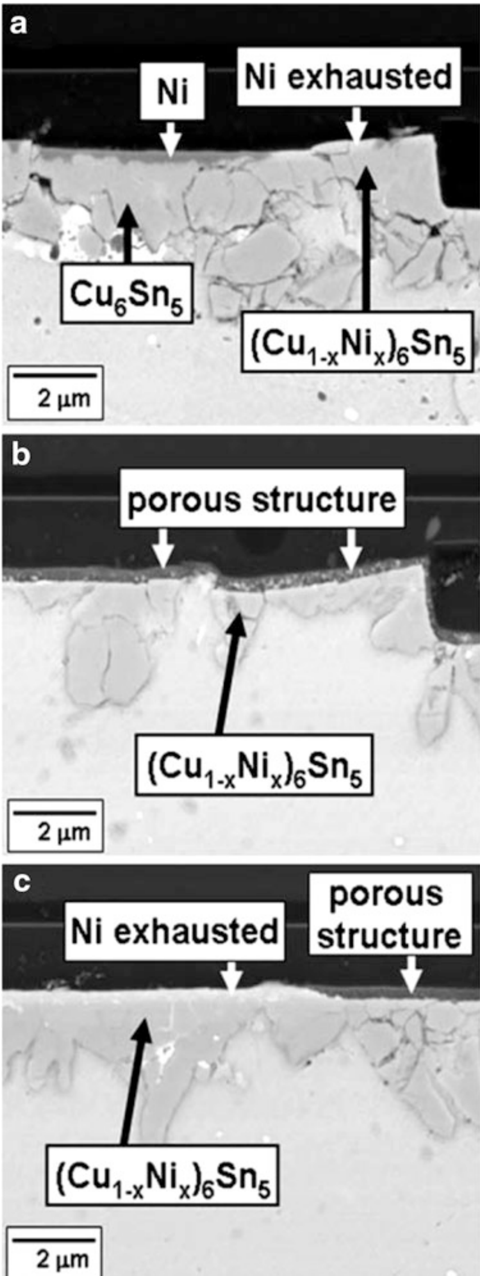


Fig. 11.31 (**a**) Higher-magnification micrograph of region A in Fig. 11.29a. (**b**) Higher-magnification micrograph of region B in Fig. 11.29b. (**c**) Higher-magnification micrograph of region C in Fig. 11.29b [147]

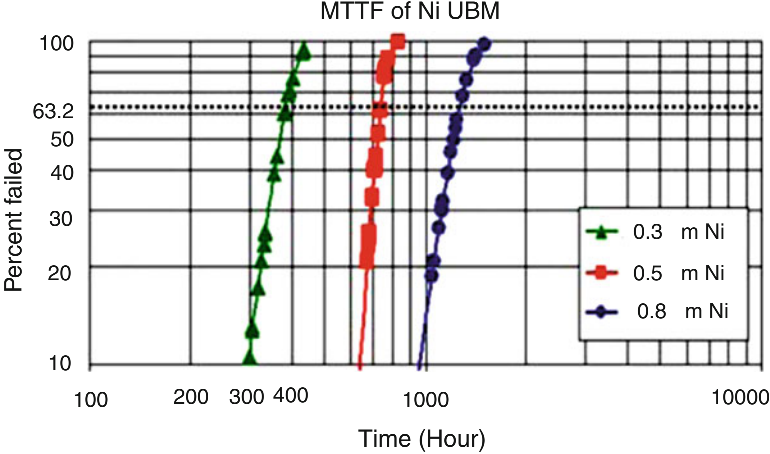


Fig. 11.32 Weibull plots of the cumulative failure of samples with different Ni thickness [147]

formation-and-propagation will be the dominant mechanism. Conversely, if the Ni UBM is consumed before the void nucleates, then the Ni UBM consumption will be the dominant mechanism. If the UBM design and the materials used are kept the same, the stressing conditions, which include the applied current density, the environmental temperature, the heat dissipation capacity of the package, and the chip/package circuit design, which determines the amount of Joule heating, will decide the dominant mechanism.

11.3.3.3 Temperature Effect on the UBM Dissolution

Equation (11.7) shows that J_{EM}^{A} becomes relatively important when x_A increases. The value of x_A depends on how fast the UBM metal can dissolve into the solder. Metal dissolution is a thermally activated process, and becomes faster as temperature increases. Therefore, an understanding of the temperature distribution for flip-chip solder joint is necessary.

The resistance of a cubic piece of solder of 100 × 100 × 100 μm (the size of a solder joint) is about 1 mΩ. The resistivity of Sn and Pb is 11 and 22 μΩ-cm, respectively. The resistance of an Al or Cu line 100 μm long with a cross section of 1 × 0.2 μm is about 10 Ω. The solder joint is a low-resistance conductor, but the interconnect is a high-resistance conductor and becomes the source of joule heating. The simple calculation in the above shows that the resistance of the interconnect will be very sensitive to the design of its dimension and to a slight microstructure change and damage. Yet the resistance of the solder bump is not, e.g., not even by the inclusion of a large void within the bulk of the solder bump. Often, the matrix of a solder bump may contain a few very large spherical voids due to residue flux from

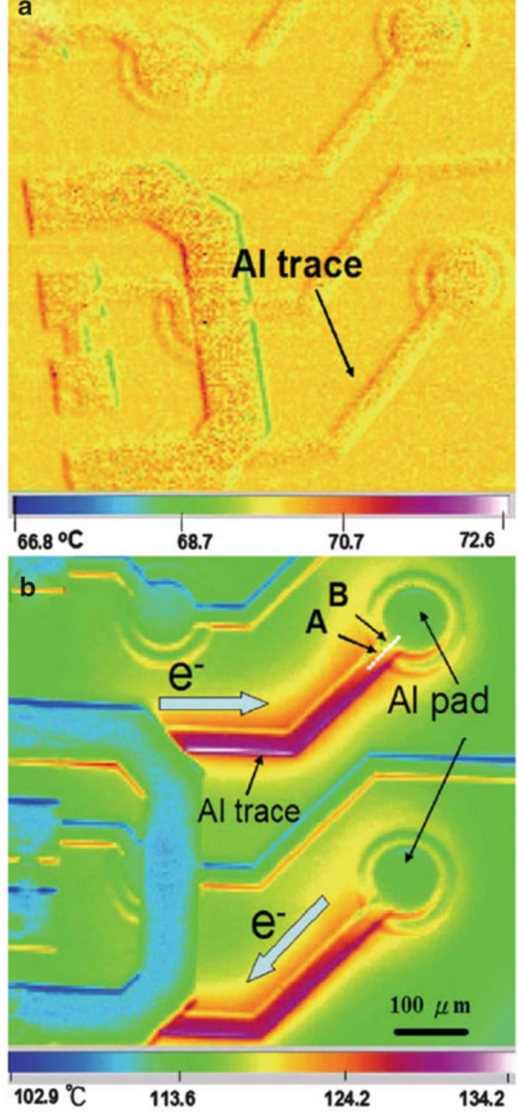


Fig. 11.33 (**a**) Temperature distribution in the package before current stressing, showing a uniform temperature in the package; (**b**) Temperature distribution in the Al trace measured by the IR microscope when powered by 0.59 A [149]

the solder paste, especially the Pb-free solder paste, yet the voids have little effect on the resistance of the solder bump. Figure 11.33 shows the infrared microscope images of solder joints before and during current stressing [149]. The temperature distribution without current stressing is shown in Fig. 11.33a. There was very little temperature variation before the current was applied. Figure 11.33b shows the temperature distribution during current stressing. The Al interconnects exhibited the highest temperature. The two solder bumps were located directly below the

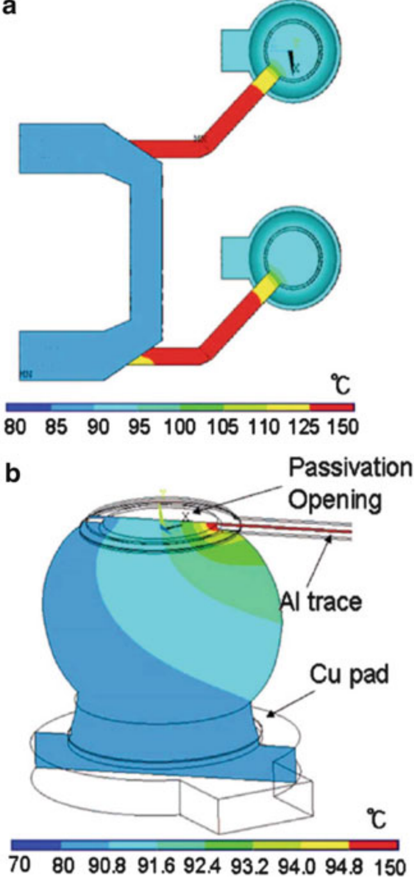


Fig. 11.34 (**a**) Simulated temperature distribution in the stressing circuit when powered by 0.59 A (**b**) Temperature distribution inside the solder for one of the cross sections near the Al trace. A hot spot was found in the entrance point of the Al traces [149]

two circular Al pads/UBMs, as labeled in the figure. The Al trace had much higher temperature than the circular Al pads, which were directly connected to the UBM and the solder bumps. The maximum temperature was as high as 134 °C, which occurred approximately at the middle of the Al trace, whereas the temperature was only about 105 °C for the Al pads above the solder bumps. Figure 11.34 shows the simulated temperature distribution in the Al trace and in the solder joints when stressed by 0.59 A [149]. The temperature distribution inside the solder in one of the cross sections near the entrance of the Al trace is shown in Fig. 11.34b. The highest temperature inside the solder was 95.6 °C. In summary, the Al interconnects near the electron entrance into the flip-chip solder joint have the highest temperature. For the solder joint itself, the current crowding region has the highest temperature.

11.3.4 Mean-Time-to Failure of Flip-Chip Solder Joints

The electronic industry uses the mean-time-to-failure (MTTF) analysis to predict the lifetime of a device. In 1969, Black provided the following equation to analyze failure in Al interconnects caused by electromigration [150]:

$$\mathrm{MTTF} = A \frac{1}{J_{\mathrm{e}}^{n}} \exp\left(\frac{Q}{kT}\right) \tag{11.10}$$

The derivation of the equation was based on an estimate of the rate of mass transport resulting in the formation of a void across an Al interconnect. The most interesting feature of the equation is the dependence of MTTF on the square power of current density, i.e., $n = 2$.

In the subsequent studies of the MTTF equation, whether the exponent n is 1, 2, or a larger number has been controversial, especially when the effect of joule heating is taken into account. However, assuming that mass flux divergence is required for failure and the nucleation and growth of a void requires vacancy supersaturation, Shatzkes and Lloyd have proposed a model by solving the time-dependence diffusion equation and obtained a solution for MTTF in which the square power dependence on current density was also obtained [151]. Nevertheless, whether Black's equation can be applied to MTTF in flip-chip solder joints deserves a careful examination.

To determine the activation energy, accelerated tests at high temperatures are performed. Attention must be paid to the temperature range in which lattice diffusion might overlap grain boundary diffusion and also grain boundary diffusion might overlap surface diffusion. For eutectic Sn-Pb solder, it is more complicated because of the change of dominant diffusion species between Pb and Sn above and below 100 °C.

The formation of a void requires nucleation and growth. In the case of a flip-chip solder joint, the bulk part of the time to failure is controlled not by the growth of a void across the contact interface, but by the incubation time of void nucleation. The latter takes about 90 % of the time of failure. The propagation of the void across the entire contact takes only about 10 % of the time. Furthermore, as shown in the earlier sections in this chapter, the effect of current crowding on failure is crucial and cannot be ignored in the analysis of MTTF. Black did point out the importance of current gradient or temperature gradient on interconnect failure, although he did not take them into account in his equation explicitly [150]. On the basis of the unique failure mode of a flip-chip solder joint by the pancake-type void mechanism, the major effects of current crowding are to increase greatly the current density at the entrance of the solder joint and also to increase the local temperature due to joule heating. Furthermore, solder joint has IMC formation at both the cathode and the anode interfaces, electromigration affects IMC formation, and in turn IMC formation affects failure time and mode. This was not considered in Black's original model of MTTF. Therefore, we cannot apply Black's equation to predict flip-chip solder joint lifetime without modification.

Table 11.4 Mean-time-to-failure of eutectic Sn-Pb flip-chip solder joints [10]

	1.5 A (1.9×10^4 A/cm^2)		1.8 A (2.25×10^4 A/cm^2)		2.2 A (2.75×10^4 A/cm^2)	
	Calculated (h)	Measured (h)	Calculated (h)	Measured (h)	Calculated (h)	Measured (h)
100 °C	...	...	380	97	265	63
125 °C	108	573[a]	79.6	43	55.5	3
140 °C	46	121	34	32	24	1

[a]Not failed

Brandenburg and Yeh used Black's equation with $n = 1.8$ and $Q = 0.8$ eV/atom, without taking into account the effect of current crowding. The equation, with $n = 1.8$ and $Q = 0.8$ eV/atom, has been found to have greatly overestimated MTTF of flip-chip solder joints at high current densities. Table 11.4 compares the calculated and measured MTTF of eutectic Sn-Pb flip-chip solder joints at three current densities and three temperatures. At the low current density of 1.9×10^4 A/cm^2, the measured MTTF is slightly longer than the calculated, but at 2.25×10^4 and 2.75×10^4 A/cm^2, the measured MTTF is much shorter than the calculated. This is also true for the eutectic Sn-Ag-Cu flip chip solder joints. These findings show that MTTF of flip-chip solder joints is very sensitive to a small increase of current density; the MTTF drops rapidly when the current density is about 3×10^4 A/cm^2. Also, the Pb-free solder has a much longer MTTF than the Sn-Pb solder. For example, at 2.25×10^4 A/cm^2 at 125 °C, the MTTF is 580 h for the Pb-free versus 43 h for the Sn-Pb.

Black's equation can be modified to include the effect of current crowding and joule heating [152]:

$$MTTF = A\frac{1}{(cJ_e)^n}\exp\left(\frac{Q}{k(T + \Delta T)}\right) \tag{11.11}$$

where c is due to current crowding and has a magnitude of 10 and ΔT is due to joule heating and may be higher than 100 °C. Both parameters c and ΔT will reduce the MTTF from Black's equation, i.e., make the solder joint fail much faster. Since ΔT depends strongly on j, the modified equation is much more sensitive to the change of current density than Black's equation. We recall that the value of ΔT will depend on the design of flip-chip solder joint and interconnect, because of heat generation and heat dissipation.

11.3.5 Mitigation Strategy Against Electromigration

To improve reliability against electromigration, the current crowding must be reduced. This can be achieved by improving the design of the flip-chip configuration and materials. Since the basic principle in current distribution is that electric current will take the least resistive path, there are options in the design of a flip-chip

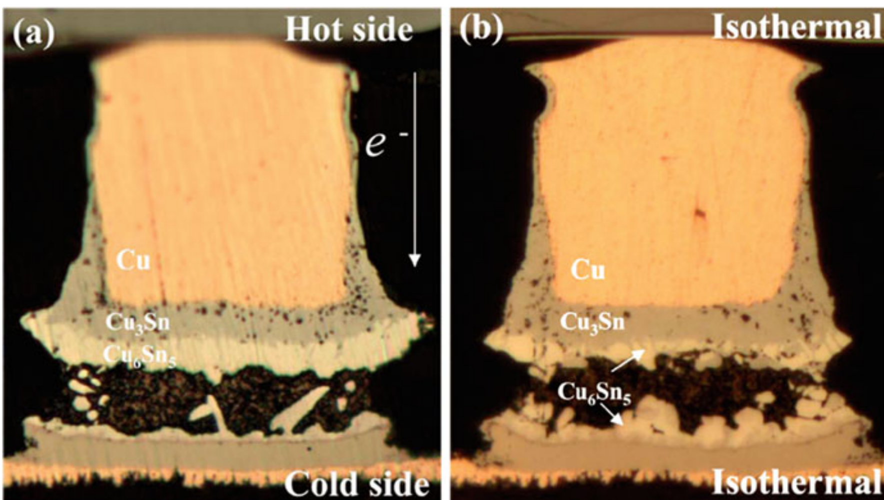


Fig. 11.35 Optical microscope images of the cross-sectional flip-chip joints (**a**) After current stressing for 720 h at 100 °C with current density of 10^4 A/cm^2. (**b**) After aging for 720 h at 150 °C with no current stressing [4]

to reduce current crowding. Using finite element analysis, current distribution in a flip-chip solder joint can be studied as a function of geometry and resistance of all the conducting elements associated with a solder joint, including the Al or Cu interconnect, the UBM, and the solder bump itself. The factors that affect current distribution the most have been found to be the thickness and resistance of under-bump metallization. This observation leads to the design of Cu column bump (or Cu pillar bump). Figure 11.35 shows the cross-sectional view of the flip-chip joints of Cu column bumps with eutectic Sn-Pb solder bumps after current stressing for 1 month at 100 °C with current densities of (a) 10^4 A/cm^2, (b) 0 (as a reference) [4]. This results show that Cu column bump has a strong resistance against electromigration induced failure because of the uniform current distribution in the solder region. The reduction of current crowding in the solder region by using thick Cu column bumps increased the reliability against electromigration induced failure. However, microvoid formation was found to be much serious and enhanced by electromigration at the Cu/Cu_3Sn interface due to the large Cu/Sn ratio. Since this is a system of a limited amount of Sn and an infinite supply of Cu, the Cu_6Sn_5 transforms to the Cu_3Sn after all the Sn content in the solder bump is consumed and the Cu_3Sn can grow very thick; the vacancy flux that opposes the Cu flux will condense to form microvoids.

In addition to the microvoid formation, it was found that electromigration in Cu column accelerated the consumption rate of copper column and converted almost the entire solder joint into intermetallic compound, as shown in Fig. 11.36 [153]. Mechanically, drop impact test indicates a brittle fracture failure in the intermetallic.

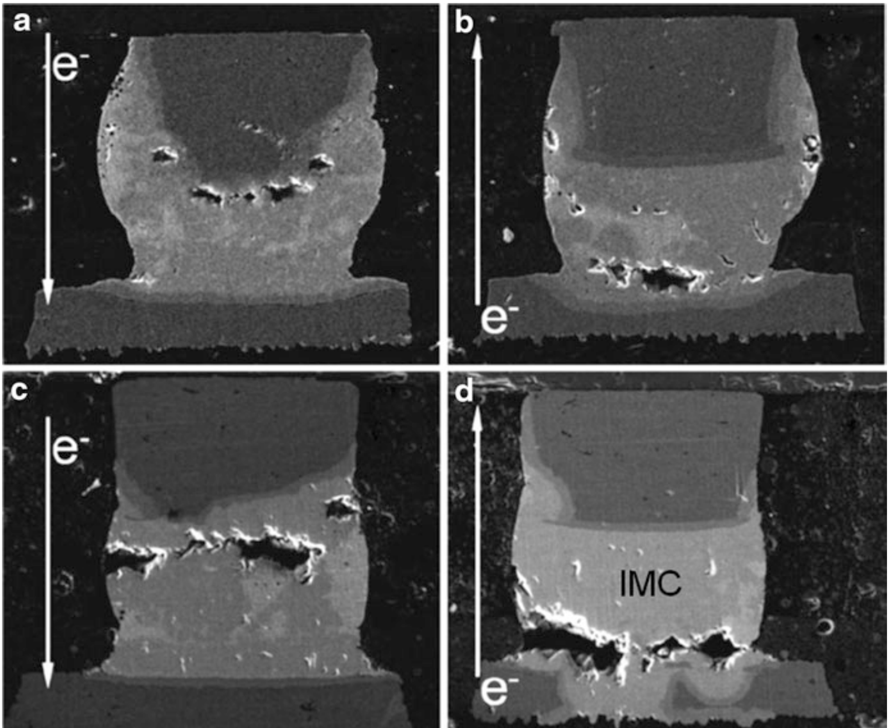


Fig. 11.36 Formation of voids in Cu pillar solder joint after electromigration at current density of 2×10^4 A/cm^2 [153]

To overcome the two issues associated with the Cu column bumps, a layer of Ni was coated over the Cu column, as illustrated in Fig. 11.37 [154]. Electromigration reliability of such solder joints was reported to be much enhanced compared to a bare Cu column joints [154]. The Ni layer was much more resistant to dissolution and offered protection against microvoid formation and excessive intermetallic formation.

11.4 Emerging Issues

The incessant trend towards smaller and lighter consumer electronics necessitates the continuous shrinking of flip-chip solder joints. At present, 50 μm diameter flip-chip solder joints are under active development. Micro-bumps of even smaller sizes are also not far from the horizon. Accordingly, in future flip-chip solder joints, the amount of solder per joint will decrease. As solder volume decreases, several chemical and physical processes become ever more threatening toward the

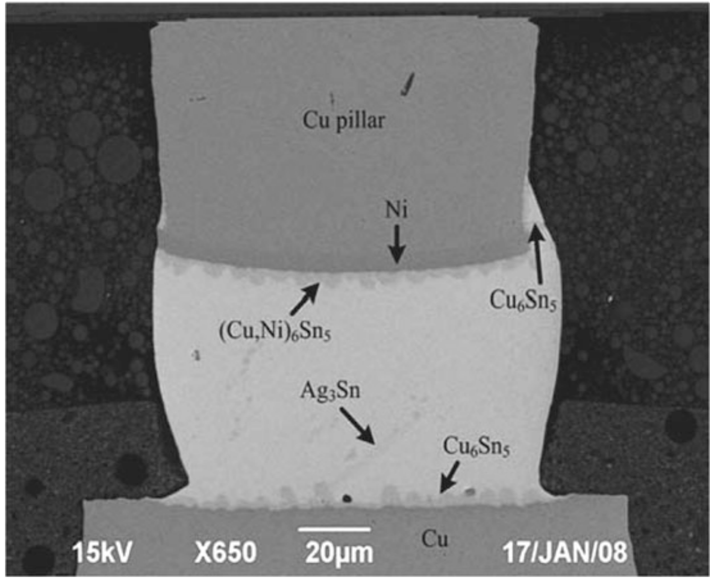


Fig. 11.37 Cross-sectional view of an as-assembled Cu pillar solder joint [154]

reliability of flip-chip joints. These processes include chemical reactions, metal dissolution, diffusion driven by chemical potential gradient, electromigration, Joule heating, thermomigration, and stress migration. There are two threatening issues arise from the combined effect of these process: (1) excessive intermetallic formation and (2) excessive UBM consumption. Bulk of the solder may be consumed and converted into brittle intermetallics, and high volume percentage of a solder joint may be occupied by brittle intermetallics. At this stage, little is known about the mechanical properties of such joints and their effects on the long-term reliability. There is an urgent need to initiate such study.

References

1. Hashino E, Shimokawa K, Yamamoto Y, Tatsumi K (2001) Micro-Ball wafer bumping for flip chip interconnection. In: Proceeding of the IEEE of Electronic Components and Technology Conference, May 2001, pp 957–964
2. Ruhmer K, Laine E, Gruber P (2006) C4NP—IBM manufacturing & reliability data for lead free flip chip solder bumping. International Microsystems, Packaging, Assembly Conference Taiwan (IMPACT), Taipei, 19 Oct 2006
3. Feger C, LaBianca N, Gaynes M, Steen S, Liu Z, Peddi R, Francis M (2009) The over-bump applied resin wafer-level underfill process: process, material and reliability. IBM research report, 31 Aug 2009
4. Nah JW, Suh JO, Tu KN, Yoon SW, Rao VS, Kripesh V, Hua F (2006) Electromigration in flip chip solder joints having a thick Cu column bump and a shallow solder interconnect. J Appl Phys 100:123513

5. Nah JW, Suh JO, Tu KN, Rao VS, Yoon SW, Kripesh V (2006) Electromigration in Cu column flip chip joints. TMS Annual Meeting Report, 2006
6. Kloeser J, WeiBbach EA (2006) High-performance flip chip packages with copper pillar bumping. Global SMT & Packaging, May 2006, p 28
7. Tu KN (2003) Recent advances on electromigration in very-large-scale integration of interconnects. J Appl Phys 94:5451
8. Kim BJ, Lim GT, Kim J, Lee K, Park YB, Joo YC (2008) Intermetallic compound and Kirkendall void growth in Cu pillar bump during annealing and current stressing. In: Proceeding of the 58th Electronic Components and Technology Conference (ECTC), 2008, pp 336–340
9. Tu KN, Zeng K (2001) Tin-lead (SnPb) solder reaction in flip chip technology. Mater Sci Eng R34:1
10. Tu KN (2007) Solder joint technology: materials properties, and reliability. Springer, New York
11. Chan YC, Yang D (2010) Failure mechanisms of solder interconnects under current stressing in advanced electronic packages. Prog Mater Sci 55:428
12. Lin YH, Hu YC, Tsai CM, Kao CR, Tu KN (2005) In-situ observation of the void formation-and-propagation mechanism in solder joints under current-stressing. Acta Mater 53:2029
13. Rahn A (ed) (1993) The basics of soldering. Wiely, New York
14. Glazer J (1995) Metallurgy of low temperature Pb-free solders for electronic assembly. Int Mater Rev 40:65
15. Abtew M, Selvaduray G (2000) Leadfree solders in microelectronics. Mater Sci Eng R27:95
16. Suganuma K (2001) Advances in lead-free electronics soldering. Curr Opin Solid State Mater Sci 5:55
17. Zeng K, Tu KN (2002) Six cases of reliability study of Pb-free solder joints in electronic packaging technology. Mater Sci Eng R38:55
18. Tu KN, Gusak AM, Li M (2003) Physics and materials challenges for lead-free solders. J Appl Phys 93:1335
19. Wu CML, Yu DQ, Law CMT, Wang L (2004) Properties of lead-free solder alloys with rare earth element additions. Mater Sci Eng R44:1
20. Kang SK, Lauro PA, Shin DY, Henderson DW, Puttlitz KJ (2005) Microstructure and mechanical properties of lead-free solders and solder joints used in microelectronic applications. IBM J Res Dev 49:607
21. Laurila T, Vuorinen V, Kivilahti JK (2005) Interfacial reactions between lead-free solders and common base materials. Mater Sci Eng R49:1
22. Ho CE, Yang SC, Kao CR (2007) Interfacial reaction issues for leadfree electronic solders. J Mater Sci Mater Electron 18:155
23. Laurila T, Vuorinen V, Paulasto-Kröckel M (2010) Impurity and alloying effects on interfacial reaction layers in Pb-free soldering. Mater Sci Eng R68:1
24. Huanga ML, Loeher T, Ostmann A, Reichl H (2005) Role of Cu in dissolution kinetics of Cu metallization in molten Sn-based solders. Appl Phys Lett 86:181908
25. Dybkov VI (2009) The growth kinetics of intermetallic layers at the interface of a solid metal and a liquid solder. JOM 61:76
26. Bader WG (1969) Dissolution of Au, Ag, Pd, Pt, Cu and Ni in a molten tin-lead solder. Weld J 48:551
27. Chang CC, Lin YW, Wang YW, Kao CR (2010) The effects of solder volume and Cu concentration on the consumption rate of Cu pad during reflow soldering. J Alloys Compd 492:99
28. Lau JH (ed) (1996) Flip chip technology. McGraw Hill, New York
29. Yang SC, Ho CE, Chang CW, Kao CR (2006) Strong Zn concentration effect on the soldering reactions between Sn-based solders and Cu. J Mater Res 21:2436
30. Baker H (ed) (1992) ASM handbook: volume 3 alloy phase diagram. ASM International, Materials Park, OH

31. Ahat S, Sheng M, Luo L (2001) Microstructure and shear strength evolution of SnAg/Cu surface mount solder joint during aging. J Electron Mater 30:1317
32. Chiu TC, Zeng K, Stierman R, Edwards D, Ano K (2004) Effect of thermal aging on board level drop reliability for Pb-free BGA packages. In: Proceedings of 2004 I.E. Electronic Components and Technology Conference (ECTC), p 1256
33. Vianco PT, Rejent JA, Hlava PF (2004) Solid-state intermetallic compound layer growth between copper and 95.5Sn-3.9Ag-0.6Cu solder. J Electron Mater 33:991
34. Mei Z, Ahmad M, Hu M, Ramakrishna G (2005) Kirkendall voids at Cu/solder interface and their effects on solder joint reliability. In: Proceedings of 2005 I.E. Electronic Components and Technology Conference (ECTC), p 415
35. Zeng K, Stierman R, Chiu TC, Edwards D, Ano K, Tu KN (2005) Kirkendall void formation in eutectic SnPb solder joints on bare Cu and its effect on joint reliability. J Appl Phys 97:024508
36. Oh M (1994) Doctoral Dissertation, Lehigh University
37. Li M, Lee KY, Olsen DR, Chen WT, Tan BTC, Mhaisalkar S (2002) Microstructure, joint strength and failure mechanisms of SnPb and Pb-free solders in BGA packages. IEEE Trans Electron Packag Manuf 25:185
38. Bader S, Gust W, Hieber H (1995) Rapid formation of intermetallic compounds by inter diffusion in the Cu-Sn and Ni-Sn systems. Acta Metallurgica Mater 43:329
39. Gur D, Bamberger M (1998) Reactive isothermal solidification in the Ni-Sn system. Acta Mater 46:4917
40. Moon KW, Boettinger WJ, Kattner UR, Biancaniello FS, Handwerker CA (2000) Experimental and thermodynamic assessment of Sn-Ag-Cu solder alloys. J Electron Mater 29:1122
41. NEMI (National Electronics Manufacturing Initiative) (2001) Workshop on Modeling and Data Needs for Lead-Free solders, New Orleans, LA, 15 Feb 2001
42. Soldertec-ITRI (1999) Lead-free alloys-the way forward. (http://www.lead-free.org). Oct 1999
43. IDEALS (International Dental Ethics and Law Society) (1994) Improved design life and environmentally aware manufacturing of electronics assemblies by lead-free soldering. Brite-Euram contract BRPR-CT96-0140, project number BE95, 1994
44. JEITA (Japan Electronics and Information Technology Industries Association) (2002) Lead-Free Roadmap 2002, vol 2.1
45. Seeling KF, Lockard DG (1994) Lead-free bismuth free tin alloy solder composition. US Patent 5352407, 4 Oct 1994
46. IPC Roadmap (2000) Assembly of lead-free electronics, 4th draft. IPC, Northbrook, IL
47. Ho CE, Lin YL, Kao CR (2002) Strong effect of Cu concentration on the reaction between lead-free microelectronic solders and Ni. Chem Mater 14:949
48. Jang JW, Frear DR, Lee TY, Tu KN (2000) Morphology of interfacial reaction between lead-free solders and electroless Ni-P under bump metallization. J Appl Phys 88:6359
49. Ho CE, Tsai RY, Lin YL, Kao CR (2002) Effect of Cu concentration on the reactions between Sn-Ag-Cu solders and Ni. J Electron Mater 31:584
50. He M, Chen Z, Qi G (2004) Solid state interfacial reaction of Sn-37Pb and Sn-3.5Ag solders with Ni-P under bump metallization. Acta Mater 52:2047
51. Hong SM, Kang CS, Jung JP (2004) Plasma reflow bumping of Sn-3.5 Ag solder for flux-free flip chip package application. IEEE Trans Adv Packag 27:90
52. Jang GY, Duh JG (2005) The effect of intermetallic compound morphology on Cu diffusion in Sn-Ag and Sn-Pb solder bump on the Ni/Cu under-bump metallization. J Electron Mater 34:68
53. Liu CM, Ho CE, Chen WT, Kao CR (2001) Reflow soldering and isothermal solid-state aging of Sn-Ag eutectic solder on Au/Ni surface finish. J Electron Mater 30:1152
54. Shiau LC, Ho CE, Kao CR (2002) Reaction between Sn-Ag-Cu lead-free solders and the Au/Ni surface finish in advanced electronic packages. Soldering Surf Mount Technol 14:25

55. Kang SK, Choi WK, Yim MJ, Shih DY (2002) Studies of the mechanical and electrical properties of lead-free solder joints. J Electron Mater 31:1292
56. Kang SK, Shih DY, Fogel K, Lauro P, Yim MJ, Advocate GG Jr, Griffin M, Goldsmith C, Henderson DW, Gosselin TA, King DE, Konrad JJ, Sarkhel A, Puttlitz KJ (2002) Interfacial reaction studies on lead (Pb)-free solder alloys. IEEE Trans Electron Packag Manuf 25:155
57. Alam MO, Chan YC, Tu KN (2003) Effect of 0.5 wt % Cu in Sn-3.5%Ag solder on the interfacial reaction with Au/Ni metallization. Chem Mater 15:4340
58. Hiramori T, Ito M, Yoshikawa M, Hirose A, Kobayashi KF (2003) Sn-Ag based solders bonded to Ni-P/Au plating: effects of interfacial structure on joint strength. Mater Trans 44:2375
59. Lee CB, Yoon JW, Suh SJ, Jung SB, Yang CW, Shur CC, Shin YE (2003) Intermetallic compound layer formation between Sn-3.5 mass%Ag BGA solder ball and (Cu, immersion Au/electroless Ni-P/Cu) substrate. J Mater Sci Mater Electron 14:487
60. Lee KY, Li M (2003) Interfacial microstructure evolution in Pb-free solder systems. J Electron Mater 32:906
61. Torazawa N, Arai S, Takase Y, Sasaki K, Saka H (2003) Tranmission electron microscopy of interfaces in joints between Pb-free solders and electroless Ni-P. Mater Trans 44:1438
62. Hwang CW, Suganuma K, Kiso M, Hashimoto S (2004) Influence of Cu addition to interface microstructure between Sn-Ag solder and Au/Ni-6P plating. J Electron Mater 33:1200
63. Sharif A, Islam MN, Chan YC (2004) Interfacial reactions of BGA Sn-3.5%Ag-0.5%Cu and Sn-3.5%Ag solders during high-temperature aging with Ni/Au metallization. Mater Sci Eng B113:184
64. Jeon YD, Paik KW, Ostmann A, Reichl H (2005) Effects of Cu contents in Pb-Free solder alloys on interfacial reactions and bump reliability of Pb-free solder bumps on electroless Ni-P under-bump metallurgy. J Electron Mater 34:80
65. Kumar A, He M, Chen Z (2005) Barrier properties of thin Au/Ni-P under bump metallization for Sn-3.5Ag solder. Surf Coating Technol 198:283
66. Yu DQ, Wu CML, He DP, Zhao N, Wang L, Lai JKL (2005) Effect of Cu contents in Sn-Cu solder on the composition and morphology of intermetallic compounds at a solder/Ni interface. J Mater Res 20:2205
67. Chen WT, Ho CE, Kao CR (2002) Effect of Cu concentration on the interfacial reactions between Ni and Sn-Cu solders. J Mater Res 17:263
68. Kao ST, Duh JG (2005) Interfacial reactions and compound formation of Sn-Ag-Cu solders by mechanical alloying on electroless Ni-P/Cu under bump metallization. J Electron Mater 34:1129
69. Luo WC, Ho CE, Tsai JY, Lin YL, Kao CR (2005) Solid-state reactions between Ni and Sn-Ag-Cu solders with different Cu concentrations. Mater Sci Eng A396:385
70. Ho CE, Luo WC, Yang SC, Kao CR (2005) Copper concentration effect and solder volume effect on the soldering reactions between Sn-Ag-Cu lead-free solders and Ni. In: Proceedings of IMAPS Taiwan 2005 International Technical Symposium, Taipei, Jun 2005, p 98
71. Ho CE, Lin YW, Yang SC, Kao CR (2005) Volume effect in the soldering reaction between SnAgCu solders and Ni. In: Proceedings of the 10th International Symposium on Advanced Packaging Materials: Processes, Properties and Interface, IEEE/CPMT, Irvine, Mar 2005, p 39
72. Ho CE, Lin YW, Yang SC, Kao CR, Jiang DS (2006) Effects of limited Cu supply on soldering reactions between SnAgCu and Ni. J Electron Mater 35:1017
73. Yoon JW, Jung SB (2005) Interfacial reactions between Sn-0.4Cu solder and Cu substrate with or without ENIG plating layer during reflow reaction. J Alloys Compd 396:122
74. Jeon YD, Nieland S, Ostmann A, Reichl H, Paik KW (2003) A study on interfacial reactions between electroless Ni-P under bump metallization and 95.5Sn-4.0Ag-0.5Cu alloy. J Electron Mater 32:548

75. Alam MO, Chan YC, Tu KN, Kivilahti JK (2005) Effect of 0.5 wt % Cu in Sn-3.5%Ag solder balls on the solid state interfacial reaction with Au/Ni/Cu bond pads for ball grid array (BGA) applications. Chem Mater 17:2223
76. Kim DG, Kim JW, Jung SB (2005) Effect of aging conditions on interfacial reaction and mechanical joint strength between Sn-3.0Ag-0.5Cu solder and Ni-P UBM. Mater Sci Eng B121:204
77. Kim KS, Huh SH, Suganuma K (2003) Effects of intermetallic compounds on properties of Sn-Ag-Cu lead-free soldered joints. J Alloys Compd 352:226
78. Ha JS, Oh TS, Tu KN (2003) Effect of super-saturation of Cu on reaction and intermetallic compound formation between Sn-Cu solder and thin film metallization. J Mater Res 18:2109
79. Yoon JW, Kim SW, Jung SB (2004) Effect of reflow time on interfacial reaction and shear strength of Sn-0.7Cu solder/Cu and electroless Ni-P BGA joints. J Alloys Compd 385:192
80. Wang CH, Chen SW (2006) Sn-0.7 wt.%Cu/Ni interfacial reactions at 250 °C. Acta Mater 54:247
81. Kang SK, Choi WK, Shih DY, Lauro P, Henderson DW, Gosselin T, Leonard DN (2002) Interfacial reactions, microstructure and mechanical properties of Pb-free solder joints in PBGA laminates. In: Proceedings of 2002 I.E. Electronic Components and Technology Conference (ECTC), p 146
82. Zeng K, Vuorinen V, Kivilahti JK (2002) Interfacial reactions between lead-free SnAgCu solder and Ni(P) surface finish on printed circuit boards. IEEE Trans Electron Packag Manuf 25:162
83. Zheng Y, Hillman C, McCluskey P (2002) Intermetallic growth on PWBs soldered with Sn3.8Ag0.7Cu. In: Proceedings of 2002 I.E. Electronic Compounds and Technology Conference (ECTC), p 1226
84. Cheng MD, Chang SY, Yen SF, Chuang TH (2004) Intermetallic compounds formed during the reflow and aging of Sn-3.8Ag-0.7Cu and Sn-20In-2Ag-0.5Cu solder ball grid array packages. J Electron Mater 33:171
85. Pang JHL, Low TH, Xiong BS, Luhua X, Neo CC (2004) Thermal cycling aging effects on Sn-Ag-Cu solder joint microstructure, IMC and strength. Thin Solid Films 462:370
86. Yoon JW, Kim SW, Koo JM, Kim DG, Jung SB (2004) Reliability investigation and interfacial reaction of ball-grid-array packages using the lead-free Sn-Cu solder. J Electron Mater 33:1190
87. Mattila TT, Kivilahti JK (2005) Failure mechanisms of lead-free chip scale package interconnections under fast mechanical loading. J Electron Mater 34:969
88. Yoon JW, Kim SW, Jung SB (2005) Interfacial reaction and mechanical properties of eutectic Sn-0.7Cu/Ni BGA solder joints during isothermal long-term aging. J Alloys Compd 391:82
89. Lee CB, Jung SB, Shin YE, Shur CC (2002) Effect of isothermal aging on ball shear strength in BGA joints with Sn-3.5Ag-0.75Cu solder. Mater Trans 43:1858
90. Zribi A, Clark A, Zavalij L, Borgesen P, Cotts EJ (2001) The growth of intermetallic compounds at Sn-Ag-Cu solder/Cu and Sn-Ag-Cu solder/Ni interfaces and the associated evolution of the solder microstructure. J Electron Mater 30:1157
91. Ho CE (2002) Doctor Dissertation, National Central University, Taiwan
92. Ho CE, Shiau LC, Kao CR (2002) Inhibiting the formation of $(Au_{1-x}Ni_x)$ Sn_4 and reducing the consumption of Ni metallization in solder joints. J Electron Mater 31:1264
93. Lehman LP, Athavale SN, Fullem TZ, Giamis AC, Kinyanjui RK, Lowenstein M, Mather K, Patel R, Rae D, Wang J, Xing Y, Zavalij L, Borgesen P, Cotts EJ (2004) Growth of Sn and intermetallic compounds in Sn-Ag-Cu solder. J Electron Mater 33:1429
94. Lin CH, Chen SW, Wang CH (2002) Phase equilibria and solidification properties of Sn-Cu-Ni alloys. J Electron Mater 31:907
95. Li CY, Duh JG (2005) Phase equilibria in the Sn-rich corner of the Sn-Cu-Ni ternary alloy system at 240 °C. J Mater Res 20:3118
96. Oberndorff P (2001) Doctoral Dissertation, Technical University of Eindhoven

97. Chen SW, Wu SH, Lee SW (2003) Interfacial reactions in the Sn-(Cu)/Ni, Sn-(Ni)/Cu, and Sn/(Cu, Ni) systems. J Electron Mater 32:1188
98. Wang SJ, Liu CY (2003) Study of interaction between Cu-Sn and Ni-Sn interfacial reactions by Ni-Sn3.5Ag-Cu sandwich structure. J Electron Mater 32:1303
99. Shao TL, Chen TS, Huang YM, Chen C (2004) Cross interactions on interfacial compound formation of solder bumps and metallization layers during reflow. J Mater Res 19:3654
100. Wang SJ, Liu CY (2006) Kinetic analysis of the interfacial reactions in Ni/Sn/Cu sandwich structures. J Electron Mater 35:1955
101. Xia Y, Lu C, Chang J, Xie X (2006) Interaction of intermetallic compound formation in Cu/SnAgCu/NiAu sandwich solder joints. J Electron Mater 35:897
102. Chang CW, Yang SC, Tu CT, Kao CR (2007) Cross-interaction between Ni and Cu across Sn layers with different thickness. J Electron Mater 36:1455
103. Chen HT, Wang CQ, Yan C, Li MY, Huang Y (2007) Cross-interaction of interfacial reactions in Ni (Au/Ni/Cu)-SnAg-Cu solder joints during reflow soldering and thermal aging. J Electron Mater 36:26
104. Kim JY, Sohn YC, Yu J (2007) Effect of Cu content on the mechanical reliability of Ni/Sn-3.5Ag system. J Mater Res 22:770
105. Hong KK, Ryu JB, Park CY, Huh JY (2008) Effect of cross-interaction between Ni and Cu on growth Kinetics of intermetallic compounds in Ni/Sn/Cu diffusion couples during aging. J Electron Mater 37:61
106. Wu WH, Chung HL, Chen CN, Ho CE (2009) The influence of current direction on the Cu-Ni cross-interaction in Cu/Sn/Ni diffusion couple. J Electron Mater 38:2563
107. Tsai CM, Luo WC, Chang CW, Shieh YC, Kao CR (2004) Cross-interaction of under-bump metallurgy and surface finish in flip-chip solder joints. J Electron Mater 33:1424
108. Kao CR, Ho CE, Shiau LC (2003) Solder point with low speed of consuming nickel. R.O.C. patent 181410, 2003
109. Chung CM, Lin KL (2003) Effect of microelements addition on the interfacial reaction between Sn-Ag-Cu solders and the Cu substrate. J Electron Mater 32:1426
110. Tsai JY, Hu YC, Tsai CM, Kao CR (2003) A study on the reaction between Cu and Sn3.5Ag solder doped with small amounts of Ni. J Electron Mater 32:1203
111. Anderson IE, Harringa JL (2006) Suppression of void coalescence in thermal aging of Tin-Copper-X solder joints. J Electron Mater 35:94
112. Gao F, Takemoto T, Nishikawa H, Komatsu A (2006) Microstructure and mechanical properties evolution of intermetallics between Cu and Sn-3.5Ag solder doped by Ni-Co additives. J Electron Mater 35:905
113. Nishikawa H, Piao JY, Takemoto T (2006) Interfacial reaction between Sn-0.7Cu (-Ni) solder and Cu substrate. J Electron Mater 35:1127
114. Yu H, Vuorinen V, Kivilahti J (2006) Effect of Ni on the formation of Cu_6Sn_5 and Cu_3Sn intermetallics. In: Proceeding of the 2006 Electronic Component and Technology Conference (ECTC), p 1204
115. Gao F, Nishikawa H, Takemoto T (2008) Additive effect of Kirkendall void formation in Sn-3.5Ag solder joints on common substrates. J Electron Mater 37:45
116. Wang YW, Chang CC, Kao CR (2009) Minimum effective Ni addition to SnAgCu solders for retarding Cu_3Sn growth. J Alloys Compd 478:L1
117. Wang YW, Lin YW, Kao CR (2009) Kirkendall voids formation in the reaction between Ni-droped SnAg lead-free solders and different Cu substrates. Microelectron Reliab 49:248
118. Wang YW, Lin YW, Tu CT, Kao CR (2009) Effect of minor Fe, Co, and Ni additions on the reaction between SnAgCu solder and Cu. J Alloys Compd 478:121
119. Anderson IE, Harringa JL (2004) Elevated temperature aging of solder joints based on Sn-Ag-Cu: effects on joint microstructure and shear strength. J Electron Mater 33:1485
120. Kang SK, Leonard D, Shih DY, Gignac L (2005) Interfacial reactions of Sn-Ag-Cu solders modified by minor Zn alloying addition. IBM Research Report, 9 Mar 2005

121. Kang SK, Leonard D, Shih DY, Gignac L, Henderson DW, Cho S, Yu J (2006) Interfacial reactions of Sn-Ag-Cu solders modified by minor Zn alloying addition. J Electron Mater 35:479
122. He M, Acoff VL (2008) Effect of Bi on the interfacial reaction between Sn-3.7Ag-xBi solders and Cu. J Electron Mater 37:288
123. Chung CM, Shih PC, Lin KL (2004) Mechanical strength of Sn-3.5Ag-based solders and related bondings. J Electron Mater 33:1
124. Garner L, Sane S, Suh D, Byrne T, Dani A, Martin T, Mello M, Patel M, Williams R (2005) Finding solutions to the challenges in package interconnect reliability. Intel Technol J 9:297
125. Ohriner EK (1987) Intermetallic formation in soldered copper-based alloys at 150 degrees to 250 degrees celcius. Weld J Res Suppl 7:191
126. Wang YW, Lin YW, Kao CR (2010) Inhibiting the formation of microvoids in Cu_3Sn by additions of Cu to solders. J Alloys Compd 493:233
127. McCormack M, Jin S, Kammlott GW, Chen HS (1993) New Pb-free solder alloy with superior mechanical properties. Appl Phys Lett 63:15
128. Kang SK, Shih DY, Leonard D, Henderson DW, Gosselin T, Cho S, Yu J, Choi WK (2004) Controlling Ag3Sn plate formation in near-ternary-eutectic Sn-Ag-Cu solder by minor Zn alloying. JOM 56:34
129. Kivilahti JK (1995) Modeling joining materials for microelectronics packaging. IEEE Trans Compon Packag Manuf Technol B18:326
130. Huang CW, Lin KL (2004) Interfacial reactions of lead-free Sn-Zn based solders on Cu and Cu plated electroless Ni-P/Au layer under aging at 150 °C. J Mater Res 19:3560
131. Chou CY, Chen SW (2006) Phase equilibria of the Sn-Zn-Cu ternary system. Acta Mater 54:2393
132. Jang JW, Ramanathan LN, Lin JK, Frear DR (2004) Spalling of Cu3Sn intermetallics in high-lead 95Pb5Sn solder bumps on Cu under bump metallization during solid-state annealing. J Appl Phys 95:8286
133. Ramanathan LN, Jang JW, Lin JK, Frear DR (2005) Solid-state annealing behavior of two high-Pb solders, 95Pb5Sn and 90Pb10Sn, on Cu under bump metallurgy. J Electron Mater 34:L43
134. Wang KZ, Chen CM (2005) Intermetallic compound formation and morphology evolution in the 95Pb5Sn flip-chip solder joint with Ti/Cu/Ni under bump metallization during reflow soldering. J Electron Mater 34:1543
135. Yang SC, Ho CE, Chang CW, Kao CR (2007) Massive spalling of intermetallic in solder-substrate reactions due to limited supply of the active element. J Appl Phys 101:084911
136. Brandenburg S, Yeh S (1998) Electromigration studies of flip chip bump solder joints. In: Proceeding of Surface Mount International Conference and Exposition, SMTA, Edina, MN, pp 337–344
137. Chen SW, Chen CM, Liu WC (1998) Electric current effects upon the Sn/Cu and Sn/Ni interfacial reactions. J Electron Mater 27:1193
138. Liu CY, Chih C, Liao CN, Tu KN (1999) Microstructure-electromigration correlation in a thin stripe of eutectic SnPb solder stressed between Cu electrodes. Appl Phys Lett 75:58
139. Lee TY, Tu KN, Kuo SM, Frear DR (2001) Electromigration of eutectic SnPb solder interconnects for flip chip technology. J Appl Phys 89:3189
140. Yeh ECC, Choi WJ, Tu KN, Elenius P, Balkan H (2002) Current crowding induced electromigration failure in flip chip technology. Appl Phys Lett 80:580
141. Huang AT, Gusak AM, Tu KN, Lai YS (2006) Thermomigration in SnPb composite flip chip solder joints. Appl Phys Lett 88:141911
142. Hu YC, Lin YL, Kao CR, Tu KN (2003) Electromigration failure in flip chip solder joints due to rapid dissolution of Cu. J Mater Res 18:2544
143. Gan H, Tu KN (2005) Polarity effect of electromigration on kinetics of intermetallic compound formation in Pb-free solder V-groove samples. J Appl Phys 97:063514

144. Warburton WK, Turnbull D (1975) Fast diffusion in metals, Chapter 4. In: Nowick AS, Burton JJ (eds) Diffusion in solids: recent developments. Academic, New York, pp 171–229
145. Huntington HB, Grone AR (1961) Current-induced marker motion in gold wires. J Phys Chem Solids 20:76
146. Zhang L, Ou S, Huang J, Tu KN, Gee S, Nguyen L (2006) Effect of current crowding on void propagation at the interface between intermetallic compound and solder in flip chip solder joints. Appl Phys Lett 88:012106
147. Lin YL, Lai YS, Lin YW, Kao CR (2008) Effect of UBM thickness on the mean-time-to-failure of flip chip solder joints under electromigration. J Electron Mater 37:96
148. Lin YL, Lai YS, Tsai CM, Kao CR (2006) Effect of surface finish on the failure mechanisms of flip-chip solder joints under electromigration. J Electron Mater 35:2147
149. Chiu SH, Shao TL, Chen C, Yao DJ, Hsu CY (2006) Infrared microscopy of hot spots induced by joule heating in flip-chip SnAg solder joints under accelerated electromigration. Appl Phys Lett 88:022110
150. Black JR (1969) Electromigration failure model in aluminum metallization for semiconductor devices. In: Proceeding of the IEEE Electronic Components Conference, Washington DC, 30 April–2 May, vol 57, p 1587
151. Shatzkes M, Lloyd JR (1986) A model for conductor failure considering diffusion concurrently with electromigration resulting in a current exponent of 2. J Appl Phys 59:3890
152. Choi WJ, Yeh ECC, Tu KN (2003) Mean-time-to-failure study of flip chip solder joints on Cu/Ni(V)/Al thin film under-bump metallization. J Appl Phys 94:5665
153. Xu L, Han JK, Liang JJ, Tu KN, Lai YS (2008) Electromigration induced high fraction of compound formation in SnAgCu flip chip solder joints with copper column. Appl Phys Lett 92:262104
154. Lai YS, Chiu YT, Lee CW, Shao YH, Chen J (2008) Electromigration reliability and morphologies of Cu pillar flip-chip solder joints. In: Proceeding of the 58th Electronic Components and Technology Conference (ECTC), 2008, pp 330–335
155. Rönkä KJ, van Loo FJJ, Kivilahti JK (1997) The local nominal composition-useful concept for microjoining and interconnection applications. Scr Mater 37:1575
156. Rönkä KJ, van Loo FJJ, Kivilahti JK (1998) A diffusion-kinetic model for predicting solder/conductor interactions in high density interconnections. Metallurgical Mater Trans A 29:2951