IEC-PAS 62084

Edition 1.0
1998-12

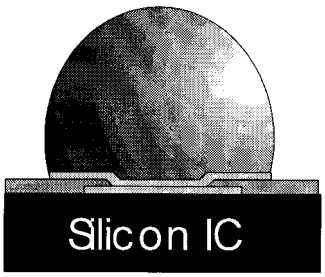
Implementation of Flip Chip and Chip Scale Technology

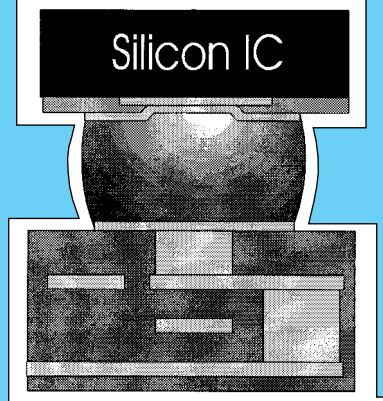
PUBLICLY AVAILABLE SPECIFICATION

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

Reference number
IEC/PAS 62084

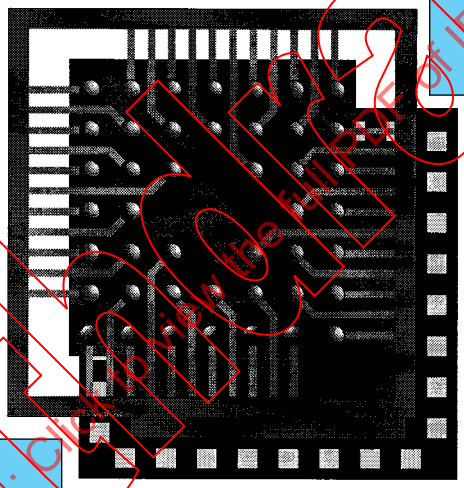
J-STD-012
JANUARY 1996

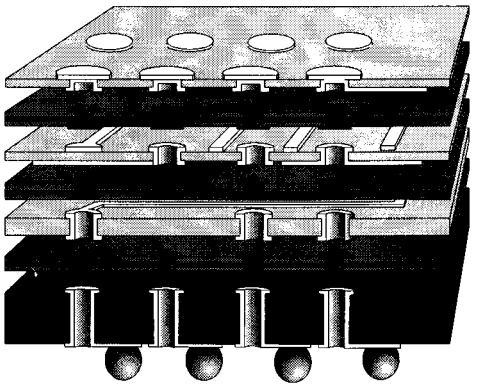




JOINT INDUSTRY STANDARD

IMPLEMENTATION OF FLIP CHIP AND CHIP SCALE TECHNOLOGY





COORDINATED BY THE SURFACE MOUNT COUNCIL

ELECTRONIC INDUSTRIES ASSOCIATION EIA EST. 1924

IPC

SOLID STATE PRODUCTS JEDEC ENGINEERING COUNCIL

MCNC

SEMATECH

INTERNATIONAL ELECTROTECHNICAL COMMISSION

IMPLEMENTATION OF FLIP CHIP AND CHIP SCALE TECHNOLOGY

FOREWORD

A PAS is a technical specification not fulfilling the requirements for a standard, but made available to the public and established in an organization operating under given procedures.

IEC-PAS 62084 was submitted by the IPC (The Institute for Interconnecting and Packaging Electronic Circuits) and has been processed by IEC technical committee 91: Surface mounting technology.

The text of this PAS is based on the following document:	This PAS was approved for publication by the P-members of the committee concerned as indicated in the following document:
Draft PAS	**Report on voting**
91/139/PAS	91/153/RVD

Following publication of this PAS, the technical committee or subcommittee concerned will investigate the possibility of transforming the PAS into an International Standard.

The following statement has been made by IPC (The Institute for Interconnecting and Packaging Electronic Circuits):

The IPC has the leadership position on this publication, as suggested by the Surface Mount Council. Any input or suggestion from other persons or organizations, not a part of the IPC membership, has been coordinated by the IPC during the development process.

The IEC and its members are authorized to exploit the following document:

J-STD-012 Implementation of Flip Chip and Chip Scale Technology

under the PAS procedures for the purpose of international standardization.

1) The IEC (International Electrotechnical Commission) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of the IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, the IEC publishes International Standards. Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. The IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
2) The formal decisions or agreements of the IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested National Committees.
3) The documents produced have the form of recommendations for international use and are published in the form of standards, technical reports or guides and they are accepted by the National Committees in that sense.
4) In order to promote international unification, IEC National Committees undertake to apply IEC International Standards transparently to the maximum extent possible in their national and regional standards. Any divergence between the IEC Standard and the corresponding national or regional standard shall be clearly indicated in the latter.
5) The IEC provides no marking procedure to indicate its approval and cannot be rendered responsible for any equipment declared to be in conformity with one of its standards.
6) Attention is drawn to the possibility that some of the elements of this PAS may be the subject of patent rights. The IEC shall not be held responsible for identifying any or all such patent rights.

Material in this standard was voluntarily coordinated by the Surface Mount Council (SMC), and established by Technical Committees of IPC, EIA and JEDEC. Committee members of the three organizations contributed their time, knowledge and expertise to blend a cohesive report on the topic covered by this document. Representatives from Sematech and MCNC also participated and contributed to the development of this document. Proposals were sent to key individuals in each of the individual standard development organizations for consensus. Meetings were held to resolve differences or conflicts prior to documenting the information in the final released version.

The material contained herein is advisory and its use or adaptation is entirely voluntary. IPC, EIA and JEDEC disclaim all liability of any kind as to the use, application, or adaptation of this material. Users are also wholly responsible for protecting themselves against all claims or liabilities for patent infringement.

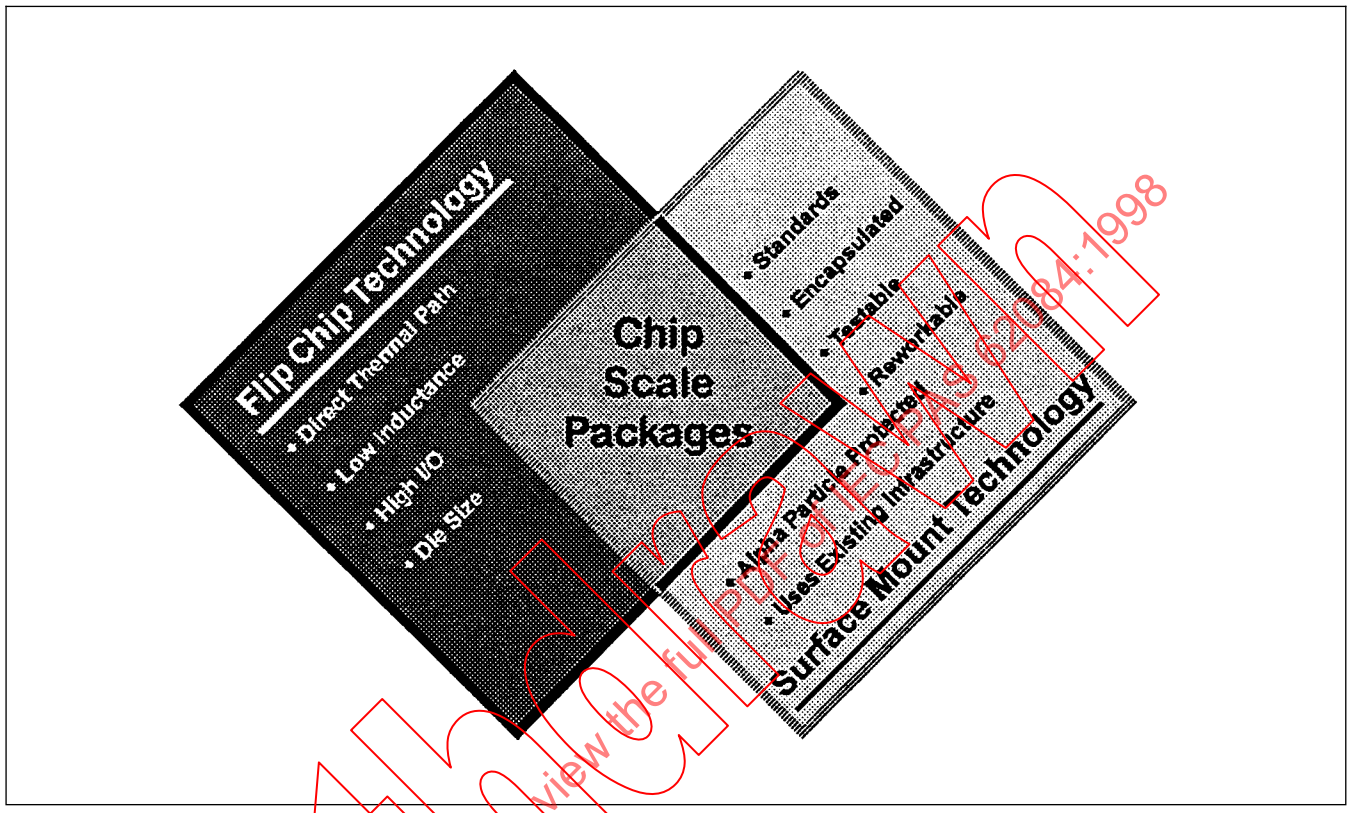
Comments Welcome

The J-STD-012 is intended to serve as a roadmap for flip chip and chip scale implementation.

In order to keep the document current, the Surface Mount Council welcomes comments from individuals reading the document, or implementing the suggested concepts.

Comments may be sent to the EIA, IPC or JEDEC. All comments will be organized and sent to representatives of the Ad Hoc Committee for a yearly review and updating procedures.

J-STD-012
IMPLEMENTATION OF FLIP CHIP AND CHIP SCALE TECHNOLOGY



About this Document

This document is intended to report on the work being done by a variety of organizations concerned with surface mounting of bare die in flip chip or chip scale configurations. The details were developed by companies who have implemented the processes described herein and have agreed to share their experiences. Readers are encouraged to communicate to the appropriate trade associations or societies any comments or observations regarding details published in this document, or ideas for additional details that would serve the industry.

Section 10 of the document represents a listing of standards that are being developed, being updated, or need to be created in order to provide for the orderly implementation of flip chip or chip scale technology. Members of industry are invited to participate in the ongoing standardization process.

For additional information regarding material published herein or inquiries regarding the status of standardization activities, we urge you to contact the organization listed below.

IPC
The Institute for Interconnecting and Packaging Electronic Circuits
2215 Sanders Road
Northbrook, IL 60062-6135
Telephone: (847) 509-9700
Fax: (847) 509-9798

EIA
Electronic Industries Association
2500 Wilson Blvd.
Arlington, VA 22201-3834
Telephone: (703) 907-7552
FAX: (703) 907-7501

JEDEC
Joint Electron Device Engineering Council
2500 Wilson Blvd.
Arlington, VA 22201-3834
Telephone: (703) 907-7560
Fax: (703) 907-7501

Acknowledgment

In response to the need for a document covering flip chip and chip scale technology, J-STD-012 was developed. Because EIA, IPC, and JEDEC have members worldwide who participated in this effort, J-STD-012 represents a comprehensive report on the subject of flip chip and chip scale design rules. While the principal members of the Task Group are listed below, a special note of thanks goes to everyone who assisted in the development of this document.

Abbas I. Attarwala, Hewlett Packard
Walt Ackerman, Universal
Bernie Aronson, EIA
Pamela Ashcraft, Delco Electronics
Mark Baker, National Semiconductor
Roger Benson, Grace Specialty Polymers
Dieter Bergman, IPC
Ruben Bergman, HDP User Group
Rao Bonda, Motorola
Kimera Cho, Rockwell International
Dave Corisis, Micron Technology
Richard Cooley, AMKOR Electronics
Brian Conaghan, DuPont
Ben Cranston, AT&T Bell Labs
Philip Deane, MCNC
Keith DeHaven, Motorola
Werner Englemaier, Englemaier Associates
Joe Fjelstad, Tessera
Garry Green, Cypress Semi Conductor
Bill Hargis, Coors Electronic Packaging Co.
Matthew Heydinger, Hewlett-Packard
Gary Hill, IBM
John Jackson, Sematech
Donald Johnson, Sandia National Labs
William Kenyon, Global Center for Process Change
Mohan Kirloskar, Shinko
Young Kwon, Philips Semiconductor
Helen M. Low, Celestica
Nick Lycoudes, Motorola
Rob Lyn, Celestica (IBM Canada Sub.)
Alan Lyons, AT&T Bell Laboratories
Susan Mack, Delco Electronics
Paul A. Magill, MCNC
Phil Marcoux, Chip Scale Inc.
Charles Martin, Hewlett Packard
Mark McLeod, IBM
Kevin Moody, Hewlett Packard
Archie Mones, E.I. DuPont
Charles Pagel, NSWC-CRANE
Tze W. Poon, Sematech
David Read, NIST
Cathy Reber, Sandia National Labs
Glenn Rinne, MCNC
Ray Rinne, IBM
Paul Roeder, Universal Instruments
Paul Totta, IBM East Fishkill
Iwona Turlik, Motorola
Dan Ward, Delco Electronics
David Wright, NSWC - Crane Division
Frank S. Stein - Consultant
James Young, Chip Scale Inc.

A special note of thanks goes to the representatives of MCNC, who hosted several of the meetings during which this document was developed. Also a word of appreciation to Bernie Aronson, EIA, and Joe Fjelstad, Tessera, for their participation in the special Ad Hoc Committee that assembled the input for this document.

The IPC is also grateful to representatives of IBM and Motorola, who provided information on design of flip chip assemblies for incorporation into this document.

Table of Contents

Figures

Tables

Implementation of Flip Chip and Chip Scale Technology

1 SCOPE

This document describes the implementation of flip chip and related chip scale semiconductor packaging technologies. The areas discussed include: design considerations, assembly processes, technology choices, application, and reliability data. Chip scale packaging variations include: flip chip, High Density Interconnect (HDI), Micro Ball Grid Array (µBGA), Micro Surface Mount Technology (MSMT) and Slightly Larger than Integrated Circuit Carrier (SLICC).

1.1 Purpose

This document is intended to provide general information on implementing flip chip and chip scale technologies for creating single chip or multichip modules (MCM), IC cards, memory cards and very dense surface mount assemblies.

1.2 Categorization

Flip chip is categorized as versions of a tin-lead (SnPb) solder bump process, and alternative solutions that use other forms of chip bond site bumping.

Chip scale technology is categorized as semiconductor chip structures that have been made robust to facilitate ease of chip handling, testing and chip assembly. The chip scale technologies have common attributes of minimal size, no more than 1.2X the area of the original die size, and are direct surface mountable.

2 TECHNOLOGY OVERVIEW

Flip chip and chip scale technology relate to methods used to provide an efficient technique for interconnecting semiconductor die to a substrate. Over the years flip chip technology has grown in stature and expanded into the more generic area of chip scale technology. In addition, many new techniques for mounting bare die were also established. Today, enhancements of the semiconductor die, which ruggedize the product to facilitate ease of handling and testing, have established another level of packaging technology in which many companies can participate.

2.1 History of Flip Chip

Flip chip solder bump interconnection technology is over 30 years old. It was first conceived and developed in 1960-61 by IBM for the Solid Logic Technology (SLT) hybrid electronic circuitry in IBM's System 360 computers introduced in April 1964.

At that time, standard transistor packaging used hermetically sealed metal cans with glass sealed wires emerging from a header upon which the germanium or silicon chip was metallurgically backbonded. Manual thermocompression wire bonding to the chip was the common technique.

Although transistor technology was much more reliable than vacuum tube technology that preceded it, the packaging and interconnection technologies were weak. Faulty manual wire bonds, purple plague (gold-aluminum intermetallic formation), and aluminum corrosion of thin film interconnections on the chip, even in hermetic packages, were all reliability concerns. In addition, manufacturability and productivity were deficient.

SLT transistors and diodes were glass passivated at the wafer level to protect aluminum wiring from the environment. Glass frits of borosilicate glass were fused on the surface of transistor wafers after the aluminum wiring was formed. The glass film obviated the need for a hermetic enclosure because the transistor was sealed at the chip level.

Consistent with maintaining a hermetic seal was the idea of etching via holes in the glass where electrical connections were to be made, and hermetically resealing the holes with an overlapping, larger pad of metal films. The first film, typically chromium, was used to create a good bond to glass and aluminum.

Subsequent metal films, typically copper and gold, were deposited to provide wettability and solderability for a ball of metal alloy that was to be melted on the pad. Originally, the ball was an alloy of Au-Sn eutectic which melted at 360°C; therefore, the pad was called the ball-limiting metallurgy because the melted ball would not wet the glass beyond the area of the pad. Later the ball or bump became a solder bump with an embedded copper ball ''stand-off,'' shown in figure 2-1.

The packaging concept required that thick film electrodes or wires be screened and fired on alumina substrates. Originally, the thick film was glass fritted Au-Pt paste; later it became Ag-Pd. The wiring was hot dip tinned with 90Pb-10Sn solder to improve conductance and to interconnect the active and passive components. For each transistor, three of the screened wires were to terminate in a spot that would be the mirror image of the chip bumps.

After each chip site was fluxed with a droplet of sticky water white rosin flux, the chip was picked up and placed on the solder coated electrodes face down or ''flipped'' compared to normal chip placement. Subsequently, the substrate was heated in a furnace to remelt the solder coating,

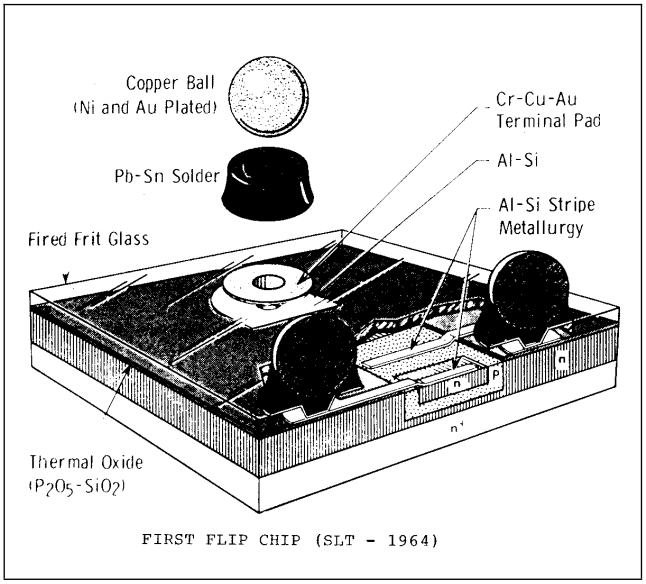


Figure 2−1¬ Basic Metallurgy/Glass Design for SLT Transistors

thereby reflow soldering the device bump asperities and joining the chip to the substrate.

However, there were problems with the original gold alloy balls in that they would alloy and melt with the high lead solder at a relatively low temperature (Pb-Au eutectic melts at 215°C). The ball would alloy and flow into the solder lands thereby ‘‘collapsing’’ and allowing solder to short to the raw edges of the chip (Figure 2-2). There were three solutions to the problem.

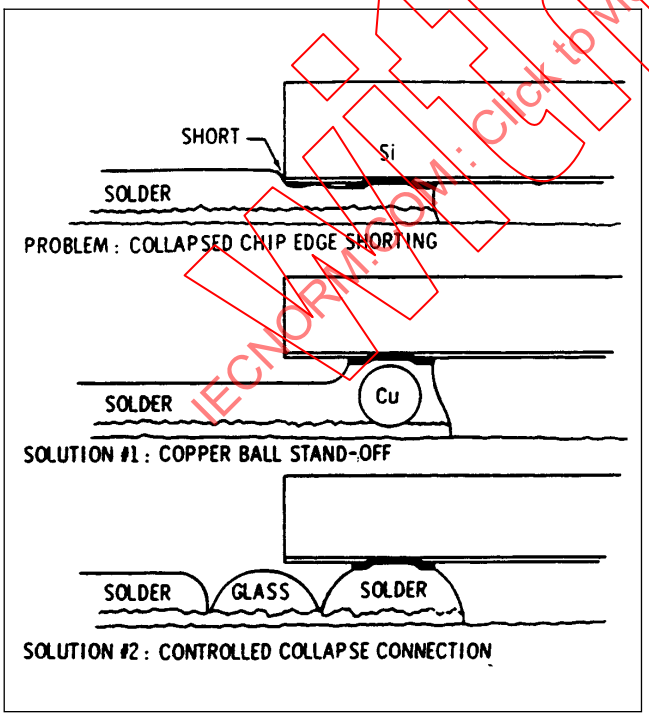


Figure 2−2¬ Chip Collapse and Edge Shorting Problem During Solder Reflow Joining and Two Solutions

The first was to embed copper balls in the solder bumps. This positive stand-off solution was used for all of the SLT and SLT extension designs. The second solution, developed by Delco Electronics (General Motors) in the late 60s, was similar to embedding copper balls except that the design employed a rigid silver bump. The bump provided a positive stand-off and was attached to a thick film ceramic substrate by means of solder that was screen-printed onto the substrate, as shown in figure 2-3.

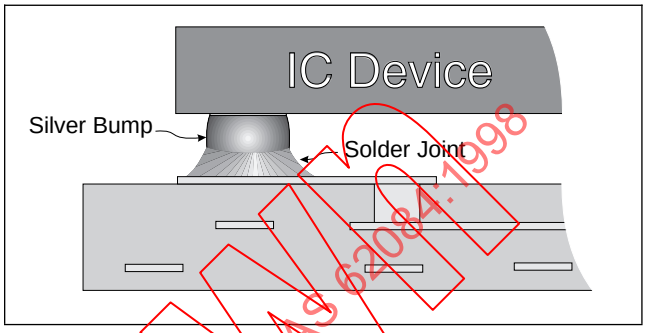


Figure 2−3¬ Flip Chip with Silver Bump Stand-Off

The third solution was to use a screened glass dam near the electrode tips to act as a ‘‘stop-off’’ to prevent the ball solder from flowing down the electrode. The silver-palladium paste under the glass provided sufficient conductivity for circuit continuity. By then the Ball Limiting Metallurgy (BLM) with a high-lead (Pb) solder system and a copper ball had proven to work well. Therefore, the ball was simply removed and the solder evaporation process extended to form pure solder bumps which were approximately 125µm high.

This system, shown in figure 2-4, became known as the controlled collapse chip connection (C^4, or C4). For integrated circuit fabrication where the I/O count would grow from 3 pads to 20 or 30, the process for fabrication and the uniformity of C4 bumps was superior to previous IBM copper ball design.

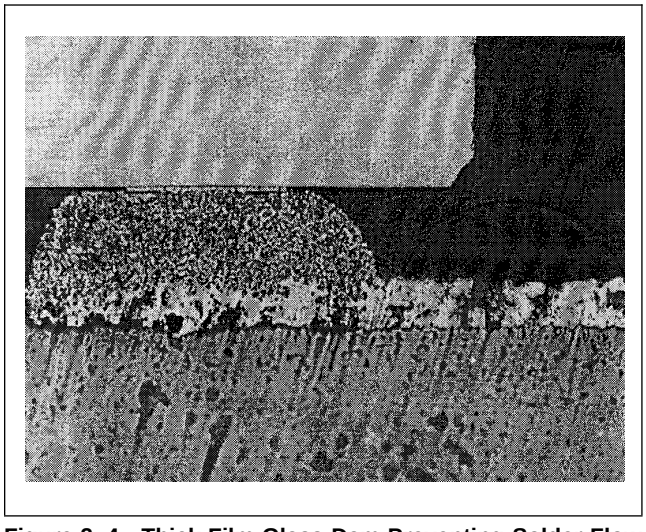


Figure 2−4¬ Thick Film Glass Dam Preventing Solder Flow and Collapse in First C4 Application

There was a long era of thick film hybrid technology from the early 60s to late 70s at IBM using thick film screened and tinned conductors and thick film trimmable resistors with flip chips. Discrete transistor and diode chips used copper ball solder bumps, while early integrated circuits had C4 solder balls. Sometimes both were mixed on the same substrates, shown in figure 2-5.

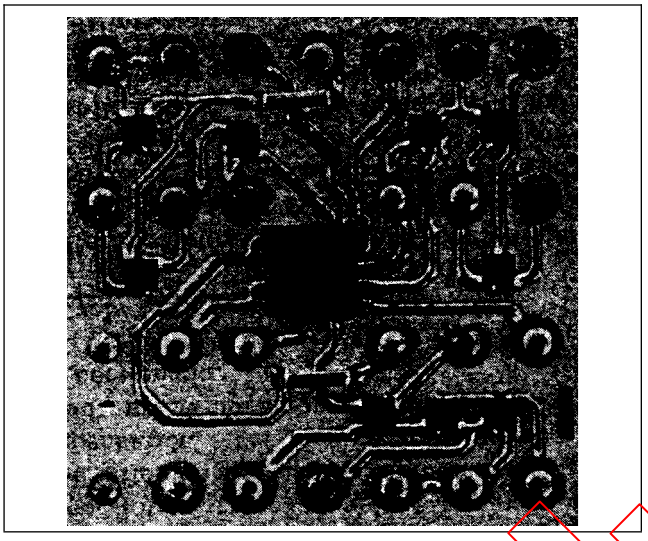


Figure 2−5¬ Early IBM Hybrid Thick Film Module Mixing Copper Ball SMT Transistors and C4 Integrated Circuit. (12 mm)

Contemporary with this, Delco Electronics' silver bumps did not require tinning of the conductors. By the mid 80s, Delco Electronics was using a reflowable indium-lead solder bump with alternative methods for controlling solderability on the substrate. This bump was compatible with low temperature solder processing without requiring deposition of solder on the substrate. Motorola's efforts in this time period included evaporated (and later electroplated) tin-lead bumps.

As the number of circuits on a chip and the number of interconnections on logic and memory IC's grew, it became necessary to increase the number of I/O pins on substrates. This led from 12.5 mm (0.5 inch) SLT substrates to 25 mm (1 inch) substrates of thin film metallized ceramic (MC). The thin film metallization was evaporated, later sputtered, Chromium-Copper-Chromium. The upper chromium film was left intact except at pin interconnect site heads and C4 finger tips. The chromium (or chromium oxide surface) then replaced the glass dam as ''stop-off'', shown in figure 2-6.

Later, using cofired multilayer alumina substrates, electrical connections were brought to the surface through a molybdenum stud-capped nickel/gold vias. Since solder does not wet to the alumina, the need for glasss or chromium dams disappeared, as shown in figure 2-7.

Early integrated circuits had solder bumps around the perimeter of the chip because of concerns that the passivation vias opening process or the bump stresses would affect underlying transistors. Later, it was discovered that there was no problem putting solder bumps directly over integrated circuits, transistors and diodes. This permitted the staggering of bumps around the perimeter with an occasional inboard power pad for single level substrates of thick or thin film designs. Thin film substrates enabled finer pitches and more escape channels between pads so that depopulated area arrays could be accommodated. Multilayer ceramic substrates with buried redistribution wires allowed full area arrays and higher I/O capability than TAB or wire bond peripheral connections, shown in figure 2-8.

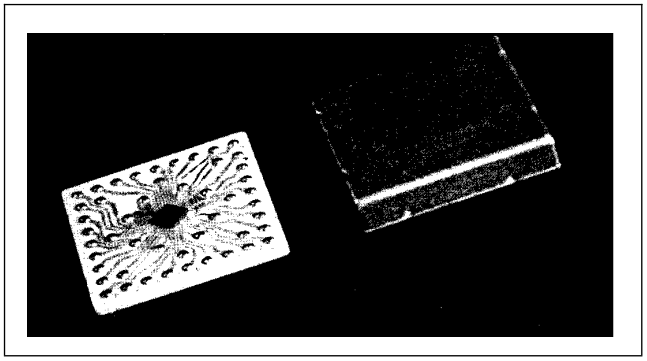


Figure 2−6¬ Thin Film Chromium-Copper-Chromium on Ceramic (MC)

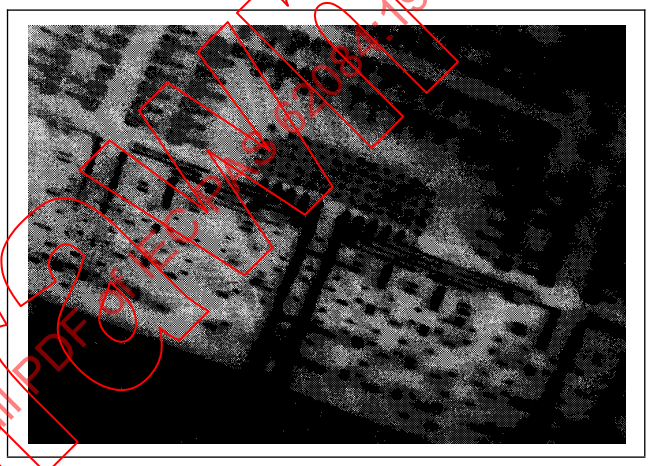


Figure 2−7¬ Cross-Section of Cofired Alumina Multilayer Ceramic Package

2.1.1 Hybrid Thick Film Packages

Early solder bump flip chip packages used thick films of glass fritted precious metals tinned with 90% lead-10% tin solder. Cost reduction favored silver-palladium paste but reliability issues with silver caused later use of ternary pastes of silver-palladium-gold. A typical package, see figure 2-5, shows the bottom-side view of thick film resistors that were dynamically trimmable with abrasive jets and later with lasers.

Equally important to the elimination of hermetic packaging and manual wire bonding was the profound improvement in manufacturability of circuits. Glass passivated transistors

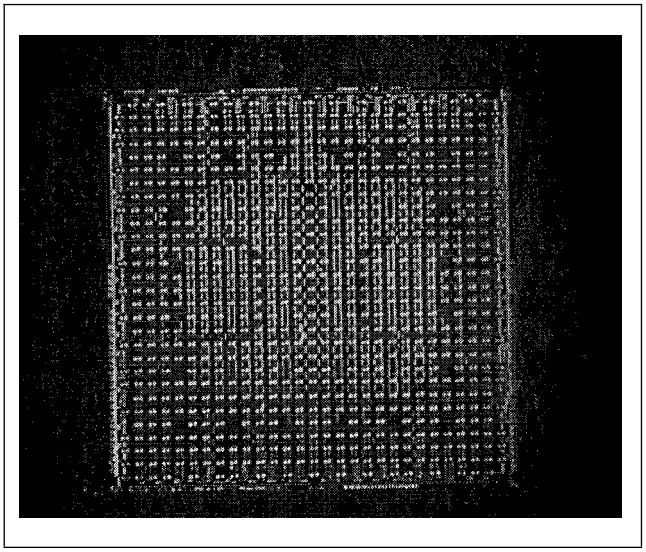


Figure 2−8¬ Full Area Array C4 Configuration Microprocessor with 762 Solder Bumps in a 29x29 Array.

and diode chips were so robust they could withstand random mechanical handling in vibratory bowl feeders. Testing and chip placement were highly automated, attaining process rates of 3 to 6 chips per second. Chip joining to substrates was done with thousands of solder joints being made in reflow furnaces simultaneously.

Low cost and high productivity were closely associated. Those factors along with the mechanical ruggedness made flip chip attractive to the automotive industry as well. General Motors (through Delco Electronics) became an early high volume user of hybrid flip chip technology for voltage regulators and ignition modules.

Completion of a thick film module typically required silicone gel around the chips to protect against moisture, a metal can for mechanical protection and RTV™ silicone rubber for a backseal.

Solder bump integrated circuit chips brought other changes. Fired frit glass films were replaced by sputtered SiO_2 films that were compressive in residual stress and lower in pinhole defect density. These films were ultimately good enough to support two and three levels of chip metallization.

As the number of circuits grew from three to four to tens of circuits, the number of I/O bumps on logic chips grew as did the demand for pins in the substrate. Interstitial pins were put in the pin grid array to increase the I/O count. Rent's rule was found to apply to logic chips as well as logic cards (empirical observations showed that card I/O terminal count is directly related to the number of logic circuits).

2.1.2 Hybrid Thin Film Packages

Due to the continued of logic chips, large ceramic substrates were required. These were approximately 25 mm^2 with higher pin capacity and grew in size as the need developed. Also, thick film screening was supplemented with thin film chromium-copper-chromium depositions (Metallized Ceramic or MC). These films were subtractively etched, using photolithographic processes, to form finer lines and pitches than their thick film counterparts (20 µm lines, 60 µm pitches, minimum).

Thin film processing maximized the wireability of the substrate and allowed more wires per channel between pins or C4 bumps. Subsequently, polyimide thin films were used in combination with metal thin films to allow a buried ground plane and a two level wiring capability with vias and crossovers. [Metallized Ceramic with Polyimide (MCP)]

As the number of logic circuits on a chip continued to grow, the peripheral solder bump pad arrangement became a constraint to circuit count. It was soon discovered that solder bumps could be put over active circuits with no negative effect. Stresses from testing or joining were not transferred to active devices unlike other chip joining technologies such as wire bonding. This led to in-board solder bumps, at first peripherally because chip testers were best suited to such configurations, then in staggered multi-row peripheral patterns. Later, with MC and MCP capabilities to fan out the interconnection of arrays and new chip testers, patterns known as depopulated arrays were possible as shown in figure 2-9. Solder bumps over the active area greatly relieved the I/O constraint. Chips could be smaller or could accommodate a greater number of circuits, as the designer desired.

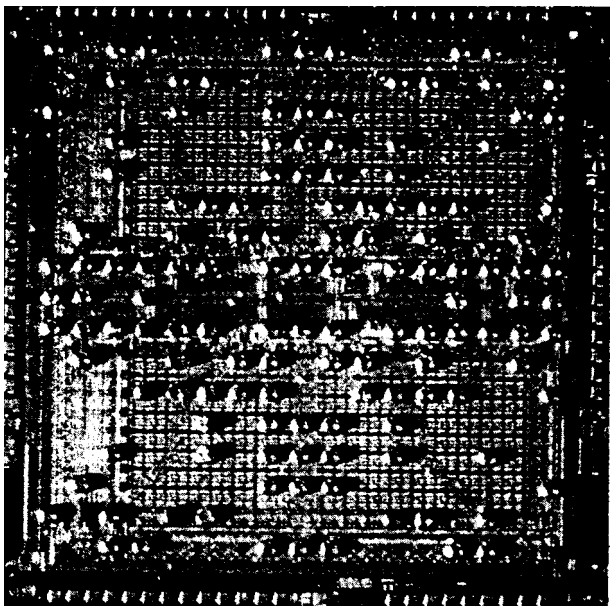


Figure 2−9¬ Depopulated C4 Array on Chip

2.1.3 Multilayer Ceramic Substrate (MLC)

A full area array C4 chip footprint required the high I/O redistribution capability of multilayer ceramic technology.

For example, the top five layers of green sheet in the cofired ceramic packages introduced by IBM in 1978 were necessary to fan out an 11x11 array of C4 pads, shown in figure 2-10.

The package body was alumina with glass binder and

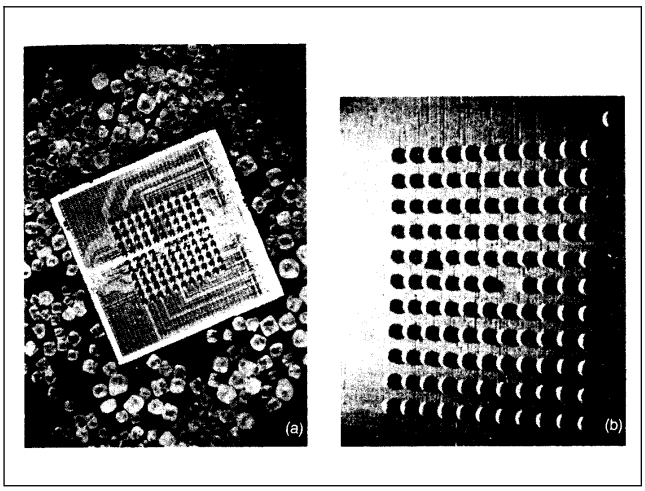


Figure 2−10¬ Area Array C4 Configuration (a) 11x11 Full Array with Cantilevered Silicon, (b) SEM View

molybdenum conductors. The surface finish of the molybdenum studs emerging from the top and bottom of the package was nickel and gold film for solderability and brazeability.

Other manufacturers, such as Hitachi and NTT, began to use flip chip and co-fired MLC not only for silicon chips but also for thin film flip chip resistors and capacitors as shown in figure 2-11. The module illustrated was for a telecommunications application.

Because of the coefficient of thermal expansion (CTE) mismatch in expansivity between alumina (CTE=6 PPM/°C) and silicon (CTE=3 PPM/°C) much of the silicon was cantilevered beyond the pad array to minimize distance from the neutral point (DNP), shown in figure 2-10. Chips had grown sufficiently large that the mismatch in expansion rate could plastically deform the most remote solder pads during thermal cycling. This caused fatigue fracture and electrical discontinuity of those pads in the corners.

Multilayer ceramic technology also allowed the use of Multiple Chips on the Module (MCM). Figure 2-12 shows an uncapped flip chip multichip module. Each chip had 120 I/O pads, supported a maximum of 704 available bipolar circuits and dissipated a maximum of 1.5 watts to an air cooled cap.

IBM developed a series of Thermal Conduction Modules (TCMs) starting in 1981 where the chip count on the substrate grew to 100-130. Each chip was backed by a spring loaded piston in a helium atmosphere and could dissipate 3 watts/chip, as shown in figure 2-13. The area array pads started as 11x11 arrays but grew to 17x17 arrays in subsequent models.

Improved thermal cycle solder joint performance, by reducing distance-to-neutral-point, was short lived as chip circuit density was growing rapidly and larger chips were imminent. The fatigue life vs. CTE relationship is shown in figure 2-14.

It is clearly shown that fatigue life is not a problem if the expansion rate of the chip could be matched with a silicon substrate (see figure 2-15) or a man-made substrate such as glass-ceramic (cordeirite-like) which matches silicon.

2.1.4 Thin Film on Multilayer Ceramic

The glass ceramic MLC substrate which IBM developed for System/390 also had a redistribution layer of thin film wiring on its surface to reduce the number of buried layers that would have been necessary for a 28x28 array of C4's (648 bumps) on each chip.

The heat dissipation through each piston was raised to a maximum of 28 watts using an oil heat transfer fluid in place of helium, see figure 2-16.

2.1.5 Thin Film on Organic Epoxy/Glass Laminate

Chip under fill development improved the solder joint reliability of flip chips by at least an order of magnitude by encapsulating the solder joints in filled epoxy, as shown in figure 2-17. The strain level is reduced and reliability is therefore extended. Chips can be mounted to a printed board as a surface mount component without a plastic or ceramic package. This technology is an example of Direct Chip Attach (DCA). Smaller, lighter, less expensive electronic applications therefore are attainable.

Another example of DCA is IBM's SLC technology. By using layers of photosensitive epoxy and copper thin film wiring on the epoxy glass laminate board surface, it is possible to create chip bonding sites that match solder bump footprints similar to ceramic packages, shown in figure 2-18. By pretinning the land pattern with tin lead (SnPb) eutectic, high-lead bumps can be attached to the substrate.

Application of chip underfill enhances reliability performance. Epoxy underfill reinforcement allows use of lower-compliance eutectic SnPb bumps, which eliminates the need to pretin or deposit solder on the substrate thereby reducing cost.

2.1.6 Chip on Flexible Printed Board

Chip on flex is another direct chip attach technology. Most flexible circuits are polyimide-based substrates with or without adhesives. This technology is used in static and dynamic applications. Viewed as a compliant substrate this DCA method is designed to reduce or eliminate the need for underfill.

2.1.7 Thin Film on Silicon

An AT&T developed process places flip chips into an alloy solder paste that has been printed onto an active silicon

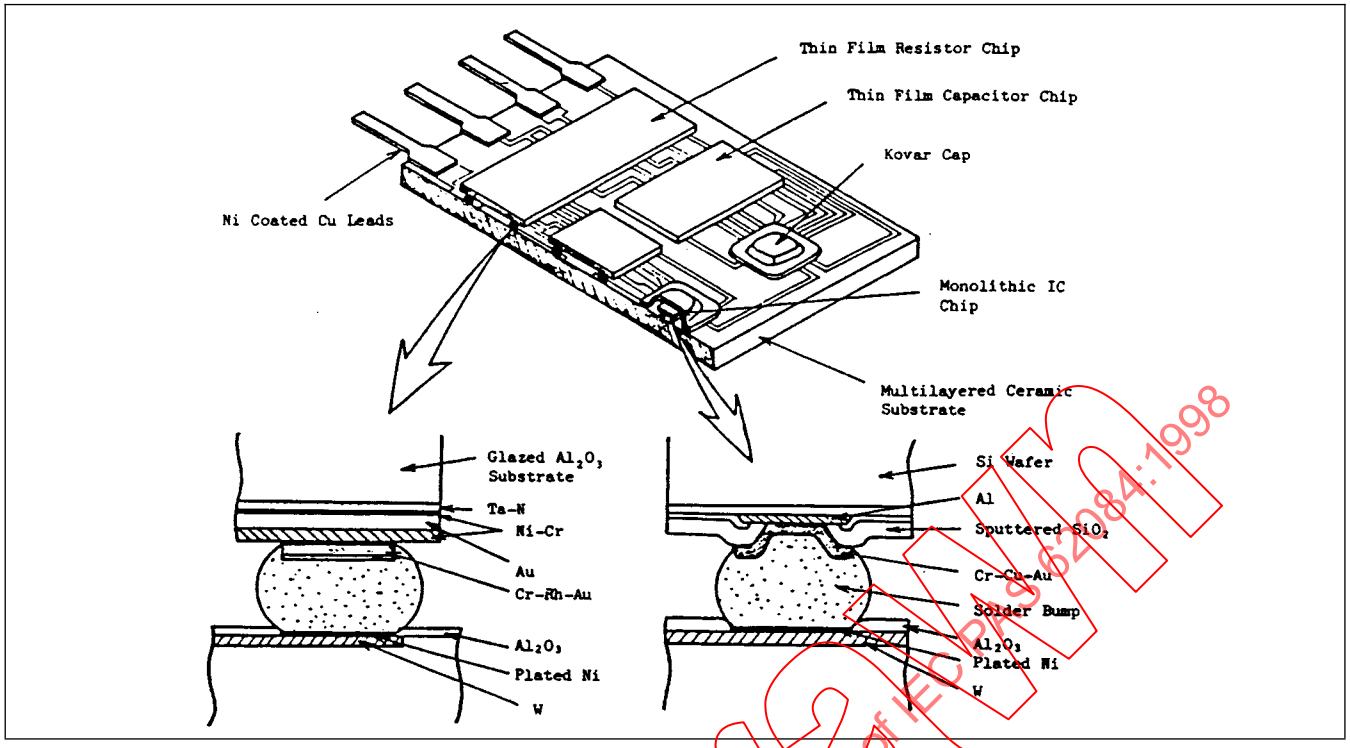


Figure 2−11¬ Structure of Hybrid IC Using Solder Bump and Cross-Sections

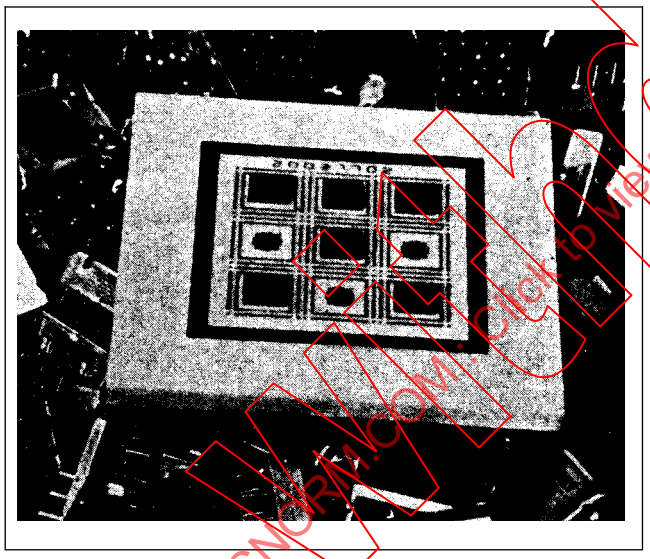


Figure 2−12¬ Uncapped 50 mm Multi-Chip Module (MCM)

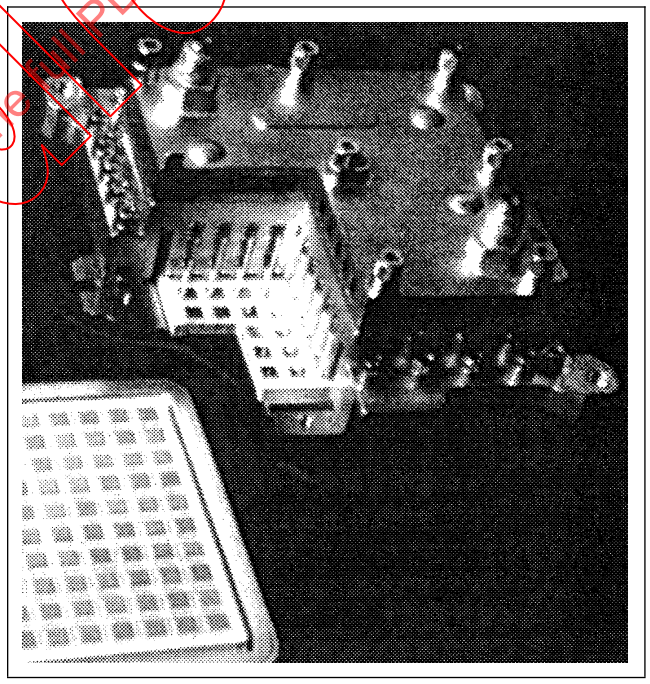


Figure 2−13¬ IBM Thermal Conduction Module with 100-130 Flip Chips and Hat with Piston Assemblies

interconnection MCM substrate. Both the die and substrate have matching patterns of solder wettable connection pads. The printing step deposits uniform volumes of solder paste on all the interconnection lands on the substrate in one step.

2.1.8 Characteristics of Flip Chip Technology

There are many advantages to the use of flip chip. Some of these are:

- Reliability without hermeticity at the chip level has been achieved successfully for many years through the use of glass, sputtered SiO_2, CVD oxide or nitride, or polyimide passivation films on the face of the chip.
- Rugged, high strength interconnections between chip and package make disturbance unlikely in severe mechanical environments of shock and vibration such as automotive, aircraft and space launch applications.
- The highest density of packaged circuits currently

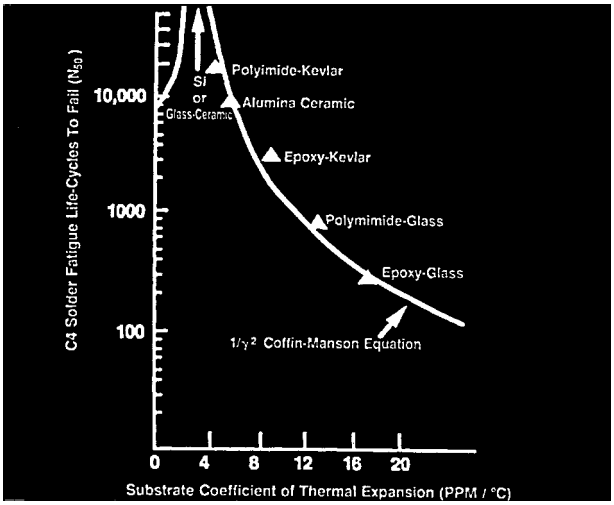


Figure 2–14¬ Effect of TCE Mismatch on Solder Fatigue Life

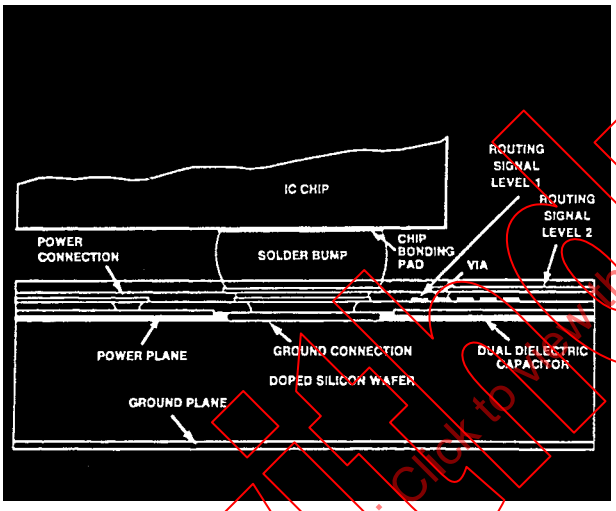


Figure 2–15¬ AT&T Silicon-on-Silicon Packaging System

achievable has been demonstrated with multichip modules of flip chips with area array interconnects on multilayer ceramic substrates.

- Self alignment of chips during reflow soldering due to surface tension forces in liquid solder allows chips to be placed relatively crudely (as much as 50% off of substrate lands). This allows thousands of joints to be made simultaneously in a reflow furnace.
- Photonic applications gain very high optical efficiency by attaining 1-2 µm overlay from solder bump self alignment, and therefore flip chips have found wide application in photonics.
- Excellent electrical properties with low capacitance, inductance and resistance exist, due to the short electrical paths enabling future application for high frequency circuitry.
- Proven field reliability is experienced in billions of shipped flip chip solder connections in a variety of packaging applications from thick film hybrids to cofired multilayer ceramic (MLC).
- Reliable replacement methodology is possible for multichip modules, which makes repair and engineering change possible and practicable.
- Low profile and small footprint enable very dense applications.
- Direct chip attach on organic rigid material and flex with lower melting temperature solders is SMT compatible and has successfully demonstrated opportunities for more consumer goods applications.

Flip chip packaging technology has some limitations. Users have determined some areas where flip chip technology does not apply. Some of the limitations and concerns are:

- The required infrastructure for manufacturing, testing and handling of flip chips can be an obstacle. Commitment to fixed footprints of bump patterns and matching substrates can be constraining.
- Availability and cost of flip chips from semiconductor vendors have been a problem, and contract bump facilities are in their infancy.
- CTE mismatch between silicon and substrate expansivity causes such limitations as thermal cycle fatique failure of solder and thermal shock failure. Where possible the strain level in the solder may be greatly reduced by using underchip encapsulants and adhesives, but presently at the cost of throughput rework capability.
- Inability to optically inspect hidden area array bumps may be a consideration in the quality control of soldered interconnections. However, process optimization can eliminate the need for bump inspection.
- Heat dissipation (0.5-2 watts) is limited in flip chips without thermal enhancements.
- Alpha particle radiation from ordinary solder can cause soft error events.

2.2 Introduction to Chip Scale Packaging

Chip scale packaging or CSP is an emerging branch of semiconductor packaging that seeks to bridge the gap between flip chips and conventional surface mount packages. The technology is developing in response to some of the limitations of flip chip technology and to address the concerns and perceived risks associated with handling and assembling bare die while still maintaining most of the volumetric packaging and performance advantages that flip chip technology offers. See figure 2-19.

The terms chip scale package and CSP became popular in

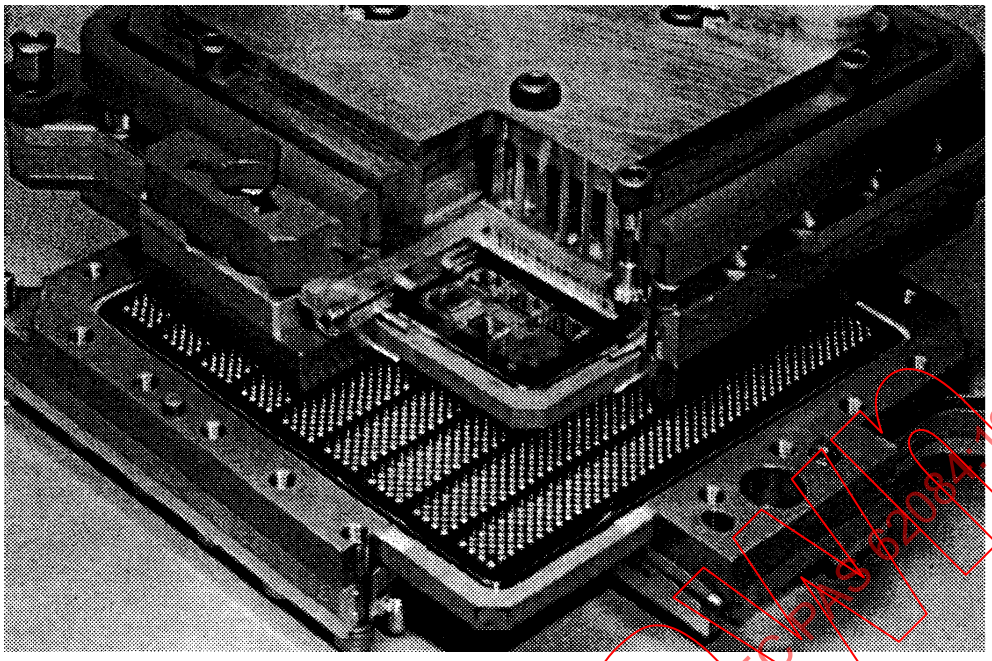


Figure 2−16¬ Cut-away of IBM Glass/Ceramic TCM with Polyimide/Thin Film Surface Redistribution Layer

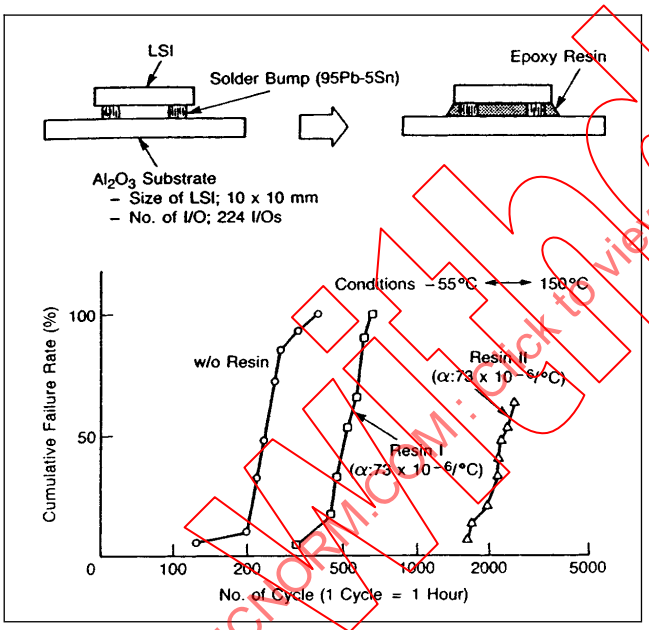


Figure 2−17¬ C4 Life Extension by Use of Filled Epoxy Resins with Matching Expansivity (Hitachi)

1994. The term was used in several presentations to indicate comparison of the economics of multichip modules vs. denser SMT assemblies using smaller surface mountable packages that were called CSPs. The conclusion was that CSP's were far more economical than MCMs.

Chip scale packages parallel standard surface mount packages in form factor, being found in both peripherally leaded and area array formats. As with their larger counterparts, chip scale packages are designed to facilitate test and burn-in of the semiconductor before committing them to the assembly process, thus effectively providing and answer to the current problem of known good die (KGD). CSPs generally use the same assembly process as larger SMT packages. Presently there are a number of CSPs being developed, evaluated, or offered as solutions.

The original chip scale package was the beam lead package developed by AT&T in the 1960s. This package was proven very reliable and economical. However, it was viewed as more difficult to handle than the dual in-line package, or DIP.

The common advantage of these differing approaches is the ability to offer packaged chips that will facilitate movement toward smaller, lighter, high performance systems at lower cost, using the current SMT assembly infrastructure.

The need for smaller devices, the developments in flip chip bumping processes, and advancements in the packaging of microwave devices have lead to the development of chip scale package concepts. CSPs are extensions to the flip chip concept as the packaging of the chip provides robust handling and attachment without the need for additional materials, such as underfill epoxies.

Chip scale packages are available in two basic configurations;

1. Chip Scale Grid Array Packages (CSP-A)
2. Chip Scale Peripheral Leaded Packages (CSP-P)

Chip scale packages are attached to the PCB using the basic fine pitch surface mount technology assembly process. The fine pitch process starts by stenciling a thin layer of fine particle solder paste onto the PCB land areas. Next, the bumps, balls, or leads of the chip scale and other SMT packages are placed in contact with the solder paste. The assembly is heated in a controlled oven to a temperature

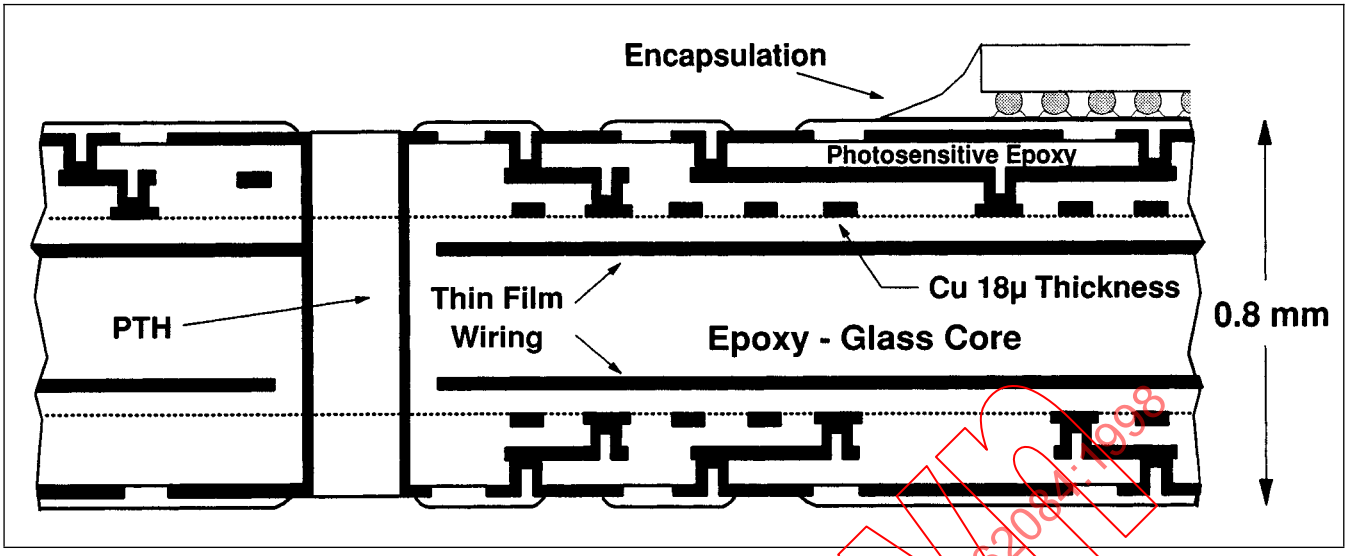


Figure 2−18¬ IBM SLC Chip-on-Card Technology

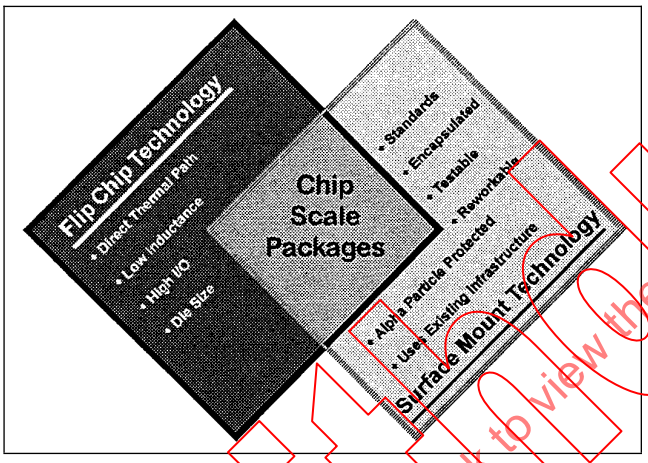


Figure 2−19¬ Bridging the Gap

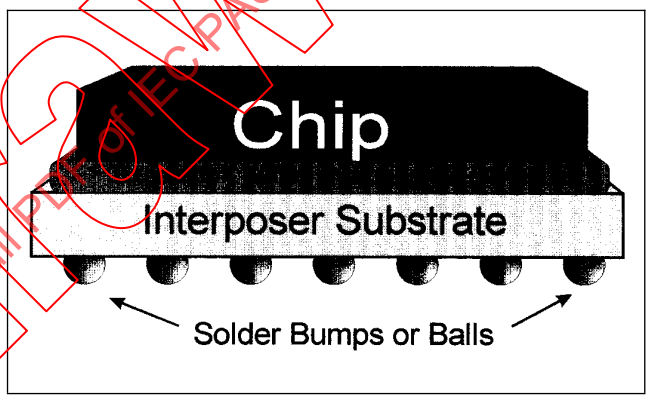


Figure 2−20¬ Chip Scale Grid Array Package (CSP-A)

above the melting temperature of the solder. The assembly is cleaned, if required. As with flip chip, an underfill epoxy may be injected under and around the chip scale devices to enhance their solder joint reliability, however, like conventional SMT, certain chip scale packages do not require the underfill epoxy. After curing the underfill epoxy, the assembly is tested.

2.2.1 Chip Scale Grid Array Packages

Chip Scale Grid Array Packages (CSP-A) are produced in different variations. These include the mini and micro ball grid arrays. These packages are similar in appearance to the larger Ball Grid Array or BGA packages. The pitch of the solder balls on BGAs is 1.0 to 1.5 mm; on CSP-As the pitch can be from 0.5 to 0.65 mm. The basic structure of a CSP-A is shown in figure 2-20. An example CSP-A is shown in figure 2-21.

In one chip scale array the chip is attached to an interposer that contains an interconnection method designed to translate bonding sites on the chip to a large array. The interposer is connected to solderable balls, posts or planar lands on the bottom. The chip interconnection methods to the interposer include wire bonds, flip chip bumps and TAB style ribbon leads.

2.2.1.1 Micro BGA (µBGA)

In constructing a micro BGA, a flexible circuit similar to a TAB circuit is attached to the surface of the IC by means of a semiconductor grade elastomer. This flexible circuit structure forms the basic interposer. Thin flexible ribbon like bond leads of metal such as gold, gold plated copper, gold plated nickel etch are then bonded directly to the gold or aluminum pads on the IC. (See figure 2-22). While the illustration shows a single metal layer construction, micro BGAs can and have been fabricated with two metal layers for power and/or ground distribution and controlled impedance designs for highest level performance.

The elastomer or compliant polymer layer serves to decouple the differential expansion of the silicon and from that of the interconnection substrate. This compliant layer together with the ‘‘S’’ shaped bond lead ribbon effectively decouple the devices from the strains of thermal expansion. The novelty is contained in the ability to create chip size

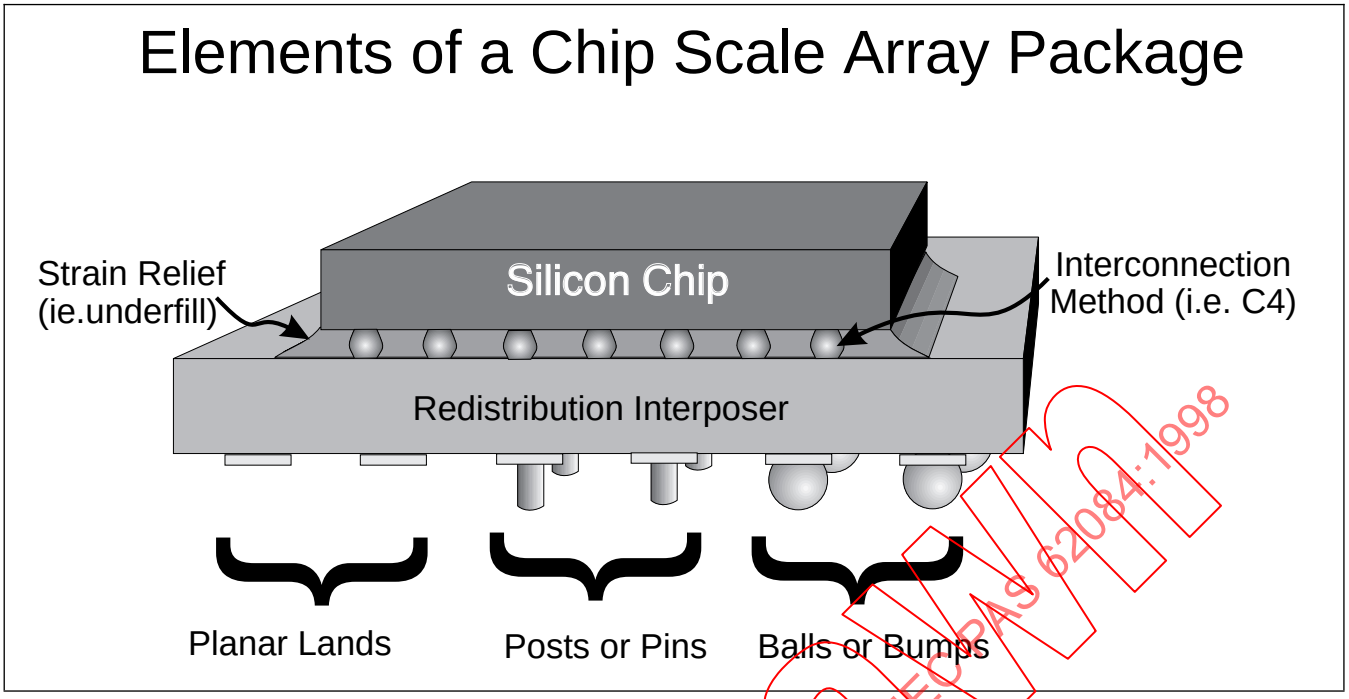


Figure 2−21¬ Example of Chip Scale

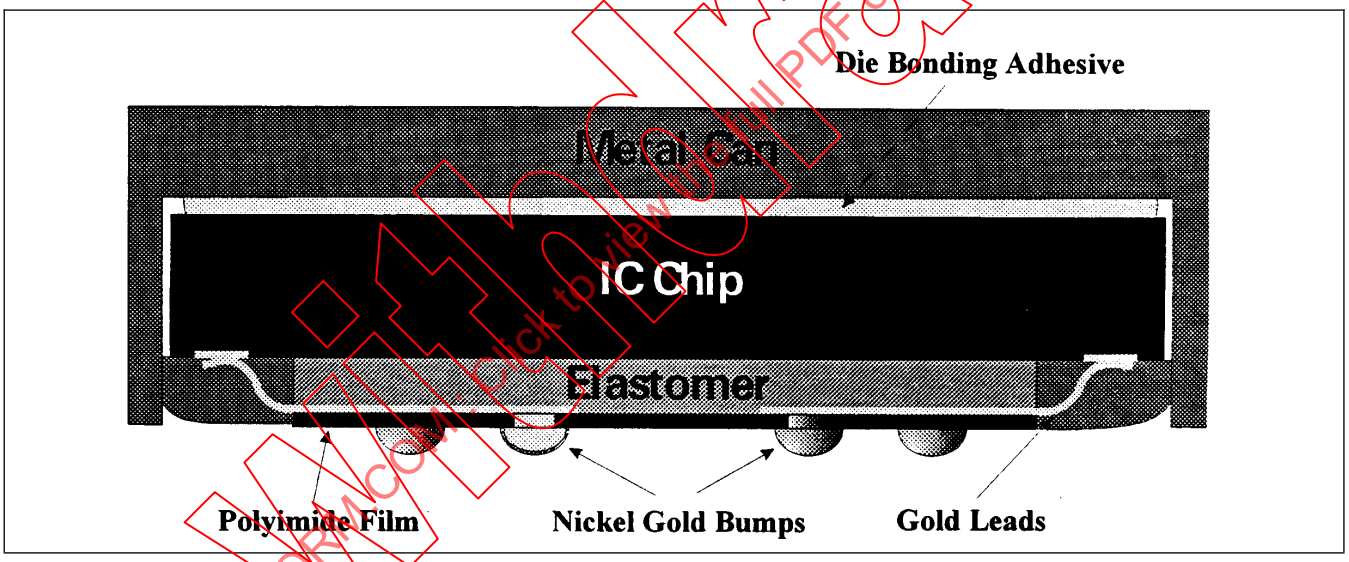


Figure 2−22¬ Micro BGA

packages that are compliant in X, Y, and Z direction, facilitating both test and assembly while enhancing reliability.

The package is interconnected to the substrate by means of either nickel-gold bumps or solder balls. The I/O grid pitch for the micro BGA are typically spaced at bump pitches which are increments of the international Electrotechnical Commission's Publication 97 recommended base pitch of 0.5 mm. Depending on the lead count and bump pitch, the micro BGA area can vary from die size to several times the chip area. Smaller chips may require a fan-in/out approach to layout due to the area required by the bumps.

Finally, the micro BGA is ameable to mass processing at the wafer level, thus making it possible to move the packaging and assembly process to the end of the wafer fab process, affording a very low cost method of packaging ICs.

2.2.1.2 Mini BGA

The mini BGA is very similar to the flip chip described previously, but it utilizes different metals than flip chip (see figure 2-23)

The chip is coated with polyimide resin. Vias are etched into the resin to access the chip's bonding pads. Metal is deposited and patterned from the via to the polimide surface. Solderable balls are attached to the via connection sites.

The lead count of the mini BGA is limited to the chip area

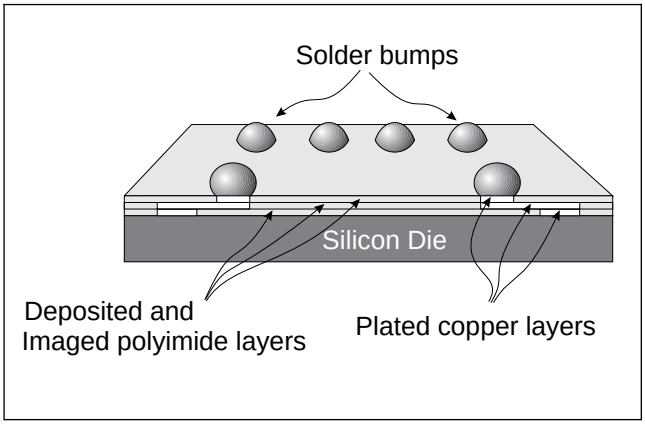


Figure 2–23¬ Mini BGA

and the assembly limits of the bump pitch.

2.2.1.3 SLICC

SLICC stands for Slightly Larger than IC Carrier. Its construction uses flip chip mounting on a rigid organic interposer to create the chip scale package with redistributed I/Os. An underfill epoxy is injected around and under the chip and the bumps. The purpose of the underfill is to constrain the CTE of the organic interposer so that it is close to the low CTE of the silicon chip. The interposer is connected to the PCB via solderable Cu-Ni balls.

As with the micro BGA, the SLICC area is dependent on the number and pitch of the balls. (See figure 2-24)

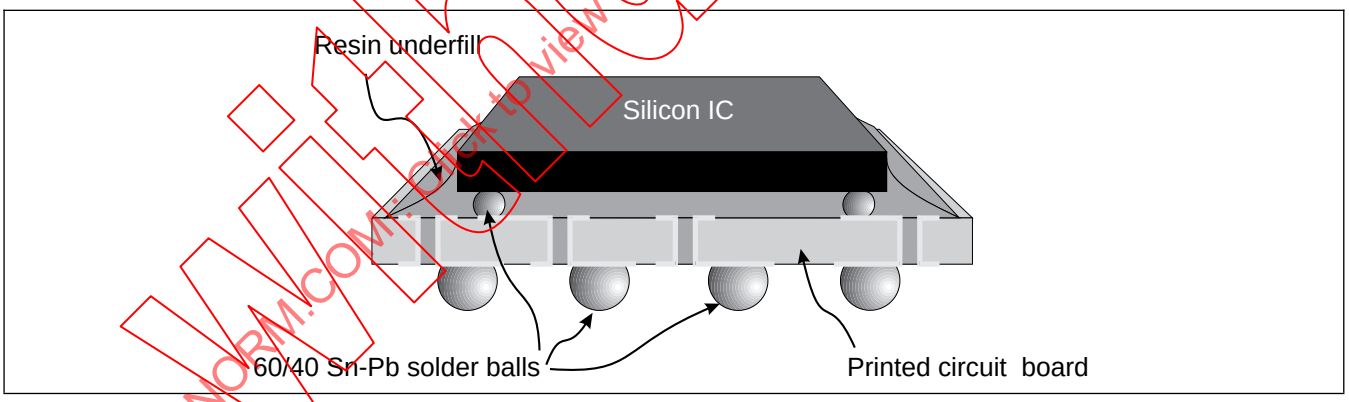


Figure 2–24¬ SLICC Chip Scale Grid Array

2.2.1.4 Chip Scale Package

A flip chip like package called a ‘‘CSP’’ has been reported. This CSP is constructed of an LSI semiconductor chip, a thin plastic resin and electrode bumps.

The package has neither leadframe nor any bonding wires (figure 2-25). The chip surface is insulated from the chip wiring pattern and conductor paths are redistributed on the face of the die and are electrically connected to external electrode bumps (figure 2-26). Since the internal/external pad locations can be adjusted by those wiring conductor patterns, this package does not impose any additional restrictions on LSI designers.

Several process improvements involving material selections and reliability studies have been executed in the package development.

2.2.1.5 Peripheral to Area Array Convertor

The Peripheral to Area Array Converter (PAAC) technology provides standard peripheral array chip pad ICs with redistribution and voltage planes made from organic dielectrics and copper, for low resistivity, and an area array of solder bumps. The PAAC employs a thin film transfer process either at the wafer or bare die level. This allows for yield testing of the completed thin film add-on prior to transfer and minimizes IC exposure to processing. The solder bumps are simultaneously electroformed to make electrical contact from the IC chip pad to the redistribution lines and to produce the area array at the surface. The chip pad contact method is the photolithographic ‘‘Reach Through Via’’ (RTV) made by thin film patterning, dielectric etch and metal deposition processes. (See figure 2-27)

As few as one (simple fan-in redistribution wiring) and as many as four layers may be transfered in a single step. The more complex structure is comprised of intra-chip ‘‘net list’’ redistribution which translates and optimizes the chip I/O to the system interconnect requirements. This significantly reduces the wiring burden and resulting complexity and thereby cost in the inter-chip substrate.

Voltage layers provide for low inductance power supply and impedance control of the interconnect line from the point of the output of the chip driver, to the substrate and into the receivers. An elastomeric mechanical buffer layer at the IC/PAAC interface enables reliability from thermal cycle induced solder connection fatigue failure due to chip and substrate CTE mismatch. A Multi-chip Module or Chip on Board system utilizing PAAC may be reworked by means of solder reflow of the failed IC die. PAAC high volume production costs for bare die transfer are estimated to be comparable to wire bonding. Wafer level transfer significantly reduces cost.

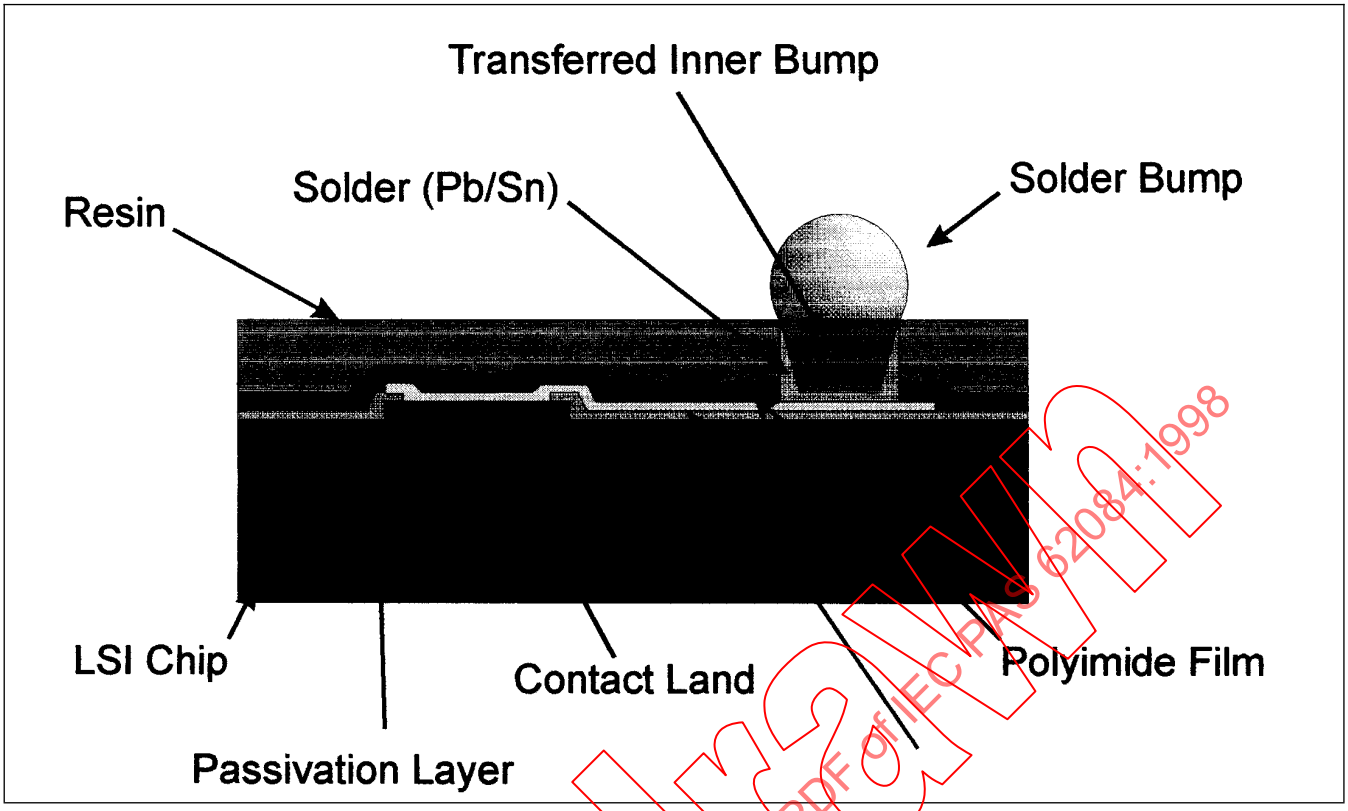


Figure 2–25¬ Chip Scale Package

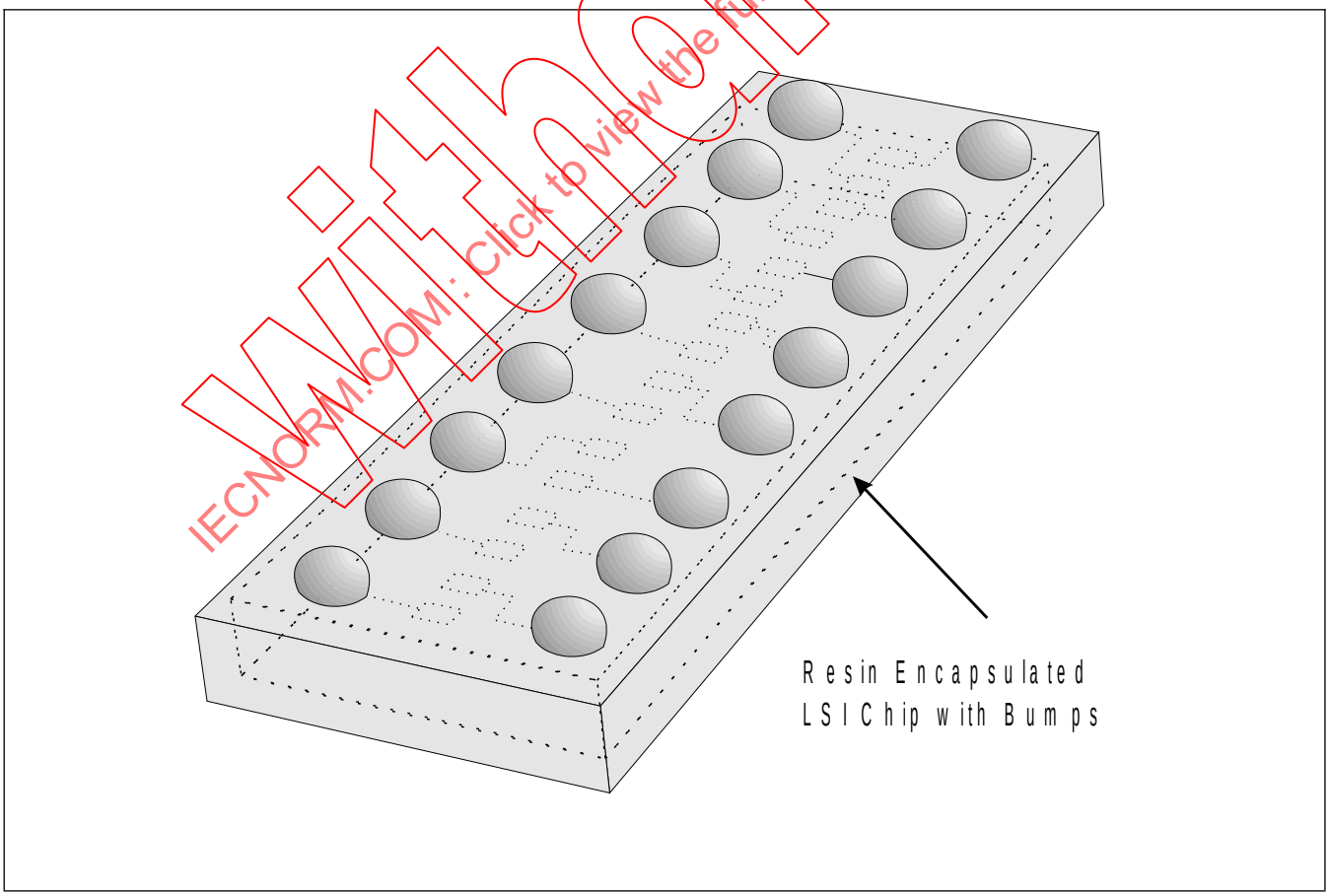


Figure 2–26¬ Resin Encapsulated LSI Chip with Bumps

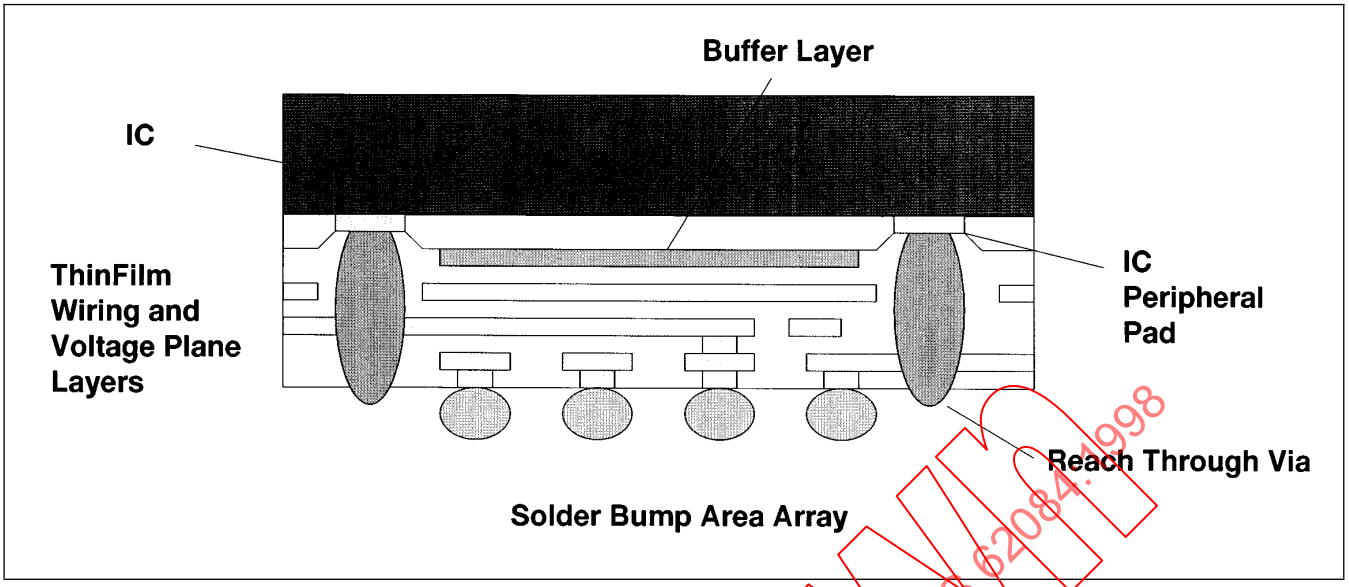


Figure 2–27¬ Peripheral Area Array Converter

2.2.2 Peripheral Leaded Chip Scale Packages (CSP-P)

Peripheral leaded chip scale packages are those that make use of, or replace the existing bonding pads on the semiconductor chip. TAB is one form of peripheral package and qualifies as a chip scale package if the TAB area is within the area defined for chip scale packages. (More information is available on TAB in the document SMC-TR-001 available from the IPC).

Beam lead technology is another chip scale peripheral approach that is still used for microwave devices.

New ideas for chip scale packages have developed. One of these is the Micro SMT or MSMT package. This package consists of metallized silicon or GaAs posts and metal beams. The chip, the posts, and the beams are encapsulated in an epoxy or similar compound.

2.2.2.1 Micro SMT or MSMT Package

The micro SMT or MSMT package evolved from the microwave industry as a replacement for the beam lead package. The MSMT package was invented in 1990. The chip scale grid array packages evolved from the combined concepts of flip chip and ball grid array technologies.

MSMT packages, like flip-chips, are created while the semiconductor chips reside in wafer form. Most chip scale grid arrays are attached to the chip after it has been singulated from the wafer.

The MSMT package is formed by micro-machining the semiconductor, while in wafer form to create posts. The micro-machining starts by depositing and patterning metal from the junction to the top of the post site, forming a beam. An encapsulant covers and protects the chip, posts, and beams. A cut away drawing of a MSMT packated IC is shown in figure 2-28.

The post site may be in the saw lane, or where the bonding pads reside. Locating the posts in the saw lane, illustrated in figure 2-29, adds silicon area to some devices, which may reduce cost savings, size reduction, and gross die per wafer. A more optimal application of the MSMT package is to design the post sites over the bonding pads, as illustrated in figure 2-30.

A cross-sectional view of an MSMT package is shown in figure 2-31, showing the use of a cap. For larger chips, a cap of silicon, glass or ceramic is placed on the epoxy and the wafer. This cap provides an additional heat removal path, and it helps to planarize the posts to within ±3 microns of each other.

The wafer is thinned and flipped over and the remaining silicon that holds the post is etched away. Metals are deposited on the sides and bottom of the post. The solderable metals provide a low resistance path from the bottom of the post through the beam and to the chip.

MSMT posts are located on the periphery of the chip (see figure 2-32). They add little area to the chip. The primary lead count range covered by MSMT packages if from 2 to 100. Higher lead counts are limited primarily by the printed board fabrication and assembly pitch limitation rather then packaging limitations. MSMT posts for 2 to 100 leads are fabricated with areas of 0.1 mm square on 0.30 to 0.65 mm pitch. Package sizes depend on the chip size, but range from 1005 (1 mm by 0.5 mm) to 20 mm by 20 mm.

2.2.2.2 Chip Scale Peripheral Packages

Another form of CSP with peripheral leads has been developed using a planar lead frame. The lead frame provides the redistribution of the chip bonding sites from the corner to the peripheral.

The chip is wire bonded to the lead frame and is then encapsulated to protect the wiring and the active silicon

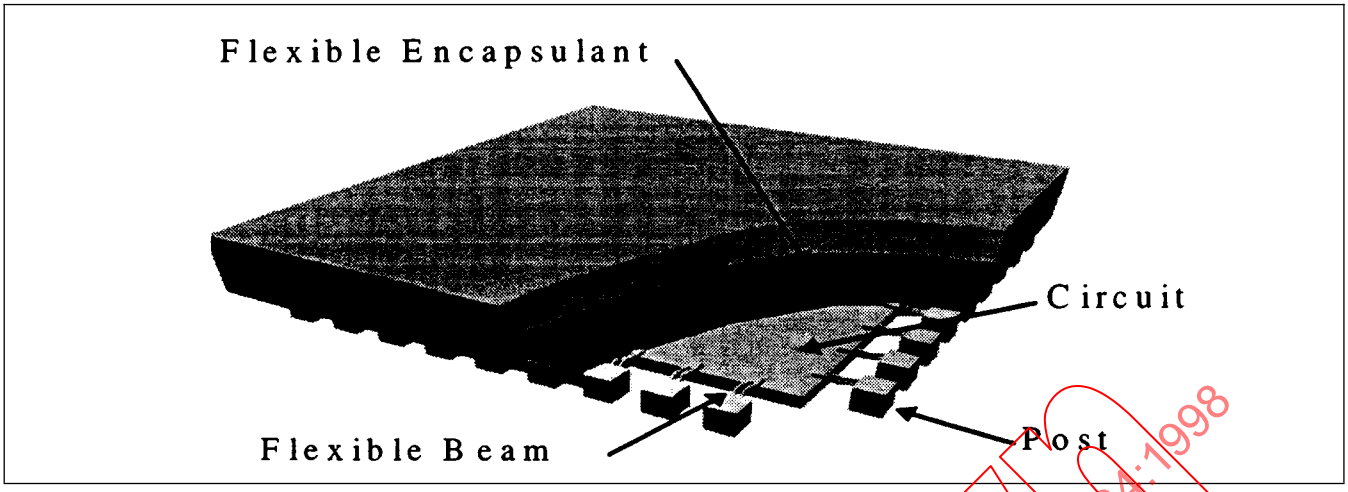


Figure 2–28¬ Cutaway View of an MSMT Packaged IC

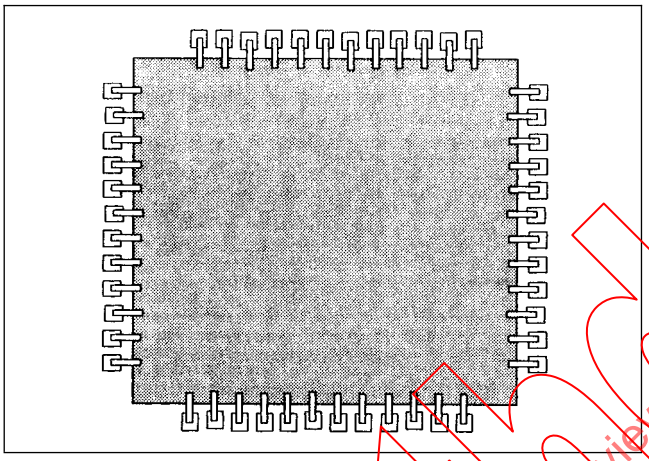


Figure 2–29¬ MSMT Posts in Saw Lane

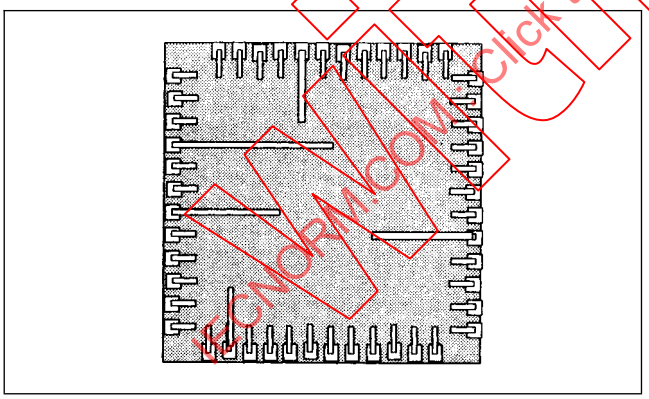


Figure 2–30¬ MSMT Posts in Bonding Pad Area

(see figure 2-33). As in many of the chip scale packages the back of the die is accessable for heat transfer.

2.2.3 Advantages and Disadvantages of Chip Scale Technology

There are many characteristics of flip chip and chip scale technology that need to be compared in order to select the most appropriate package technology. Advantages of flip chip has been stated in 2.1.8. Some of the advantages of chip scale are:

- Packaged semiconductors are very close to the size of unpackaged die
- Packaging the chips allows for easier handling, testing, and assembly than bare die or flip chips
- Chip scale packages assemble to PCBs without the need for additional materials, such an underfill epoxy and processes than currently used for SMT assembly
- Chip scale packages add minimal electrical and thermal parasitics
- Preliminary reliability data for several of the chip scale packages shows results matching current SMT package levels
- Some of the chip scale package technologies produce less expensive packages than conventional SMT packages

There are also several limitations and concerns with Chip Scale Packages. These include:

- Availability of most chip scale packages is very limited
- Reliability data for most technologies is preliminary
- Useability data for all chip scale packages is limited

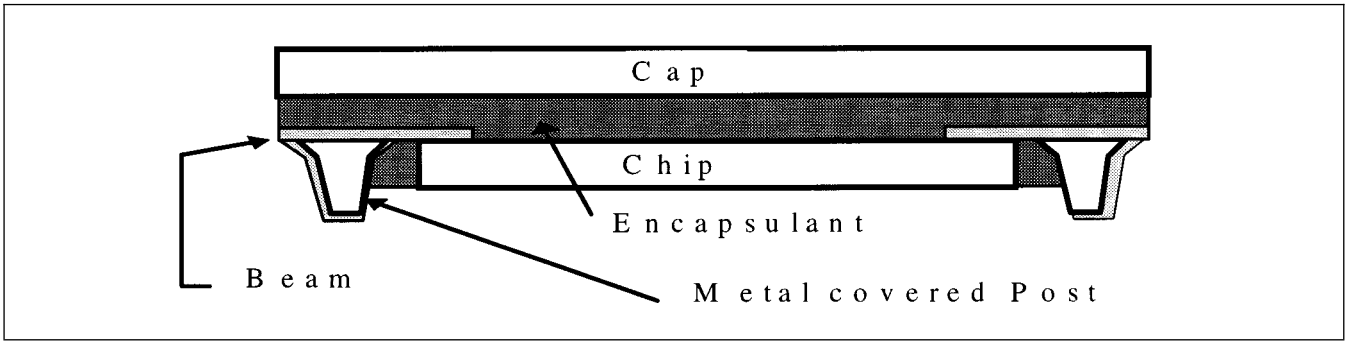


Figure 2−31¬ Cross-Sectional View of an MSMT Package

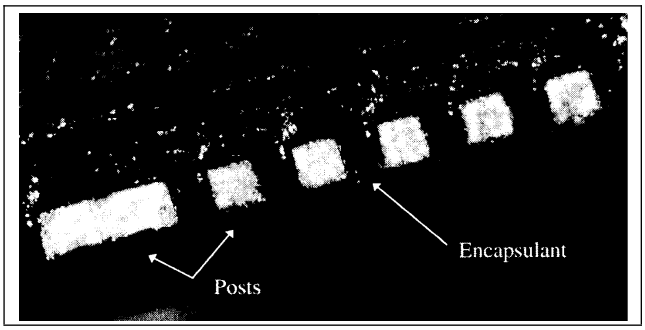


Figure 2−32¬ Close-up Photo of MSMT Posts, Encapsulant and Bottom of the Chip

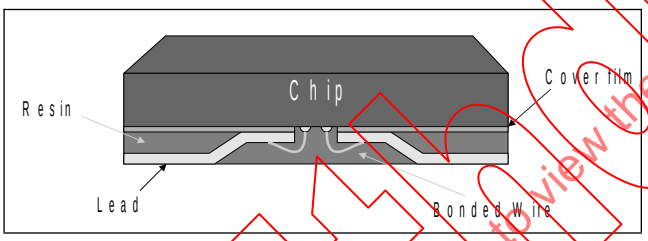


Figure 2−33¬ Chip Scale Peripheral Package

3 APPLICATIONS OF FLIP CHIP AND CHIP SCALE

Over the years there have been many applications of flip chip technology. Chip scale technology is now becoming a viable solution for those companies that prefer a more ruggedized approach to bare chip mounting. Table 3-1 illustrates some of the applications for flip chips.

Table 3-2 provides a comparison of different commonly used technologies including flip chip, micro-BGA and MSMT. Section 4 provides more detailed descriptions of each of the cited chip scale technologies.

Table 3–1¬ Commercial Flip Chip and Chip Scale Applications

Application or Market	Bump Type				Number of I/O's	Organic Substrates	Bonding Method
	Material	Fabrication Method¬	Layout¬	Pitch µm			
Automotive¬	SnPb on Cu InPb	Electroplate, print	Peripheral Array	250- 400¬	10-100¬	Ceramic¬	Mass reflow
Computer¬	SnPb¬	Evaporate, Electroplate	Array¬	210-250¬	500+¬	Ceramic¬	Mass reflow
LEDs, MCMs¬	Au, SnPb¬	Plate¬	Peripheral Array	63.5-100¬	110-126¬	Glass, silicon¬	mass reflow UV cure resin
IC Cards¬	SnPb¬	Electroplate¬	Peripheral¬	140-160¬	100¬	Laminate¬	Mass reflow
Computers,Telecom. (cellular) IC Cards	SnPb¬	Evaporate Electroplate	Array¬	250¬	300-500¬	Laminate¬	Mass reflow
Telecommunications including pagers	SnPb¬	Lift-off evaporation	Peripheral¬	200¬	20-40¬	Ceramic Laminate	Mass reflow
		Print¬	Array¬	150-200¬	8-40¬	Silicon¬	Mass reflow
		Plate or Evaporate	Array¬	225-350¬	84-400¬	Laminate, ceramic	Mass reflow
Power amplifier for cordless phone	Au¬	Electroplate¬	Array, peripheral	Varies¬	<20¬	Ceramic¬	Thermocomp. bond
Watch Module¬	SnPb on Cu¬	Electroplate, Print	Peripheral¬	200¬	13-60 (15 typ)	Laminate¬	Compression bond
LCD¬	Au or Au-plated	Plating¬	Peripheral¬	216¬	10-20¬	Glass¬	Conductive adhesive
LCDs (camcorder view finder/pocket TV	Au¬	Stud bump bond	Peripheral¬	100-140¬	110-126¬	Glass¬	Conductive adhesive
Military (FPA)¬	In¬	Electroplate¬	Array¬	100¬	4096¬	3D Si stack¬	Reflow
Military¬	In, Au, SnPb¬	Evaporate¬	Array¬	50-120¬	100-262, 144¬	Si, saph. cer.¬	Cold weld
Military LCD¬	Au¬	Microballs (LETI-proc.)	Peripheral¬	50-100¬	100-200¬	Si, glass¬	Hot gas

Note:¬ Quite often conductive adhesives are used to connect bumped chips to various substrates .

Table 3–2¬ Comparative Table of Various Technologies for a 100 Lead 10x10 mm Die

Package Parameter	Bare Die (with wire bonds and pads)¬	Flip-chip¬	0.3QFP¬	µBGA¬	MSMT
Pitch, mm¬	**0.15¬**	**0.25¬**	**0.30¬**	**0.50¬**	**0.30**
Footprint Area, mm^2	125¬	120¬	785¬	150¬	110
Package/Chip Area¬	1¬	1¬	6.30¬	1.20¬	0.90
Height, mm¬	0.4 to 0.6¬	0.5 to 0.7¬	1.4 mm¬	0.84¬	0.5 with cap 0.3 without
Inductance nH (cirucit length)	1-2 (0.75 mm wire)	0.1-0.2 (0.5 mm bump)	1 to 7 (0.7 to 3 mm wire and lead)	0.5-2.1 (1 mm bump)	0.1 to 0.2 (0.5 mm post and bridge)
Capacitance pF¬	0.2¬	0.03¬	0.5-1¬	0.05-0.2¬	0.02 to 0.03
PCB Attachment¬	wire bond & epoxy die attach encapsulant	solder & underfill epoxy	solder¬	solder¬	solder

4.0 DESIGN CONSIDERATIONS

Flip chip technology began as an interconnection technology developed by IBM during the 1960's as an alternative to manual wire bonding. In this methodology the chip is attached to circuitry facing the substrate. Solder bumps are deposited onto a wettable chip pad that connects to matching wettable substrate lands as shown in figure 4-1.

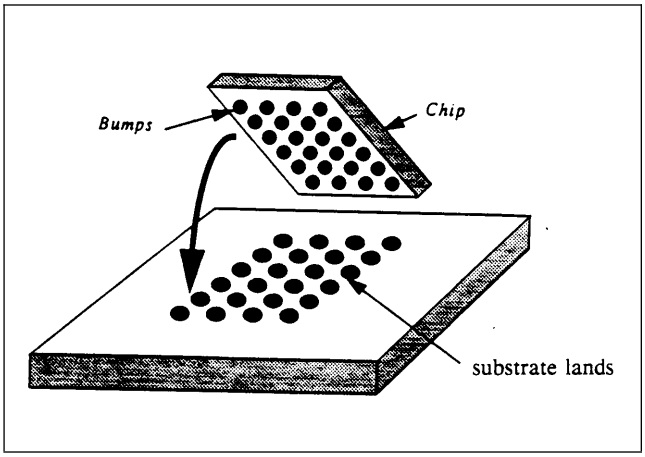


Figure 4–1¬ Flip Chip Connection

Flipped chips are aligned to corresponding substrate metal patterns. Interconnections are formed by reflowing the solder bumps as shown in figure 4-2 simultaneously forming the electrical and mechanical connections. The joining process is self-aligning, i.e., the wetting action of the solder will align the chip bumped pattern to the corresponding substrate lands. This action compensates for chip-to-substrate misalignment incurred during chip placement.

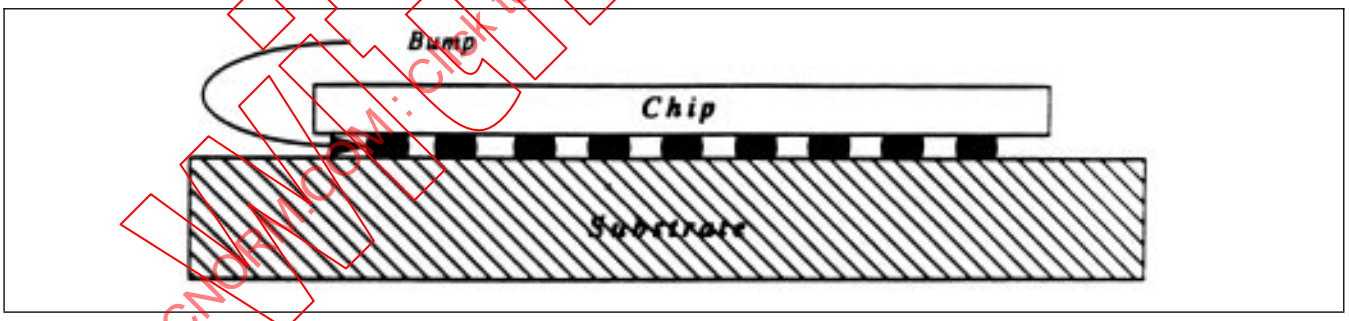


Figure 4–2¬ Mechanical and Electrical Connections

An added feature of the flip chip process is the potential to rework. Several techniques exist that allow the removal and replacement of chips without scrapping the chip or substrate. Depending on substrate material, rework can be performed numerous times without degrading the quality or reliability provided that the mounting substrate can tolerate the rework temperatures.

Injection of chip underfill, as illustrated in figure 4-3, improves reliability most notably in cases of high mechanical stress. It should be noted that currently any rework must be performed prior to the application of chip underfill.

An alternate joining technology to reflowable bump flip chip is one that has low melting solder bumps attached directly to a printed board. The bump on the semiconductor die is not reflowed, instead the lower melting solder on the printed board wets the bump on the die to form the interconnect.

4.1 Chip Size Standardization

Standardizing the size of semiconductor chips may appear to be counter to the objectives of the chip designer and supplier. Many chip suppliers rely on the ability to shrink the size of a chip as it matures. This results in more good chips per wafer. This strategy is sound if the chip is the dominant cost item for the product.

Today many products are dominated by packaging, as opposed to chip cost. Prime examples are discrete diodes and transistors, low lead count logic devices and general purpose analog chips. These products benefit from the cost reduction and performance improvements possible with chip scale packaging.

Chip suppliers and their customers will benefit if when converting their die to chip scale packages they adhere to area size standards. These standards are established by the EIA/JEDEC JC-11 Committee. The committee develops standards for outlines for semi-conductor devices as their counterpart in the EIA Committee develops standards for passives.

Accepting chip size standards will reduce the need for many different tape stocks for the tape and reel packaging of these devices. Standards will also allow users multiple sources for the same device.

4.1.1 Bump Site Standards

The chip scale bump grid array packages give chip designers significant freedom to chose the bump location and the signal type transmitted on the bump. Chip suppliers should observe standards for grid pitch, bump size, and bump location.

Optical assemblers will use placement machines that will place the package based on the edge dimensions. This is because the machine can't see the bump locations. The

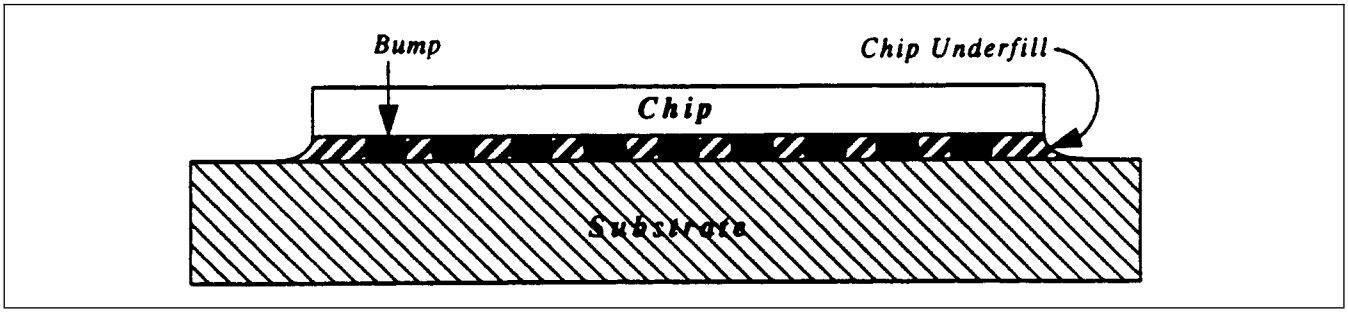


Figure 4–3¬ Joined Chip with Chip Underfill

placement accuracy and assembly yield will depend on the assumed location of the bumps relative to the distance from the edge of the package and the bump grid.

4.1.2 Peripheral Lead Standards

The lead pitch standards for peripherial leaded chip scale packages, such as MSMT should follow the existing standards set by the JEDEC JC-11 Committee. Applicable pitch standards are 0.2 mm, 0.3 mm, 0.4 mm, 0.5 mm, 0.63 mm, 0.65 mm. Refer to JEDEC publication JEP-95 for detailed dimensions.

4.1.3 Package Size Standards

The use of package sizes and tolerances as published by JEDEC in the JEP-95 publication is highly recommended. For smaller sizes, such as those applicable to MSMT and other packages, the use of the size dimensions published in the EIA pubication PDP-100 is recommended. However, be sure to refer to the suppliers package drawings in all cases, where there may be deviations.

4.2 General Considerations for Flip Chips

Flip chip technology encompasses preparation of the chip surface, deposition and shaping of the bumps and joining the bumped chips to a substrate. The main physical attributes consist of four primary components. These are:

Final Metal
Passivation
Metals/Under Bump Metals
Bump

Each component plays a key role and all components work together to form the appropriate connection as shown in figure 4-4. The design of the bump, the land pattern, and the fabrication are essential to forming a strong, reliable and high performance interconnect.

4.2.1 Final Metal

Final Metal is the top layer of metallization (typically aluminum) on a chip. It may be the first layer or the fourth of n^{th} depending on the device complexity. The solder bump connects downward through the Final Metal to the chip. All electrical paths to and from the chip travel through the Final Metal.

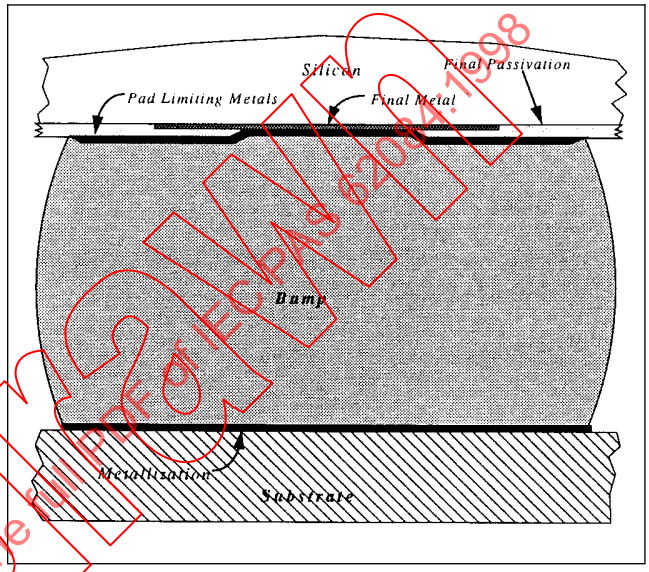


Figure 4–4¬ A Solder Bump Flip Chip Connection

Redistribution, illustrated in figure 4-5, is the relocation of original terminal locations to preferred bump sites. This requires a new metal layer that becomes final metal as illustrated in figure 4-6.

4.2.2 Passivation

Passivation consists of a layer of protective films that coat the Final Metal and chip. These films can vary with semiconductor technology, e.g., CMOS, bipolar, BiCMOS, etc. and with the wafer fabrication facility. These films are deposited and imaged with standard wafer processing techniques.

There are many types of passivation in practice, some of which are silicon nitride, silicon oxy-nitride, spin-on-glass, polyimide, and sputtered SiO_2. Figure 4-7 illustrates passivation for a CMOS product showing cured polyimide covering inorganic chip passivation. It must be noted that sidewall profile is critical.

The cross-section of passivation shown in figure 4-7. Polyimide is recessed from the chip passivation because the same mask is used to pattern the chip and the passivation layer plus different processing is used for each layer.

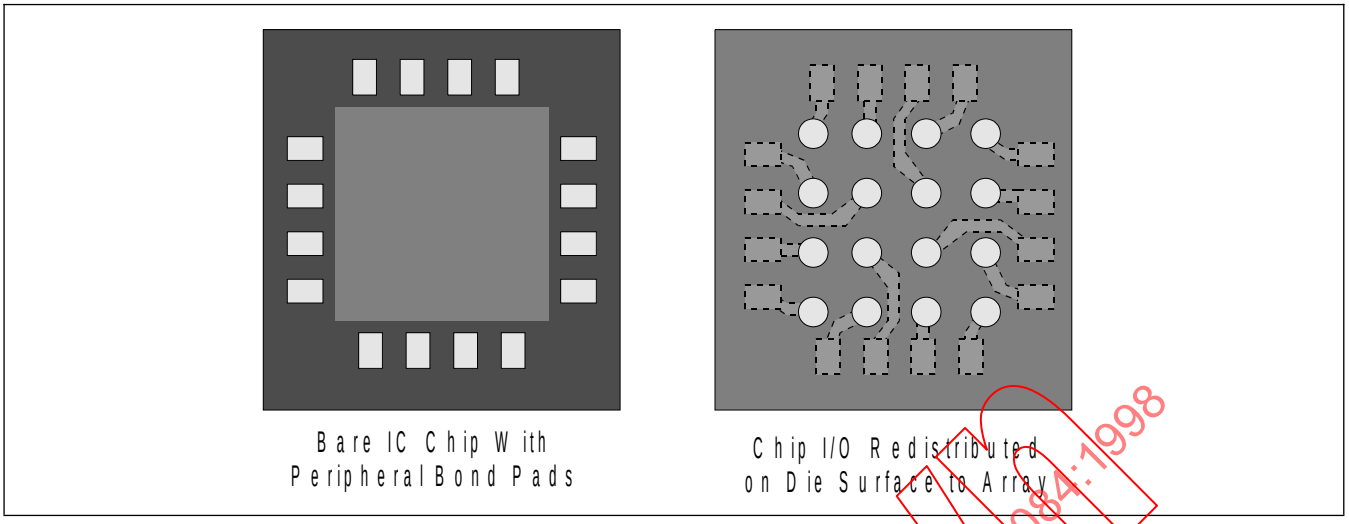


Figure 4–5¬ Two Simple Chips, Showing Original Pad Locations and Rerouted Bumps

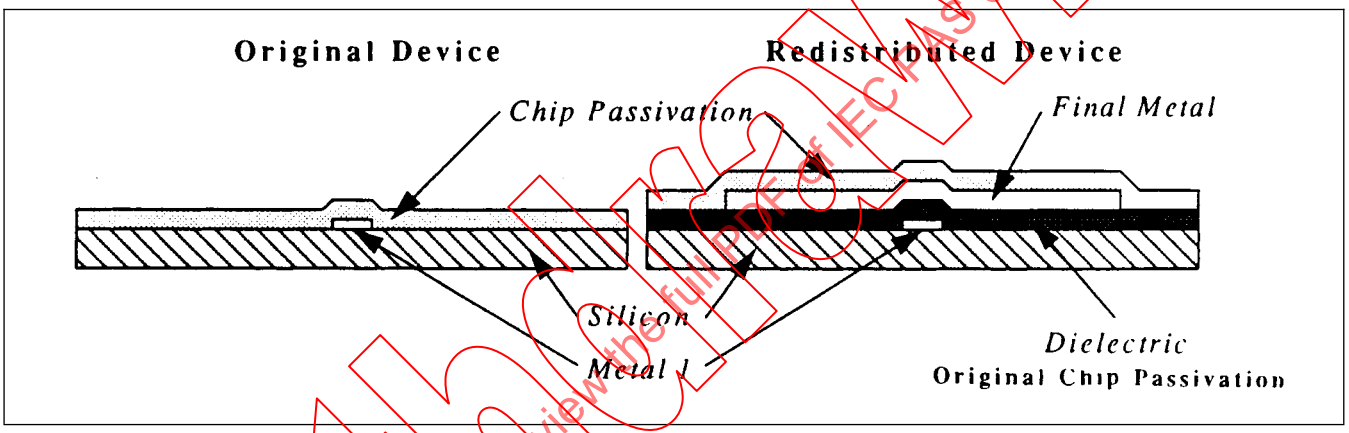


Figure 4–6¬ Redistribution of a Single Metal Layer Device

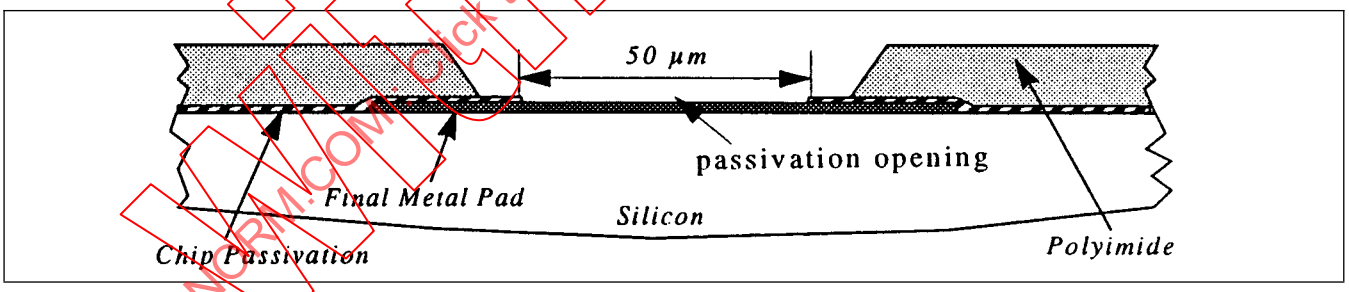


Figure 4–7¬ Passivation (Cross-Section)

Chip passivation provides moisture, ionic and physical protection to the chip, adhesion, and electrical insulation to the pad limiting metal.

Openings in the passivation are centered over the Final Metal pads to allow contact of the pad limiting metal (PLM) to the Final Metal pad as shown in figure 4-8.

4.2.3 Pad Limiting Metals

The Pad Limiting Metal (PLM) is a solder wettable terminal metallurgy that defines the size and area of a soldered connection, limits the flow of the solder bump, and provides adhesion and contact to chip wiring. Requirements of the PLM are:

- Low electrical contact resistance
- Good adhesion to passivation and aluminum
- Reliable diffusion barrier
- A wettable metal
- Good adhesion between PLM layers
- Protection of final metal from the environment

Table 4-1 shows some commonly used PLM systems for each of the desired properties.

PLM systems are deposited after oxide removal of the final metal. Common deposition methods include evaporation and sputtering. Figure 4-9 shows an example of pad limiting metals.

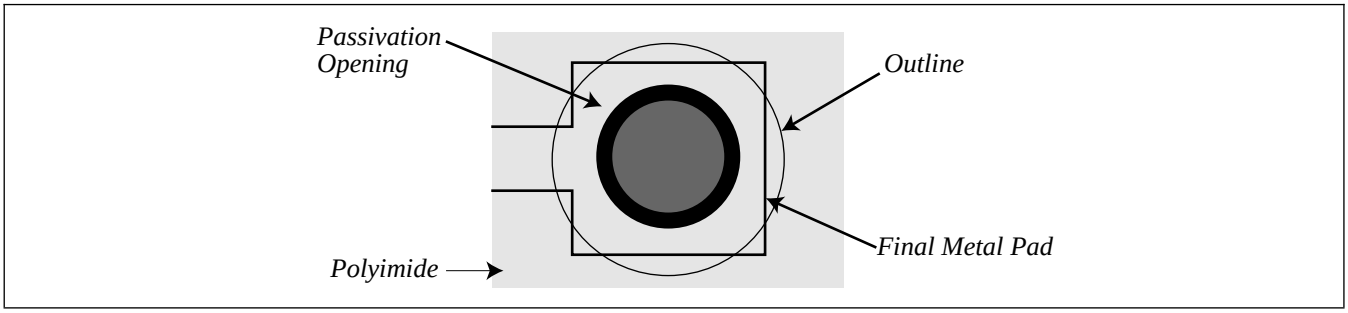


Figure 4–8¬ Schematic Plan View of Bump

Table 4–1¬ Commonly Used PLM Systems

Characteristic¬	Pad Limiting Metal Combinations				
Adhesion¬	Ti¬	Al¬	Cr/Cu¬	TiW¬	TiW
Diffusion¬	Pd¬	Ni¬	Cr¬	TiW¬	TiW
Wettability¬	Au¬	Cu¬	Cu¬	Au¬	Cu
Oxide Preventor¬	—¬	—¬	Au¬	—¬	Au

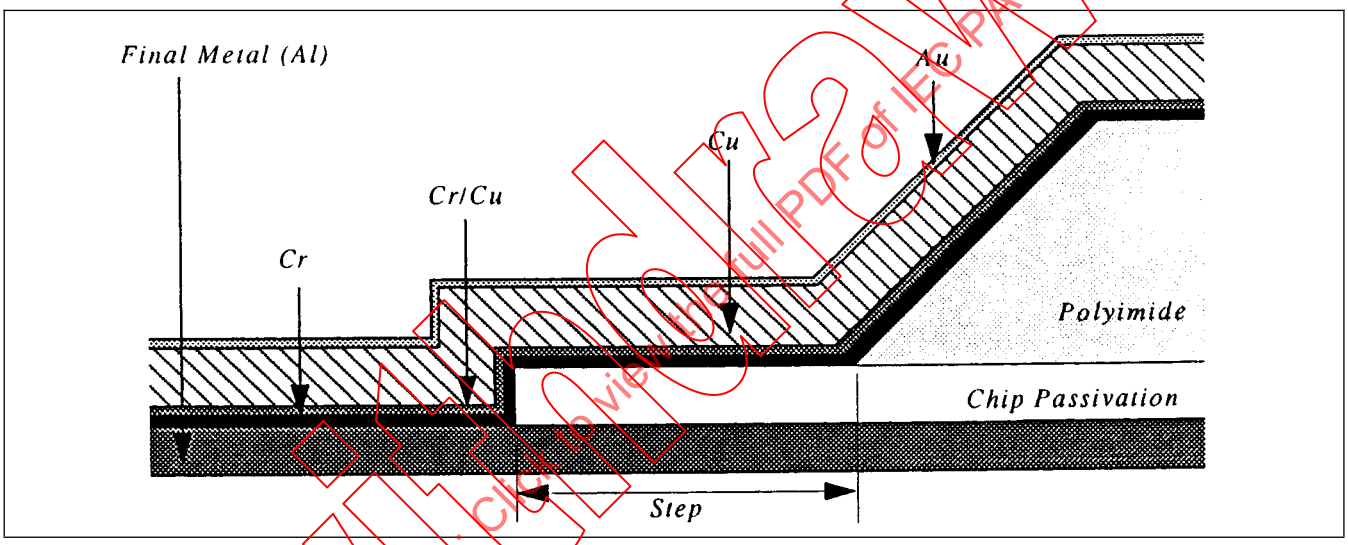


Figure 4–9¬ Example of Pad Limiting Metal

The IBM process requires a sputter clean to prepare the Final Metal. Chrome is deposited as a barrier and adhesion promoter to the aluminum and polyimide. Next, copper and chrome are co-deposited to enhance inter-metal adhesion. Then a thick layer of copper is deposited as the primary wetting surface for the bump. This is followed by a flash of gold to inhibit copper oxidation.

4.2.4 Solder Bump

The solder bump forms the electrical and mechanical bridge between the chip and next level assembly. It absorbs the stress between the chip and next level of assembly caused by variations in their relative thermal expansion rates.

The solder composition of the flip chip bumps varies according to required mechanical and thermal properties. Common bump compositions include:

- 90 PbSn
- 97 PbSn
- 63 SnPb
- 50 InPb
- 95 PbSn

In addition, conductive polymer bumps are also used.

With the IBM process, an evaporated shadow mask or bump mask is used to define the solder bumps. Once the bumps are deposited, the Bump Mask is removed, and the bump resembles a truncated cone as depicted in figure 4-10. A subsequent non-oxidizing reflow process yields the final round C4 Bump illustrated in figure 4-11.

Electrodeposition offers a wider range of SnPb composition than evaporation, and thus offers greater control over mechanical and thermal properties. Electrodeposition also produces smaller diameter bumps and is more difficult to control uniformity.

A variety of solder paste compositions can be screen printed. Delco Electronics printed InPb which combines

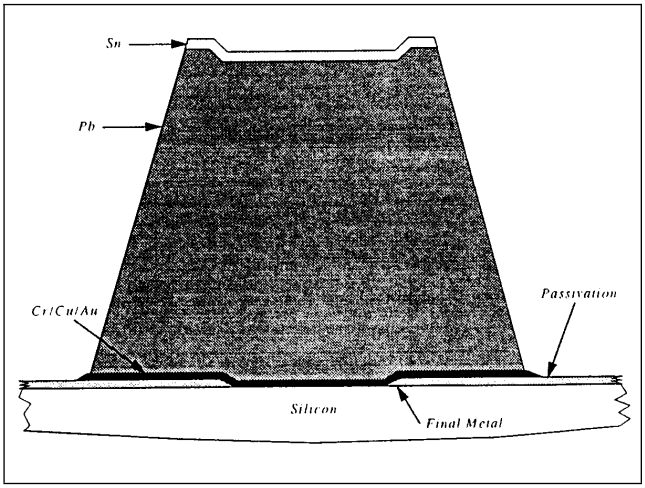


Figure 4−10¬ Initial C4 Bump

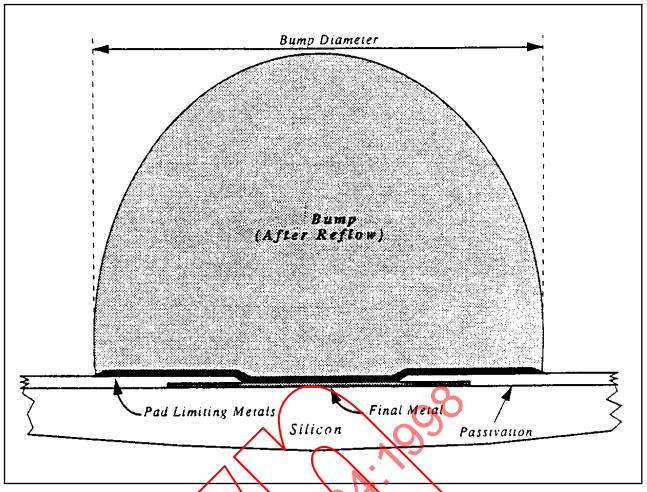


Figure 4−11¬ A C4 Bump After Reflow

lower processing temperatures and provides improved compliance and fatigue life.

Common bump diameters and minimum pitches are shown in Table 4-2. Minimum pitch requirements are based on radial distances (center to center) between bumps. Bump diameters are measured at the widest point of the ball.

Table 4−2¬ C4 Bump Diameter and Minimum Pitch Options

Deposition Types¬	Diameter (µm)	Minimum Pitch (µm)
(A) Evaporation or Electrodepositio	150 µ 125 100	300/350 250 225
(B) Screen Printing	200 150 125	375 250* 200*

*peripheral only

It should be noted that 250 µm pitch is the most commonly used for the evaporative and electrodeposited bump process, due to its ease of manufacture and smaller pitch capability; 225 m pitch is used for very high interconnect density applications. Generally the larger bumps provide improved reliability.

Today, for flip chip applications, minimum pitch requirements are larger to account for printed board wiring capabilities. An example of typical DCA bump design is shown in figure 4-12.

4.2.5 Existing Chip Designs

To take complete advantage of flip chip technology, the chip designer must comprehend substrate and system requirements. However, the bumping process may be implemented on existing chips without circuitry redesign. An advantage of wafer fabrication is that wafers can be built to the point where they can be split for bumping, or packaging.

If the existing perimeter wirebond or TAB pads do not violate the minimum bumping ground rules for pitch and location, bumps can be located on or near these existing sites. If the existing perimeter pads violate the minimum ground rules, an extra layer of redistribution metal (final metal covered with additional passivation) will be required.

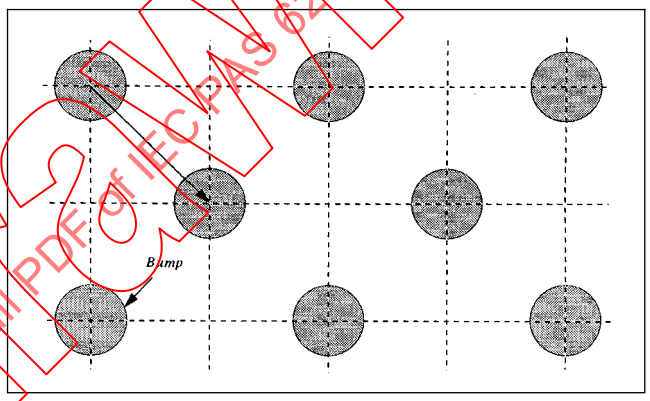


Figure 4−12¬ Recommended DCA Grid Pitch (250 µm Grid, 150 µm Bumps)

In either case, a wire bond or TAB chip can be converted to a bumped or posted chip. This is appealing to those trying to implement flip chip or reduce module to package area but cannot afford the time and/or cost to design of a new chip.

4.2.6 I/O Capability

With the chip bumping process nearly the entire chip surface can be utilized for interconnect pad locations. In fact, over 2500 bumps on a chip has been demonstrated and chips with over 1500 bumps are in production.

Chip bumping enables increased interconnect density. Signal, clock and power connections can be placed almost anywhere on the chip and redundancy means distributions can be optimized for minimum noise and skew, current density and line length. Additionally, on-chip wiring can be reduced since the z-axis escapes are available where needed.

Figure 4-13 compares single row wire bond and bump chips. Each chip is 8 mm square. Wire bond pad size is 76 µm with pads on 100 µm centers. Bump size is 100 µm bumps on a 230 µm center. In this example, interconnect density is increased over 2.4X using bumping technology.

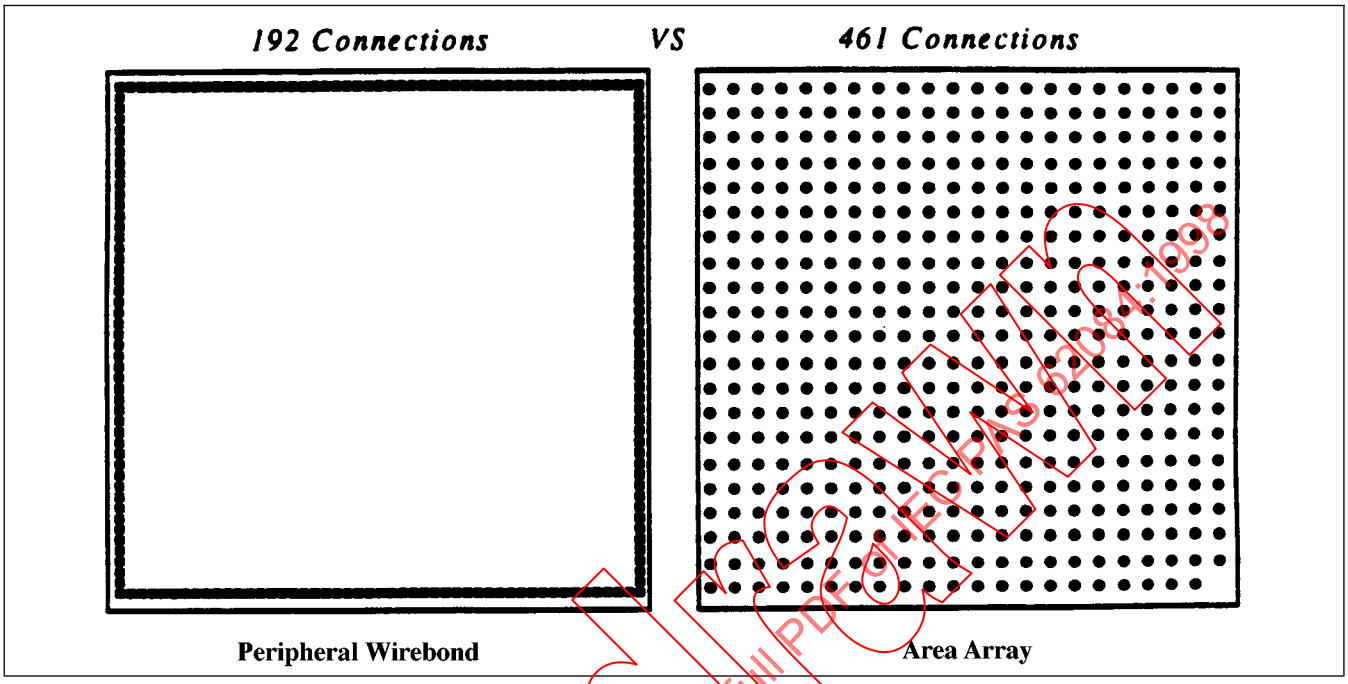


Figure 4–13¬ Interconnect Density (Peripheral vs. Area Array)

4.2.7 Alpha Particle Emissions (Soft Errors)

Alpha particles may cause soft errors in semiconductor devices. Circuit sensitivity to alpha particle emissions and the allowed soft error rate must be clearly understood for each chip design.

Lead contains traces of thorium and daughter elements, e.g., polonium, which emit alpha particles. Circuits resistant to soft errors are designed to withstand alpha particle emission energies.

Circuits sensitive to soft errors are DRAMs, closely followed by SRAMs and some flip-flop circuitry. Sensitivity increases as chip geometries shrink and device (node) critical charge diminishes.

The primary concern of the particle emission is the energy that can be imparted to the device circuitry. This is quantified and illustrated in figure 4-14.

- Alpha Particle = He nucleus = 2 protons + 2 neutrons
- Up to 8.8 MeV of kinetic energy
- Energy converted to electron/hole pairs (3.6 eV per E/H Pair)
- Highest charge density near end of emission track.

The resulting effect of an alpha particle can be a distortion of the depletion region's electric field near the alpha particle track as shown in figure 4-15.

Many materials used in semiconductor packaging emit low levels of alpha particles. This is illustrated in table 4-3.

Emission rates vary for solder and chip underfill depending on the material source. Certain solders may emit a higher rate of alpha particles relative to other materials. These emissions may not be a problem depending on the device circuitry, sensitivity and the allowable soft error rate.

Table 4–3¬ Alpha Particle Emissions of Semiconductor Materials

Material¬	Activity [a/(cm^2-hr)]
High Lead Solder (3/97)¬	0.05 - 10.0
Alumina¬	0.1
Chip Underfill (cured)¬	0.002 - 0.020
Plastic¬	0.04
Silicon Wafer¬	<0.004

If sensitive circuits are present, their proximity to a bump, the bump diameter, the chip structure, and the alpha activity rate must be considered in order to estimate the soft error rate.

Because distance and materials attenuate alpha particles, emission effects are localized to a bump. Hence, only the bump near the sensitive circuits warrants concern. The total number of bumps on a chip does not necessarily increase or decrease the risk to a sensitive circuit. Consequently, chips utilizing bump solder must meet Guide 2 as shown in 4.4.2.

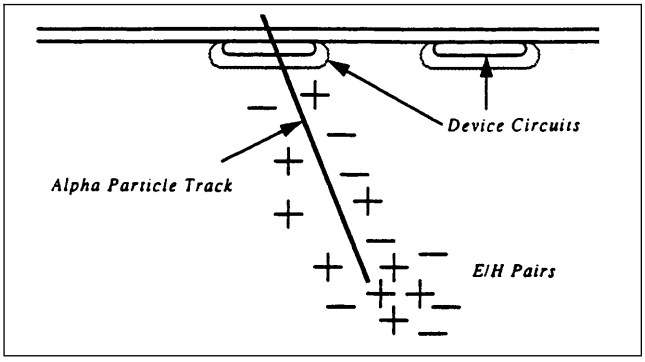


Figure 4–14¬ Alpha Particle Emission Track and E/H Pairs

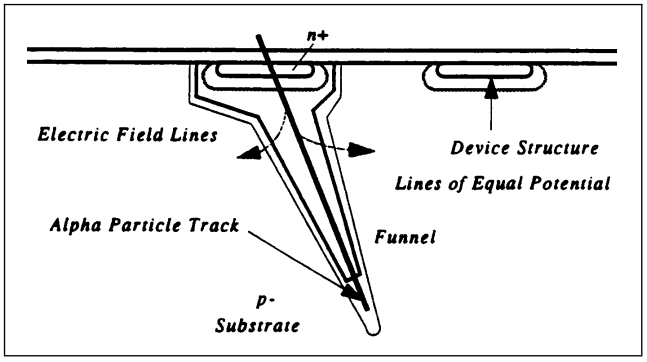


Figure 4–15¬ Distortion of Depletion by Alpha Particles

4.2.8 Edge Seal Design

The chip edge seal is typically associated with CMOS devices. Edge seal considerations resulting from passivation and Final Metal design must meet existing chip design rules. Other than noted, solder bumps do not alter or supersede any existing chip design rules for chip edge closure.

Table 4-4 defines the typical dimensions for chip edge seal. As illustrated in figure 4-16, the polyimide seal is drawn back from the chip edge seal. The cross-sectional view of these chip edge enclosures relates to those detailed dimensions shown in Table 4-3.

scribe is a ratio required by the particular wafer fabrication. Although the same reticle/mask is used for the chip passivation and polyimide layers, due to etch bias, the polyimide edge seal is recessed by 7µm (typical).

4.3 General Consideration for Chip Scale

Developments in flip chip bumping processes have lead to the development of concepts identified as Chip Scale interconnection. Chip Scale Packages are extensions of that concept that have been enhanced to provide more robust attachment technology.

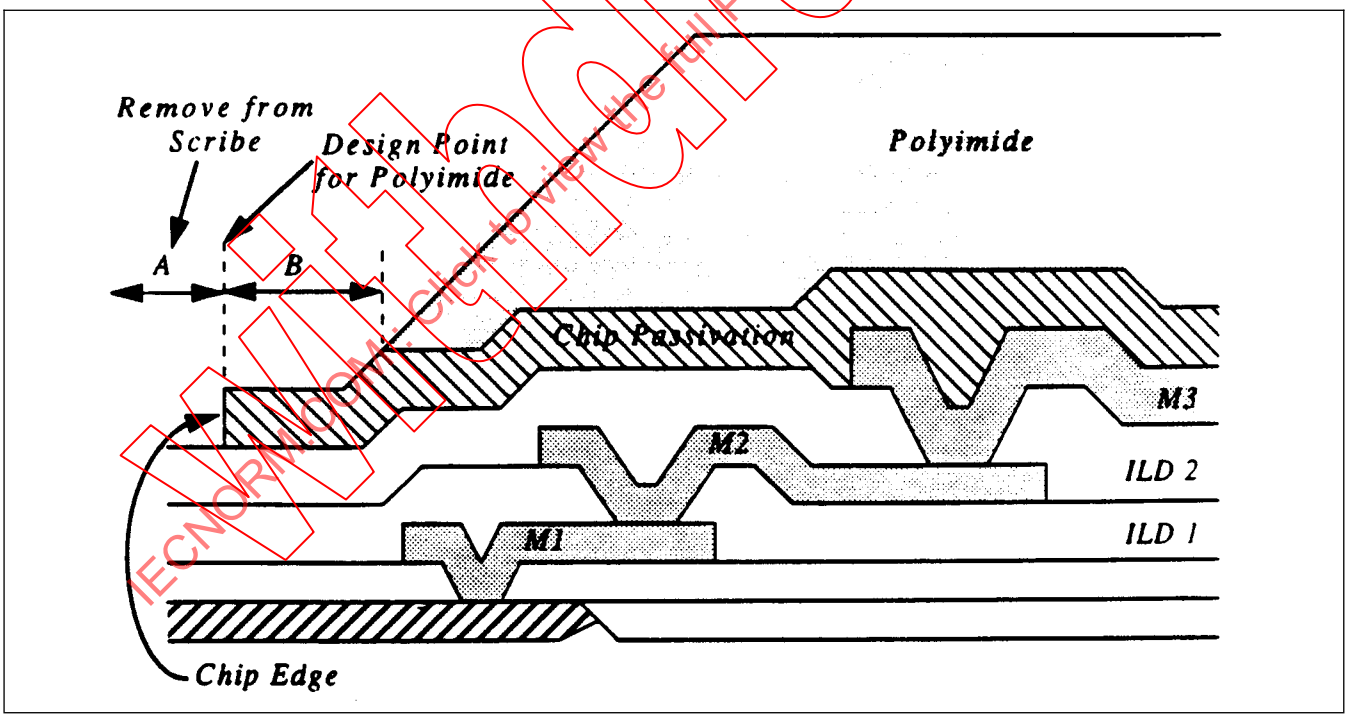


Figure 4–16¬ Chip Edge and Polyimid Seal

Table 4–4¬ Chip Edge Seal Dimensions (Typical)

Description¬	Dimension¬	Width
Chip Passivation Edge Seal	A¬	A.R
Polyimide Edge Seal¬	B¬	7 µm

It should be noted that edge seal design width to the kerf/

4.3.1 Chip Scale Grid Arrays (CSP-A)

Chip Scale Grid Arrays are produced by a few different fabrication methods.

The CSP-A configurations developed to date include:

- Micro BGAs
- Mini BGAs

- SLICC
- Other chip scale packages

Each has its unique method of manufacture, however, most of the devices contain an interconnection method for the chip on the top side in an array format similar to flip chip.

In the case of µBGA and SLICC a flexible or rigid interposer serves at the redistribution dielectric. In the case of mini BGA and CSPs the dielectric is a polyimide which is deposited on the active part of the chip.

4.3.2 Peripheral Leaded Chip Scale Packages (CSP-P)

One of the new chip scale package types is the Micro SMT (MSMT). This package consists of metallized silicon posts and metal beams. The chip is encapsulated in an epoxy or similar compound. A second CSP-S is the system that uses a lead frame described in 2.2.2.2.

The major difference between the two peripheral packaging systems is that the MSMT is accomplished at the wafer level.

The MSMT process utilizes four mask steps that are applied to the wafer after the bondable metal is complete and covered with a passivation, such as silicon nitride or silicon dioxide. The process sequence, including the masking steps is:

1. Deposit the Beam metals
2. Mask the Beam metal to form the Beams
3. Apply the top trench mask and etch the trench
4. Apply epoxy and centrifuge
5. Apply cap, if required
6. Thin wafer and cap
7. Mask the bottom Post forming trench and etch the trench
8. Deposit Post metal
9. Mask the Post metal
10. Singulate the chips

To summarize the masks used to form an MSMT package, the masks include:

1. The Beam mask
2. Top trench mask
3. Bottom Post forming mask
4. Post metal mask

The Posts of the MSMT package can be designed in a variety of configurations. These include face-wrapped, inside-wrapped, side wrapped, and non-wrapped posts. The wrap style determines the silicon area needed to fabricate the post as shown in figure 4-17.

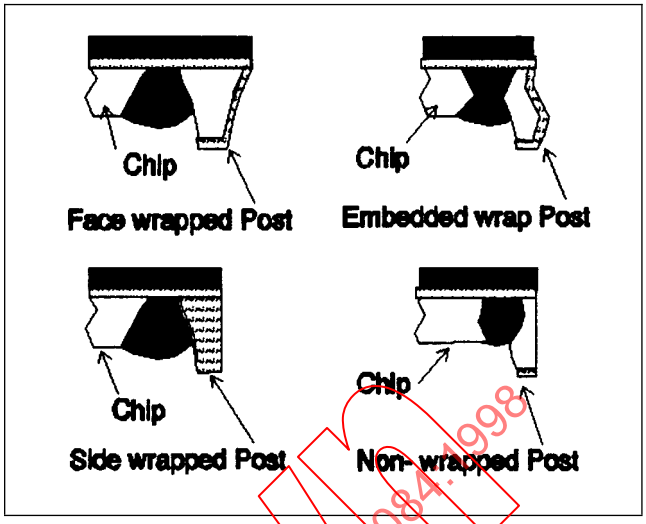


Figure 4–17¬ MSMT Post Configurations

4.4 Substrate Structure Standard Grid Evolution

More recent entries into the arena of substrates designed to meet the demanding requirements of flip chip and chip scale technologies appear well suited to the task. These technologies allow for the construction of shortest path routing for the IC device. Typically based on the concept of standard grids, laminated substrates will allow for the production of inexpensive, high performance systems.

These technologies are fundamentally elegant in concept and are predicated on the notion that standard grids will be required to create economically tomorrows' most advanced systems. The grid pitch that appears most attractive is the one forwarded by the International Electrotechnical Commission (IEC) in their Publication 97.

The pertinent citation from the IEC publication is as follows:

5.0 Preferred Grid System

5.1 For positioning connections on a printed circuit board, a grid with a nominal spacing in the two directions of 0.5 mm shall be used.

5.2 Where a grid with a nominal spacing of 0.5 mm is not adequate, a grid with a nominal spacing in the two directions of 0.05 mm shall be used.

This seemingly simple standard will ultimately prove to be extremely important because it facilitates implementation of common grid systems that will be necessary for future high performance systems. This was recognized by the IEEE's Computer Society Technical Committee on Packaging. The committee appointed a special task force in the fall of 1990 "to seek early consensus on the need to standardize MCM sizes and to propose some possible sizes." The task force recommended 0.5 mm center lines for peripherally leaded packages and PGA standards for area array packages.

Thus IEC-97 with a base pitch of 0.5 mm with a subdivision of 0.05 mm would replace the familiar 0.100″, 0.050″, 0.025″, and 0.005″ grid that has been outdated by the wholesale move to metric based measurement systems for electronics. A common grid will facilitate common design

Table 4–5¬ Design rules for substrates for Chip Scale Technology

Feature	0.5 mm Grid (2 lines between pads)	1.0 mm Grid (2 lines between pads)	1.5 mm Grid (2 lines between pads)
Line Width (L)¬	50 µm (0.002″)¬	125 µm (0.005″)¬	200 µm (0.008″)
Space (Line to Line) (S)¬	75 µm (0.003″)¬	175 µm (0.007″)¬	300 µm (0.012″)
Space (Line to Pad) (SC)¬	50 µm (0.002″)¬	125 µm (0.005″)¬	200 µm (0.08″)
Hole Size (H)¬	125 µm (0.005″)¬	200 µm (0.008″)¬	250 µm (0.010)
Pad Size (P)¬	225 µm (0.0009″)¬	326 µm (0.013″)¬	400 µm (0.026″)
Wire Routability¬	40 cm/cm^2 (100 in/in^2)¬	20 cm/cm^2 (50 in/in^2)¬	13 cm/cm^2 (33 in/in^2)

rules for routing. Table 4-5 illustrates the design rule concept.

An example of a structure that supports the standard grid concept is shown in the figure 4-18. The use of interposers and substrates designed on a common grid allows for the construction of substrates that can offer ‘‘Manhattan routing’’ of signals. This eliminates the need for redistribution wiring, which normally consumes large amounts of valuable board real estate while limiting higher performance opportunities.

4.4.1 Footprint Design

A flip chip or chip scale footprint design is the arrangement of bumps on the chip surface. When laying out the array of bumps, forethought and planning are required. Bump footprints can be arranged in peripheral, array, or interstitial array formats. Examples of these are shown in figure 4-19.

The size and population of the bumps affects the attachment reliability. The use of the design guide checklist shown in 4.4.2 will minimize reliability problems.

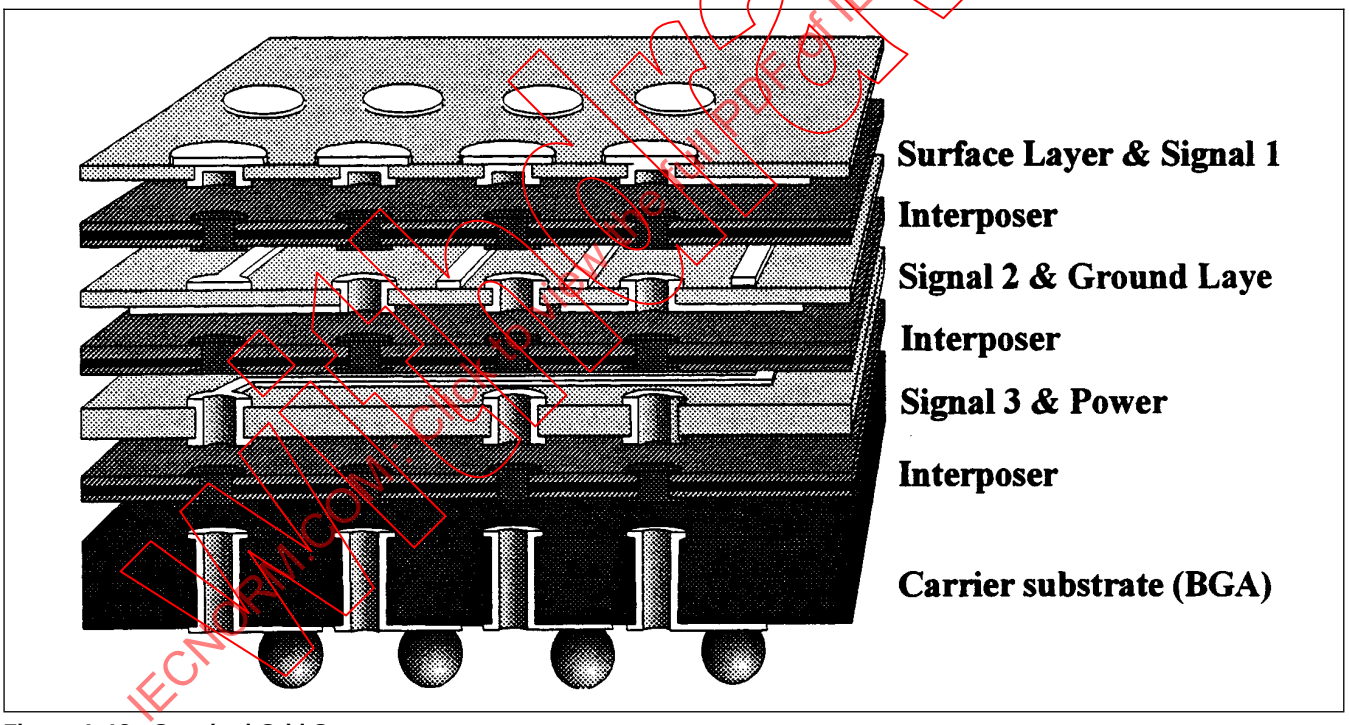


Figure 4–18¬ Standard Grid Structure

In the illustrated format the interposer joins the innerlayers and interconnects them in a single step process. Because the inner layers are thin, the plated through-holes have very small aspect ratios for plating and are relatively easy to produce. These innerlayers can then be electrically tested before committing them to the finally assembled laminate stack, thus providing greater assurance of a high yield.

4.4.2 Design Guide Checklist

Various design guidelines are applicable in the development of the chip, the CSP, or the attachment process. The guidelines in some instances relate directly to the chip manufacturing conditions. In other instances they also apply to chip scale I/Os. In any event, the guidelines are for very specific reasons and designers should consider the following issues:

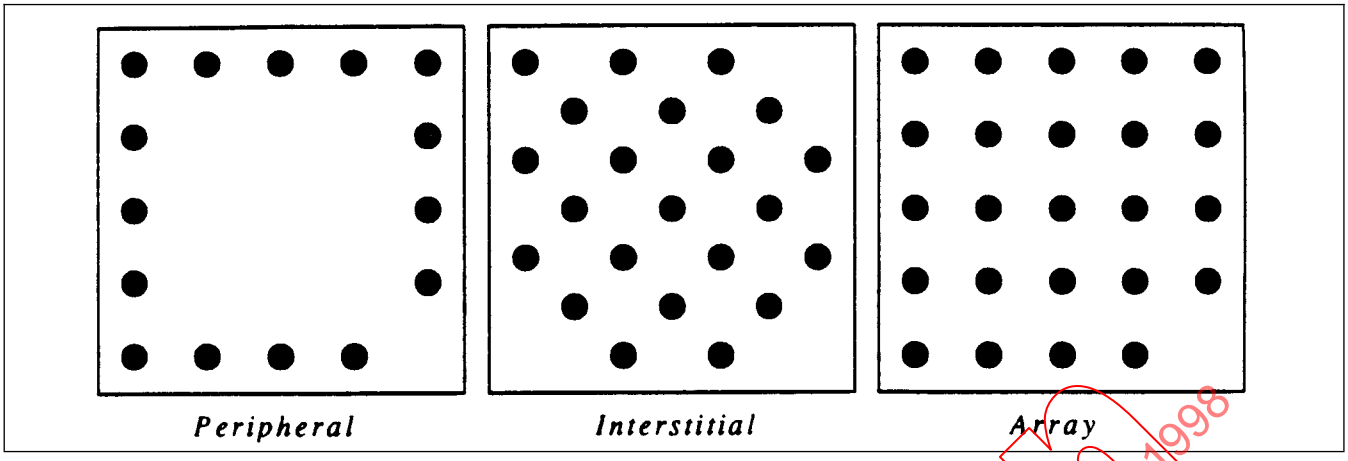


Figure 4–19¬ Bump Footprint Planning

Guide 1¬ Operating temperature must be considered when determining interconnect reliability. Devices with underfill may operate reliably at higher temperature than devices without.

Reason¬ Solder creep at elevated temperatures is not well characterized. Applications will require evaluation and testing at operating temperature to determine expected interconnect reliability. There are five basic thermal and mechanical factors that affect solder bump fatigue. These are:

- Strain
- Thermal Activation
- Thermal cycling frequency
- Bump footprint
- Chip underfill (once underfilled, the performance of the underfill is the key, not the solder)

Guide 2¬ Solder and other alpha emitting materials must be kept an appropriate distance from alpha particle sensitive chip circuitry. For example, a separation of 150 µm is required between C4 bumps (97 PbSn) and sensitive circuitry on polyimide passivated devices.

Reason¬ Alpha particle emission rate of various materials may be too high for reliable operation of alpha sensitive chip circuitry within close proximity. Measures such as thicker passivations or metal shields can mitigate the effects of proximity.

Guide 3¬ Redistribution designs must minimize the inductance and resistance to reduce sensitivity to ESD.

Reason¬ Increased inductance and resistance of the redistribution line reduces the effectiveness of ESD circuits.

Guide 4¬ Maximum current should be modeled for the interconnect system used.

Reason¬ Current density and therefore reliability, varies with materials, construction, cross-sectional area, operating temperature, etc.

Guide 5¬ Designs should conform to all applicable chip design and wafer fab specifications.

Reason¬ Design for manufacturability.

Guide 6¬ The bump footprint cannot be symmetrical with respect to the X or Y axis unless there are other features to define orientation.

Reason¬ To identify chip orientation for subsequent processing.

Guide 7¬ At least two bumps are required on each edge of a chip positioned near the chip corner. Dummy Bumps can be used to satisfy this requirement.

Reason¬ Needed for visual inspection and mechanical support of the chip required during joining to maintain parallelism to the substrate.

Guide 8¬ It is recommended that bump diameters conform to the mandates of the chosen bump pitch to assure reliable connection without risking shorting between bumps. Some desirable bump sizes are 100, 125, 150, and 250 µm.

Reason¬ Conforming with individual suppliers' processes and anticipated industry standards (e.g., ANSI, JEDEC, etc.)

Guide 9¬ Some minimum number of bumps may be required for manufacturability and reliability.

Reason¬ To withstand dynamic and steady state mechanical forces.

Guide 10 In most applications only one bump diameter per chip/wafer should be used.

Reason¬ Varying bump diameter may impair chip attach. Specialized applications for multiple bump sizes per chip/wafer do exist.

Guide 11 The minimum bump pitch should be held at 250 µm, if possible to facilitate the production of standardized substrates. (Multiple minimum pitches will cause confusion).

Reason¬ Minimum pitches prevent bump to bump, bump-to-trace, and trace-to-trace shorting and provide the distance needed to effectively clean chip joining flux (if applicable).

Guide 12 Location of bumps near areas requiring visual recognition will depend on final bump size as illustrated in figure 4-20.

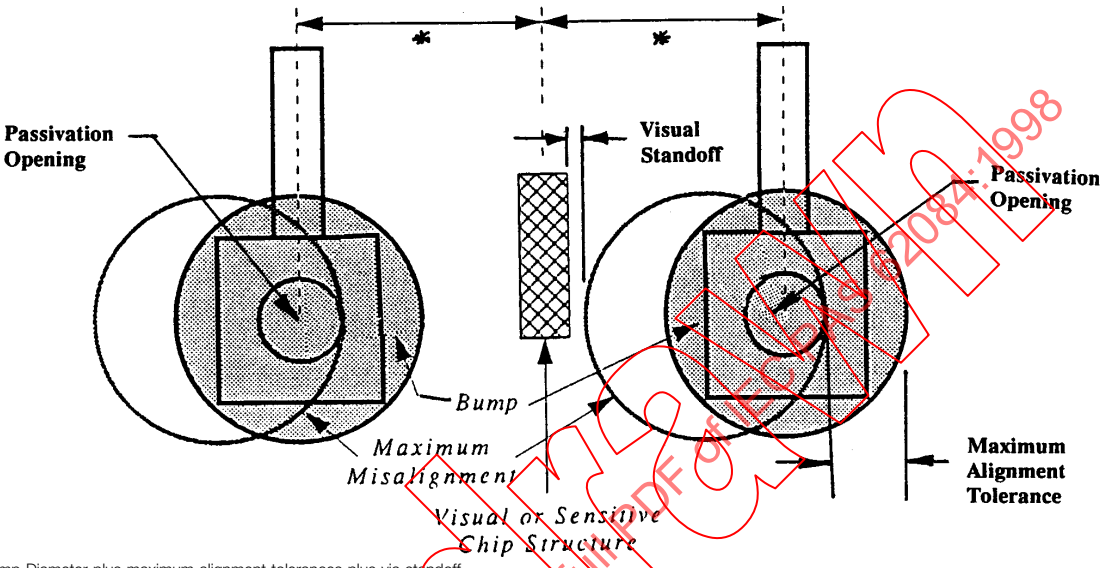


Figure 4–20¬ Alignment to Visual/Sensitive Chip Structures

Reason¬ Allows for alignment tolerance such that misaligned bumps do not block or cover areas needed for visual or optical recognition.

Guide 13 The minimum distance from the center of the passivation opening to the passivation seal at the edge of the chip is 1/2 the final bump diameter, plus the maximum alignment tolerance, plus the desired minimum standoff between bump and chip seal. (figure 4-21)

Reason¬ Minimum pitch required for Bump alignment tolerance.

Guide 14 Every bump must be connected to a corresponding substrate metal pad.

Reason¬ Metallurgical attachment of every bump to both chip and substrate is required for flip chip interconnect integrity.

Guide 15 Vias under final metal wiring must be spaced away from the edges of the passivation opening. Examples of minimum spacing values used by IBM and Motorola are shown in table 4-6.

Reason¬ Intersection of the edge of the passivation opening with the underlying via topography results in steep features which are difficult to cover with PLM metallurgy and may induce stress related defects.

Guide 16 The edge of the PLM as patterned should not transect an underlying step feature such as a via edge. The importance of this guideline increases with increasing step height and slope.

Reason¬ Pad limiting larger metal can cause stresses when positioned over irregular surfaces.

Guide 17 Practical test cell bumps should be identical to bumps on product die.

Reason¬ To provide correlation between test cell bumps and active product bumps.

Table 4–6¬ Terminal Via and Final Metal Via Pitch

Descriptions¬	Size¬	Tolerance
Passivation Opening Dimensions	50 µm¬	±1 µm
Passivation Opening Edge vs Stacked Wiring Via	2 µm¬	Minimum
Passivation opening edge Vs. Common Wiring Via	10 µm¬	Minimum

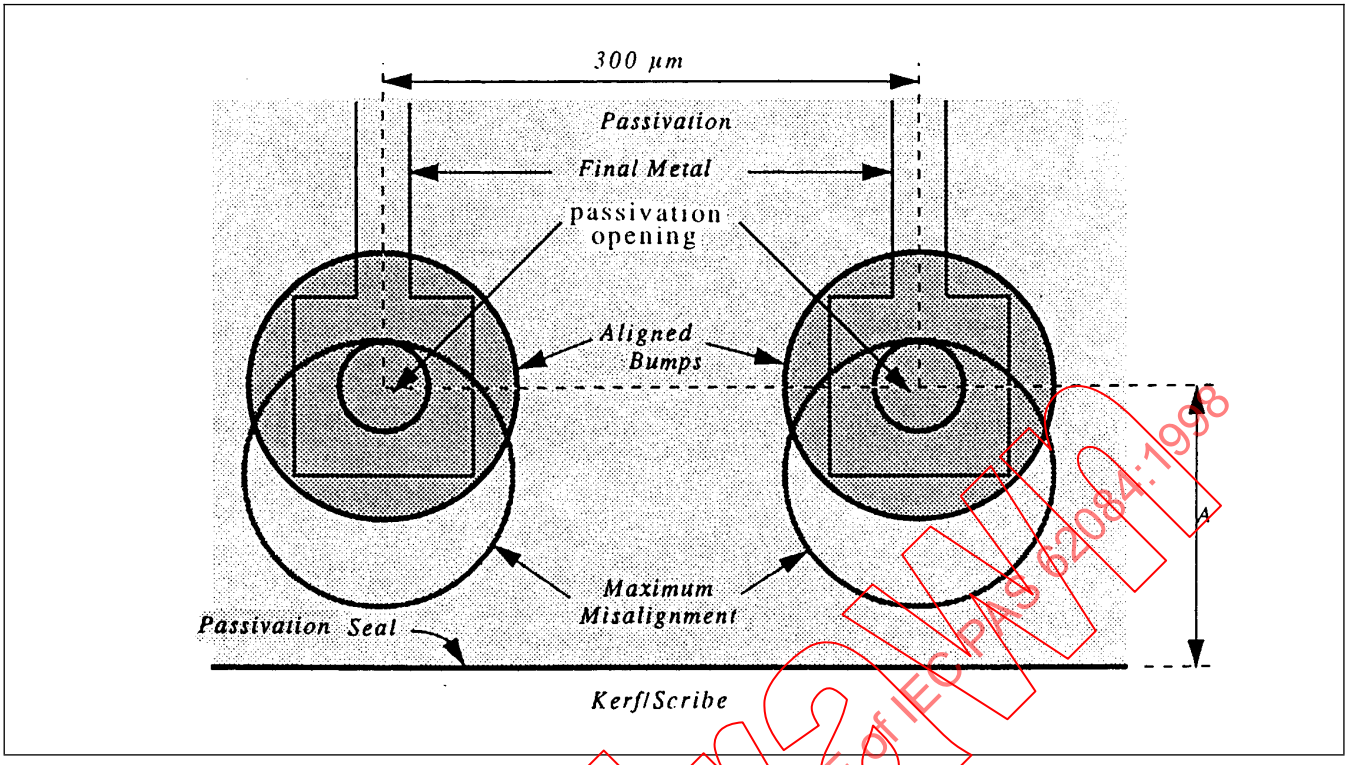


Figure 4–21¬ Minimum Pitch from Bump to Passivation Seal ("A"-½ finished bump diameter, plus alignment to tolerances, plus desired minimum)

4.4.3 Footprint Population

The number of bumps needed varies with electrical, thermal, and mechanical requirements. Besides electrical connections, bumps may be included for heat dissipation, mechanical support (outriggers, chip orientation, or future design migration).

Each application has its own unique characteristics. This makes it impossible to establish a general methodology for footprint population; however, there are various strategies that can be employed depending on the application.

The following are some examples of several footprint population design issues and suggested approaches for solving them.

4.4.3.1 Redundant Bumps

Redundant bumps are designed for power and critical signals. The chip area will support a larger footprint but the resulting distance from the neutral point (DNP) may negatively impact reliability. In this approach the outermost rows of bumps join chips and endure the most stress. As the DNP increases, the stress in the outer rows increases. By placing redundant bumps in the outer rows and critical bumps in the inner rows, if bumps in the outer rows fail, chip functionality will not be compromised. Include a sufficient number of redundant bumps to allow for potential loss. Figure 4-22 shows how this might be accomplished.

An alternative and supplemental solution would be to use chip underfill after the chip is attached to a substrate. This might be necessary if the DNP is sufficiently large. This does not involve additional wafer level processing.

4.4.3.2 Chips That Will Incur Design Shrinks

Designing a footprint for chips that incur design shrinks requires additional considerations.

The approach for this chip design is to locate the critical bumps in the chip's interior. Use outer rows for redundant and mechanical bumps. When the shrink occurs, the outer, less essential rows will be lost, but the critical bumps will not be affected. As the chip shrinks so does the DNP - making the loss of redundant and mechanical bumps less of an issue. This is illustrated in figure 4-23.

It should be noted that with this approach, the package is not affected by the bump footprint as the critical connections do not change. Redundant connections to the package simply become open pins. Caution should be exercised when the chip area shrinks so that the bump footprint does not shrink. In addition, provision must be made for orientation and alignment of the shrunk die.

4.4.3.3 I/O Drivers on the Periphery

Some chip designs require that the I/O drivers are on the periphery with other circuits internal to the chip. Redundant connections for power and signals are desired.

When using the approach that positions I/O drivers on the chip periphery, shorter interconnection lengths are always desirable in order to minimize parasitic effects. Bumps

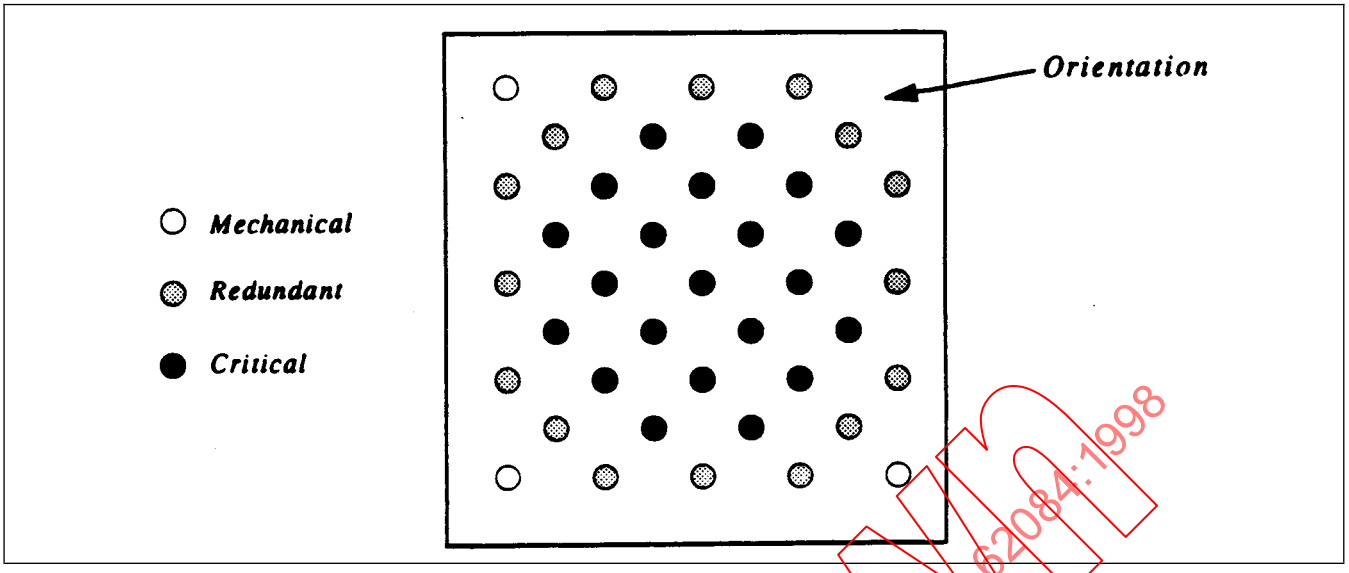


Figure 4–22¬ Redundant Footprint

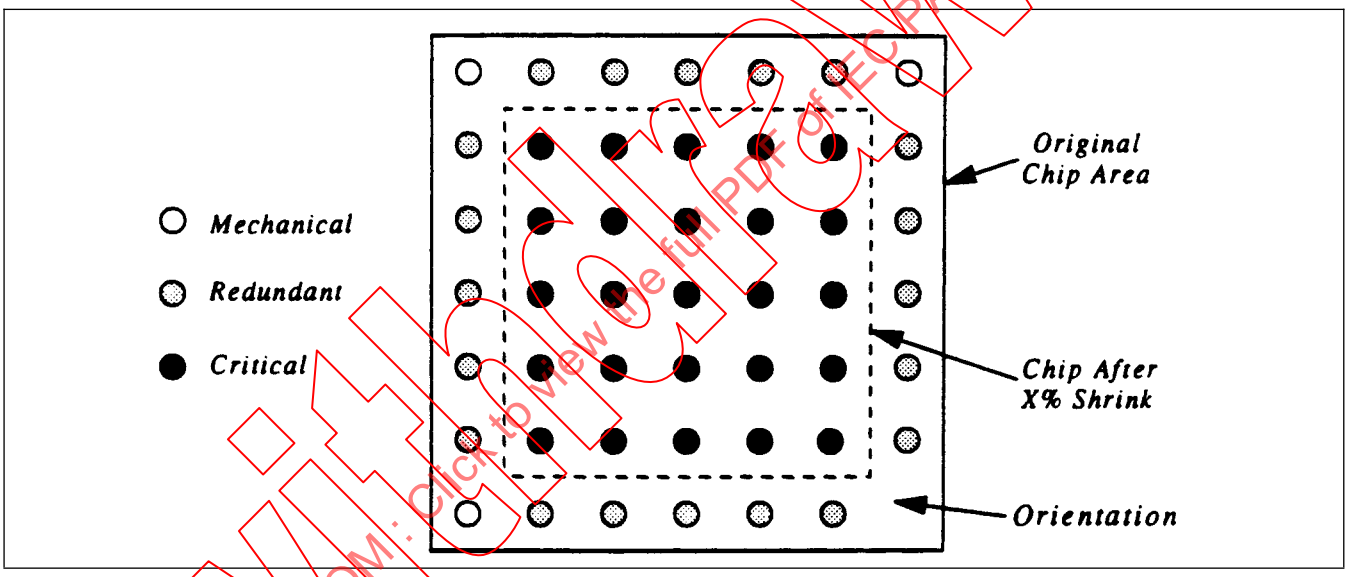


Figure 4–23¬ Design Shrink Footprint

should be located as close to the pertinent circuitry as possible. Figure 4-24 shows redundant power and ground bumps which are located above the internal circuitry to minimize distribution lengths. Input bumps are placed near the input circuitry and output bumps are placed near the output drivers.

4.4.4.4 Isolating Sensitive I/Os

In some chips, it is important to electrically isolate a sensitive input/output position from crosstalk or other noise.

The approach to doing this is to surround the area with power and ground bumps as shown in Figure 4-25. Caution should be exercised when placing bumps of different electrical potentials in close proximity, as they could cause shorting problems with dense wafer probes. Bump pitch should be maximized to minimize this hazard.

4.5 Design Output Requirements

There is a minimum of information that should be supplied to a designer in order to produce a bumping design. The information, at a minimum should include:

• Final metal mask¬ • Passivation mask¬ • Unit cell plan

It should be noted that the mask requirements are Wafer Fab dependent. One should consult with a particular wafer fabrication group for specific information. In addition, each plan is typically a computer-aided design (CAD) file and may include specific drawings. The most prevalent format in use is GDSII though others are possible.

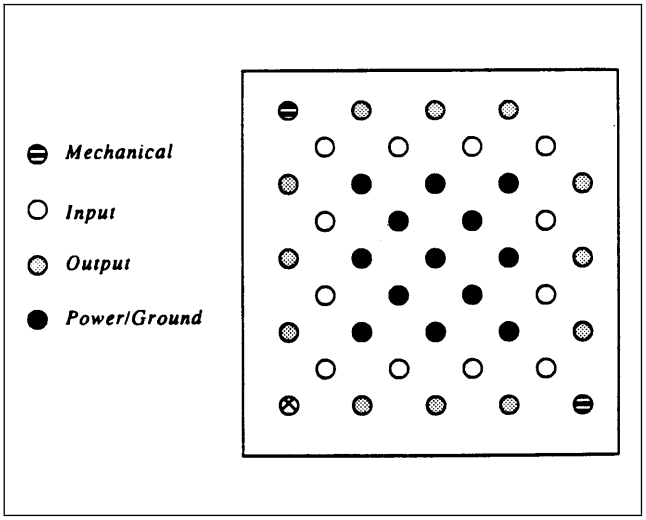


Figure 4–24¬ Signal and Power Distribution Position

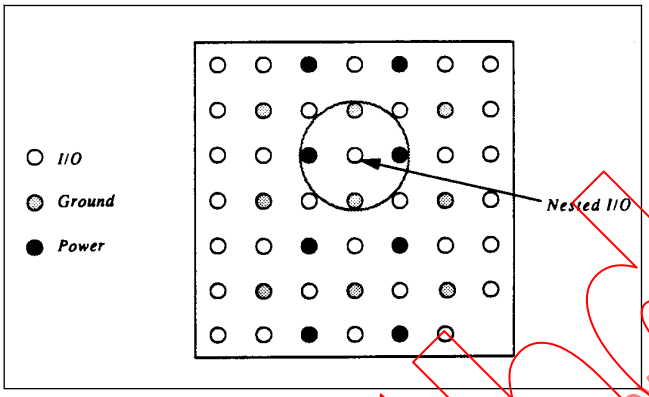


Figure 4–25¬ Nested I/O Footprints

4.5.1 Final Metal Reticle Mask

This mask conforms with the particular wafer fab reticle or mask requirements.

4.5.2 Passivation (Terminal Via) Reticle Mask and Plan

A suggested format is shown in figure 4-26. In this example the fiducial-marks represent the chip passivation edges. A non-wettable metal appears everywhere except the terminal via sites. The as-drawn dimensions of the terminal vias are wafer fab dependent. Positive photoresist is used with this type of reticle mask.

Terminal via sizing may require compensation for etch bias depending on the particular wafer fab etching process. One should consult with a particular wafer fabricator in order to determine the specific design requirements.

The bump passivation plan is a CAD file containing the terminal via location data needed for the bump mask design.

4.5.3 Unit Cell Design

A drawing and/or CAD file is required for each type of Unit Cell Plan (Product, CMA and Test) as illustrated in figure 4-27. This information is needed for generating the Wafer Step Plan and Bump Mask design.

4.5.4 Printed Board Land Pattern Design

The Printed Board Land Pattern for flip chips and grid arrays is simply a circle of coated copper whose diameter is the same as the bump's circle. The coating on the copper is one of the common solderability preservatives, such as solder, gold flash, or an organic protective coating (OSP). For MSMT packages, the land pattern may be a soldermask defined opening over the conductive area.

A top and cross sectional view of the printed board land pattern is shown in figures 4-28 and 4-29.

4.6 Electrical Design

This section addresses electrical modeling, analysis considerations and design guidelines for the chip bumping processes. Substrate and package for direct chip attach, electrical models and combined noise analysis of the system are also covered.

Bumping technology offers significantly less inductance than wire bonding or TAB interconnects. Although wire bond and TAB chips can be converted to bumped flip chip parts, to take full electrical advantage, chips should be initially designed for the bumping process. This allows optimization of internal chip wiring, circuit placement and bump locations.

Whether using a common solder bump layout or designing a new footprint, the chip power, ground, signal, and clock pinout must coincide with the substrate pad designations. This is especially true when existing substrates are utilized. Also the location and proximity of sensitive signals must be accommodated. Because bump pitch affects coupling noise, separation and isolation of sensitive signals must be considered in the chip and substrate footprint design.

4.6.1 Equivalent Circuitry

Figure 4-30 depicts the physical bump electrical path for a peripheral pad chip redistributed for bumping technology. The electrical path consists of a wiring via from the peripheral pad to the Final Metal (redistribution trace), the final metal trace to the pad, the terminal via to the bump.

The substrate metallization and associated wiring are part of the substrate electrical path. The bump electrical equivalent circuitry can be derived as shown in Figure 4-31 for first order approximation.

The wiring via and final metal (trace and pad) are modeled separately. Final metal consists of multiple sections (1...n). The bump and the passivation opening are represented by equivalent lumped circuits with both fringing and area capacitance accounted for in the circuits.

It should be noted that the bump-to-chip and the final metal

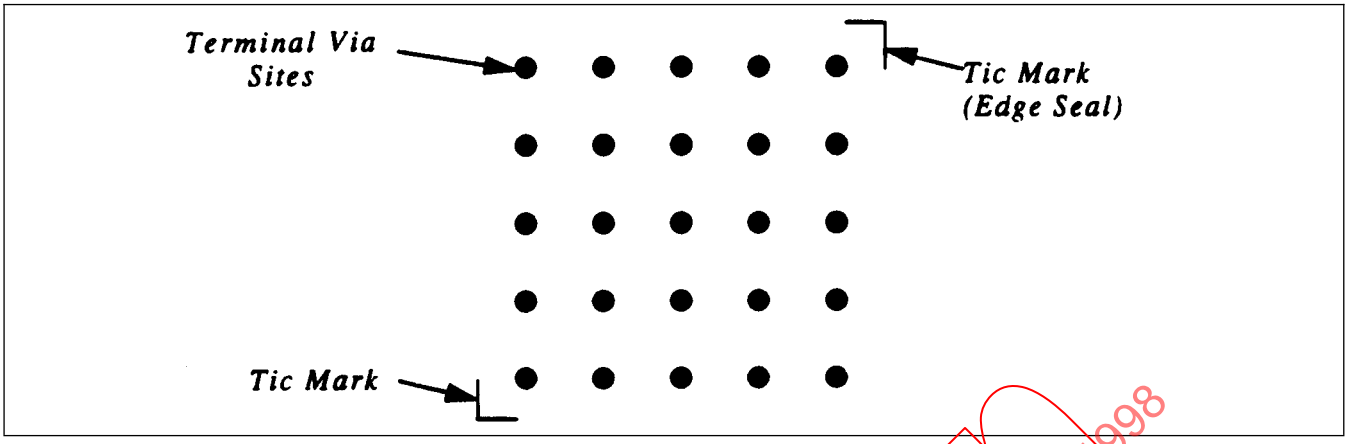


Figure 4–26¬ Typical Bump Passivation Reticle Mask Format

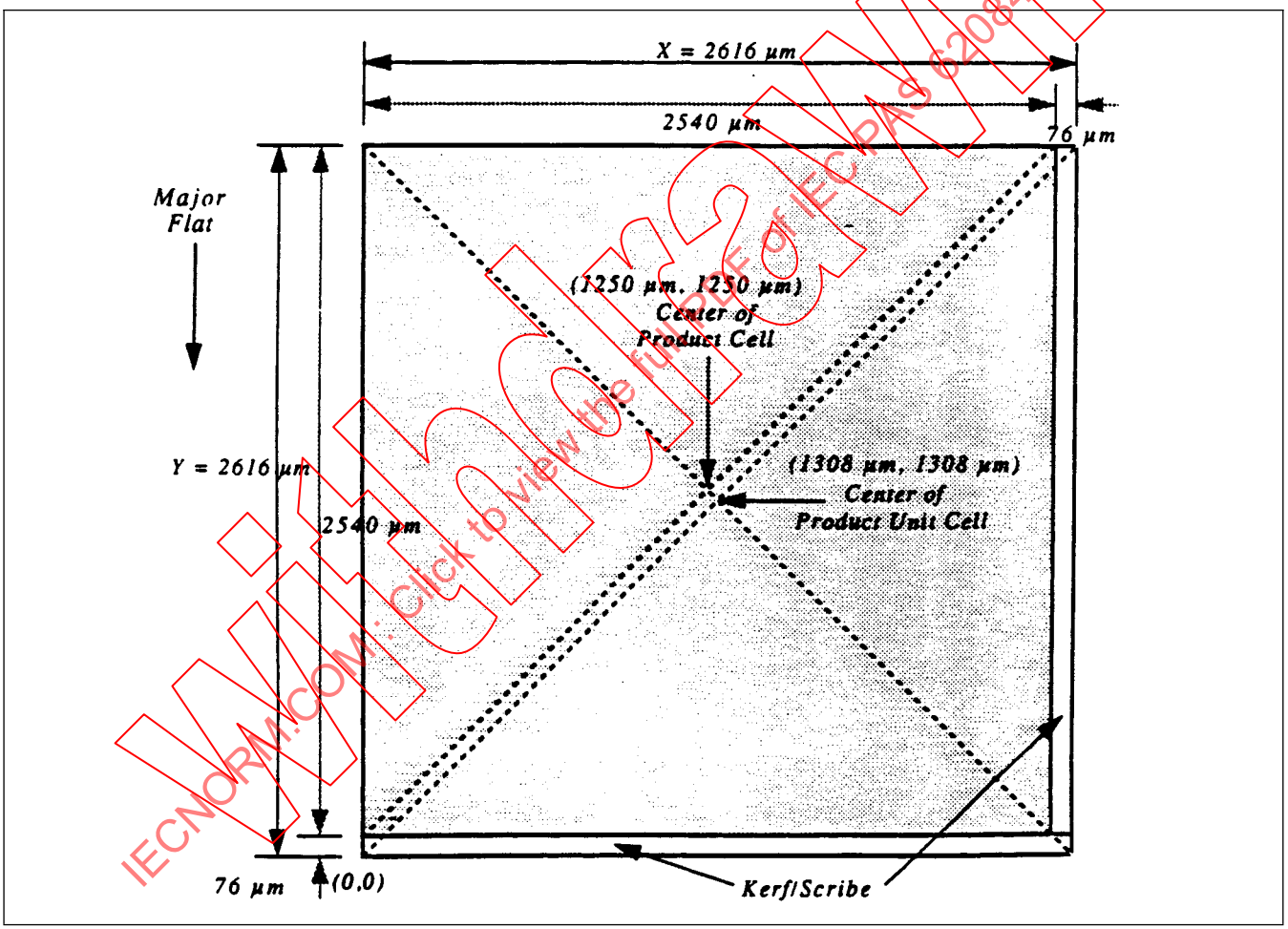


Figure 4–27¬ Product Unit Cell Plan (example)

pad-to-chip area capacitances overlap. This coincidence needs to be accounted for in the model. One way to do this is to combine the effect of the two capacitors.

Generally, transmission line modeling is not required unless rise times become significantly short and/or chip traces become significantly long. Since the bump is relatively short compared to wire bonds or TAB interconnects, lumped element analysis can be used because distributed effects are minimal under most conditions.

4.6.2 Final Metal Traces

The final metal power ground, clock and signal traces are a concern of any chip design and especially bump redistribution design. Metal thickness has a major impact on final metal resistance (Rfm). It should be noted that metal thickness is limited by the wafer process capability for traces on flip chips.

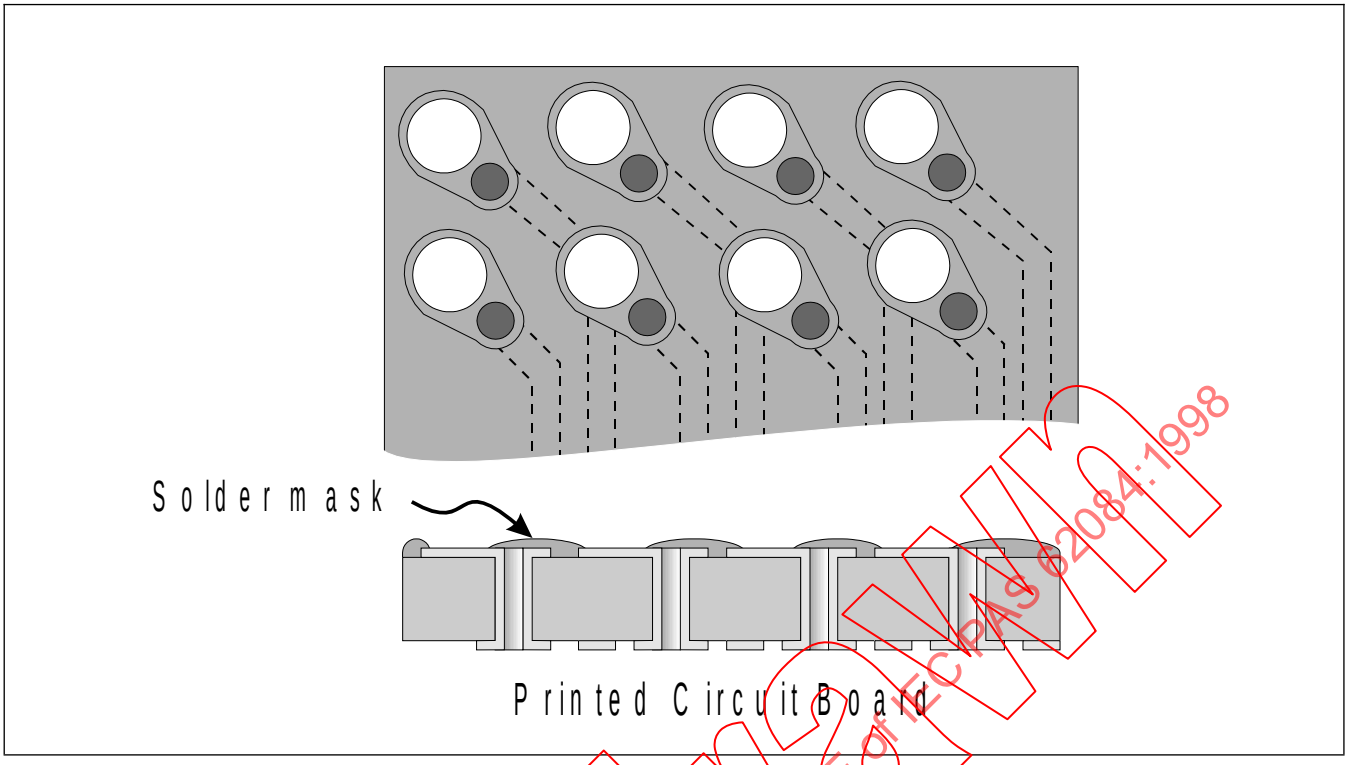


Figure 4–28¬ Printed Board Flip Chip or Grid Array Land Patterns

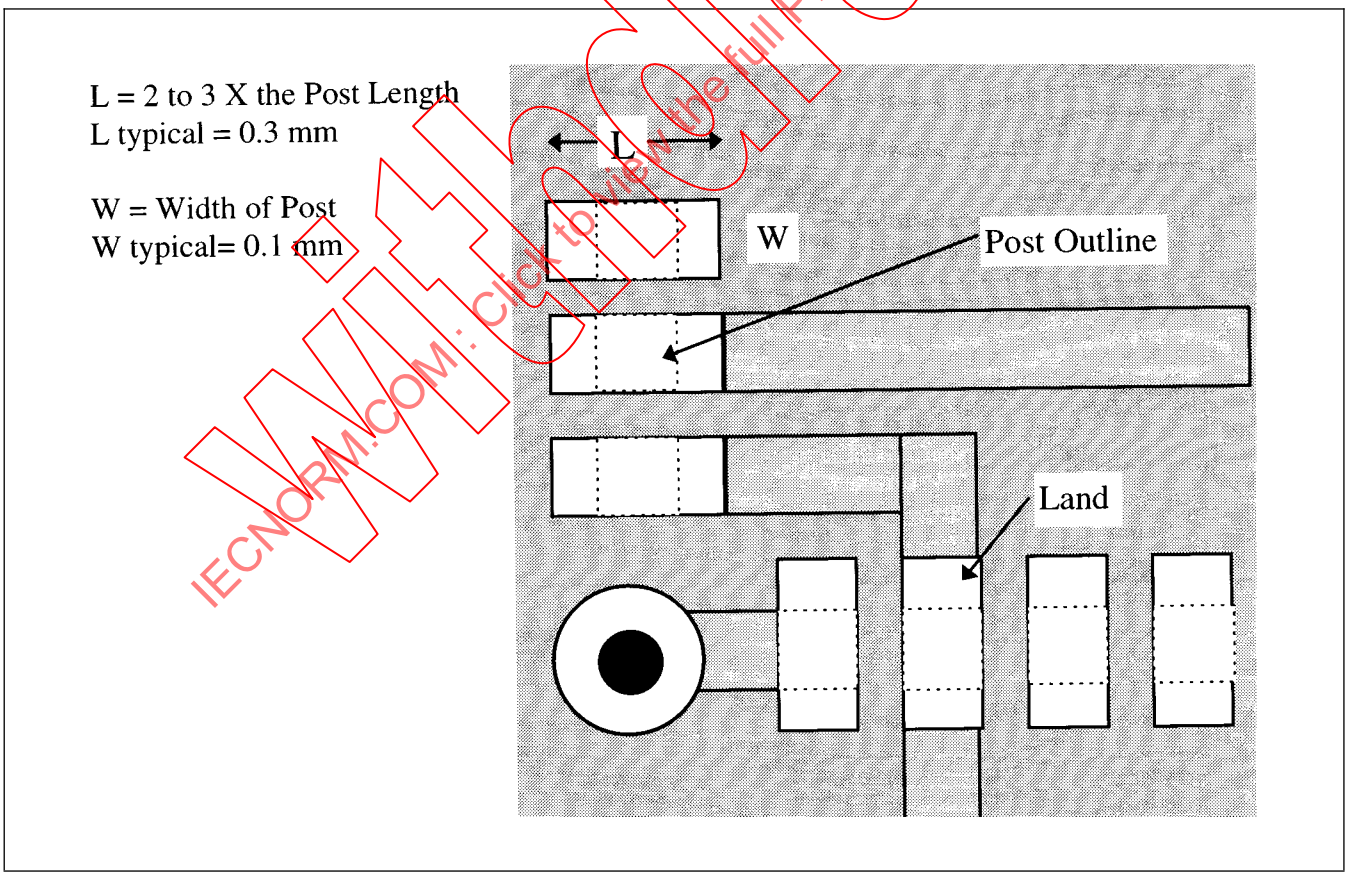


Figure 4–29¬ MSMT Land Drawing and Dimensions

In the following example, typical dimensions are used to illustrate variations in final metal resistance. For this example the final metal is doped aluminum with the following parameters:

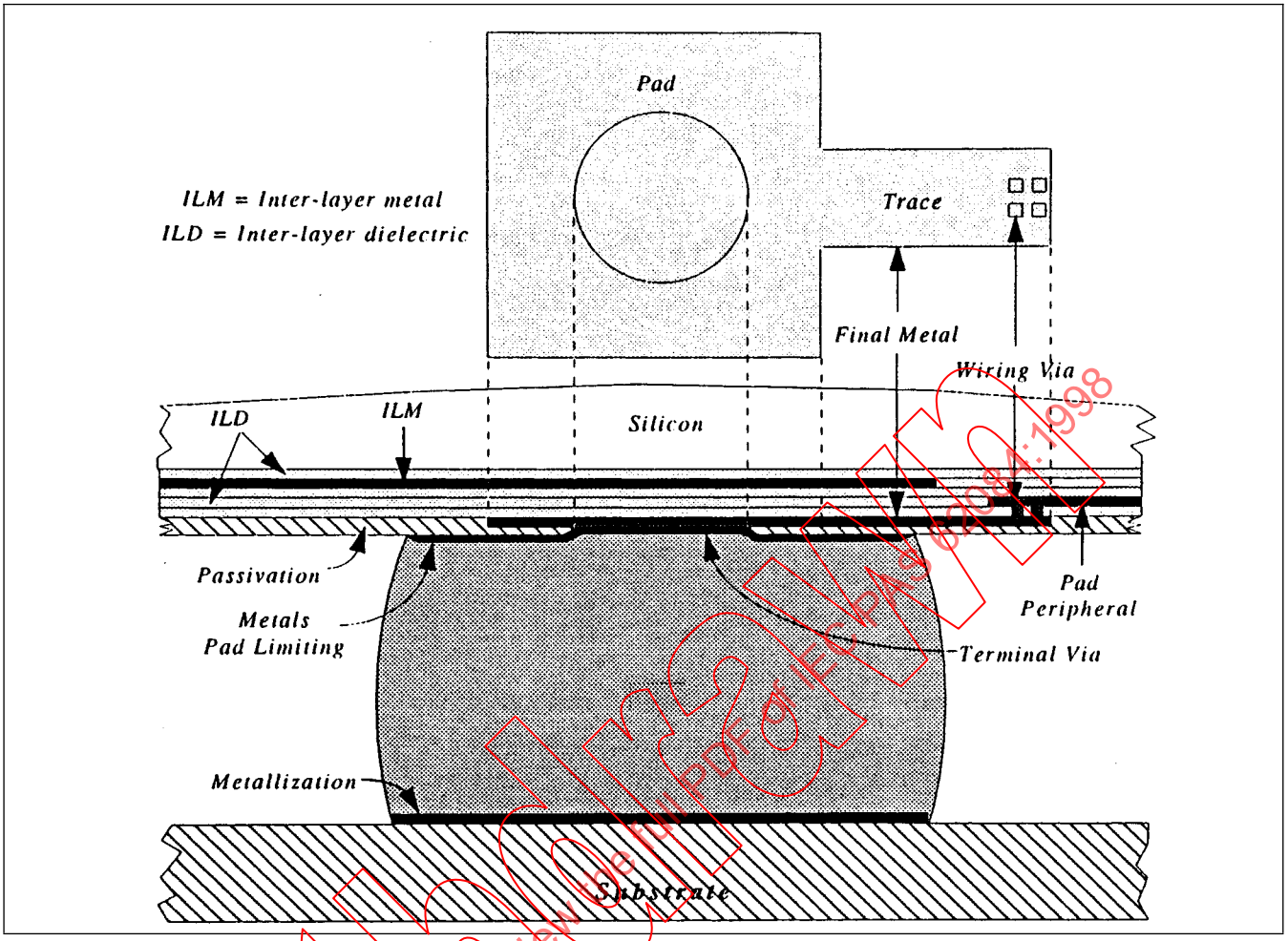


Figure 4–30¬ Bump Electrical Path (Redistributed Chip)

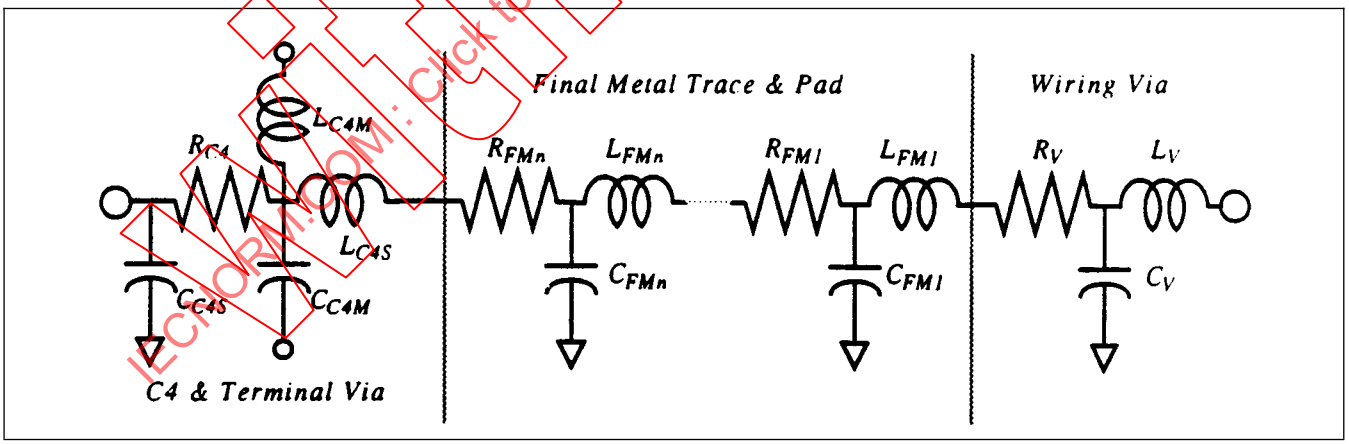


Figure 4–31¬ Bump Equivalent Circuit (Redistributed Chip)

Surface Resistivity (Rs)
= 0.037 ohm/square area (1.0 µm thick)
= 0.028 ohm/square area (1.5 µm thick)
= 0.018 ohm/square area (2.0 µm thick)
Signal Trace = 30 µm
Power Trace = 60 µm
Trace¬ = 700 µm (minimum)
= 3350 µm (average)
= 6000µm (maximum)
Trace¬ = 1µm, 1.5µm, and 2µm

the largest final metal trace resistance (length = 6000µm and width = 30µm) can be calculated from Ohm's law.

$$R_{FM} = \frac{R_S \cdot \text{Length}}{\text{Width}} = \frac{0.037\ \Omega\ 6000\ \mu m}{30\ \mu m} = 7.4\ \Omega$$

Using the value from above, Table 4-7 and 4-8 provide a

quantitative view of how final metal resistance varies according to thickness and length for signal and power traces.

Table 4–7¬ Final Metal Signal Trace (30 µm) Resistances (example)

Final Metal Trace	Final Metal Signal Trace Resistance(Ω)		
	1.0 µm¬	1.5 µm¬	2.0 µm
700 µm¬	0.9¬	0.6¬	0.4
3350 µm¬	4.0¬	3.1¬	2.1
6000 µm¬	7.4¬	5.6¬	3.7

Table 4–8¬ Final Metal Power Trace (60 µm) Resistances (example)

Final Metal Trace	Final Metal Power Trace Resistance (Ω)		
	1 µm¬	1.5 µm¬	2.0 µm
700 µm¬	0.4¬	0.3¬	0.2
3350 µm¬	2.1¬	1.6¬	1.0
6000 µm¬	3.7¬	2.8¬	1.8

4.6.3 Inductance and Capacitance

Calculation of final metal trace inductance and capacitance requires determination of self and mutual effects. These effects will be unique for each chip design. As shown in figure 4-32, the orientation (parallel, orthogonal or in between) of the Final Metal trace to underlying inter-layer metal and ground planes will determine self and mutual capacitance.

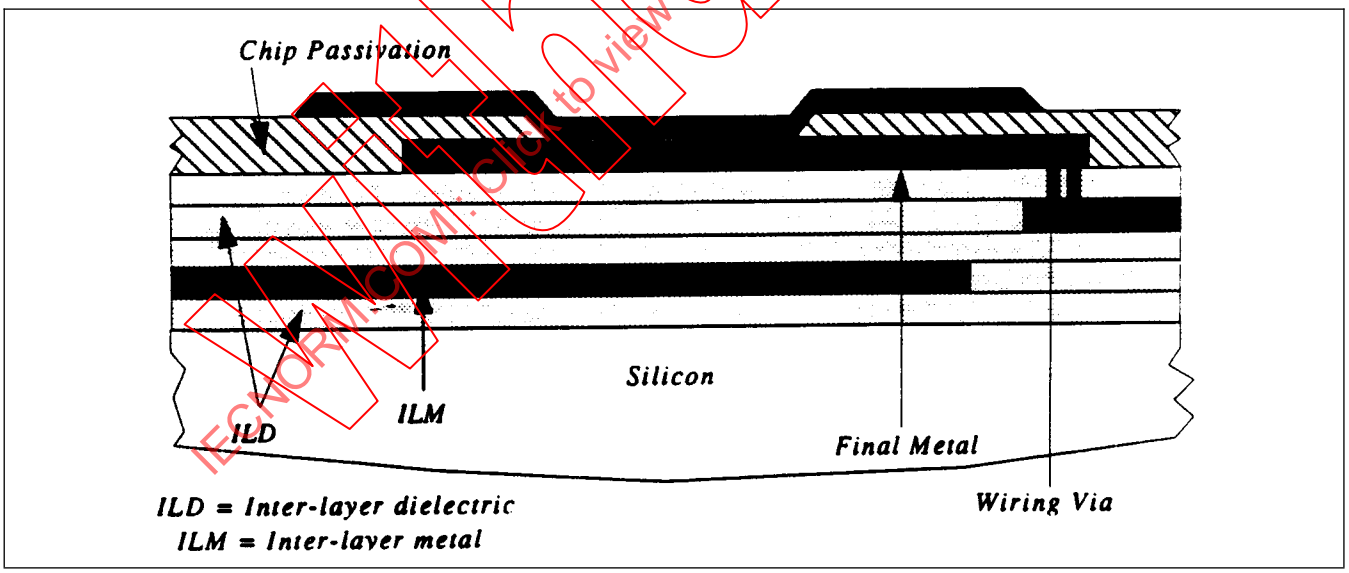


Figure 4–32¬ Final Metal Trace and Underlying Traces (Cross Section)

Inductance (L_{FM}) is also influenced by orientation with respect to underlying metal and other Final Metal traces. Direction of current flow in these metal structures must be known to determine mutual inductance. The appropriate current return path is needed to determine self inductance.

Final Metal trace capacitance (C_{FM}) consists of self and fringing capacitance. A first order approximation of final metal total capacitance is shown below:

$$C_{FM} = (C_A \cdot L \cdot W) + (C_F \cdot L) \times 2$$

where, C_A = Area Capacitance (F/m^2)
C_F = Fringing capacitance (F/m)
L = Trace (m)
W = Trace (m)

For accurate modeling, coupling between Final Metal traces and underlying traces should be modeled using 2D/3D CAE programs.

4.6.4 High Frequency Performance

Because the chip active surface is face down or near the active substrate, special consideration in trace routing on the chip, interposer, and substrate must be given for flip chips and grid arrays. Changing signal levels on either the chip or the substrate traces close to the chip are coupled. This coupling creates cross-talk, noises, EMF, interference, etc., in the chip or substrate. Controlling the distance, the dielectric between the chip and substrate traces, and characteristic impedance, helps reduce the coupling.

4.7 Thermal Design

The primary thermal path with bump technologies is through the silicon to the chip backside. To this end, packaging becomes the major component of thermal consideration.

Thermal aspects of bump packages provide front and/or backside thermal path. Bump technology offers an added dimension of thermal management. Depending on the application, the designer has the option of selecting the thermal path.

The degree of consideration given to thermal design is dependent upon three factors:

- The amount of energy or heat that must be dissipated
- The desired operating junction temperature of the chip
- The ambient temperature of the surrounding environment

Conduction, convection, and radiation are the models for which heat can be dissipated. For bump interconnect conduction is the primary heat dissipation mode. Convection is a small contributor.

As illustrated in figure 4-33, thermal and electrical analysis are analogous.

4.7.1 Bump Interconnect Thermal Model

The bump thermal interconnect can be modeled as two cylinders between the chip power source and the substrate (see Figure 4-34). Cylinder 1 represents the inner-layer materials on the chip. Cylinder II represents the Pad Limiting metals and the bump size. The heat source is the device in the chip.

Because there are three bump interconnect diameter options, cylinder sizes vary accordingly. Assuming that Cylinder I is SiO_2 (k = 1.01 W/m.K) and Cylinder II is tin/lead (3/97) solder (k = 36 W/mK) then the approximate Interconnect thermal resistance can be calculated for each option. See Table 4-9.

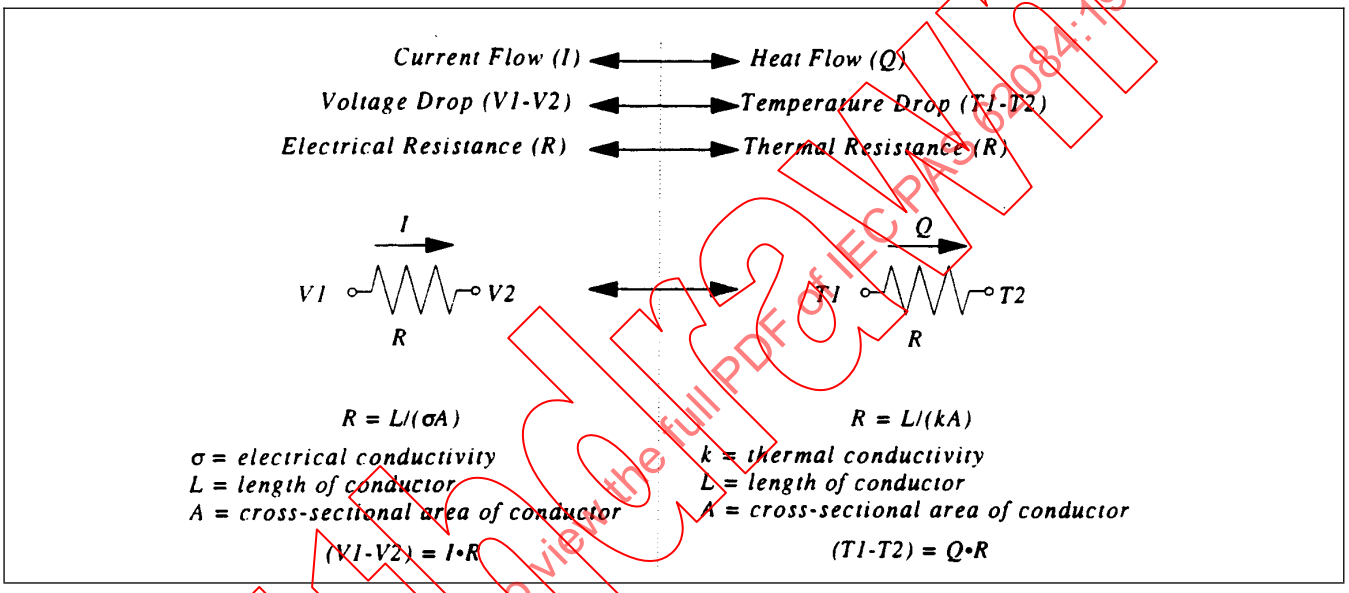


Figure 4−33¬ Thermal/Electrical Analogy

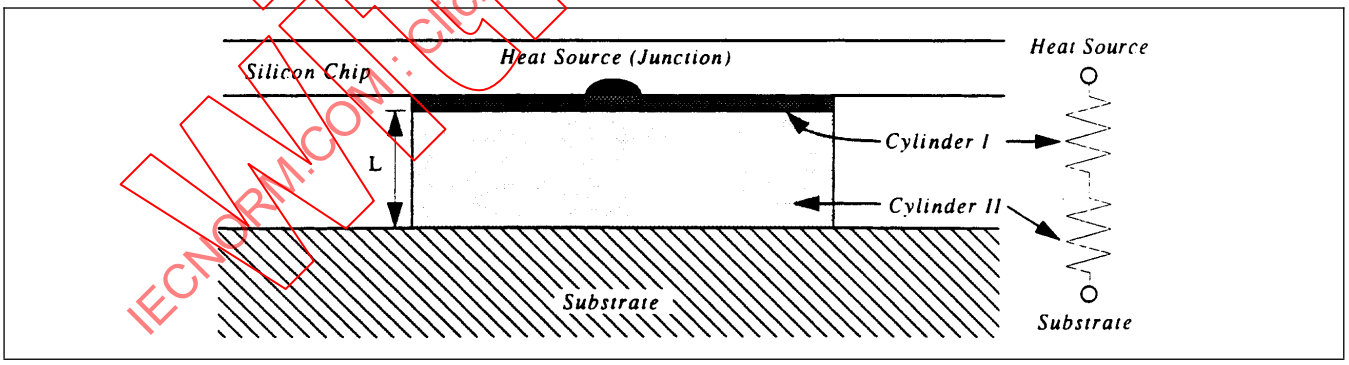


Figure 4−34¬ Bump Interconnect Equivalent Model

Table 4−9¬ Typical Thermal Resistance for Variable Bump Options (Triple Layer Chip)

Interconnect Option	Diameter (m)	Cylinder I¬			Cylinder II			Total R (°C/W)
		L (m)	A(π^2) (nm^2)	R (°C/W)	L (µm)	A(π^2) (nm^2)	R (°C/W)	
Option A¬	150¬	10¬	18¬	550¬	75¬	18¬	116¬	666
Option B¬	125¬	10¬	13¬	762¬	63¬	13¬	135¬	897
Option C¬	100¬	10¬	8¬	1238¬	50¬	8¬	174¬	1412
Option D (DCA)	150¬	10¬	18¬	550¬	≥75¬	18¬	≥116¬	≥666

It should be noted that Cylinder I thermal resistance varies according to the number and specifically, the thickness (L) of innerlayers on the chip. As the main contributor to thermal resistance, it should be modeled accurately. Furthermore, interface and bulk resistances have been combined to simplify calculations and illustrations.

Table 4−10¬ Typical Bump (150 µm) Thermal Resistances Multi-Layer Metal Chips

	Single Layer¬	Double Layer¬	Triple Layer
Maximum (°C/W)	550¬	750¬	950
Nominal (°C/W)	450¬	650¬	850
Minimum (°C/W)	350¬	550¬	750

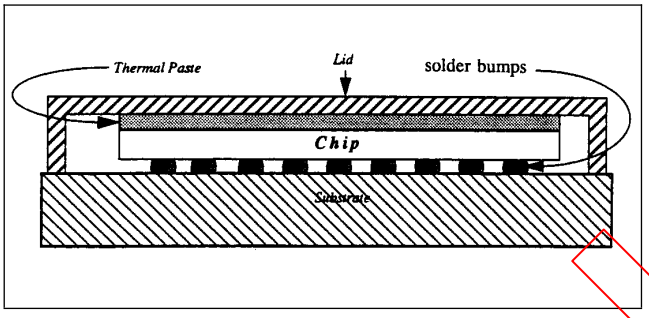


Figure 4−35¬ Thermal Paste Example

Packaging options can range from hermetic, metallized ceramic modules to printed circuit boards for direct chip or chip scale packaging for highly specialized thermal conduction packages. Supplemental packages features like thermal paste can be added to enhance thermal performance. Other materials like chip underfill used to improve reliability also alter heat dissipation. All these components influence the junction to case thermal models.

4.7.2.1 Thermal Paste Model

The thermal paste is applied to the chip backside to reduce the thermal resistance between the chip, package lid as shown in figure 4-35 or heat sink. The approximate thermal model is shown in figure 4-36.

4.7.2.2 Chip Underfill

Chip underfill is used to enhance reliability for flip chip and some grid array packages. It is dispensed between the chip and substrate surrounding the bumps as shown in figure 4-37. Because it completely fills the gap between the chip and substrate, it does change the thermal model. The approximate model is shown in figure 4-38.

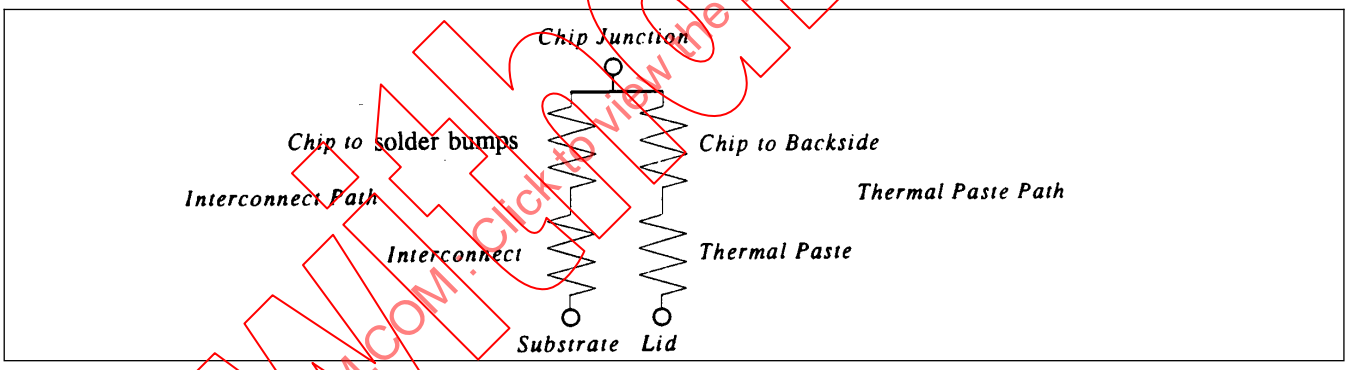


Figure 4−36¬ Approximate Thermal Model for Thermal Paste

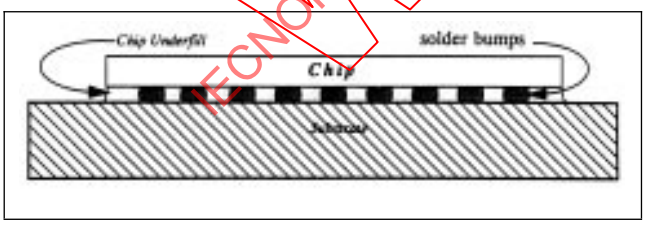


Figure 4−37¬ Chip Underfill Example

Bump interconnect thermal resistance for a typical single, double and triple layer metal device are shown in table 4-10. It should be noted that these values are based on 150 µm bump on silicon where SiO_2 inter-layer dielectric is modeled..7.2 Junction to Case Thermal Models

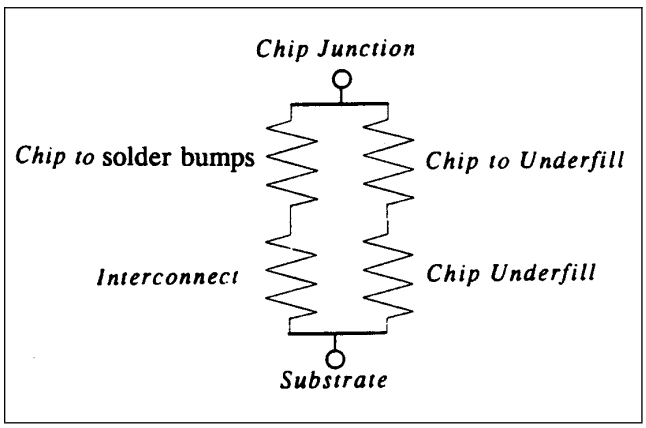


Figure 4−38¬ Approximate Thermal Model for Chip Underfill

5 MATERIAL PROPERTIES AND PROCESSES

This section deals with the materials and processes currently used for flip chip and chip scale manufacturing. Subjects addressed are the primary techniques for bump deposition and various types of chip to substrate attach methods currently used and the materials necessary for each type of attach. Also included are the various options for bump and post metal evaluation.

5.1 Solder Bumping

5.1.1 Solder Evaporation

In the evaporative process the necessary metal structure under the bump and the solder itself are sequentially evaporated. This structure is deposited in an array of pads on the wafer surface. This is a batch process in which the solder and its underlying metal pads are deposited onto many wafers at the same time.

Prior to evaporation of this metal structure, an in-situ sputter clean is commonly performed. This is done to remove undesirable oxides and photoresist prior to metal deposition. This step generally assures low contact resistance, and the surface roughening enhances the adhesion of the metal to the organic or inorganic insulating layer.

As a part of this process it is necessary to provide a mask that controls the shape and formation of the pattern formed on the wafer. The most commonly used mask for this particular process is a metal mask. This metal mask/wafer assembly typically consists of five parts: 1) a backing plate, 2) a spring, 3) the wafer to be bumped, 4) the metal mask and 5) a clamp ring. These typically have been assembled and aligned manually. With this manual procedure it is possible to align to within 25 µm.

Of primary concern in any bumping process is the ability to control the uniformity of the bumps. The uniformity of the bumps formed in an evaporative process is primarily controlled by the geometry of the evaporator. This is then a function of the source and object areas and the distance separating them.

Baffles may be added to enhance the uniformity of the evaporated material. The height of the bump as determined by the volume of the deposited material is a function of the standoff of the metal mask and the opening in the mask. The material as deposited is conical in shape. This is controlled by the opening, which will gradually close off as material builds up on the mask during deposition.

5.1.2 Solder Electroplating

The second commonly used method for the formation of solder bumps is electroplating. This is usually done by plating through a prepatterned photoresist mask.

This process begins with the blanket evaporation of an Under Bump Metallurgy (UBM) layer consisting typically of an adhesion layer such as chromium or titanium and then a thicker wettable layer such as copper.

Prior to deposition of this layer it is necessary to clean the via surface in order to remove any contaminants such as oxides or organic residues. Typically a sputter clean has been used; however, wet processing to clean these surfaces has also been used.

After deposition of the UBM two methods are generally used for forming the solder limiting layer or solder dam. The first consists of blanket depositing a layer of Chromium or some other non-wettable metal on top of the UBM. Next a layer of photoresist is coated and patterned. At this point the chromium is etched open and the solder is then plated on top of the copper.

The second method consists of patterning the nonwettable layer, e.g., chromium. This layer can be patterned in photoresist; the metal can then be deposited and lifted off. After this step a second photopattern is formed to plate up the solder. The second method offers greater control of the size and dimensionality of the bump. An example of an array of plated bumps prior to reflow is shown below in figure 5-1.

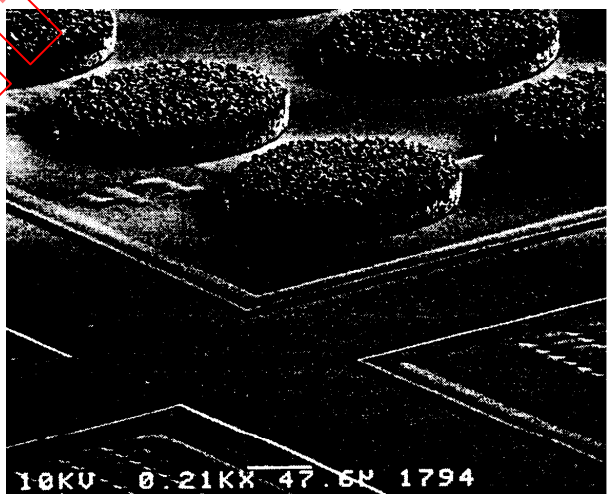


Figure 5−1¬ Photomicrograph of As-Plated Solder Pads

At the conclusion of the plating process the resist is stripped and the blanket metal is ready to be removed. This metal must of course be removed completely in order to prevent shorting between the bumps.

This removal process can be handled in either of two ways. It is possible with large solder bumps to wet etch the metal in the intervening areas using the solder as a mask, after which the solder would reflowed. This generally works better with larger bump sizes as some undercutting may result.

The second method would involve reflowing the solder

first. In this case the intermetallics formed under the bump during the reflow process may be used to provide a barrier to the etchants. No undercutting would be expected, and the size of the bump would be controlled by the thin film patterning of the solder dam and the plating height. An example of solder bumps formed by electroplating after reflow and removal of the UBM in the intervening areas is shown in figure 5-2.

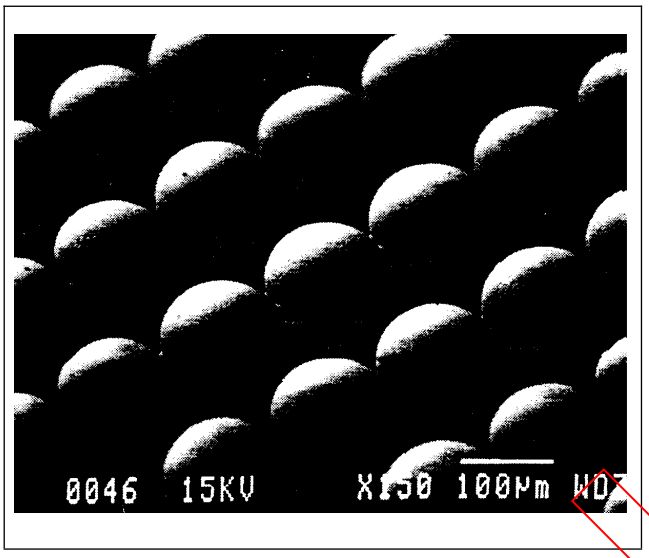


Figure 5–2¬ Photomicrograph of Solder Bumps

5.1.3 Solder Paste Deposition

Solder paste is most often used as the medium for bump formation when screen printing is the deposition method. Solder pastes are slurries of differing viscosities depending on the solids loading.

Spheres of different diameter solder balls, typically 20 micrometers, are suspended in flux which gives the paste its reflow properties and tackiness that aids in the placement retention during reflow.

The solder paste screened bump has the self-aligning properties during reflow that the solder bumps from other deposition methods exhibit. This key feature is one of the largest advantages of using solder, and it applies to screened-on bumps. Screen printing and stencil transfer are similar in that they provide a physical means for patterning a paste onto a substrate.

5.1.3.1 Solder Screen Printing

Screen printing generally refers to a woven mesh that is covered by an emulsion to provide a pattern for transfer. Screen printing of conductive materials for bump formation is most often used for lower lead count devices. With these devices the pad pitch is generally large enough to allow for the screen printing of paste without shorting bump to bump with the paste.

Where different devices are to be mounted on the same substrate, bumps of differing sizes may be desired. Through etching the appropriate mask material, various size openings can be put into the mask to give the corresponding bump sizes.

The major advantage of using screen printing techniques for bump formation involves costs. The equipment set is much less expensive, by an order of magnitude, than that for either bump plating or evaporation techniques. A screen printing machine runs $50k to $150k for production volumes.

Drawbacks include the pitch limitations alluded to earlier and volume control. The screen printing process technique tends to have the worst volumetric dispense repeatability of bump formation techniques; however, in many cases screen printing of paste gives acceptable results.

5.1.3.2 Stencil Transfer

Stencil refers to a metal mask that has been cut with a laser or photochemically etched. Stencils are made of metallic sheets that have openings etched or cut into the metal. Those openings allow paste to be dispensed to the mounting surface. The size of the opening and stencil thickness controls the volume of paste deposited on the mounting structures for bump reflow.

To control solder volume and provide different bump geometries, stencils may be etched thinner in local areas. Figure 5-3 provides an example that shows the local etched area of the stencil and the resulting reduced solder volume.

5.1.4 Conductive Paste Method

Conductive paste methods have been implemented in mass production at Matsushita and Citizen.

In this method the LSI driver chips are plated to form 50 µm copper bumps with a thin overcoat of Au. Then either the bumps or panel electrode are coated with the Pd-Ag paste. Next, the coated bumps are aligned with the panel to electrodes and joined under pressure. This is followed by setting the paste and coating the chip with a protective resin.

5.2 Conductive Adhesives

Screened-on conductive epoxies are available, but their reliability is yet to be proven. The conductive material is usually silver or some other corrosion/oxidation resistant metal. Part of the function of the conductive filler is to scratch through the native oxide formed on the aluminum bond pads of the integrated circuit during bonding.

Conductive particles dispersed into insulating adhesive provide an interconnect wedge between pads or posts on the chip to lands on the interconnection structure. There are basically two types:

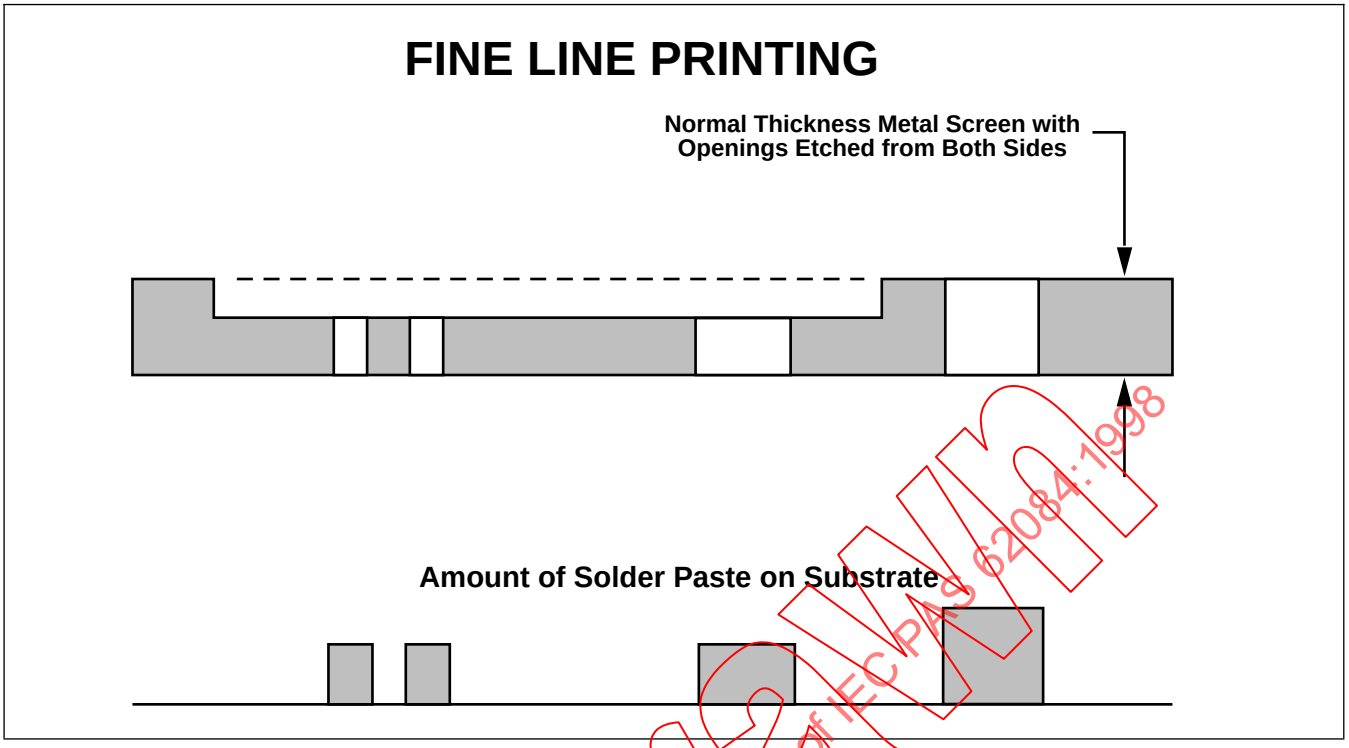


Figure 5−3¬ An Example of Fine Line Stencil Printing

1. Isotropic conductive adhesives in which current is conducted in all directions by the particle-to-particle contacts.

2. Anisotropic conductive adhesives containing a concentration of particles too low to conduct electricity in any direction. During the assembly process, excess adhesive is squeezed out between the pads on the chip and the substrate leaving single particles to bridge the chip substrate gap. The gap is established by the diameter of the conductive particle. The advent of flat panel displays has greatly accelerated the use of these techniques. Anisotropic adhesives eliminate the need for underfill, however the planarity requirements are stringent.

Both techniques offer the following advantages over solder systems:

- Allows for connection to non-solder-wettable metallic conductor (i.e., ITO), as well as polymer thick film
- Lower joining temperatures (i.e., <180°C or UV curable at room temperature). This makes the use of low cost substrate materials possible
- Avoids use of flux and subsequent cleaning

On the other hand, there are disadvantages and concerns as follows:

- Long term high temperature exposure is limited today to <130°C
- Current carrying capability can be limited by particle material construction and concentration
- Repairability is difficult or impossible
- There is a lack of large amounts of reliability data. However, today thermosetting polymer systems are offering greater reliability
- Not self aligning

5.2.1 Design Issues

For both types of adhesives, the contact metallization must be resistant to oxidation to ensure long term reliability. The interface materials need to be compatible with the adhesive.

The pad pitch for the isotropic version is controlled by the process capability to define the conductive adhesive volume sufficient to make the electrical connection without shorting to adjacent pads.

In the case of the anisotropic, the pad pitch has been demonstrated in production to <0.1 mm [.004].

A design consideration in the anisotropic application is the topology of the chip and substrate. Both chip and substrate lands must be at least level with the surrounding dielectric or above. The adhesive must completely fill the gap between the chip and the substrate while allowing good contact for the conductive particles. Care must be taken to ensure that there is sufficient coplanarity between the chip and substrate to ensure reliable connection. The use of a compliant metalized polymer particle can help compensate for small deviations in coplanarity. Compliant substrates also aid in this regard. Isotropic adhesive systems are more tolerant of variations in topology.

5.2.2 Anisotropic Adhesive Details

An alternative to the use of solder for flip chip and chip scale applications is the use of an anisotropic conductive adhesive.

Anisotropically conductive adhesives can provide electrical as well as mechanical interconnections between integrated circuit silicon chips and substrates. The conductivity of these materials is restricted to the Z-direction while maintaining electrical isolation in the X-Y plane. In addition, materials act as an encapsulant and seal the under surface of the chip, thus eliminating the need for underfilling.

Most anisotropically conductive materials consists of electrically conductive particles that are dispersed in an adhesive matrix. The concentration of these particles is such that enough particles are present to provide a reliable connection between a chip and substrate while there are too few particles to allow conduction within the X-Y plane. It is important to control size, size distribution and concentration of particles to prevent the formation of short circuits or leakage paths.

There are several criteria that must be evaluated when selecting an anisotropic conductive adhesive. The material must possess adequate adhesion between the surfaces of the chip and the substrate. The adhesive may be in contact with SiO_2, Si_3N_4, polyimide, gold copper or aluminum.

Some surfaces may require chemical treatment before the adhesive is applied in order to achieve adequate bond strength. The anisotropic adhesive must not contain any ionic impurities that could degrade the electrical performance of the chip.

There are a variety of materials that can be used as conductive particles. Gold has the best properties, but its cost could be prohibitive for some applications. Silver offers high conductivity at a moderate price, but problems with electromigration can occur. Nickel is lower in cost than either gold or silver, however, corrosion of the nickel can occur over time. The use of nickel particles that are overplated with silver or gold offers an alternative solution. Other particles that have been used as conductive fillers include metallized polymer spheres and metallic particles that are coated with thin layers of dielectric material.

The particles form electrical interconnections through several different mechanisms. Metal core particles can physically penetrate the soft metallization on either the chip or the substrate. Polymer core particles are deformed by the pressures applied during curing and relax chip substrate coplanarity requirements. Dielectric-coated particles mitigate the formation of short circuits; their insulating coatings are only removed or penetrated by bump contacts.

The electrical requirements of the assembly must be carefully evaluated before using an anisotropic conductive adhesive. Like other flip chip techniques, this technique also provides low inductance due to its short interconnection length. The current carrying capability of the interconnection can vary widely, depending upon the size, type, and loading of the conductive particles used.

The organic adhesive matrix can be either a thermoset or thermoplastic material. Thermoset adhesives maintain their strength at high temperatures and generally have better mechanical properties than thermoplastic adhesives. The primary advantage of the use of a thermoplastic adhesive is that chips can be removed relatively easy in order to be reworked.

5.3 Solder Bump Evaluation

Solder bumps should be evaluated in terms of their electrical, mechanical and material properties. This should include a time zero evaluation (quality) and a time greater than zero evaluation (reliability).

From an electrical point of view the bump and post plus their interface should provide a low resistance electrical path. The actual resistance may vary depending on the cross sectional area of the bump or metal surface area of the post, the under bump metals (UBM) used and the cleanliness of the via/ metal interface. This last part can be eliminated through proper cleaning before metal deposition.

For a standard bump configuration (0.125 mm on a 0.250 mm pitch nominally 85 µm high after reflow) the bump resistance should be less than 10 milliohms. Stress testing of the bump can lead to a gradual increase in this bump resistance without causing a catastrophic failure. A limit as to the amount of change allowed during stressing is necessary. Examples in the literature have indicated the use of a 300% increase as one cutoff point.

From a mechanical point of view the bump interface and UBM interface must form a bond of sufficient strength in order to maintain integrity during stress developed under normal operating conditions. Typical values for this bond strength are on the order of 45 - 50 grams per solder bump for a high lead solder bump. The actual value this number takes will strongly depend on the solder composition and ball diameter and will increase as the tin content increases.

Along with the bump strength another important mechanical property is the uniformity of the bump geometry across a wafer or more importantly across a chip. For some uses such as chip attach to a ceramic substrate, it may only be necessary to limit the bump variation to 30%.

It is also possible to obtain knowledge about bump formation and strength from shear measurements of a single bump (ball shear). While this is a destructive test it has the advantage of needing only one or a handful of bumps for the evaluation. It would of course be possible to add these

bumps in a test pattern. The tensile pull on the other hand requires the joining and subsequent separation of the chip and substrate. This can lead to the destruction of one or both.

The final area for evaluation would be the material analysis. This would include analysis of the bump interface and the bump itself. For the standard bump previously defined, the deposited or plated solder composition will generally be equivalent before and after reflow.

For very small bumps having a high lead content it is possible to see large fractional changes in the tin content as it is consumed by the underlying copper during reflow. If the tin content drops too low the protective environment of the SiO_2 tin oxide will degrade leaving the lead more susceptible to environmental attack. Lead oxides are formed protecting the bumps.

5.4 Other Bumping Techniques and Materials

Most bumping techniques require some form of under bump metallurgy.

5.4.1 Board Process Attach

Eutectic solder is used as a joining metallurgy in direct chip attach of high lead solder balls (e.g., 97/3) to organic substrates. Most organic substrates cannot survive the temperatures required to join high lead solder. Therefore, eutectic solder is used to provide a joining metallurgy for the high lead to organic substrate joint. During the reflow process, the eutectic solder melts and joins to the high lead solder. The design rules are essentially the same as those for the high lead rules, except ball pitch. The current technique for applying the eutectic to the organic substrate is such that the finest ball pitch is 350 µm that will allow for practicable joining yields is 350 µm. Alternatives to chip eutectic ball technologies are being pursued. These would allow the use of high lead design ground rules and simplify the substrate process.

It should be noted that wettability of the substrate by the eutectic ball is critical for reasonable yields. Also the height of the joint is less than that of the joint using eutectic for joining high lead balls.

5.4.2 Gold Bumping

Gold bumps are fabricated using either electroplating or wire bonding. They have been used in area TAB and flip chip applications, e.g., GaAs. They are chosen for thermal and electrical properties. Gold attachment to other materials can cause embrittlement.

There has been some limited use of gold on the bump where the process for the production of the bump is simply implemented into the wafer process. Gold bumps lend themselves to the gallium arsenide process well and as a result are finding use. In this case, the joining is done by thermocompression bonding.

5.4.3 Gold Bump Flip Chip

In this method, an ultraviolet light curable adhesive is used for the attachment of a gold bumped LSI driver chip to a glass substrate (Figure 5-4). The adhesive is cured while a bonder is holding the die against the glass substrate. The contacts are in compression during the bond cycle. This process relies on the stress induced by the shrinkage of the resin to weld the Au bumps to the panel. The use of UV curable adhesives results in very short cycle times during assembly.

5.4.4 Indium-Based Bumps

Fine pitch bonding systems which utilize indium-based flip chip technology have been developed for liquid crystal displays to minimize connection pitch. Figure 5.5 shows the chip on glass (CoG) structure in which an LSI chip is connected to a glass substrate using flexible indium bumps and a photo-thermosetting resin.

The process begins with blanket sputtering of Au/Ti-W. Next a layer of photoresist is coated and patterned. Then, indium bumps are formed by electroplating followed by the removal of the metal layers. This is followed by spin-coating of the photo-thermosetting resin solution on the wafer with indium bumps. The resin is prebaked at 90 C for a few minutes to form a 9 to 16 micrometers thick film.

The film thickness is greater than the indium bump height. The negative-type photo-thermosetting resin is patterned and the resin is selectively removed from the indium pads. The chip is first bonded to the glass substrate at low temperature (less than 150°C), then inspected (and repaired if necessary). The final bonding is performed at 100 C during exposure to ultraviolet light at intensity of 2J/sq cm for about 20 seconds.

Another bonding technique developed for use with LCD panels is the fine-pitch chip-on-glass technique illustrated in Figure 5-6. An IC chip with gold bumps is immersed in an indium alloy bath heated to 200-230°C for a few seconds, in a nitrogen atmosphere without flux. The average height for the Au bump is 18.5 micrometers while that for the Au core indium alloy bump is 23.8 micrometers with a minimum pitch of 50 micrometers. The chips can be mounted directly to Mo, Ni, Au, and ITO patterns without flux at low pressures (30 g/bump or less) and low temperature (110°C or less).

5.4.5 Bump Transfer

This technique involves the transfer of solder from one substrate to another, e.g., chip-to-chip, wafer-to-wafer, wafer-to-chip. Advantages for this method include cost improvement, easy application for delicate substrates, and different size structures of substrates.

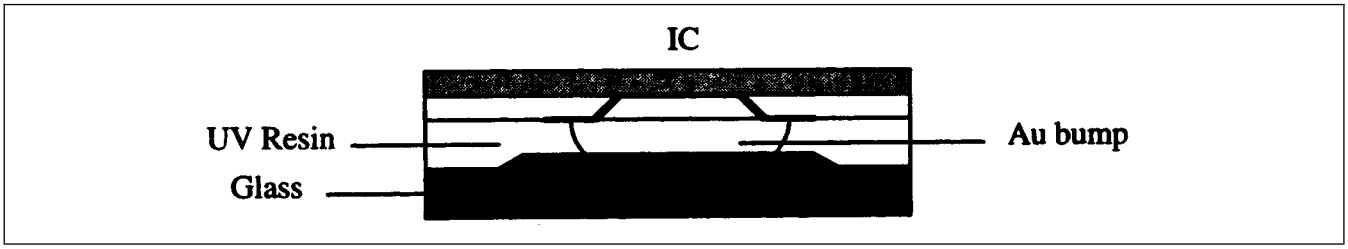


Figure 5−4¬ Gold Bumped IC Adhesively Attached to a Glass Substrate

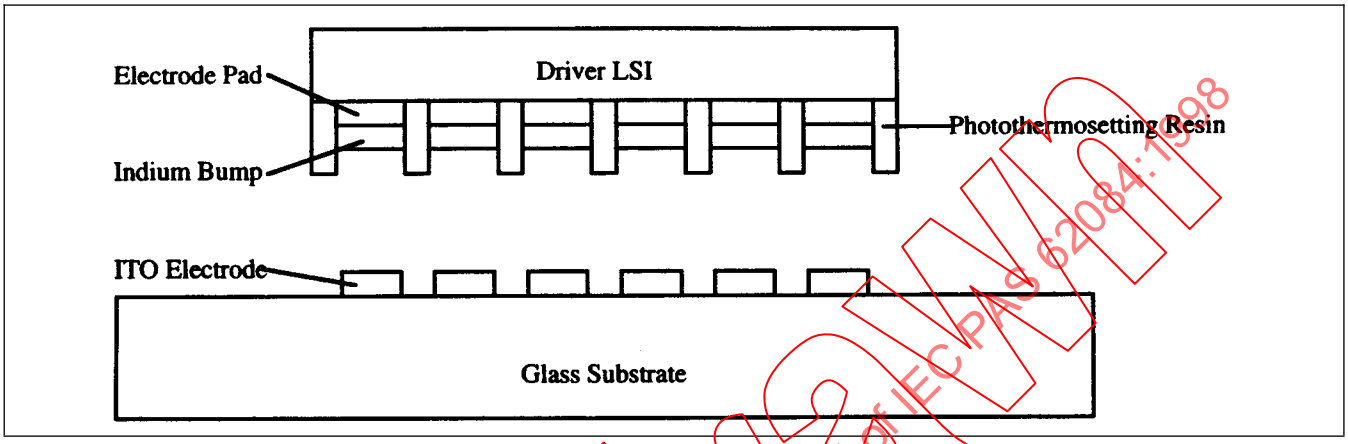


Figure 5−5¬ Newly Developed Chip on Glass Technology

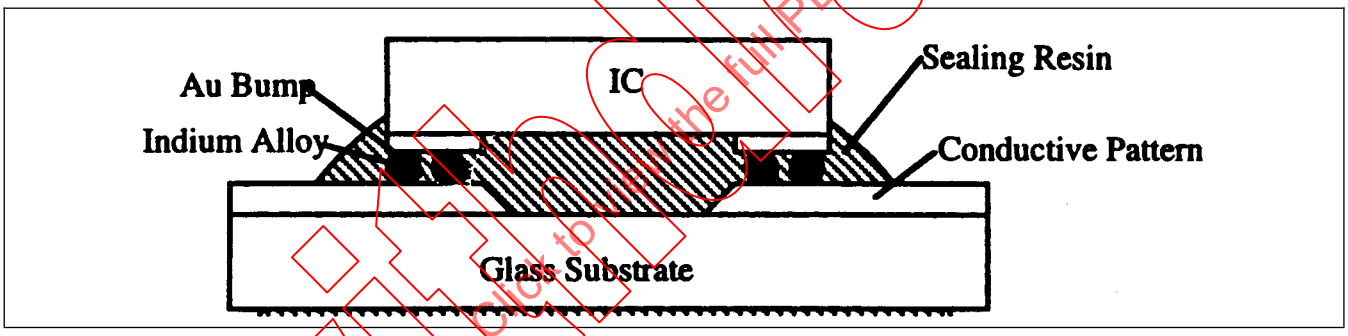


Figure 5−6¬ Schematic Cross-Sectional View of Connection for Chip on Glass

5.4.6 Stud Bumping (Ball Bonding)

Stud bumping is a variant of traditional wire bonding in which the ball is formed on the bond pad and the wire is broken off to leave a short tail. The tail is then flattened or coined to create uniform bump height. Bumps can be made from gold or solder wire. They cannot be used over active circuitry.

5.5 Chip Materials

From the chip point of view, the primary materials of importance are those concerned with protecting the chip (the dielectric interface) and those involved with making the electrical connection (the terminal metals).

The dielectric must provide a material barrier to the solder. Typically this would mean that the dielectric is not wettable to the solder. Since the solder will enter a liquid phase, it is also important that the dielectric be relatively pin-hole free. The dielectric must be stable at the solder reflow temperature and should provide low stress environment after this reflow. In addition, the passivation must be compatible with underfill, glob top, or other assembly-level passivants.

Because the lead will have some portion contaminated with alpha-particle-producing isotopes, it is desirable that the dielectric interface also provide a barrier to alpha emissions from these isotopes.

The terminal metals must provide an electrical and mechanical interface for the solder. The three requirements for determining the proper terminal metallurgies would be:

1. Adhesion to the chip
2. A diffusion barrier
3. A wettable layer

The terminal metals must provide adhesion to both the underlying chip metallurgy and the dielectric. The adhesion layer generally has a higher resistance than the underlying

metal, and therefore its thickness is kept to a minimum.

It is important that the interface metals be wettable but not be entirely consumed during the reflow and assembly processes. This requirement may be satisfied with the use of a diffusion barrier. While this diffusion barrier material is wettable, it is generally much less reactive with the tin than the wettable layer. Therefore, some of this diffusion layer remains in order to accommodate further reaction with the tin or other materials, during subsequent high temperature cycles should rework be required.

5.6 Other Bumping Process Considerations

5.6.1 Redistribution

In an ideal world, chips would be designed from the ground up for flip chip and chip scale attachment. Dice that have not been designed for solder bumping require additional processes that reconfigure the existing I/O geometries and/or relocate them from the perimeter to a centrally located area array. The latter process is typically referred to as a redistribution. Redistribution allows the use of existing wire bond or TAB products in flip chip applications. Chip scale grid array technology is one means of achieving this. Figure 5-7 illustrates how the peripheral leads can be redistributed inwardly to create a chip scale area array package.

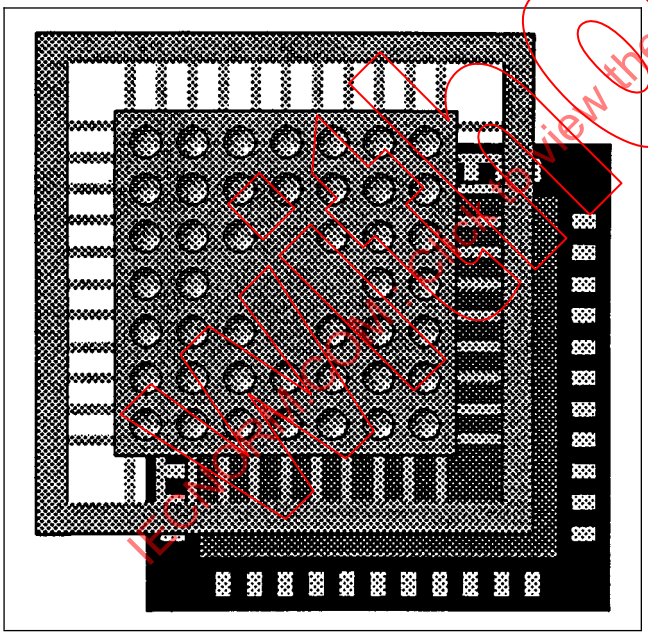


Figure 5−7¬ Redistribution of a Peripheral Pattern

5.6.2 Electrostatic Discharge (ESD)

While the idea of redistribution or the physical processing necessary to provide redistribution is not difficult, there are some fundamental problems associated with its implementation that need to be addressed.

Foremost among these would be the problem of Electrostatic Discharge (ESD). ESD would be of special importance where long redistribution metal traces would provide a high impedance path as compared to the capacitive path provided by the intervening dielectric. A schematic diagram of this is shown in figure 5-8.

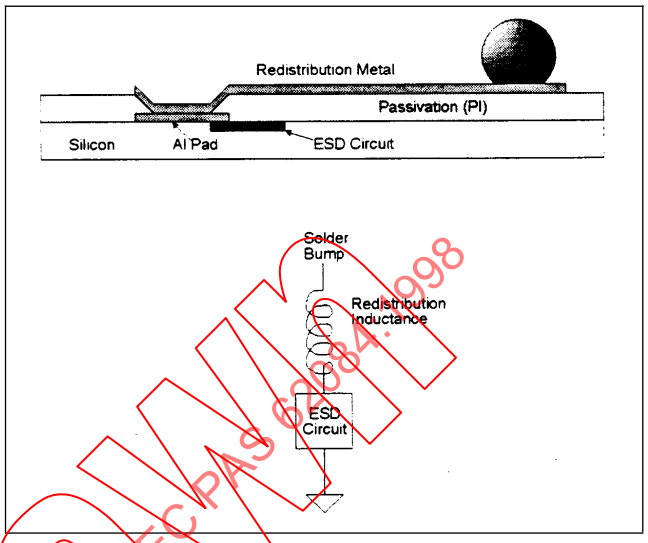


Figure 5−8¬ ESD Path to Ground

5.7 Handling, Shipping and Storage

Bumped dice are shipped using tray, waffle packs, tape and reel, and wafer frames depending on flip chip assembly method. A temperature-controlled dry air environment is necessary for proper storage and handling.

6 Mounting and Interconnection Structures

The steady advance of electronic packaging technology continues to have a significant technological impact on the substrates used to interconnect today's electronic assemblies. Mounting and interconnection structures for flip chip and chip scale packaging have special demands and requirements placed on them that in many cases exceed those placed on such structures used in more common applications such as traditional through hole and surface mount.

The expanding interest in flip chip and the other newer packaging approaches has created a technological need for finer circuit lines and spaces, smaller interconnection vias and plated thru holes, lower dielectric constant materials, and more strenuous thermal management requirements, as well as many other considerations. The interconnection structures supporting these advanced, first level electronic packages have become key limiting elements to the realization of the package's maximum performance potential.

Presently, the cost and performance demands of electronics are trending in opposing directions. That is, the price pressures of the expanding global economy continue to drive prices down, while the demand for ever higher levels of performance maintains a steady upward pressure . Flip chip

and chip scale packaging technologies hold out the promise of meeting such demand by minimizing the cost of electronic packaging by reducing packaging costs to levels that may ultimately be only slightly higher than the cost of the chip itself. This combination of factors has created new pressure on the next level of interconnection to facilitate interconnection of such densely packed chips. Figure 6-1 illustrates the general capabilities of varying substrate manufacturing technologies in terms of their area efficiency relative to the utilization of active silicon.

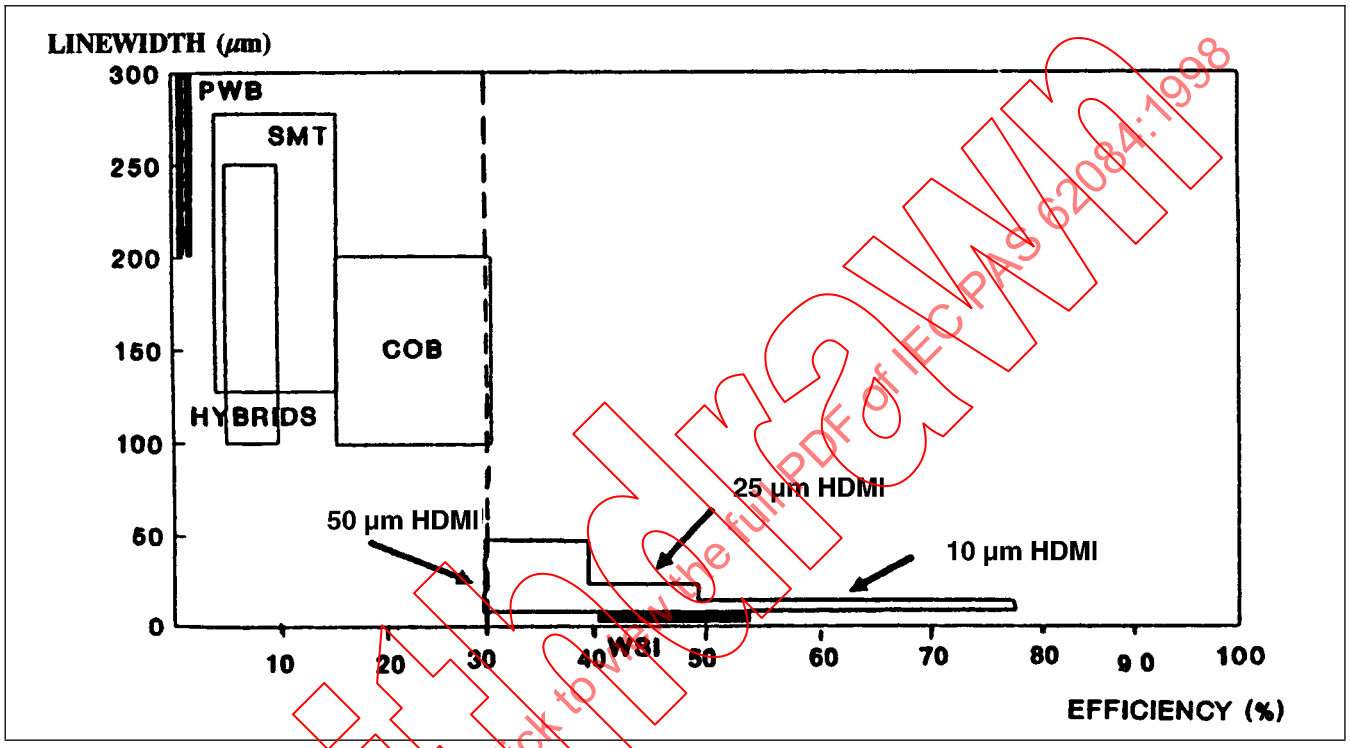


Figure 6–1¬ Packaging Efficient

Today, it is rapidly becoming apparent that a significant barrier to implementation of these new chip scale packaging technologies is located with the substrates to which they must be interconnected. Such advanced chip packaging technologies make across-the-board demands on the substrate, from design through assembly. There are, however, several areas of overriding importance.

Some of these areas of importance are related to the mechanics of assembly, such as surface topography, planarity, flatness, as well as to the properties of the finish that make it best suited to making reliable interconnections. Others relate to the wiring of the substrate itself and what methods can be employed to create interconnections within those substrates that do not sacrifice valuable area unnecessarily and with it the loss of some measure of improved performance. Still others relate to the thermal demands that could be placed upon the substrate when closely packed ‘‘tiling’’ of silicon results in excessive heat generation on the substrate. This section will discuss these and other concerns related to the creation of substrates for flip chip and chip scale packaging.

6.1 Background

As stated in Section 2, it was IBM that introduced flip chip technology as part of Solid Logic Technology (SLT), the component technology for System 360. The flip chip was a single transistor device with three small copper balls encased in eutectic, solder mounted on a half-inch square hybrid pin-grid ceramic substrate.

Several such devices were used on each substrate to form one logic circuit. The conductors on the substrate were screen-printable thick-film with glass dams to prevent the solder on the chip connection from wicking down the conductors. The substrate was dry-pressed and fired alumina with staked pins for I/Os. On some substrates, wiring crossovers were accomplished by printing a line of glass dielectric over the lower conductor.

In the late 1960s, as integrated circuits were introduced, the copper ball was eliminated, and the surface tension of the molten solder controlled the chip stand-off. This was the first real controlled collapse chip connection (C-4).

The substrate technology became the key to advances in the flip chip technology for two reasons. First, as the demand for more I/Os off the chip grew, the ability to bring wiring out from the bump array (so-called escape) became critical. Second, as the chips increased in size, the issue of the difference in coefficient of thermal expansion (CTE) between the ceramic substrate and the chip could limit the

number of chip I/Os and chip size.

Single layers of thick film conductors quickly became limited in the ability to ‘‘escape’’ wiring from under the chip. This fostered the advent of a thin film wiring technology using chromium-copper-chromium metallization deposited by evaporation or sputtering and photolithographically defining conductors down to 25 µm (.001″) lines and spaces. The substrate itself remained dry press alumina and continued to use staked pins for its pin grid array.

The top chromium layer provided the solder-stop similar to the original thick film glass dams. This technology allowed for a significant increase in bumps per chip by allowing wiring to ‘‘escape’’ lands on the substrate top surface.

As the number of chip I/Os and the size of the chip increased, alumina ceramics continued to be the substrate of choice over organic because of the closer match of CTE between¬ alumina¬ ceramic¬ ($7x10^{-6}$/°C)¬ and¬ silicon ($3.5x10^{-6}$/°C).

The next substrate advance allowing essential ‘‘escape’’ routes was the multilayer ceramic (MLC) technology at the end of the 1970s. By stacking layer upon layer of screen printed ceramic green-sheet, it became possible to wire out any bump or flip chip footprint. The top surface of the MLC substrate had no conductors to the lands, which were all connected to the internal wiring by vias within the land configuration, thus eliminating the solder wicking. The practicable limit was the increasing cost of adding layers to the substrate.

In the late 1980s, low temperature cofired ceramics (LTCC) using mixtures of glass and ceramic gave the ability to closely match the substrate CTE to that of the silicon chip. This allowed for larger bump arrays. Later, photolithographically defined thin films, usually on top of the ceramic MLC, provided even more ‘‘escape’’ routes, reducing the number of layers of ceramic required. One disadvantage of LTCC substrates is the low thermal conductivity as compared to alumina ceramic.

Beyond ceramic substrates, there are a number of other viable approaches to creating interconnection substrates for flip chip and chip scale packages, most of which are based on organic material . Each of the alternate choices comes with its own set of advantages and disadvantages, and each substrate has its own set of constraints with respect to the assembly method that can be employed. Which interconnection structure-base material combination is the correct one is very much a matter of individual product requirements and the final choice is normally arrived at by considering the trade-offs of conflicting needs.

On first glance, it appears that the cost of such fine featured structures will be greater due to the limited manufacturing base capable of producing product with such advanced requirements. While it is generally true that such substrates may cost more per unit area, the area reduction allowed by the minimal form IC packages should pave the way for reduced overall costs of electronics. The balance of this section will address specific demands and requirements that will need attention before such advantage can be gained.

6.2 Mounting Structures General Considerations

The mounting structure or substrate for flip chip and chip scale packaging must be capable of simultaneously meeting the demands of several different design requirements. These include conductor routing needs, via structure chosen, layer count requirements, thermal management needs, etc.

Each of these must be woven effectively into the overall design in order for the product to be successful. Beyond these considerations, there are the mechanical requirements for the substrate as well. Data indicates that the thickness/rigidity of the structure can impact the reliability of flip chip interconnections through thermal cycling; thicker boards appear to fail sooner.

In the final analysis, the board design is driven by the design and layout of the chip scale package itself. The two cannot exist independent of each other, and therefore there must be common ground at which they can interface.

While ceramic substrates were among the first substrates to be successfully used for flip chip applications, and are still by far the highest volume in production, the use of organic substrates for flip chip and chip scale packaging has been facilitated by the addition of an underfill epoxy between the chip and the substrate. The underfill epoxy extends the fatigue life of the bumped joints by approximately an order of magnitude.

The substrate material chosen for chip scale packaging can be any of the normal printed board materials such as FR-4, BT resin, Dryclad™ or polyimide. Since most organic base laminates have glass transition temperatures well below the reflow temperature of the solder used for the chip bump, eutectic solder is used on the board to reflow around the solder bumps. Underfill is required in all cases, with the possible exception of flexible substrates, which are compliant in nature, or when there is sufficient compliancy in the chip package itself.

The most common structure for PWBs is one where vias are formed by drilling a hole into the board and plating the insides of the hole to make interconnections. This via, or plated through-hole (PTH), cannot be used as a reflowable solder bump mounting land, as the solder will wick into the hole and destroy the reliability of the joint. Therefore, a separate land is required that has no PTH, and is connected to a PTH land by a surface wiring layer. An alternative approach is the construction as is shown in figure 6-2.

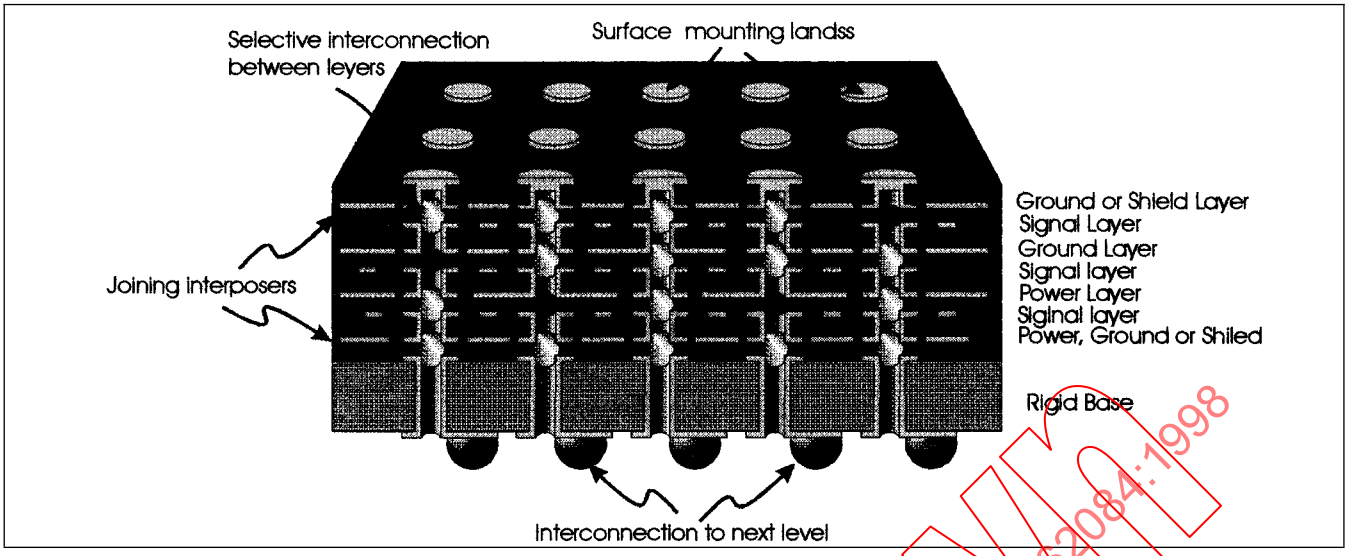


Figure 6−2¬ Exploded View of a Substrate for Flip Chip or Chip Scale Applications

Without such a construction, solder mask is required to prevent wicking.

Since the current practical limit for mechanical drilling capability is in the 0.03 mm diameter range, and PTH lands approximately 0.5 mm, fan-out or wire escapes becomes a very important consideration. Nonsquare arrays and perimeter land arrangements are often utilized. This can be the gating factor in the number of I/Os that can be wired out as the fan-out area increases with increased I/Os, and the land pitch needs to increase to allow more wire escape channels.

Because of practical printed board dimensions, the flip chip pitch is often extended to 350 µm. For chip scale grid arrays this pitch may be extended from 0.5 to 1.0 mm. A technique called photolithographically defined vias and fine lines, using additive techniques, is being used by certain manufacturers to allow wiring to more chip I/Os (more escapes).

As stated earlier, organic materials can be used provided that the chip is underfilled. Although underfill can have a dramatic effect, the CTE of the substrate still must be considered. For example, the amount of metallization, usually copper, affects the CTE of the substrate.

Multilayer thin film technology does allow for fewer escape layers, as more wiring can channel between bumps from the I/O array underneath. These techniques are often combined with more conventional and less costly interconnection techniques, e.g., power planes in MLC, to allow area array connections to the next level.

All else being equal, thin film conductors have higher electrical performance than their thick film counterparts. However, they are limited in current handling capability because of their thinness. They should be considered for applications over 200 megahertz. The major advantages of multilayer structures is that they can further increase wiring density and can create transmission line structures for high data speed.

The layout of the flip chip pattern has a material effect on the substrate to which it is attached. Often the center of the array pattern is directed toward voltage and ground. This allows these connections to go directly to the voltage and ground planes in the substrate without the need for escape routing.

Another technique is to widen the grid at the periphery of the array to allow for more interconnections that can escape. This can be accomplished by eliminating every other land. See figure 6-3.

Layout of the bump pattern on the flip chip package can significantly affect layout of the next level substrate. The same I/O count creates different demands on the substrate. The chip on the left is routable with coarse features of one conductor between lands while the pattern on the right requires finer features using two conductors between lands to complete the route.

In an idealized case all conductor routing would take place in a single row, with rectangular layouts favoring the wiring process. From the perspective of the substrate manufacturer, this allows either more channels for routing or accommodates the use of coarser features. In reality however, die are laid out in a manner to achieve the best possible performance from the smallest possible area. Figure 6-4 illustrates that concept.

Another important consideration for the selection of the substrate technology is the thermal requirement. Ceramics have far better thermal conductivity than any organic. Alumina is some five times higher in conductivity than the LTCCs, and significantly higher than most organics.

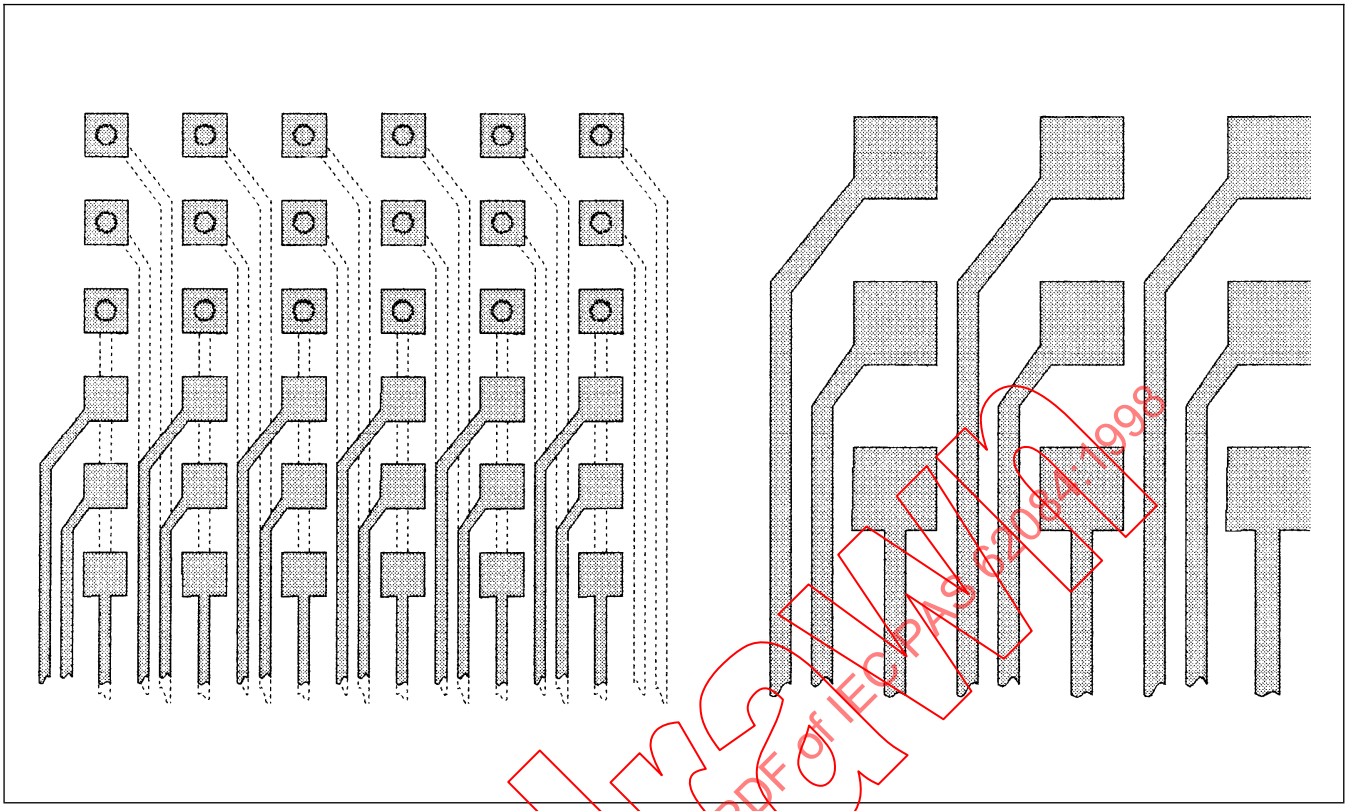


Figure 6−3¬ Wire Routing Ability of Different Design Rules for One and Two Sided PWBs

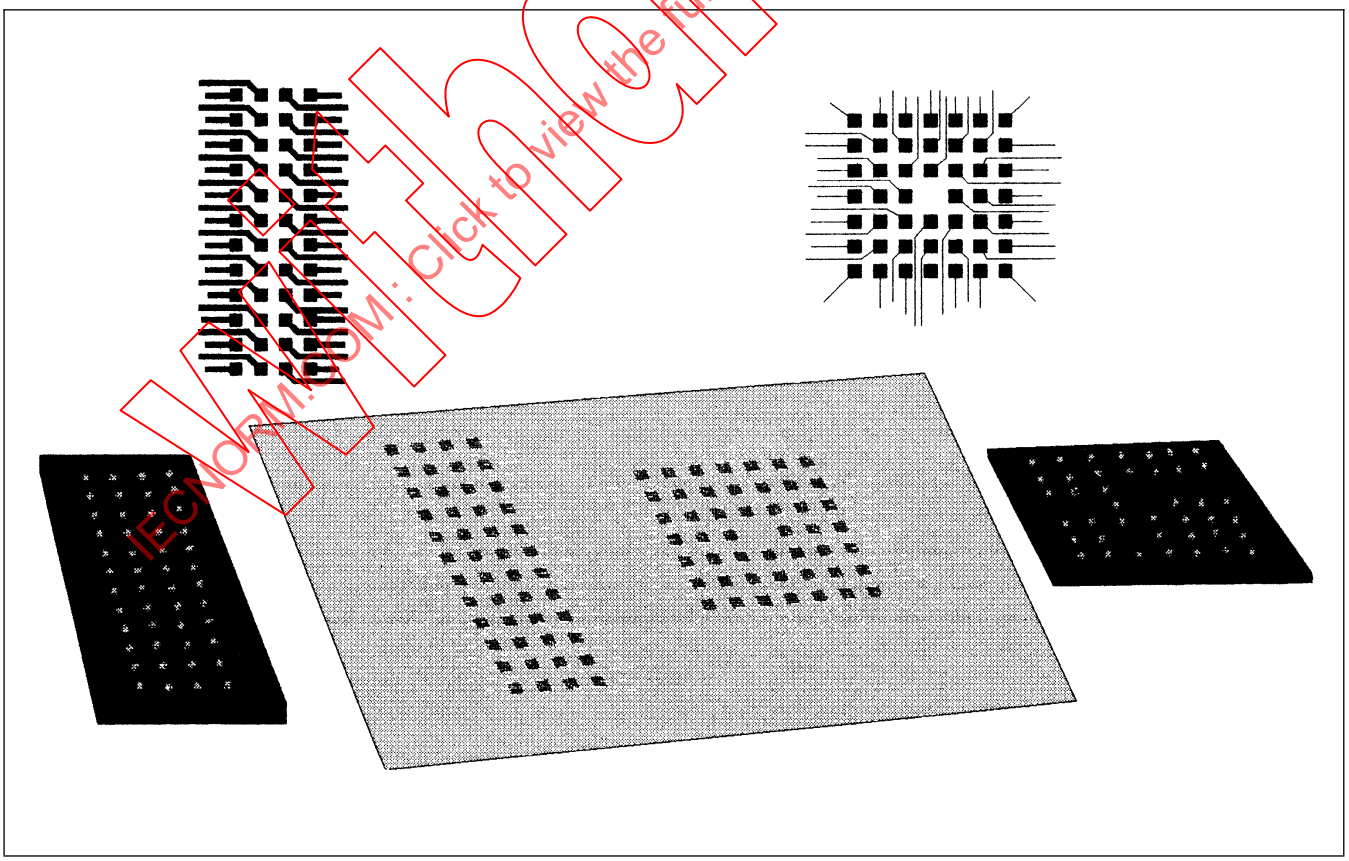


Figure 6−4¬ Conductor Routing Comparison

6.3 Interconnection Substrate Material Choices

As indicated in the preceding introduction and background discussions, there are numerous potential choices available, for use as substrates for minimal form chip packages. Following is a review of some of those choices.

6.3.1 Rigid Organic Substrates

Rigid organic substrates are the main type of material used in the construction of PWBs. There is a large and well established base of manufacturers available, and thus the cost effectiveness of these types of structures is expected to remain high. This will be important if flip chip and chip scale packaging technologies are going to maximize their potential influence.

Rigid organic substrates are the product of a marriage of organic resins and some type of reinforcement system. The resin system provides important electrical and mechanical properties of the finished laminate and serves as the glue that holds the laminate together. The reinforcement material, on the other hand, is the ‘‘backbone’’ of a laminate structure, providing the strength and dimensional stability required to make the laminate a viable interconnection structure. The reinforcement also contributes to the electrical properties of the laminate and can influence manufacturability if not selected with care. Table 6-1 provides a comparison of some selected properties, e.g., glass transition temperature (Tg), of some of the available candidate materials available for use with flip chip and chip scale package assemblies.

6.3.1.1 Epoxy glass

Flame retardant epoxy (FR-4) has been the workhorse of printed board resins for many years. A thermoset resin (requiring heat to cure and harden), flame retardant epoxy has an excellent blend of properties that make it well suited to most electronic applications. For traditional flip chip applications these materials are only capable of meeting the demands of certain applications, and then only with underfill.

Most laminates use a fiberglass cloth as the reinforcement material. E-glass or electrical glass is the most frequently employed for this purpose, due to its very low content of soluble ionic components. E-glass is available in variety of weave styles and thicknesses in order to meet the varied requirements of multilayer printed wiring products.

In addition to E-glass, another type, S-glass or high strength glass, is also available. S-glass is lower in dielectric constant (values of 4.5 - 5.2 have been reported for S-glass versus values of 5.8 - 6.3 for E-glass) and may prove useful for certain high speed applications.

Table 6−1¬ Comparison of Selected Material Properties

Base Material (No Cladding)	Coefficient of thermal (x-y) expansion (ppm/°C)	Dielectric Constant	Glass Transition Temperature °C
Epoxy-Glass¬	16 - 19¬	4.2 - 5.0¬	120 - 130
High Temperature Epoxy-Glass	17 - 18¬	4.3 - 4.6¬	150 - 180
BT/ Epoxy-Glass	15¬	4.0 - 4.2¬	185 - 195
Cyanate-Ester Glass	12 - 14¬	3.5 - 3.7¬	240 - 260
Epoxy-Aramid¬	6 - 9¬	3.7 - 3.9¬	120 - 130
Polyimide-Glass	14¬	4.0 - 5.0¬	240 - 250
Polyimide-Quartz	6-8¬	3.4 - 3.5¬	240 -250

6.3.1.2 Polyimide glass

Polyimide glass based materials offer the highest temperature resistance (Tg 260°C) of all normally available resins for printed circuit manufacture. Polyimide has been traditionally most favored for military applications where there is an anticipation that a product must go through several repair cycles, with the attendant thermal stresses, over the life of the product. This same feature makes it an attractive alternative to meet the needs of traditional flip chip assembly when compared with other resin systems. However, underfill is still advisable. While polyimide resin tends to be more costly and more demanding to process than other resins, it is well suited for thermosonic wire bonding of devices. The only major detraction of polyimide is that it has a greater tendency to absorb water than some of the other resins, causing changes in electrical properties, but its advantages tend to outweigh this deficiency.

6.3.1.3 BT-Epoxy Glass (Dryclad variant)

BT or bismalimide triazine/epoxy is another high temperature resin system with the added advantages of a relatively low coefficient of thermal expansion (approximately 70% of epoxy) and a lower dielectric constant when compared to normal epoxies (3.5 vs 4.0). The material is intermediate in cost between epoxy and polyimide and may also serve the needs of flip chip with underfill processing.

6.3.1.4 Cyanate Ester

Cyanate Esters, though somewhat costly, appear well suited to serve in applications where good electrical properties and good thermal performance are sought simultaneously. This is evidenced by the resin’s low coefficient of thermal expansion (which is roughly equivalent to polyimide) and its low dielectric constant (2.9 vs. 4.0 for epoxy and 3.7 for polyimide). Cyanate ester resins also tend to be tougher, easier to process than some of the alternatives, and offer better adhesion.

6.3.1.5 Teflon

Fluoropolymers, such as duPont's Teflon® , offer the lowest dielectric constant values (2.0) of all normally available resin systems for printed wiring laminates. The material is also, however, quite soft, has a low glass transition temperature (Tg) and also has a fairly large coefficient of thermal expansion which may require some extra effort in circuit design to accommodate its weaknesses. This may cause some problems also when flip chip is involved because the underfill may not be sufficiently well bonded, limiting its efficacy as a strain-relief mechanism.

6.3.1.6 Polyimide Quartz

Polyimide resin has been successfully mated with quartz cloth for certain high performance applications. The quartz cloth was used as a reinforcement material when tightly controlled dimensional tolerances and a low coefficient of thermal expansion were sought. It may prove of some value in flip chip applications, but there is limited historical application data available. Another difficulty is that quartz cloth material is very expensive and extremely difficult to drill, normally requiring diamond drill bits to process.

6.3.1.7 Corelam™ (Epoxy or Polyimide/Aramid)

Corelam is a substrate created by laminating aramid fibers cloths such as duPont's Kevlar® with a chosen resin. The finished product is one that has a low thermal expansion rates in the X and Y axes. The low dielectric constant of the material and light weight when compared to glass cloth based laminates makes it an attractive alternative for those applications where such a mix of properties is advantageous.

A disadvantage of aramid fiber reinforced laminates is that they have a fairly high rate of thermal expansion in the Z direction below glass transition (90 ppm/°C vs. 60 ppm/°C for FR-4); consequently, plated through hole reliability could be at risk if care is not exercised in design, manufacture and assembly. In addition, drilling and routing of the laminate can be difficult and thus costly due to the toughness of the aramid fibers, which leave small fibrils on cut surfaces.

6.3.1.8 High Performance Epoxy

High performance epoxy laminates were developed to fill a niche between the more expensive high performance resins and the lower cost regular epoxies. These laminates are often employed when higher temperature performance is required, but cost is still important as well. The operational temperatures of these resins are intermediate between epoxy and polyimide; yet their cost does not usually justify the added expense associated with the increased temperature capability of polyimide. The electrical properties of these resins are nearly identical to those of normal epoxy.

6.3.1.9 Solder Resist

While technically not a substrate material, solder resist can play a substantial role in defining the properties of the finished product. Solder resist, or solder mask as it is also commonly known, is applied to the circuit both to protect the circuit traces from being soldered as the name implies and also to physically protect the circuits from environmental damage after the assembly process has taken place. The material is available in both liquid and dry film forms. Circuit manufacturers differ in manufacturing approach and thus the type of solder resist used can vary between vendors. Liquid, screen printed types are the ones most commonly specified and used. However, when the feature size requirements become more demanding, such as for chip scale packaging, the use of photoimagable-type solder masks is often mandated.

6.3.2 Flexible Laminates

While there are a number of available choices for flexible laminates, polyimide films are normally viewed as the material of choice for high-end/high performance flex circuit applications. Polyimide films are used in diverse applications as disk drives, cameras, missile systems, video tape players and computer mice. They have also been the subject of much interest for flip chip and chip scale packaging due to their natural compliance.

Polyimide films for flex circuit applications normally consist of base film and adhesive (although adhesiveless forms are also available). Many adhesive choices are available, such as epoxy, acrylic, and polyimide. The most commonly used are acrylic-based thermosetting adhesives. Such adhesives allow polyimide flexible circuits to be soldered with relative ease.

Polyimide films offer excellent performance over a wide range of temperatures, from cryogenic to soldering temperatures and beyond. While the material tends to absorb more moisture than polyester, polyimide is at least equivalent in nearly every other performance category, from flexural endurance to tear strength.

In addition polyimide films offer a unique advantage, in that they can be chemically milled, an advantage well understood by manufacturers of TAB tape. This means that it is possible to create the equivalent of TAB bonding sites directly in the circuit. Admittedly, other methods can be used to create the same effect, most notably eximer or CO2 lasing, but such methods normally require a cleaning step to remove charred edges.

6.3.3 Inorganic Substrates

Inorganic substrates were the early choice for flip chip applications due largely to the good match in CTE. There are a number of different inorganic materials that have been developed over the years for electronic applications. The choice of which material to use is based on a number of

Table 6−2¬ Inorganic Substrate Characteristics

Material	Tensile Modulus MPa	Coefficient of Thermal Expansion (ppm/°C)	Tensile Strength Pa	Thermal Conductivity Units	Dielectric Constant (1 MHz)¬	Dissipation Factor
Alumina (90%)¬	324.1¬	6.7¬	317.2¬	16.7¬	9.4¬	*
Alumina (96%)¬	324.1¬	7.1¬	317.2¬	25.1¬	9.9¬	0.0001-0.0002
Alumina (99.6%)¬	344.8¬	6.3¬	448.2¬	37.4¬	10.0¬	0.0001-0.0002
Aluminum Nitride¬	291.7¬	4.3¬	367.5¬	170.0¬	8.8¬	*
Beryllia¬	319.4¬	8.0¬	*¬	250.0¬	7.0¬	0.0001
Quartz¬	72.4¬	0.7¬	*¬	0.8¬	3.8¬	0.0002
Sapphire¬	344.8¬	7.7¬	0.0¬	25.0¬	0.0¬	0.0002
Porcelainized Steel¬	*¬	4.4¬	89.6¬	1.7¬	5.5¬	*

factors including thermal conductivity, electrical insulation properties, mechanical strength, thermal expansion, chemical resistance, weight, ease of metallization, and cost. Table 6-2 contains a comparison of some selected properties of the various inorganic substrates available. Following are descriptions of some of those alternatives.

6.3.3.1 Aluminum Nitride

Among inorganic substrates of the ceramic type, aluminum nitride is favored for its very high thermal conductivity. It also has a very low CTE.

6.3.3.2 Alumina

There are four common formulations for substrates broadly defined as alumina or aluminum oxide (Al_2O_3). The deciding factor is the aluminum oxide percentage in the final product. The four common percentages are 90%, 92%, 96% and 99.6%.

6.3.3.3 Co-fired Multilayer Ceramic

This class of substrate is commonly used to produce thick film circuits. The uncured alumina substrates are each printed with a different circuit pattern and then laminated into a multilayer structure. The product is generally considered to be more expensive than some of the alternative approaches.

6.3.3.4 Beryllia

Beryllia, a polycrystalline form of BeO implanted with impurities to improve physical properties, is a very lightweight substrate that also offers exceptional thermal transfer for an inorganic material. The toxic nature of the raw material, as well as its cost, has limited its acceptance for use in a broader array of applications.

6.3.3.5 Silicon

Silicon is the ideal substrate for flip chip applications in terms of matched properties, especially CTE, which is commonly viewed as the greatest concern. It is also a very expensive and very delicate alternative, limiting its appeal to the broader class of users.

6.3.3.6 Metals

There are several potential candidate metals that can serve the needs of flip chip and chip scale packages. Porcelainized steel, for instance, has very high thermal conductivity and can provide a reasonably close match for the CTE of a silicon IC.

6.4 Surface Finish Properties

The surface properties of the substrate chosen for flip chip and chip scale packaging can be quite critical to success. The finish chosen, the base metallurgy, the surface topography, substrate flatness, wettability of the finish, surface cleanliness as well as methods of production, are factors that must be considered. In some areas, the requirements for each of these will vary with the packaging method chosen. In other areas, the requirements will be very consistent.

6.4.1 Chemical plating finishes (electrolytic and electroless)

Plated metals are applied to the surface of interconnection substrates for various reasons, the most obvious being to provide a pathway for the electronic signals. However, there are many secondary reasons such as providing a noncorroding, separable contact surface or allowing for an easily joinable metallurgical finish. While many metals can be employed in a finished circuit, only a few are used universally. They are copper, nickel, gold and tin-lead (solder).

6.4.1.1 Copper

Copper plate is applied to interconnect the printed wiring from side to side on double sided boards and to internal layers in the case of multilayer boards. The plating is normally set down in two layers, first a thin electroless copper 0.5 to 1.0 µm (20-40 microinches) seed layer and later a more robust electrolytic copper that is overplated to meet electrical and reliability requirements.

A plating thickness of 25 µm on the hole wall is a common design rule. The plating should also be capable of meeting appropriate performance requirements, such as being able

to withstand both the 10 second, 288°F solder float thermal stress test and the −55°C to +125°C thermal cycling test called out in MIL-P-55110 (the military performance specification for rigid printed wiring boards).

6.4.1.2 Nickel

Nickel is normally plated over the copper circuit and under the gold finished surface. Its function is basically two-fold: 1) Nickel acts as a barrier layer, preventing the interdiffusion of copper and gold that naturally occurs over time. This is to ensure that the integrity of the gold, either as a contact or a joining surface, is maintained. 2) Nickel serves as an ‘‘anvil’’ to enhance the strength of the surface gold plate and minimize galling or fretting of the surface when it is used as a contact finish. A commonly called out requirement for nickel is 5 µm.

6.4.1.3 Gold

Gold, being a noble metal, is a near ideal contact finish. However, with different types of assembly technologies, it can also be employed as a joining finish. For instance, with COB and gold leaded TAB the devices can be bonded to the board using single-point thermosonic bonding. The plating requirements in such cases is quite different from the type used for separable contacts. Contact gold plate thickness requirements are normally on the order of 1 µm and the deposit is commonly alloyed with metals such as cobalt or nickel to increase wear resistance.

For bonding purposes, the finish thickness cited previously will serve; however, the deposit should be 24 carat pure gold in order to achieve reproducibility and reliability in bonding. Finally, when used as a solderable finish such as for flip chip or chip scale interconnection, the gold should be kept very thin (<0.25 µm) to prevent creation of brittle solder joints due to excessive gold in the alloy (>3% by weight).

6.4.1.4 Tin-Lead

Tin-lead plate is often referred to as solder plate; however, the latter name is technically not appropriate until fusing of the ‘‘grains’’ of tin and lead in the deposit has taken place. This finish is the most commonly used circuit finish employed today. Key concerns for assembly of flip chip and chip scale package devices employing tin-lead platings center on plating thickness and planarity.

Most assemblers expect plating to occur evenly over the surface of the board, thus leading to a uniform plating finish. It fact, plating can be quite uneven, and it is heavily influenced by circuit layout and design. Thus, isolated areas will plate more heavily than more densely packed design features.

6.4.1.5 Copper Foil/ Film

Copper is the most common metallic coating for printed board laminates. It is available in both foil and film forms. Among foils there are eight different recognized types of copper. These are broadly classed as either 1) wrought (rolled) and annealed 2) electrodeposited types. Special treatments are often applied to enhance the properties of the foil for such applications e.g., when high temperature ductility is required.

In addition to the foil types of copper, there are film type copper coatings. These are typically deposited by vacuum processes such as sputtering and are normally very thin (i.e., less than 1 µm). These films can then be plated up to the desired thickness with copper, either in panel or in circuit patterned form. This method also allows for the production of very fine line circuits required for chip scale packages.

6.4.2 Thick film metallic finishes

Thick film metallic finishes have the same general requirements as their plated counterparts in that they must meet the general requirements for conductivity, wettability, flatness, etc. Following are brief discussions on these materials and processes that may be called upon to serve as substrates for flip chip and chip scale packaging. One of the areas of difficulty for this branch of interconnection technology is the inability of the technology to meet the requirements for planarity and line width feature limitations.

6.4.2.1 Thick film deposition processes

Copper, silver, gold, and palladium have all been employed as conductor metals for thick film applications. Typically, these materials are screen printed in the pattern of the desired circuit onto the surface of the substrate (normally ceramic) in a paste form with a binder of some sort. Following the screening process, the ceramic base with the screened circuit pattern is fired in a kiln to create a permanent bond between the circuit and the base material.

6.4.2.2 Polymer Thick Film

Another means of applying thick film circuits to the base material is by means of a screen printable and photocurable metal-bearing polymer. The process is commonly referred to as polymer thick film (PTF) processing.

6.4.3 Sputtered (Thin Film) Conductive Finishes

This branch of interconnection technology has traditionally been limited to the realm of ceramic based materials; more recently, however, the process technology has been employed in combination with organic films. One area where this is most in evidence is in the arena of the multichip module (MCM). Basically, the process involves the dry deposition of metals from a target of the desired metal onto the substrate by means of a gas plasma.

6.4.3.1 Sputtered Metal Deposition Process

Using the equipment described above, it is common practice to first sputter an adhesion layer on in advance of the metals used to provide the required conductivity to the circuit. Adhesion-promoting metals include: nickel(Ni), molybdenum(Mo), chrome(Cr), tungsten(W), and titanium(Ti). These materials are then commonly overplated with a more conductive metal such as those used for standard PWBs. These would include copper, gold, tin and palladium

6.4.3.2 Thin Film Subtractive Process Image Patterning

Subtractive processing is performed by sputtering the entire surface with the adhesion and conductor metals, imaging the substrate with a suitable resist and then etching the exposed metal to create the circuit pattern. Feature limits normally lie with the resolution capabilities of the imaging process.

6.4.3.3 Additive process image patterning

Additive processing of the substrate is accomplished in a fashion similar to that described for subtractive except that a negative image of the circuit pattern is left on the substrate. This pattern is then plated with metal to the desired thickness. The resist is then stripped and the thin background copper is etched away, leaving the circuit pattern on the substrate. As with subtractive processing the limits of feature size are held by the imaging process. These processes can be used both on ceramic and organic substrates as with thin film subtractive.

6.4.4 Alternative Protective Coatings for Conductors

Currently, product designers are opting for bare copper circuits that are protected with organic protective coatings (OPCs). This approach leaves the soldering lands finished with a thin coating with whatever innate planarity that may exist. It also facilitates the application of solder paste which is normally applied just before assembly. There are two general categories of materials suitable for these applications. They are alkylimidazole and benzotriazole. These materials are capable of protecting the solderability of the copper plate for several months under normal storage conditions, however, high temperatures can rapidly degrade the protective coating.

6.5 Substrate Constructions

There are a number of different ways of constructing a substrate that will meet the needs of a flip chip or chip scale package. Following is a brief review of some of the options that are available to the potential user.

6.5.1 Rigid Printed Wiring Boards

Rigid printed wiring boards are the standard of the interconnection industry and as such serve as the most widely available interconnection method. These interconnection structures are available in an array of shapes, sizes and interconnection densities, from simple single-sided printed wiring boards to complex multilayer structures such as MCM-L. Printed wiring technology is by far the most commonly chosen technology for interconnecting electronic devices.

6.5.1.1 Single-Sided PWBs

The first form of printed wiring board, the single sided board is still practical today for many applications, including certain low I/O flip chip applications. The advances in imaging processing technology have made it possible to get extra use out of single metal layer designs.

6.5.1.2 Double-Sided PWBs

Double-sided PWBs have two conductive metal layers normally interconnected by means of plated-through holes. As with single-sided boards, fine line processing capabilities of today can in certain cases can allow a two layer design to preclude the need for a multilayer construction.

6.5.1.3 Multilayer PWBs

Multilayer boards have three of more layers of circuitry. These are commonly used as the primary means of electronic interconnection in today s sophisticated computers. MLBs of more than 50 layers have been created to achieve the desired degree of interconnection. Relative to the needs of flip chip technology, multilayers allow for designs that are more densely routed than might be achieved using lower layer count constructions.

6.5.1.4 Metal Core PWBs

Metal core PWBs have historically been used in cases where high power devices were employed and thermal management was of paramount importance. Such structures were commonly enameled or porcelain coated steel onto which the circuit pattern was applied. However, in the late 1970s and early 1980s and the advent of surface mount technology, the technique began to be used as a means of both extracting heat and managing the coefficient of thermal expansion of standard multilayer boards to protect the life of solder joints on the boards. Copper-clad invar and copper-clad molybdenum are two of the more commonly used core metals. They can also serve as functional ground and power layers when designed properly, increasing their utility.

6.5.1.5 MCM-L Constructions

MCM-L circuits have much in common with standard printed boards. The major differences lie in the feature sizes (0.05 mm conductor width and clearance) and the finish (typically bondable gold). MCM-L boards tend also to be small and use higher temperature laminates. They

have the potential to be the most viable MCM alternative available today because of the large installed vendor base.

6.5.2 Flexible Printed Wiring Boards

A technological variant of the standard rigid printed wiring board, flexible printed wiring boards are interconnection structures that are created on flexible base materials. Like their rigid counterparts, they can range from simple single metal layer devices to multilayer structures of great complexity. Because of the very uniform surface they possess, these materials have used for the production of very fine features (e.g., 50 µm line and space). In addition because of their natural compliance, flex circuits have been successfully used for direct flip chip attach applications.

6.5.2.1 Single-Sided Flexible Circuits

Single-sided flex circuits are the simplest form of the technology. A single metal layer is employed and interconnected from one side only. With a coverlayer in place they are most commonly used for dynamic flex applications and are well suited in low I/O count chip scale packages.

Such simple flex circuits with fine features and 0.5 mm pitch can allow for routing three rows deep for chip scale packages with 50 µm features used in circuit design.

6.5.2.2 Double Access Flex Printed Boards

Double-access flex circuits, like single sided flex, have only a single metal layer but the circuitry can be accessed from both sides due to special processing that provides opening through the base film. These substrates can be used to mount circuits back to back. They offer alternatives for chip scale packaging that might not otherwise be available.

6.5.2.3 Double-Sided Flex Printed Boards

Double-sided flex circuits allow greater density without greatly sacrificing either flexibility or adding greatly to substrate height. They can be used in certain dynamic applications if designed properly (i.e., single metal layer through the bend areas). Plated-through holes are commonly used to provide connection from side to side. Such structures when applied to chip scale area array packages, using 0.5 mm pitch and 50 µm feature design rules, allow for wiring to be routed six rows deep. Thus a 6 mm square die with 144 I/Os can be fully routed using two metal layers.

6.5.2.4 Multilayer Flex Printed Boards

Multilayer flex circuits are most commonly found in the more complex applications. Military applications are common. Special understanding of design is required to take greatest advantage of this useful technology. As with double-sided flex circuits, multilayer designs can provide some dynamic capability but are best reserved for intermittent flexing applications. Constructions that are laminated using flexible inner layers and interposers as shown in figure 6-2 should prove capable of serving the highest performance applications at relatively low cost. Multilayer flex printed boards are often used when power and ground meshes (planes are too stiff) are required for improved signal quality.

6.5.2.5 Rigid-Flex Printed Boards

Rigid-flex boards are a hybridized form of printed board and offer the best of both worlds to the designer. Like multilayer flex they are most often used in more complex constructions such as compact back panel bus systems. The military and aerospace industries have been avid users of the technology for the enhanced reliability it can offer when compared to the complex wiring harnesses they commonly replace. These constructions have also been used with chip scale packages to create miniature interconnection structures such as for hearing aids.

6.6 Thermal Requirements

For flip chip and other chip scale applications, there are two thermal paths: first, through the solder bumps on the chip to the substrate; second, through the silicon to the back side of the chip, which is exposed. The best method of heat transfer is conduction. Convection normally can make only a small contribution. For a 600 I/O chip with 125 µm solder bumps in a triple layer chip, the thermal resistance is approximately 1.5°C/W through the solder joint.

Back side cooling can be enhanced by either mechanical conduction, i.e., pistons such as are used with the IBM Thermal Conduction Module(TCM), or by a material such as a thermal grease. Both methods conduct heat to the cap or lid, where they dissipate the heat to the surrounding environment. This can be accomplished by the use of natural or forced air, with or without a heat sink, or by use of a thermal dissipator such as a water-cooled heat exchanger. Immersion in a fluid dielectric has also had some limited use. For nonhermetic applications, low cost heat sinks can be attached directly to the back side of the chip.

Back side cooling of wire bonded chips that are bonded cavity down to a thermal slug, while effective, is less efficient in terms of area use. Flip chip and other chip scale packaging approaches are more compact because the full area under the substrate is available for the I/Os. This also affords higher performance because of shorter leads.

The key to reliable and high yield flip chip joining to the substrate is process control. The substrate interconnection pads must be planar to within 1 micrometer per millimeter, and must be solder wettable. If gold plating is used to prevent oxidation of the underlying metallurgy, the thickness of the gold must be such that gold intermetallics are not

formed in the solder joint when reflow occurs, degrading joint quality.

Standard solder wettability tests should be periodically conducted. The surface should be sufficiently smooth to prevent voids from forming in the solder joint at reflow.

7 ASSEMBLY PROCESSES

A number of assembly processes have been developed to assemble chips in the flip chip interconnections format. The exact process flows are highly dependent on the specific joining technique being used. Most chip scale packages use the current fine pitch SMT assembly materials and processes.

7.1 Substrate Preparation

The incoming substrates are typically inspected to ensure that they are clean and that they meet the design specifications. Beyond these basics, surface preparation is highly process dependent. These process-dependent requirements will be discussed in 7.3.

7.2 Chip and Chip Scale Placement

Placement is frequently the rate limiting step and the most expensive in the assembly process. The factors that contribute most significantly to the cost include:

— throughput (number of placements/time)
— vision system requirements
— die presentation options
— chip-to-substrate alignment accuracy
— chip-to-substrate coplanarity requirements
— additional required features such as supplying heat and pressure during assembly

The specific process can significantly influence placement requirements, and affect costs and throughput. For example, chip scale placement with reflowed solder interconnections and >0.3 mm pitch have greatly relaxed requirements as compared to bare chip interconnections on 0.1 mm pitch attached with anisotropically conductive adhesives. The factors that influence placement accuracy include interconnection pitch and the ability of the interconnections to self-align during attachment. Adhesive systems do not self-align during cure and require placement accuracy of better than ±15 µm for 0.2 mm pitch interconnections.

7.3 Attachment Processes

7.3.1 Attachment Process for Solder

The solder connection process starts with pickup and machine vision inspection of the chip and the attachment site. Flux is applied to either the chip and/or the substrate site prior to the actual placement of the component. The flux is typically sticky enough to hold the die in place prior to formation of the solder joints, by passing the assembly through a reflow furnace.

Solder joining techniques for MSMT and CSGA packages utilize the addition of solder to the substrate before assembly as with SMT processing. For fine pitch applications, solder is deposited by electroplating, solder ink jet, solid solder deposition (SSD), or other techniques. The solder is reflowed and the solder bumps on the substrate are planarized, if necessary. Tacky flux is applied to the solder contacts either by dipping the chip into a flux reservoir or dispensing flux onto the substrate. The chip is aligned and placed into the flux and the interconnections are formed by mass reflow.

For coarse pitch applications where die contacts have been rerouted into an area array, (>0.4 mm pitch) solder paste is deposited on the substrate by stencil printing.

Chips are placed into the tacky paste and mass-reflowed. No cleaning is required when certain ‘‘no clean’’ solder pastes are used. The need for underfill requirements can be reduced by the use of a CTE matched silicon substrate. Work is ongoing to develop fluxless processes.

7.3.2 Adhesive Interconnection

There are numerous adhesive techniques that have been developed to form flip chip to substrate interconnections. These include the use of anisotropically conductive, isotropically conductive and nonconductive adhesives.

7.3.2.1 Anisotropically Conductive Adhesives

These adhesives can be assembled by basically two techniques. Regardless of the technique, all anisotropically conductive adhesives conduct only in the Z-direction (i.e., perpendicular to the plane of the substrate.) The adhesive is applied onto the substrate over the entire surface that the chip will cover. This greatly facilitates anisotropically conductive adhesive placement, as excess adhesive will not bridge adjacent contact lands. By covering the entire surface of the chip, adhesion is improved and cleaning and underfill steps are eliminated.

The first technique uses materials supplied in film form. These are cut, placed, thermally tacked onto the substrate, then placed with heat and pressure applied, sufficient to either tack or fully cure the adhesive. ‘Tacked’ interconnections are fully cured in a separate, dedicated piece of equipment; cooling is frequently done under pressure to ensure good, electrical contacts. The use of fiducials outside the contact area facilitates displacement.

The second technique uses materials supplied in paste form, which are applied to the substrate by stencil or screen printing techniques. Chips are placed into the paste by either of two techniques. The most common approach has been to maintain pressure on the back side of the chip in the alignment machine while the adhesive is cured by the

application of either heat or UV light.

An alternative approach that has recently been developed performs the curing process in a batch curing fixture freeing the placement machine from the throughput-reducing process of curing the interconnections under pressure. The anisotropically conductive adhesive paste is stencil printed onto the substrates. Flip chips are placed into the paste with no additional pressure requirements. Curing is performed in the batch curing fixture that applies both heat and pressure to cure multiple flip chip assemblies simultaneously.

7.3.2.2 Isotropically Conductive Adhesives

This assembly technique requires different dispensing methods to ensure that excess adhesive does not short adjacent contacts. One approach, developed by Epo-Tek [ref], uses precise stencil printing techniques to print isotropically conductive adhesive bumps on both the substrate and chip surfaces.

Chips with contact lands on <0.25 mm pitch have been interconnected. An alternative approach uses dice with preformed metal bumps. Conductive epoxy is transferred to the bumps by dipping the chip in a reservoir of conductive epoxy. The height and flatness of the epoxy surface in the reservoir are maintained by rotation under a doctor blade. The chip is then placed onto the substrate; curing is performed in an oven or hot plate, as pressure is not required during cure.

7.3.2.3 Nonconductive Adhesive Process

This process for making compressive interconnects between bumps on the chip and substrate conductors relies on the cure shrinkage of a nonconductive underfill adhesive to make and sustain the pressure contacts for electrical purposes. The mechanical contact is maintained by the adhesive bond over the entire chip-substrate interface. One application recently described in the literature uses this technique to mount driver chips to glass display panels using a UV-curable adhesive.

7.3.3 Thermocompression Bonding

This attachment technique utilizes heat, pressure and time, sometimes assisted with ultrasonic or thermosonic energy, to form welded interconnections from gold bumped dice to gold-plated or bumped substrates. This technique eliminates the need for flux and cleaning prior to the underfill process. High attachment forces are required for this process especially for high I/O devices.

7.4 Cleaning

For flip chip attachment processes utilizing solder, either a no-clean, leave-on flux residue or a water-soluble flux residue cleaning process is preferred. The goal of either process is to minimize or eliminate harmful flux residue contaminants remaining from the soldering process.

7.4.1 No-Clean Flip Chip Concepts and Processes

In flip chip applications where a non-corrosive rosin based flux or paste is used the flux residue may not need to be removed. Such np-clean assembly of flip chip designs will be most successful when the PWB and the flip chip package are both solderable (assuming the flip chip will be soldered to the substrate as opposed to conductive adhesive attachment) [see J-STD-002 and J-STD-003] and clean, which is defined as free of any processing residues that could lower the SIR below 1012 when measured on the IPC-B-24 comb pattern at 85°C/85% RH.The incoming PWBs can be easily tested for polyglycols which are known to degrade SIR using IPC-TM-650, Methods 2.3.38 and 2.3.39. Low residue soldering fluxes are preferred for the attachment process, with the best balance of soldering and freedom of electrical effects being achieved when the residue is equal to or less than 20% of the applied flux (See A.8).

7.4.2 Aqueous Flip Chip Cleaning Technology

As long as stand-offs remain in the 0.5 mm and higher range, aqueous cleaning will be a viable option. Any problems encountered will usually be in the rinsing and drying steps. Water soluble fluxes and solder pastes are formulated with higher levels of more active and potentially corrosive materials (generally these are halides) than rosin fluxes. These formulations provide good soldering with low defect levels, but rigorous cleaning in a powerful aqueous (or aqueous/saponifier) process, with or without neutralizers, is required. Aqueous processes may work on first generation coarse pitch flip chip designs, but attempts to clean finer pitch designs may be compromised by the rinsing and drying issues inherent in aqueous systems.

Excessive high halide ionic contamination may lead to corrosion or dendritic growth between the flip chip joints. Either degradation pathway provides opportunities for current leakage and product failure. Excess flux deposits may also cause problems with in circuit probe testing, as well as impair the underfill process, resulting in surfaces characterized by poor adhesion, voids from trapped flux outgassing and poor epoxy flow. For a water-soluble flux/paste aqueous cleaning process, it is advisable to bake the substrate prior to underfilling to remove any potential trapped or adsorbed moisture from the substrate.

7.5 Attachment Inspection

A satisfactory inspection process must be able to detect soldering defects such as voiding, dewetting, insufficient solder, misalignment, skew, small or deformed bumps, solderballs and bridging, to name a few.

Visual inspection may be used to inspect a perimeter flip

chip connection. A high magnification microscope may be required as well as specialty tooling to position the assembly for an edge view. Visual inspection, of course, is limited in that the internal connections of a full array flip chip site cannot be verified.

High speed X-ray inspection systems are able to nondestructively examine the full array of metallic flip chip joints. X-ray techniques in use today include digital tomosynthesis and scanned beam laminography. Scanning acoustic microscopy has successfully been applied in detecting flux residues, non-wetting, voiding, and other defects .

7.6 Underfill (Flip Chip Encapsulation)

Underfilling is the process of dispensing the nonconductive, filled epoxy between the chip and substrate, around the flip chip interconnections. Underfill is used to extend the joint fatigue life of flip chip and grid arrays under thermal cycle conditions; MSMT packages do not require underfill epoxy. Underfill also serves to redistribute the stress from the joints to the chip, substrate and epoxy. This stress is a result of the CTE mismatch between the die and substrate.

A typical process flow entails incoming material inspection, substrate pre-bake, epoxy dispense, epoxy cure and underfill inspection. For incoming material inspection, properties that require inspection include viscosity, ionic content, Tg, CTE, alpha emission level, filler size and filler distribution.
Frequently organic substrates are prebaked to drive off moisture prior to underfill application. Moisture may result in poor adhesion or degradation of anhydride-type epoxies. Often the epoxy is dispensed on one or more edges of the die. The material flows between the chip and substrate by means of capillary action.

The flow behavior may be enhanced by heating the substrate, and in some cases the material is heated prior to dispense. Many materials used today achieve improved flow behavior at preheat temperatures in the range of 70-100 °C. Process control of the volume of material dispensed is important and is determined primarily by the gap, chip size and fillet requirements.

The assembly is then processed through the recommended time-temperature cure schedule. Typically materials today require a batch cure process, with cure schedules in the temperature range of 110-170 °C for a duration of anywhere from 30 minutes to 4 hours.

After cure, visual inspection of the underfill process is performed. The epoxy fillet is inspected to ensure that it is positive wetting and its length is within a specified range. The top of the chip and edges are checked for the occurrence of stress-inducing epoxy-on-chip defects. Voids which are touching or between the flip chip joints are undesirable, and cannot be visually inspected; however, voiding can be monitored using glass chips, by process audits using destructive cross-sectioning. Nondestructive acoustic inspection equipment has also been developed for epoxy void and delamination detection.

7.7 Electrical Test

The selection of the proper place in the assembly process flow for electrical test depends on the maturity and yield of the process as well as the reworkability of the attachment process utilized. For those processes where rework is practical, the electrical test step should be conducted prior to underfilling. In cases where rework is not practical, where Known Good Die are used, or where the assembly yields are high, electrical test is properly placed just prior to final inspection and shipment.

7.8 Rework

Rework or replacement of a defective solder attached flip chip device is feasible, provided it is conducted prior to underfill application. One approach to removing solder bumped devices is to mechanically lift the chip off the substrate. Chips are removed using a tool that applies either a torque or a tensile force, or by using ultrasonic vibration.

Alternatively, the chip assembly can be heated to reflow the solder joints prior to removal. Heated thermodes that reflow the solder joints while removing the chip may be used. Hot gas or IR lamps may also be used to thermally remove the chips. Residual solder remaining on the substrate must be removed prior to attaching a new chip. Fresh solder and/or flux may be required before a new chip is attached.

Anisotropically conductive adhesive or nonconductive adhesive interconnections are extremely difficult to repair. Thermoplastic materials may be easier to repair than the more reliable thermosetting systems. Isotropically conductive adhesive interconnections could be reworked before underfilling much like solder interconnections.

Rework of MSMT and other chip scale packages is more feasible. The removal technique is to surround the chip with flux, heat until reflow occurs and then remove the part. Replacement requires retinning of the lands, applying flux, then placing and positioning the part using a microscope. The solder or adhesive is reflowed or cured, using low velocity hot gas or hot air.

8.0 Flip Chip Test and Burn-in Methodology

Use of high density packaging such as MCMs and PCMCIA in the semiconductor market-place is gaining support due to the advantages of decreased costs, reduced board space, and improved performance over individual single-chip modules. However the rate of die rework at the MCM

level increases significantly with the number of dice on the module. Known Good Die can help to eliminate the need for MCM rework due to die reliability.

Cost modeling can be performed to determine whether an advantage exists for using KGD in a given application. The models account for such factors as number of dice in the MCM, rework costs, and bare die test and burn-in yields. Factors must be considered such as the complexity of the die in the MCM and whether it is possible to test and burn-in at the MCM level. Additional applications for KGD exist such as use in Direct Chip Attach (DCA) applications where chip rework may be unavailable. Single-chip module use of KGD may be effective early in the life of a program when die yields are low and the die is placed in an expensive package.

The dramatic increase in available compute power has enabled application of a variety of yield improvement techniques that were previously too data intensive to be economically viable. Extensive use of Statistical Process Control (SPC), Automatic Test Generation (ATG), and fault simulation are now practical. These and other tools can resolve many of the manufacturing problems associated with flip chip production.

The three prime concerns of all manufacturing disciplines are component quality, process control, and finished product verification; each present unique challenges to the successful production of Flip Chip Technology. These challenges are inherent consequences of the flip chip advantages of high density and elimination of the first level of Integrated Circuit (IC) packaging.

The quality of any manufactured product is limited by the quality of the components and materials of which it is made. The upper limit for the yield of any collection of components is the product of the individual component yields as follows:

$$\prod_{i=1}^{n}(1 - D_i)$$

where Y is the assembly yield
Di is the component defect rate
n is the number of components

For a collection of components with similar defect rates, this relationship reduces to:

$$Y = (1 - D)^n$$

8.1 Known Good Die

In the case of ICs, many characteristics of quality such as high speed performance and long term reliability are difficult to ascertain prior to packaging the component.

IC manufacturers traditionally check the devices on a wafer for functionality, package the functional devices in single chip packages, and address performance and reliability issues at package level testing.

By eliminating the first level of packaging, the chip manufacturer must address the issues of IC performance and reliability at the bare die level because addressing these issues at the level of a MCM will result in a yield reduction to the nth power, due to the yield relationship described above. Therefore, more stringent component-level test strategies coupled with cost effective rework processes will be required for successful known good die manufacture.

The end product of a Known Good Die (KGD) process is a die with similar quality and reliability as that same die in a finished single chip module package. Several techniques are available for contacting the die while performing full at-speed functional test and burn-in stress. These include both soft connection (metallurgical) and mechanical connection of the flip chip bumps, metalized bond pads, and posts to a KGD carrier.

8.1.1 IC Quality

The ac characteristics such as propagation delay and switching speed of ICs are difficult to measure on an unpackaged device because of the parasitic resistance and reactance of the electrical interface between the tester and the device under test (DUT). Thus, to reduce the risk of marginal components affecting module yield, a method of identifying poor ac performance is needed.

One approach is to establish statistical correlations between measurable characteristics and the probability of sufficient ac performance. This is possible because ac performance is often dependent on measurable dc parameters such as resistance, voltage/current transfer functions, power dissipation, and junction leakage currents.

A practical application of this method would require defining normalized dimensionless Figures of Merit (FOM) describing the various measured parameters. Correlation of the FOM to ac performance can be predicted via simulation and validated with historical test data from packaged devices.

The parametric data required to perform these correlations would have to be obtained at the wafer level to be cost-effective. Wafer probing of solder-bump die is complicated by the need to make consistent and reliable mechanical and electrical connections to large arrays of soft solder bumps. One solution, developed by IBM uses an array of discrete probes supported by a single-bend spring . Called the Cobra Probe due to the shape of the spring element, it has been used effectively in both wafer probing and MIS probing for many years.

New developments in materials and processing have produced alternative solutions to the wafer probing problem.

One technique receiving a great deal of attention is the Membrane Probe, originally developed by Packard-Hughes. This probe system is fabricated on a flexible film with wiring printed on one side and bumps (probe tips) formed on the opposite side. Via holes in the film provide connections from the probe tips to the wiring.

In current practice, however, wafer probing and statistical correlation techniques have not achieved the IC quality and reliability goals at the wafer level. Therefore, temporary packages have been developed to support the chip during a more traditional test. These temporary packages take the form of sockets or carriers to which the IC is temporarily mounted for testing.

8.1.2 IC Reliability

In a manner similar to the wafer level correlations for quality assessment, the Constant Failure Rate (CFR) of long term component reliability might also correlate with measurable parameters. If so, components with excessive predicted CFR could be rejected at the wafer level. Unfortunately the mechanisms which contribute most to Early Life Failures (ELF) do not correlate well with measurable parameters. Latent point defects such as oxide pin holes, occluded metal lines, and substrate lattice faults must be stressed to the point where their effect on measurable parameters is detectable. The most common method of applying the stress is High Temperature Reversed Bias (HTRB) burn-in where the IC is operated at elevated temperatures of 150 °C or more for extended periods of time (8 to 168 hours).

The standard failure acceleration stress of HTRB burn-in is exclusively applied to packaged devices due to the need for electrical connections to supply the bias voltage while operating at elevated temperatures. Although development of wafer level burn-in methods is progressing, die level burn-in is the current paradigm.

Burn-in is a process that accelerates the rate of infant mortality failures to achieve user-acceptable integrated circuit (IC) reliability levels. IC failures are caused by random manufacturing defects which produce a decreasing failure rate to the point of wearout onset. By screening out many of these defect fails through burn-in prior to customer use, the resultant failure distribution can be significantly improved. For this reason, burn-in plays a critical role throughout the semiconductor industry in ensuring product reliability.

The need for a package to support the flip chip and bare die IC during burn-in has motivated the development of temporary carriers. The carriers provide the necessary electrical connections, mechanical support and protection, and thermal management during the stress and testing. Currently, there are three approaches to temporary carriers for testing die: IBM s Reduced Radius Removal (R3), MCNCs Burn-in And Test Substrate (BATS), and IBM s Dendrite Temporary Chip Attachment (TCA), and microns KGD plus.

A number of different levels of die quality are available in the industry. Some of these levels include burn-in and some do not. In order to standardize the definitions of die quality, the section below details the four different levels of die quality that are being standardized in the industry.

8.1.2.1 Standard Probe

Standard probe is the easiest quality level for vendors to provide. Die are subjected to a wafer probe test. This test is identical to the wafer probe test used for packaged parts and consists of various functional and parametric tests. This test allows vendors to offer die with minimal effort using test programs that already exist. Because no burn-in is performed on the die, infant mortalities are not removed from the population.

This level is of most interest to users who require wafers instead of die. Because most burn-in processes are performed at the die level, standard probe (and speed probe) are the only levels that test at the wafer level. These levels are also of interest to users who will perform post-processing steps to the wafer such as adding additional metal layer, or bumps for flip chip applications. Because these extra levels may add failure mechanisms that will appear as infant mortalities, users will need to do a burn-in step after the extra processing is completed.

8.1.2.2 Speed Probe

Speed probe is an enhanced probe test to standard probe. Because speedgrading of the die is usually done on packaged parts after burn-in, a standard probe test may not give a good indication of the performance of individual die. To offset this shortcoming, the speed probe level begins with the same wafer probe program used with the standard probe level. This die is then heated using a hot chuck speed probe to assure the speed performance of the die for the fastest speed grades. Critical parameters such as access times and pulse widths are temperature dependent and this probe determines functionality and allows vendors to provide a speed map of the wafer at temperature.

8.1.2.3 Burn-In

The burn-in level incorporates all of the testing done in the speed probe and standard probe levels with the addition of a burn-in step. The purpose of the burn-in step is to stress the parts until they have passed the infant mortality stage. Because the infant mortalities are removed users will experience a higher yield when packaging the die. It also allows users to skip a burn-in step.

There are several different types of burn-in which vary among manufacturers. During the process known as static

burn-in, temperatures are increased while only some of the pins on a test device are biased. No test vectors are written to the device, nor is the device exercised under stress during static burn-in. During unmonitored dynamic burn-in, temperatures are increased while the pins on the test device are biased. The device is cycled under stress, and test patterns are written to the device but not read. Hence, there is no way of knowing if the outputs of the device under test are correct. Lastly, the most sophisticated method of burn-in is an intelligent burn-in approach which combines functional, programmable testing with the traditional burn-in cycling of the device under test in the same chamber.

8.1.2.4 Known Good Die (KGD)

The highest level of quality is Known Good Die. This level incorporates all the testing of the previously described levels. It also includes additional testing after burn-in that are performed in order to ensure that the die will meet timing and power dissipation requirements. These specifications are usually taken from a standard datasheet for the packaged part equivilent.

Users must be careful when applying the specifications for timing in their packaging environment. If the die is normally packaged in an SOJ and the user packages the part in a module with extremely long signal lines of high capacitive loads, then an appropriate derating factor may need to be applied to critical timing parameters.

8.2 KGD Techniques for Flip Chip

Chip scale packages may benefit from some of the technologies described in the following paragraphs, but in general use the same test and handling methods as SMT packages.

8.2.1 IBM's R3

The Reduced Radius Removal (R3) technology involves the solder reflow of a flip chip device to a reduced solderable area on a ceramic carrier. Once the die has been attached to the carrier, the deck is processed through the same test and burn-in processes as the single chip module version of that same die. Upon completion of the test and burn-in process, the die is removed from the carrier by applying a horizontal force to the edge of the die. The flip chip bump shears from the carrier at the interface between the bump and the reduced solderable area on the carrier. The solder bumps are reflowed to their original spherical shape. The die is inspected and ready to be shipped as a fully functionally tested and burned-in KGD.

8.2.1.1 R3 Carrier Reuse

The R3 carrier can be reused multiple times. Reuse is monitored by measuring the shear force required to remove the die from the carrier. The carrier is allowed to be reused as long as the shear force does not exceed process specifications. Certain types of ceramic carriers have the ability to be reworked to allow additional reuses. This reduces the shear force at die removal below the maximum allowed in the process specification. A single use cycle for the carrier includes a solder reflow, test, and a 140 °C burn-in.

8.2.1.2 R3 Test and Burn-In

The carrier for the R3 process is similar to the ceramic substrate used in the production of finished single-chip modules, with the exception of a reduced solderable area. Both have the same test and burn-in characteristics. Therefore any test and burn-in conditions which can be applied to single-chip modules can be applied to the R3 decks. This can include dynamic, dynamic-monitored, and in-situ burn-in.

Testing that can be performed includes deterministic, weighted random patterns, JTAG interconnection, and per-pin parametrics, as well as unique patterns. Level sensitive scan design (LSSD) using logic built-in self test (LBIST) or array built-in self test (ABIST) can be incorporated in the full at-speed functional testing.

Test and burn-in are performed on the same equipment as that used for finished single-chip modules. This includes such front-end hardware as burn-in boards (BIBs) and test socket fixturing.

8.2.1.3 R3 Die Quality

The mechanical SPQL of a KGD is guaranteed through optical inspection. The inspection may be performed manually, using an automated Individual Chip Inspection System (ICIS), or using a combination of both. The inspection would consist of looking for such defects as low solder volume bumps, smeared or missing bumps, damage to the die silicon, and damage to the die passivation.

8.2.1.4 R3 Process History

The R3 process entered volume production of KGD in 1990 upon completion of a successful qualification. More than one million KGD have been shipped since the start of production. Dice produced from the R3 process have been used in applications from mainframe MCMs to single-chip laptop computer cards using direct chip attach.

8.2.2 MCNC s Burn-in and Test Substrate (BATS)

This carrier approach makes temporary solder connections by covering nonwettable metal pads on the carrier with a thin layer of wettable metal. The solder bumps are soldered to the thin wettable layer resulting in a relatively strong solder connection (~85% of normal). After test and burn-in, the solder bumps are melted (reflowed) and the remaining wettable metal is dissolved into the solder bump. This leaves the solder bump in contact with the nonwettable

metal pad and the solder dewets from this surface. The die can then be removed from the carrier without damage to the die or the bumps.

8.2.2.1 BATS Carrier Reuse

The BATS technique can be implemented on either reusable or disposable carriers. The BATS technique provides for unlimited reuse because the top surface (the interface to the chip under test) is reconditioned with the sacrificial metal after each use.

8.2.2.2 BATS Test and Burn-In

The carrier used for the BATS process can be designed for testability requirements that match the needs of the product being tested. High performance devices can be mounted on carriers with controlled impedance transmission lines.

Integral termination resistors and decoupling capacitors can provide the ideal test environment. Lower performance devices can be cost-effectively tested and burned in using low cost substrates such as organic laminates.

8.2.2.3 BATS Die Quality

The low removal forces of the BATS process result in ICs with nearly spherical solder bumps. The bottom of the bumps is flattened in the form of a truncated sphere due to the mass of the chip and the force of gravity during the dejoin reflow. As a result, no additional reflow is necessary to return the bumps to the required spherical shape.

8.2.3 IBM's Dendrite Temporary Chip Attach (TCA)12

Dendrite Temporary Chip Attach (TCA) technology uses a mechanical interconnection between the flip chip bump and KGD carrier. This technology uses a plated dendritic structure on the carrier fingers as the contact interface to the flip chip bumps.

A clamp fixture, which contains the dendritic carrier, applies a continual force to the backside of the die, maintaining intimate contact between the flip chip bump and dendritic structure during test and burn-in to 180 °C. Upon completion of the test and burn-in process, the die is removed from the dendritic carrier. The flip chip bumps are then reflowed to their original spherical shape. The die is inspected and then shipped as a fully functionally tested and burned-in KGD.

8.2.3.1 TCA Carrier Reuse

The dendrite TCA carrier has been qualified for 20 reuses in the initial process qualification. Additional hardware is being processed to extend the qualification to a minimum of 100 reuses. Feasibility data prior to the start of qualification indicates a very high probability of achieving the 100 minimum reuse target. Certain types of ceramic carriers have the ability to be replated with new dendrites for additional carrier reuse. A single use cycle for the carrier includes both test and a 140 °C burn-in.

8.2.3.2 TCA Test and Burn-In

The carrier for the TCA process is similar to the ceramic substrate used in the production of finished single-chip modules, with the exception of a clamp surrounding the carrier. Both have the same test and burn-in characteristics. Therefore any test and burn-in conditions that can be applied to single-chip modules can be applied to the TCA decks. This can include dynamic, dynamic-monitored, and in-situ burn-in. Testing that can be performed includes deterministic, weighted random patterns, JTAG interconnect, and per-pin parametrics, as well as unique patterns.

Level sensitive scan design (LSSD) using logic built-in self test (LBIST) or array built-in self test (ABIST) can be incorporated in the full at-speed functional testing. Test and burn-in is performed on the same equipment as that used for finished single-chip modules. This includes such front-end hardware as burn-in boards (BIBS) and test socket fixturing.

8.2.3.3 TCA Heat Dissipation

The TCA clamp has a heat sink incorporated into the top portion of the clamp assembly. The heatsink is capable of removing 27 watts of power at 400 feet per minute airflow at 20 °C.

8.2.3.4 TCA Die Quality

The mechanical SPQL of a KGD is guaranteed through optical inspection. The inspection may be performed manually, using an automated Individual Chip Inspection System (ICIS), or using a combination of both. The inspection would consist of looking for such defects as low solder volume bumps, smeared or missing bumps, damage to the die silicon, and damage to the die passivation.

8.2.3.5 TCA Process History

The dendrite TCA process is the next generation KGD process replacing the R3 process. A successful qualification was completed in 1994 with volume production beginning in the fourth quarter.

8.3 Known-Good Mounting and Interconnection Structure

The high density interconnections characteristic of the mounting and interconnection structure (MIS) processing often require process steps in fabrication that differ little from those of IC fabrication. Due to the nature of these processes, the yield of the MIS is inversely proportional to the surface area.

Since an MIS is typically one to two orders of magnitude

larger than an IC, process defects present a greater economic liability. Therefore, more stringent process controls and/or cost effective rework processes will be required.

The testing of the MIS for flip chip mounting is complicated by the same factors that affect IC testing, particularly the fine pitch area array pads. A number of techniques are available for MIS testing ranging from area array probes to scanning electron beam testers.

8.3.1 Testing Techniques

There are a variety of techniques for testing known good die or mounting structures intended for flip chip or chip scale mounting. The following paragraphs indicate some of the methods used in today s processes.

8.3.1.1 Area Array Probes

For flip chip pad pitches greater than about 250 µm (0.010″), spring loaded probes can be fixtured for area array contacts. IBM has long used the Cobra probe, described above, for MIS testing. Recent developments in pogo style spring probes have achieved a 250 µm pitch. While the fixturing can be fairly expensive, these methods provide the most direct and effective MIS testing.

8.3.1.2 Flying Probes

When the pad pitch is less than 250 µm, single or dual point probing offers electrical contacts at the cost of throughput. Testers are commercially available that use automatic moving probes with better than 25 µm position accuracy. Two such probes can be used to measure the resistance between any two pads on the MIS.

While this method is effective for continuity testing, it is extremely inefficient for detecting shorts due to the $O(n^2)$ nature of the detection algorithm. When these testers are used in the single point mode, a probe makes contact to a single pad on the MIS surface.

The capacitance of the conductor being tested is then measured relative to a reference plane in the tester. If the conductor has a break or open the capacitance will be lower than expected. Conversely, if the conductor is shorted to another the capacitance will be higher than expected. In either case, the tester reverts to the two-probe method for verification of the suspected defects detected by the capacitance test.

8.3.1.3 Scanning Electron Microscope (SEM)

A recently developed technique takes advantage of the charging effects of isolated conductors exposed to a beam of electrons in a Scanning Electron Microscope. By first charging a conductor through a single pad with the electron beam and then scanning the MIS surface with the SEM in voltage-contrast mode, all other pads connected to the charged pad will also show charge. By comparing the locations of detected charge with the locations of interconnected pads, open (uncharged pads the should be charged) and shorted (charged pads that should not be) conductors can be detected.

8.4 Product Verification

With multiple ICs interconnected at the first level of packaging, an exceedingly high level of functionality is contained in the Lowest Replaceable Unit (LRU). Adequate verification of the assembled module requires large numbers of complex test vectors to ensure a high probability of fault detection. Once detected, the isolation and localization of faults is necessary to enable salvage and rework of the module.

Due to the small geometries of the flip chip, direct measurement of electrical activity is problematic. Therefore, the logic design and test vectors must supplement the direct measurement with logical observability. This has resulted in the advancement of design-for-testability (DFT), built-in self test (BIST), and other design philosophies, as well as novel test access methods.

The verification of the finished MCM is complicated by the high functionality incorporated in the LRU. Effective verification strategies require that the logic be designed within testability guidelines. Examples of testability are shown in the following paragraphs.

8.4.1 Standard Test Access Port and Boundary Scan Architecture

A specification for standardizing the implementation of a test circuit architecture, IEEE 1149.1, describes the functionality of circuits to be incorporated into digital IC designs for the purpose of enabling cost effective testing.

Special test circuits called Boundary Scan Cells (BSC) are interposed between the logic of the IC and the substrate interconnects. The test circuits are designed such that they are transparent during normal device operation and become active only upon command from on-chip logic called the Test Access Port (TAP) Controller.

In the test mode, these special circuits can interrupt the data flow and either isolate the IC from surrounding circuits, supply the IC with test stimulus, or monitor the logical activity entering or leaving the IC.

The benefit of such a scheme is to allow the implementation of hierarchical structure test strategies that address fault detection in an ordered manner. For example, a typical test scenario might require first verifying the interconnections between ICs, followed by a functional check of each IC's performance, then verifying the entire assembly operation in a slow or static mode. Finally, high speed or dynamic operation of the assembly must be verified. To implement such a scenario, the BSC associated with each

IC I/O would exercise each of its potential modes.

To verify the interconnections between ICs, the BSCs at IC output pins would be loaded with appropriate test data, (i.e., alternating 1's and 0's) and instructed to force this data out the output pin. The BSCs at the IC input pins would then be instructed to report to the TAP the logic state present at their inputs. Test software would then compare the CAD routing file information against the data reported by the TAP.

To test the functionality of each IC, the input BSCs would again be loaded with appropriate test data, (i.e., test vectors for the IC) and instructed to drive the data into the IC. BSCs at the outputs of the IC would then report to the TAP the IC's response to the test stimulus. Test software would compare the IC responses to simulator predicted responses. Faulty devices are therefore directly identified.

To verify the operation of the whole assembly, the BSCs would all be placed in monitor mode, in which they simply monitor logical activity and report the logic states to the TAP. The assembly would be stepped through typical operations, and the activity of every signal on the MCM would be reported to the TAP. Again, test software would compare the activity to the simulator predicted activity.

Finally, the assembly would be operated at its full rated speed and the module performance would be monitored in a traditional manner using Automated Test Equipment (ATE) or Test Beds. If a failure is detected, the assembly would be reset and the failing operation repeated.

Operation would be interrupted at strategic intervals to verify the proper logic state of all IC inputs and outputs. Standard fault isolation software could then be employed to isolate the defect.

Although Design For Testability (DFT) techniques such as Boundary Scan can provide verification and diagnosis capability at the module level, in practice not all IC designs are conducive to BSC implementation. Critical timing signals such as clocks often cannot tolerate the delay penalties associated with the transparent mode of BSCs. In such cases, traditional test access methods must be employed.

The challenge, then, is to provide access for conventional electrical probes without unduly impacting IC packing density or affecting high frequency signal fidelity. A number of contact technologies such as metal-on-elastomer (MOE), coplanar ceramic 'blade' probes, and microstrip flex circuits can provide electrical access to a controlled impedance circuit.

Implicit in these strategies is the need for accessible contact pads for each critical signal. These pads must be located in reasonable proximity to the signal in order to minimize the electrical parasitics. Some relief from this constraint may be provided by appropriate conductor design that considers the signal propagation of the entire conductor. Substantial economy can be derived if the contact pads can be designed such that they provide the multiple roles of test access, engineering change terminals, interconnect disconnection sites, and signal terminations.

9 REQUIREMENTS FOR RELIABILITY

The previous sections have shown that there is a diverse set of materials and processing approaches used in constructing flip chip assemblies. There is also a wide range of environments that these assemblies can experience. These include mechanical handling, chemical/thermal exposures during manufacture and environmental stresses encountered in the application.

Table 9-1 shows product categories and use environments that flip chip assemblies may experience in their applications. The stresses encountered during manufacture can also be significant, and should be accounted for when selecting the appropriate flip chip process for each application. The reliability of a flip chip assembly is a function of the following factors:

1. Materials
2. Processes
3. Design/performance/use
4. Environmental Exposures
5. Acceptable Failure Probability

Each of these constituents must be thoroughly understood and controlled if the produced flip chip assembly is to be reliable in its application.

9.1 Robustness of Products to Use

The top three factors in table 9-1, materials, processes and design, are key in establishing how robust a flip chip assembly will be in its application. Each of these topics has been discussed in previous sections, but for a flip chip assembly to be robust, a balanced consideration of each of these factors is mandatory.

Flip chip mounting requires the interconnection surface of the semiconductor device to provide electrical, thermal, and mechanical attachment to its host assembly and the application. The intrinsic properties and design guidelines of the materials define how well this interface can accommodate the chemical, thermal and mechanical stresses imposed by the assembly and the environment.

Environmentally-induced thermal expansions and mechanical stresses are distributed across all interconnects as determined by the physical geometry of the assembly. Process variation of interconnect geometries can be very important to the robustness of the final assembly.

One example of how flip chip joints can be made more robust is to include an underfill material. The underfill distributes the thermal expansion mismatch stresses increasing

the life expectancy of the interconnects to reduce the strain on the interconnects.

Underfill can and should be used for larger chips in more severe environmental applications, but may not be necessary in all applications. The underfill can induce other failure modes if the material and processes are not carefully selected.

The remainder of this section deals with the factors affecting solder bump fatigue. Explanations of these factors will be presented along with fatigue behavioral models and testing. The primary model used to predict solder bump reliability is a modified Coffin-Manson relationship.

9.1.1 Chip Scale Package Robustness and Reliability

While the bulk of this chapter is devoted to reliability studies for flip chip owing to its long established history of use, the emerging families of chip scale products are also undergoing intense reliability evaluation. Much of the information and recommendations provided here regarding flip chip is directly applicable to many of the new chip scale packages as well. Certainly the basic reliability analysis methods discussed are of near universal value in evaluating electronic assemblies. However, others of these new chip scale package schemes are anticipated to have new failure modes due to the nature of their construction which can be quite different from traditional flip chip. This is illustrated in figure 9-1.

As can be seen in the figure, some of the newer chip scale packages are designed to, without need of an underfill step, provide a measure of compliance to minimize the stresses that would otherwise occur. These chip scale devices seek to follow the lessons taught by surface mount technology experience that lead compliance is of great importance. More reliability data on these package types will be provided in future revisions of this document.

9.2 Reliability Factors

The traditional 'bathtub' failure rate curve indicates that there are three distinct types of failure distributions. Initial failures, these are minimized through process control and proper screening, as discussed in earlier sections of this document.

The end of life part of the curve is typical of failure caused by gradual accumulation of damage through the life of the product. The middle part of the curve is best described as event related failures, manifested by overstresses that have occurred during the life of a product. These could be ESD, high thermal transients, high mechanical shocks due to handling, etc.

9.2.1 Wear Out Mechanisms

The primary failure mechanism that affects the reliability of any solder joint connection is damage accumulation by thermo-mechanical processes such as creep and fatigue. Failures are also caused by electro-migration and thermal-migration of metal which result in opens. Both the thermal-mechanical and electrical failures can be increased by chemical reactions or species which can cause corrosion or enhance migration of metallic ions (shorts).

A solder bump wear-out failure in the field has never been observed. This indicates both the robustness of the solder bump and conservativeness of the reliability models that will be subsequently explained.

9.2.1.1 Creep Fatigue Interaction

Solder joints can see large strains during temperature cycling. The strains are generated by the difference in the thermal expansion mismatch between the die and the substrate. The cyclic nature of the temperature variation causes the strain experienced by the solder joints to be cyclic, and therefore, the damage in the solder to be a function of the number of thermal cycles. By this definition, solder joints experience thermal fatigue. At uniform strain distributions, fatigue causes damage by the initiation and propagation of micro-cracks.

Fracturing of solder joints by the initiation and propagation of microcracks is increased during high rates of change of strain. Room temperature is high compared to Pb/Sn solders melting temperature. In general, components see thermal cycling which can range from 0.4Tm to 0.8Tm, where Tm is the absolute melting temperature of solder.

As a rule of thumb, creep is an operative deformation mechanism whenever the temperature is above half the absolute melting temperature of the material. For example, the melting temperature of 63/37 Sn/Pb solder is 456 degrees Kelvin (183.6 Centigrade), half the absolute melting temperature is 228K or −45 °C. Hence, creep deformation can occur in solder at a given load even at temperatures as low as −45 °C. Creep strain is the result of thermally activated dislocation motion and/or movement of vacancies and atoms when the material is subjected to a given load. Stress relaxation over time observed in solder joints is a direct result of creep processes. Hence the damage accumulated in solder joints during temperature cycling is a result of creep and fatigue processes. Deformation due to creep has been observed in packages which have been subjected to both cyclic loading as well as static loading conditions.

Static loading conditions can arise, for example, from the weight of a heat sink on a package. Depending on how the board is placed, that is, sideways, right side up, or inverted, the solder joints can experience shear, compressive or tensile loads. The constant load in one direction can cause failure by creep processes.

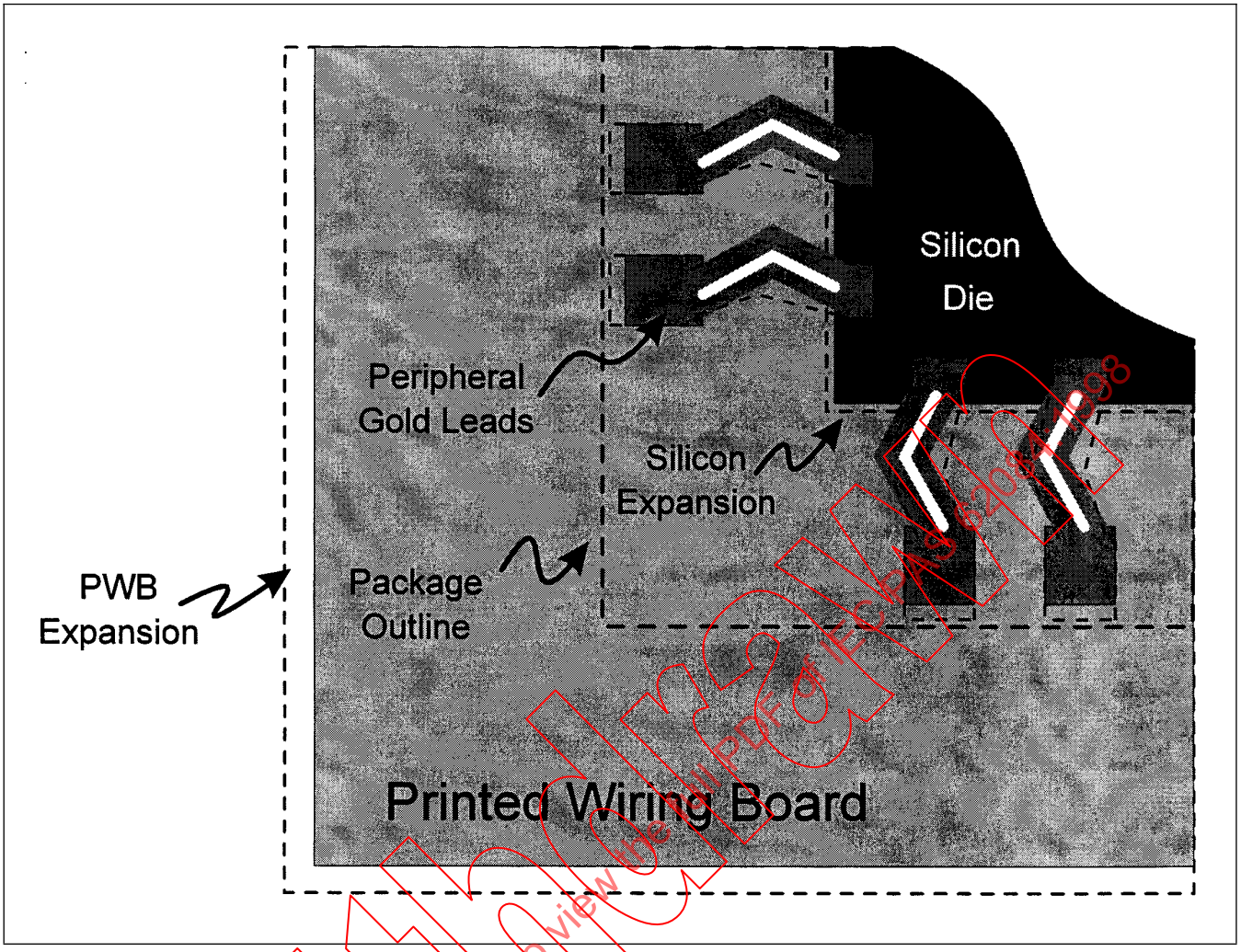


Figure 9–1¬ Chip Scale Package Lead Compliance

9.2.1.2 Electromigration

In alloy systems, the mobility of each atomic species under an applied electric field is different based on the effective valence and mass of the species.

Electromigration has been observed on chips as well as in solder joints. In a joint, migration of voids to chip or board interface has been observed. Small voids can coalesce, which can cause a mechanical reliability concern or current crowding can occur if the effective conductive cross-section is reduced. In both cases observing design rules with regards to current density and dimension can solve the problem.

9.2.1.3 Corrosion

Solder joint corrosion is usually caused by moisture and ionic contaminants. This can be prevented by proper packaging design and process control. Due to the close proximity of flip chip solder joints, the cleanliness of the solder joints is essential. Residues from corrosive processing chemicals must be minimized. Depending on use conditions, the flip chip solder joints should be either hermetically sealed, or encapsulated with a material that prevents a continuous moisture path.

9.2.1.4 Thermomigration

Thermomigration is caused by excessive thermal gradients across solder bumps, especially at higher IC junction temperatures. In the thermomigration process, atoms diffuse in the direction of, or in opposition to, the thermal gradient. Atoms in the solder bump can diffuse, leading to voiding at the interface of the solder and the UBM (Under Bump Metallurgy). The bump will eventually become electrically open. For a given bump geometry, thermomigration is a function of thermal gradient, ambient temperature, and alloy composition.

9.2.2 Solder Bump Mechanical Reliability

There are numerous factors that can affect the mechanical reliability of flip chip attachment. Some of t hese are:

Strain
Temperature Hold Times
Chip Underfill
Solder Alloy Composition

Table 9–1¬ Product Categories and Use Environments

PRODUCT CATEGORY (typical applications)	TEMPERATURE °C¬			MECHANICAL¬			ATMOSPHERIC¬			ELECTRICAL	
	PROC¬	STOR¬	OPN¬	PROC¬	STOR¬	OPN¬	PROC¬	STOR¬	OPN¬	PROC¬	OPN
I. Consumer¬	25/260¬	–40/85¬	0/55¬	Robust. (term) Th. Shock Solder	Shock. Trans Vibrtn Trans	Robust. (Act) Wear Life (Mech) Fatigue Flamma	Moisture Solvents	Moisture Corro Atm	Moisture 5000 Alt	Test¬	Endur (Elect) Overloand Surge Revemi ESD
II. Computers and peripherals	25/260¬	–40-85¬	0/55¬	Robust (Term) Th.Shock Solder	Shock Trans Vibrtn Trans	Robust (Act) Wear Life (Mech) Fatigue Flamm VIB .06 DA 1g to 500Hz	Moisture solvents	Moisture corro Atm	Mositure 5000 Alt	Test¬	Endur (Elect) Overload surge Revemi Esd
III Communica-tions	25/260¬	–40/85¬	–40/85¬	Robust (Term) Th.Shock Solder	Shock Med Vibrtn 10-500 Cycles	Same as above	Moisture Solvents	Moisture Corr0 Atm	Moisture 5000 Alt	Test¬	Endur (Elect) Overload surge Revemi Esd
IV. Industrial and transportation passenger compartment	25/260¬	–55/85¬	–40/85¬	Robust (Term) Th.Shock Solder	Shock¬	Same as above	Moisture Solvents	Moisture Corro Atm	Moisture 5000 Alt Sand/ Dust	Test¬	Endur (Elect) Overload surge Revemi Esd
V. Military, ground and shipboard; low altitude commercial aircraft	25/260¬	–40/85¬	–40/85¬	Robust (Term) Th.Shock Solder	Shock Med Vibrtn 10-500 Cycles	Robust (Act) Wear Life (Mech) Fatigue Flamm Vib .06 Da 15g to 2000 Hz Med Shock High Shock	Moisture Solvents	Moisture Corro Atm	Moisture Corro Atm Fluids 25000 Alt Sand/ dust Fungus	Test¬	Endur (Elect) Overload surge Revemi Esd
VI. Transportation under hood	25/260¬	–55/125¬	–40/125								
VII. Militay and space high altitude commercial aircraft	25/ 26040/ 84	–40/85¬	Robust (Term) Th.Shock Solder	Shock Med Vibrtn 10-2000 Cycles	Same as above	Moisture Solvents	Moisture Corro Atm	Moisture Corro Atm Fluids 25000 Alt Sand/ Dust Fungus			

9.2.2.1 Strain

The affect of strain is significant. Its impact on solder bump fatigue is dependent upon several factors. These are determined by the design, application and manufacture of the device.

The strain relationship modified for solder bump applications is shown below:

$$\Delta\delta = \frac{DNP\ (\alpha_s \Delta T_s - \alpha_c \Delta T_C)}{h}$$

Where, $\Delta\delta$ = strain on the outermost functional bumps
α = chip/substrate coefficient of Thermal Expansion (CTE) (ppm/°C)
ΔT = relative thermal excursion of the Chip/Substrate (°C)
h¬ = solder joint height (µm)
DNP = distance to the Neutral Point (µm)

Table 9-2 shows a CTE difference between silicon and alumina, a typical flip chip package combination.

Strain is directly proportional to DNP, the distance from the furthest, functional solder bump to the neutral point on the chip. The neutral point or geometric center of solder bumps remains stationary relative to the substrate during thermal excursions. Determination of the neutral point is critical for calculating DNP.

Strain is inversely proportional to joint height. Table 9-3 shows typical heights for varying solder bump diameters. Figure 9-2 illustrates bump height (h) which is dependent upon several factors including chip mass, number of bumps, bump diameter and volume, and substrate land size.

Table 9–2¬ Coefficients of Thermal Expansion for Typical Materials

Material¬	CTE (ppm/°C)
Silicon¬	2.8
GaAs¬	6.0 - 7.0
Solder (3/97)¬	28 - 29
Chip Underfill¬	18 - 35
Alumina¬	6.0 - 7.0
FR-4 for DCA¬	16 - 19

Table 9–3¬ Typical Heights (Joined)

C4 Bump Option	Diameter (µm)¬	Height (µm)¬	Range (µm)
Option A¬	150¬	82¬	64 - 100
Option B¬	125¬	77¬	64 - 95
Option C¬	100¬	70¬	64 - 85
Option D (DCA)	150¬	>82¬	>64 - 100

The resources required to perform these test are often limited. Consequently, methods have evolved that use existing solder bump footprints with known fatigue histories to predict performance of a solder bump footprint with similar characteristics and packaging but varying DNP.

A scaling factor is used to adjust the known fatigue history to predict the performance of the new bump footprint. It is a relative comparison of DNP's and is calculated as follows:

where,

$$\Psi = \left(\frac{DNP_1}{DNP_2}\right)^{1.9}$$

Ψ = Multiplier used to predict performance
DNP_1 = Distance to the neutral point of existing bump footprint
DNP_2 = Distance to the neutral point of new bump footprint

Figure 9-3 illustrates how Ψ is used to predict fatigue behavior. When the new footprint DNP is larger, Ψ will be less than 1.0 decreasing the predicted fatigue performance. The opposite is true if Ψ is equal to or greater than 1.0.

The geometrical shape of the solder joint can greatly affect the local strain. The top and bottom diameters combined with the volume of solder will determine the height of the joint, a prime influence on reliability, and the notch acuity of the joint as the joint interfaces, a second-order effect. Therein, stress "riser" factors are created which can influence crack initiation and propagation. Ideally a tall, slender solder column will distribute the strain in the solder joint and extend fatigue lifetime. A short, squat joint with sharp, reentrant shape of the solder approaching the UBM will reduce lifetime.

9.2.2.2 Temperature Cycling Frequency

Testing of the life of components under actual use conditions would take as long as the design life of the component. For this reason, packages are tested by accelerating the thermal cycles, i.e., by increasing the temperature range and decreasing the hold times at each end of the cycle. Increasing the temperature range, subjects the joints to greater strain the extent of which is determined by the thermal expansion mismatch between the different materials. Hence, increasing the temperature range should increase the damage stored in the joints if enough time is allowed for the stresses generated in the solder joints to relax out. If enough time is not allowed during ramp up/down times or during the hold times, which is typically the case with accelerated tests, then the damage stored is not equal to what it would be if the solder joints were allowed to relax completely. However, it needs to be noted that increasing the temperature range of the test much beyond the temperatures in the field will cause a confounding of multiple damage mechanisms.

The design of the temperature cycle ideally should be such

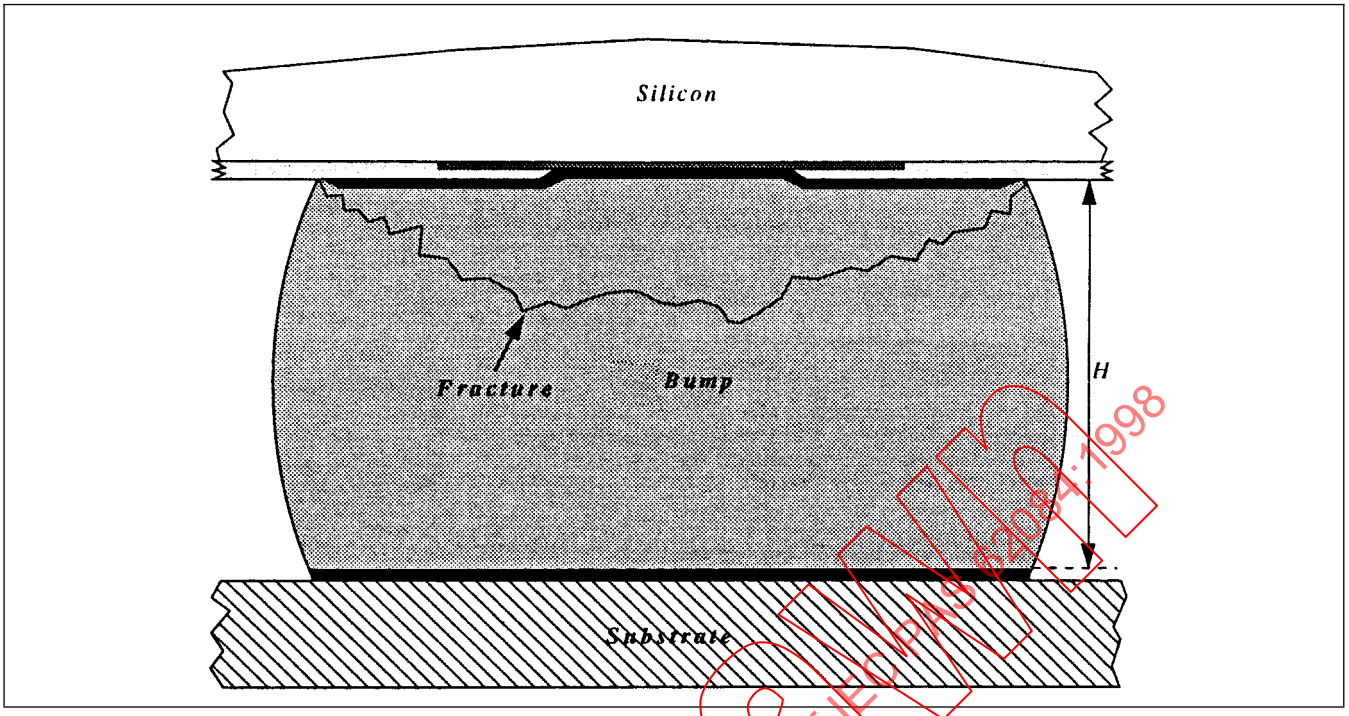


Figure 9–2¬ Fracture Due to Fatigue/Creep Interaction

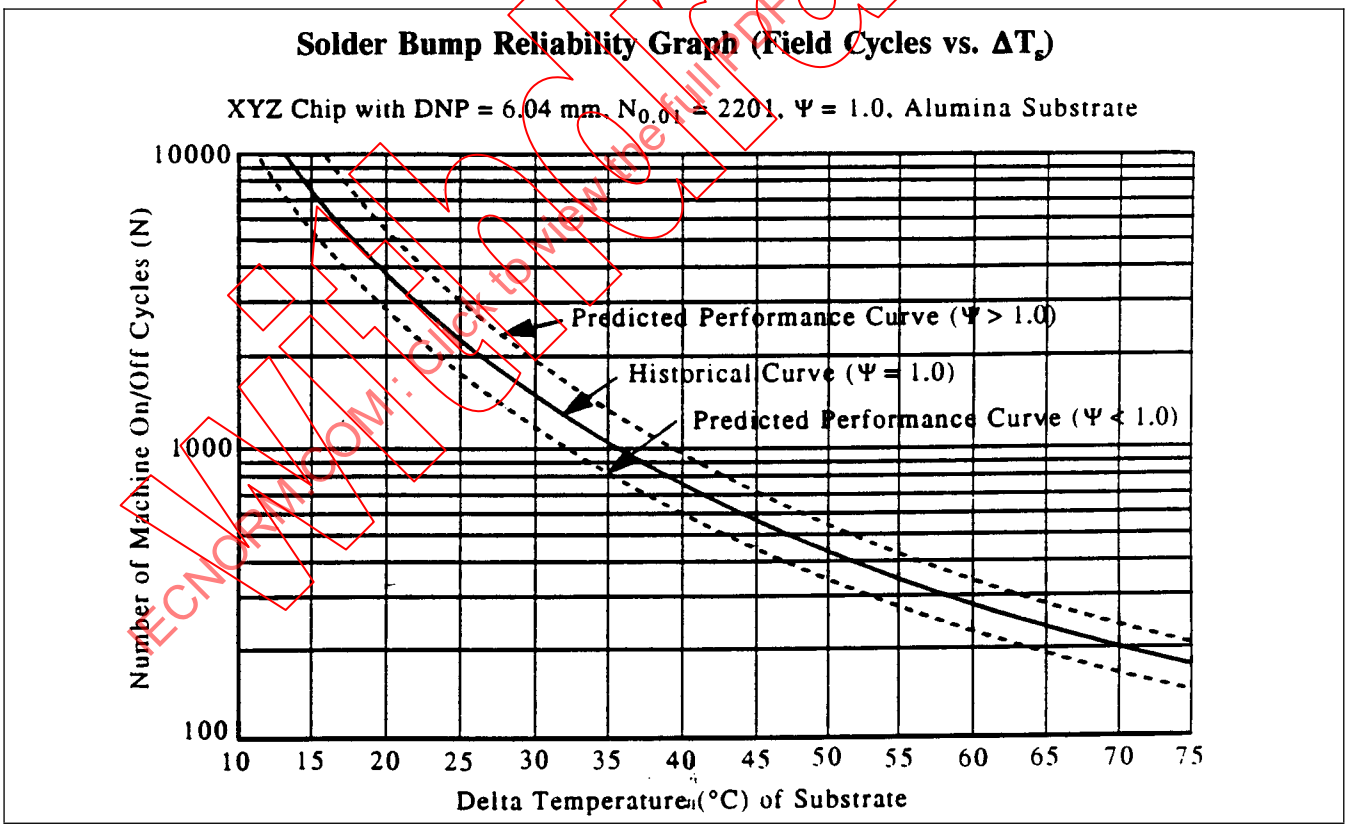


Figure 9–3¬ Example of Solder Bump Fatigue Curves

that the stresses generated because of the strain mismatch have enough time to relax out. However, this is impossible for accelerated testing. Increasing the frequency of cycling can change the failure mechanism and/or suppress the damage stored per cycle in the solder joints. For example, changing the temperature rapidly during thermal shock testing can cause high strain rates to be imposed on the solder, which changes the failure mechanism.

For eutectic solders it is necessary to have hold times ranging from 5 mins to 10 mins to relax out the stresses in the joints. The stress relaxation in high lead (90%Pb-10%Sn) solders is slower than that observed in eutectic solders and therefore the hold times at the temperature extremes need to be greater .

Thermal cycles represent the number of thermal excursions a flip chip or chip scale product will be exposed to during its lifetime. These excursions consist of power on-off cycles and environmental temperature fluctuations.

Temperature cycling tests the inelastic properties of the solder, namely stress-relaxation as a function of temperature and time. Depending on the temperature , a threshold exists below which fatigue becomes independent of cycle hold time, and there frequency are totally relieved by the flow of solder. This threshold varies between 6 and 24 cycles per day (cpd) as shown in figure 9-4.

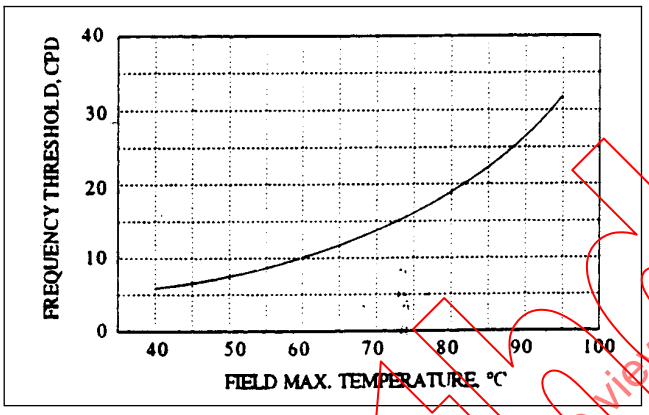


Figure 9–4¬ Thermal Cycling Frequency Threshold vs. Temperature

9.2.2.3 Chip Underfill

Chip underfill can enhance reliability by reducing solder bump fatigue. When correctly applied (e.g., all flux residues are removed from beneath the chip and the underfill resin of appropriate CTE is dispensed void free) chip underfill can significantly reduce shear strain. By physically constraining the bump, chip and substrate, chip underfill reduces strain caused by thermal excursions. This reduction, retards the onset of creep wear-out allowing the maximum operating temperature to be extended from 100°C to 125°C. Further extensions may be possible with additional testing and modeling.

DNP can also be increased with chip underfill allowing the use of bumps to be applied to larger chips and larger footprints. Again, this is due to the overall reduction in strain afforded by the chip underfill.

Present reliability test data indicates a conservative 5X improvement in solder bump fatigue life with chip underfill. Ongoing testing will eventually characterize the wear-out limits of chips with underfill. Devices currently on test have exceeded 10X the number of cycles required to induce equivalent fatigue failure rates in identical, non-underfilled devices.

Based on this data, chip underfill may have eliminated the remaining solder bump wear-out failure mode - fatigue.

Alternative solutions, to the expansion between the chip and the substrate, are provided by some of the chip scale packages. The combination of compliant lead and sufficient soldered ball height reduces the need for underfill. (See figure 9-5)

9.2.2.4 Solder Alloy Composition

Many different solder alloy compositions are used in electronic components. They can range from completely different alloy systems such as lead-indium and lead-tin to similar alloy systems but with different ratios of components such as, 90%Pb-10%Sn and 37%Pb-63%Sn.

The activation energy for atomic diffusion depends on the atomic species as well as the other atoms around oi. Therefore, the activation energies for different alloy systems and different compositions is different. For example, 95%Pb-5%Sn solder compositions have activation energies in the range of 0.122 ev/k to 0.125 ev/K, while 37%Pb-63%Sn has an activation energy around 0.9 eV/K. The activation energy is influenced by mechanical stress imposed on the system. Creep processes are dependent on the activation energies. Therefore different alloy systems under stress can experience different amounts of deformation.

The relation between stress experienced by the solder joint and the activation energy is expressed by the general creep equation:

$$\varepsilon = A\delta^{n}\left(\frac{-Q}{kT}\right)$$

Where,

- ε = strain rate
- A = constant
- δ = stress
- n¬ = stress exponent
- Q = activation energy
- k¬ = Boltzmann constant
- T = temperature in degrees kelvin

9.2.3 Reliability Modeling

The equation generally used for first-order solder bump fatigue approximation is a modified Coffin-Manson equation for low cycle metal fatigue.

$$N = \alpha(\Delta\delta^{1.9}(f)^{1/3}\exp\left[\frac{\Delta H}{kT_M}\right]$$

Where,

- N = number of on/off cycles to failure
- α = material constant
- f¬ = thermal cycling frequency

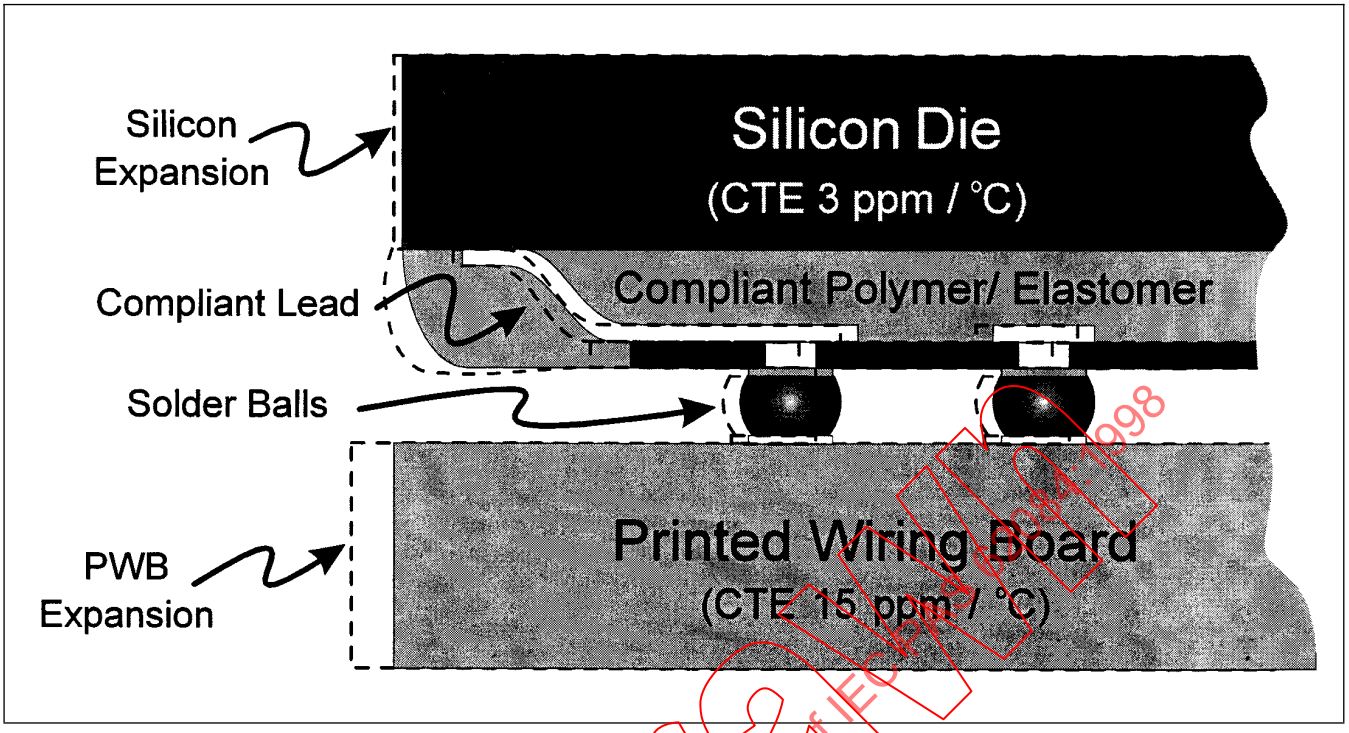


Figure 9–5¬ Compliant Lead Alternative to Underfill

ΔH = activation energy, eV
k¬ = Boltzman's Constant (8.617 x 10^{-5}eV/K)
T_M = maximum operating temperature (K)
$\Delta\delta$ = strain on the outermost functional bumps

Based on historical data, low cycle metal fatigue activation energies range from 0.122 (5% Sn) to 0.125 eV/K (3% Sn). Thus, the equation can be written as follows:

$$N = \frac{1450}{\alpha(\Delta\delta)^{1.9}(f)^{1/3}\exp} T_m \exp\left[\frac{1450}{T_M}\right]$$

Some general statements and conditions apply to this equation:

- Thermal excursions are restricted to 0-100°C for quantitative field reliability projections. The equation may be used as a ''rough'' estimate of performance outside of this temperature range.
- Alternate substrates, e.g. aluminum nitride, can be used to reduce the shear strain below that of alumina.
- This equation is not applicable for predicting reliability of bumps that do not collapse, such as a DCA bump. With DCA, a collapsed bump is not formed. Instead, the solder on the organic board wets to the bump. Thus, bump height is at least equal to the bump diameter. Efforts to establish accurate models for DCA are underway and will be included when available.

Thermal cycle testing is used to establish solder bump fatigue behavior as shown in Figure 9-4. These lognormal distributions are characterized by the cycles to reach 16% cumulative failures (N_{16}) and 50% cumulative failures (N_{50}) and an associated sigma (σ) that describes the slope or shape of the curve. From these curves, expected field performance can be predicted.

Cumulative failures only count the first solder bump failure on a device during testing. Subsequent failures on that device are not factored in the total.

Note: In this application, σ is not used to describe the standard deviation of the population of thermal cycle failures only the slope or shape of the lognormal distribution.

Based on lognormal distribution, σ is calculated as follows:

$$\sigma = 1n\frac{N_{50}}{N_{16}}$$

The primary use of σ is to extrapolate probabilities of failure and field survival rates from charts similar to figure 9-6. If $N_{0.50}$ and σ are known, mathematical methods can be used to predict fatigue performance, i.e., $N_{0.01}$. $N_{0.01}$ represents 0.01% or 100 ppm cumulative failures.

High N_{50} values are often mistaken for high reliability. This is not always the case. Figure 9-7 illustrates three curves with identical N_{50}'s but varying α. As σ varies between 0.2 and 0.4, the $N_{0.01}$ or 100 ppm point varies from 1000 to 2000 cycles, respectively. Thus, it is desired for σ to approach zero and N_{50} be high in order to increase the $N_{0.01}$ intercept.

The value of will vary based on time, materials, and manufacturing processing. It typically ranges from 0.20 to 0.40.

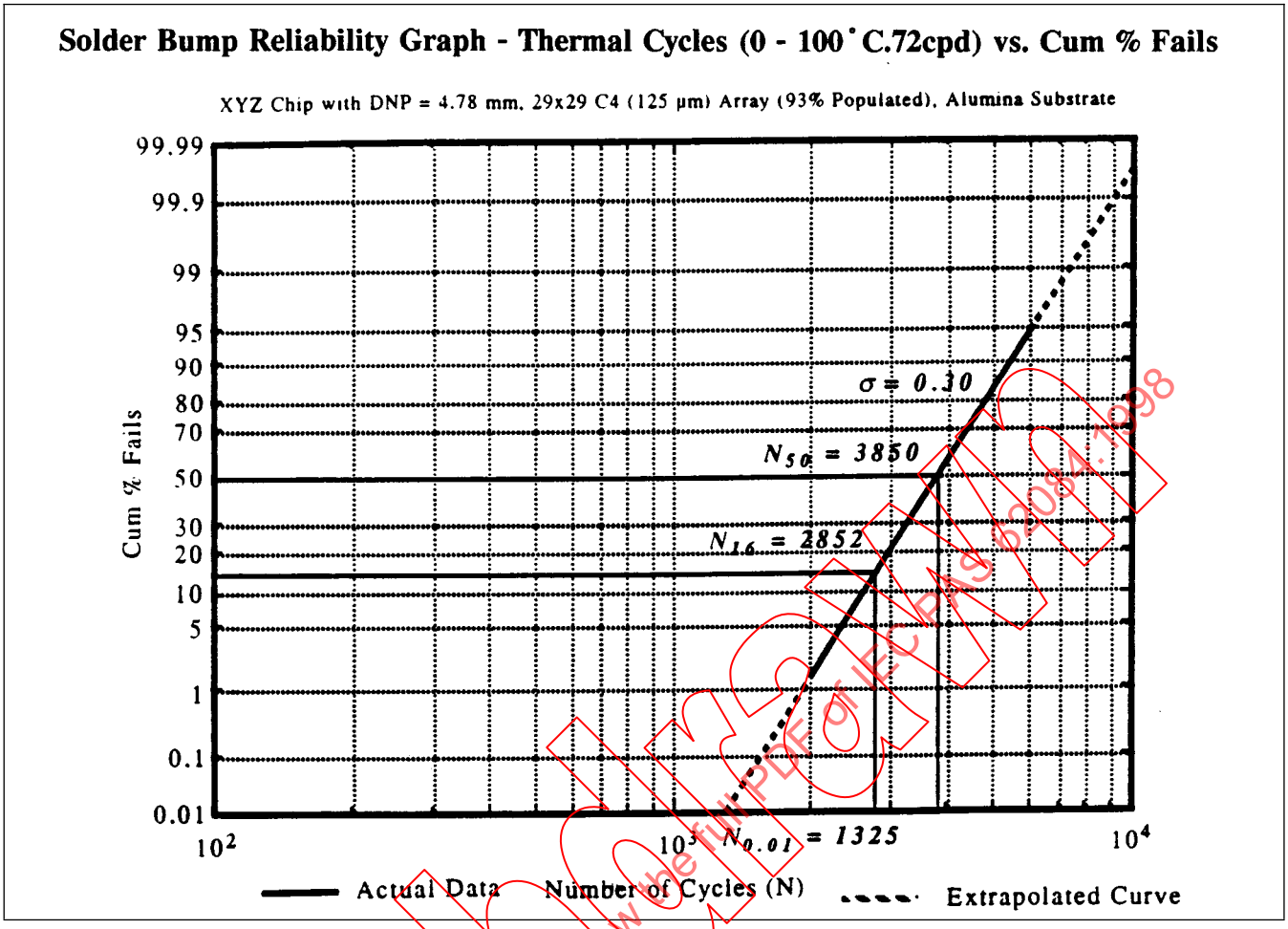


Figure 9–6¬ Lognormal Distribution of Thermal Cycle Failures

Again, the important value is the $N_{0.01}$ as this represents the onset of fatigue life failures.

Field applications often vary from laboratory test conditions. Thermal excursions, frequency, maximum temperature and DNP may all vary depending on the application. Assuming the C4 characteristics of devices are similar laboratory test results can be used to predict fatigue life as will be shown. Prediction of field fatigue life to laboratory tests is accomplished using an acceleration factor (AF) as follows:

$$N_f = AF \times N_1$$

where,

N_f = number of field on/off cycles to failure
AF = acceleration factor
N_1 = number of lab on/off cycles to failure

N_f and N_1 are the respective modified Coffin-Manson equations for field and laboratory conditions. Thus,

$$N_f = A_f(\Delta\delta_f^{-1.9})(f_1)^{1/3} \exp\left[\frac{1450}{T_{Mf}}\right]$$

$$N_1 = A_f(\Delta\delta_f)^{-1.9}(f_1)^{1/3} \exp\left[\frac{1450}{T_{Mf}}\right]$$

Solving for the acceleration factor AF,

$$AF = \frac{N_f}{N_1}$$

$$AF = \left[\frac{A_f}{A_1}\right]\left[\frac{\Delta\delta^1}{\Delta\delta_f}\right]^{1.9}\left[\frac{f_f}{f_1}\right]^{1/3} \exp\left[1450\left[\frac{1}{T_{Mf}} - \frac{1}{T_{M1}}\right]\right]$$

Variations between field and laboratory conditions affect the values of terms in the AF equation. Differences in thermal excursions and DNP changes the strain term, variations in thermal cycling frequency alters the frequency term and differences in maximum temperature impacts the exponential component.

The factor for DNP variation is included in the strain term of the AF equation. Also, for calculations, the field thermal cycling frequency factor (f_f) should be at least 6 -24 cycles per day (cpd) or greater even if field conditions are less as shown in figure 9-4. It should be noted that these equations do not include the hold times, which directly relate to frequency.

An example for determining the acceleration factor is detailed below:

Example 1¬ A new version of a chip that was thermal

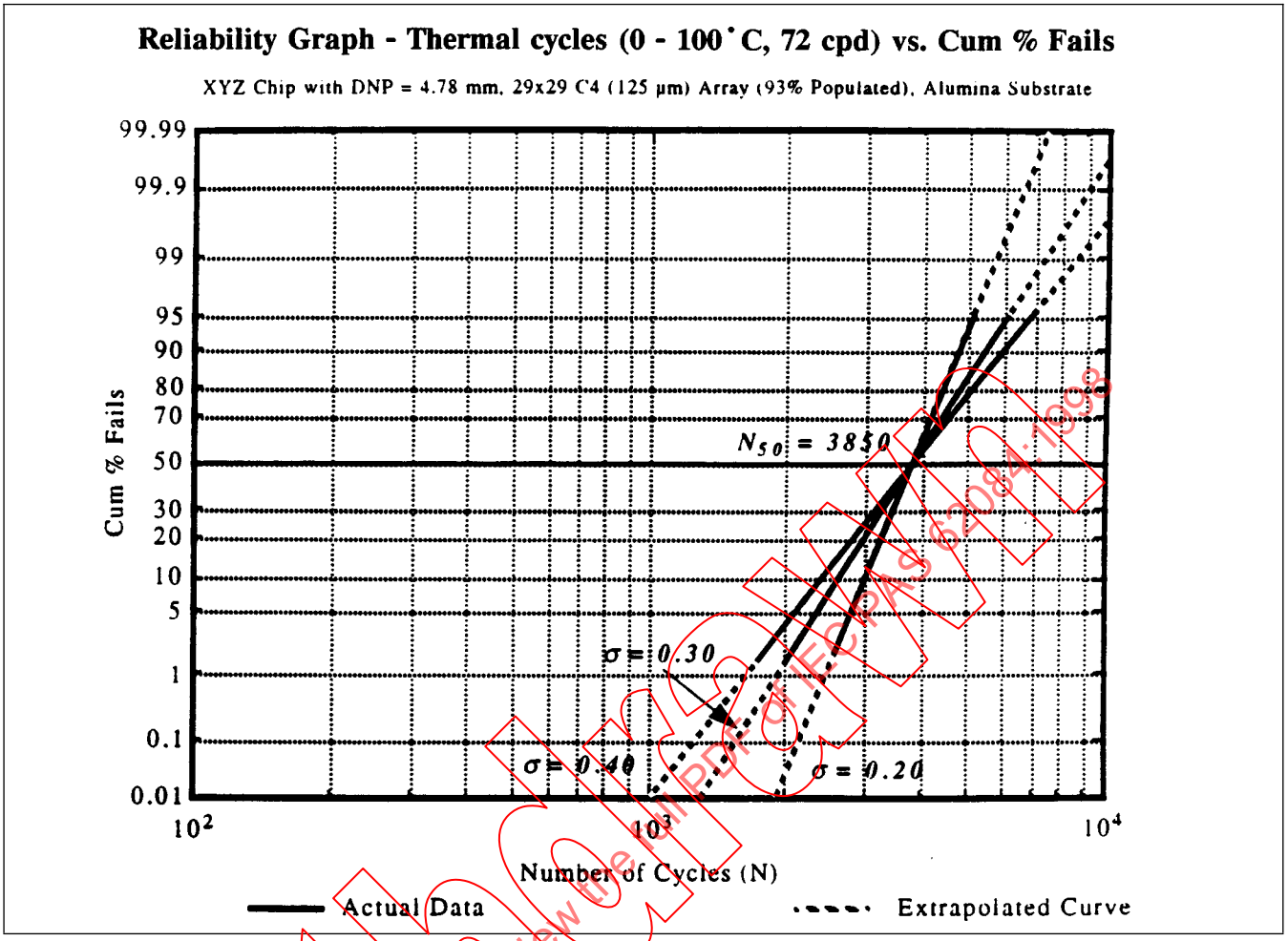


Figure 9–7¬ Effect of α on Reliability

cycle tested in the laboratory (Figure 9-4) is being proposed for field application. the N0.01 field fatigue life needs to be estimated in order to predict reliability.

Approach¬ Since laboratory thermal cycle tests were performed on a similar chip and knowing the specific differences between the chips, the acceleration factor from lab to field can be determined and used to predict the N0.01 field fatigue life of the new chip. This is illustrated below:

Table 9–4¬ Acceleration Factor Values for Example 1

Variation¬	New Chip¬	Lab Chip
DNP5.02 mm¬	4.78 mm	72 cpd
Thermal Cycle Frequency (cycles per day)	24 cpd¬	
Thermal Excursion (ΔT)	70°C¬	100°C
Thermal Activation (Max Temperature)	343 K¬	373 K

Assuming the chips and substrates are at the same temperature and the material constants are equal then, only DNP and ΔT change in the strain term. The AF equation can be simplified as follows:

$$AF = \left[\frac{DNP_1 \Delta T_1}{DNP \Delta T_f}\right]^{1.9} \left[\frac{F_f}{F_1}\right]^{1/3} \exp\left[1450\left[\frac{1}{T_{Mf}} - \frac{1}{T_{M1}}\right]\right]$$

Substituting the values in, yields:

$$AF = \left[\frac{(4.78)\,(100)}{(5.02)\,(70)}\right]^{1.9} \left[\frac{24}{72}\right]^{1/3} \exp\left[1450\left[\frac{1}{343} - \frac{1}{373}\right]\right]$$

Using this AF value and the $N^{0.01}$ value from Figure 9-8, the new chip $N^{0.01}$ value can be predicted as shown:

$$N_{0.01} = 1.75 \times 1325 \text{ cycles} = 2319 \text{ cycles}$$

The new, field chip fatigue life curve can be plotted as shown in figure 9-8.

9.3 Reliability Testing

Generally, the testing guidelines established in IPC-SM-785, Guidelines for Accelerated Reliability Testing of Surface Mount Solder Attachments should be followed.

The reliability of the flip chip assemblies will be determined by design decisions. Assumptions are made for each

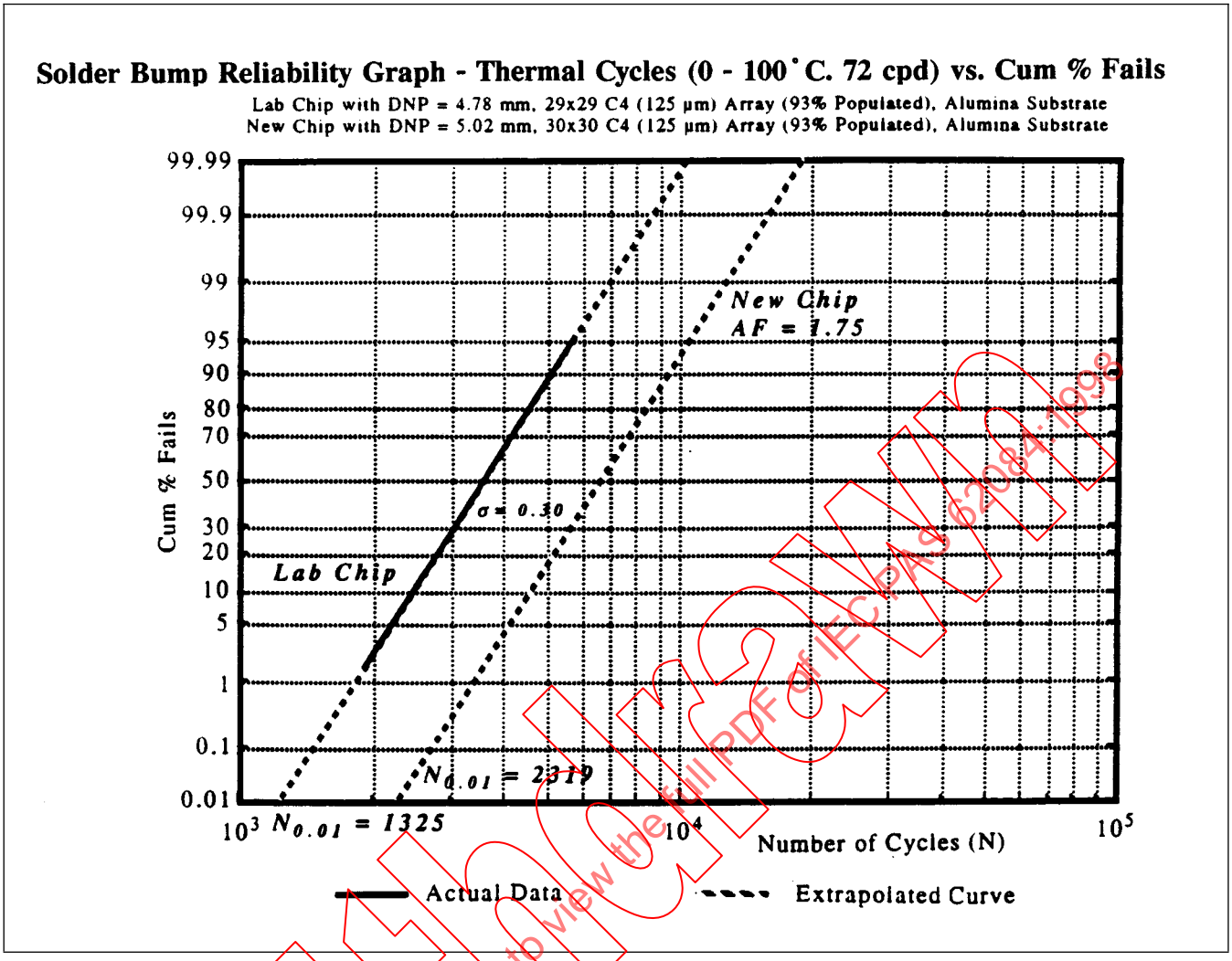


Figure 9–8¬ Acceleration Factor Shift of Lognormal Distribution

design as to how long it has to survive, in what environment the product will be deployed, and what is an acceptable failure probability. Table 9-5 shows the worst case anticipated use thermal environments, and also indicates levels of accelerated temperature testing which might correspond to these environments.

Accelerated testing is generally applied in electronics to examine product robustness to anticipated environmental exposures. Common methods include, but should not be limited to high and low temperature storage, power-temperature cycling, thermal cycling/shock and mechanical cycling/shock.

The types of tests selected for a specific product qualification should be based on actual use environmental, or a worst case composite of all intended uses. It should also be selected based on the types of materials and assembly processes used in the product. thermal and mechanical cycling are usually selected to manifest fatigue mechanisms, whereas thermal ad mechanical shock would be selected to look at fracture and interface adhesion factors.

In defining a test philosophy, a method should be chosen that will accelerate and manifest application failure mechanisms without exceeding basic design and material performance limitations. Accelerated testing should be conducted on process coupons, which include interconnects at worst case design criteria, and they should be representative of production processes.

These should include daisy chained structures to obtain data on a statistically significant number of chips, and also very precise four wire resistance measurement structures should be included to look at single interconnect degradation.

The test structures should be placed in the flip chip interconnect pattern carefully to look at geometry related affects, such as fatigue is worst at the most distant locations from the center of the thermal expansion.

Table 9-5 includes only thermal test criteria. Flip chip

assemblies will, in many situations, be packaged nonhermetically. The thermal testing will test for thermal mismatch fatigue problems, but testing should also be considered which addresses the mechanical and chemical stability when exposed to those environmental factors. Table 9-1 lists the types of use environment that could be expected.

It is important that the flip chip assembly user establish for their anticipated applications that each environmental factor is examined for their selected flip chip approach. If the flip chip assembly is to be used in a severe environment and is nonhermetic, a sequence of tests should be established to address this usage. It would be advisable to run a thermal/mechanical test sequence followed by moisture and corrosion type exposures. The long term reliability would be more completely assessed by a sequence which looks at several interacting environmental factors.

Thermal cycling is used to accelerate bump fatigue wearout. Products and application environments can be extrapolated using a modified Coffin-Manson equation approach.

Most bump fatigue data has been generated by cycling from 0°C to 100°C (at the substrate), three times per hour (72 cpd). Figure 9-9 illustrates the test conditions from which most solder bump fatigue data has been generated.

the extended thermal cycling.

These are the fundamental reasons for only testing from 0°C to 100°C and not testing at higher and lower temperatures, i.e., −55°C to + 125°C, unless these temperatures truly represent the field application.

At low temperatures, thermomechanical stresses on the bump are high. Yet, these stresses do not typically result in failures because the solder is relatively strong at the lower temperature.

However, during thermal cycling, these stresses relax with increasing temperature . At high temperature, the thermal and mechanical energy released promotes relaxation and grain growth. This changes the physical properties of the solder, lowering its fatigue strength.

In most instances, this is an invalid picture of the projected solder bump life. Unless the application demands accelerated thermal cycle testing from −55°C to + 125°C, the conditions in table 9-5 are strongly recommended followed by extrapolation to the desired operating conditions to estimate bump life.

The number of 0°C to 100°C test cycles required to establish a wear-out failure distribution varies from a few thou-

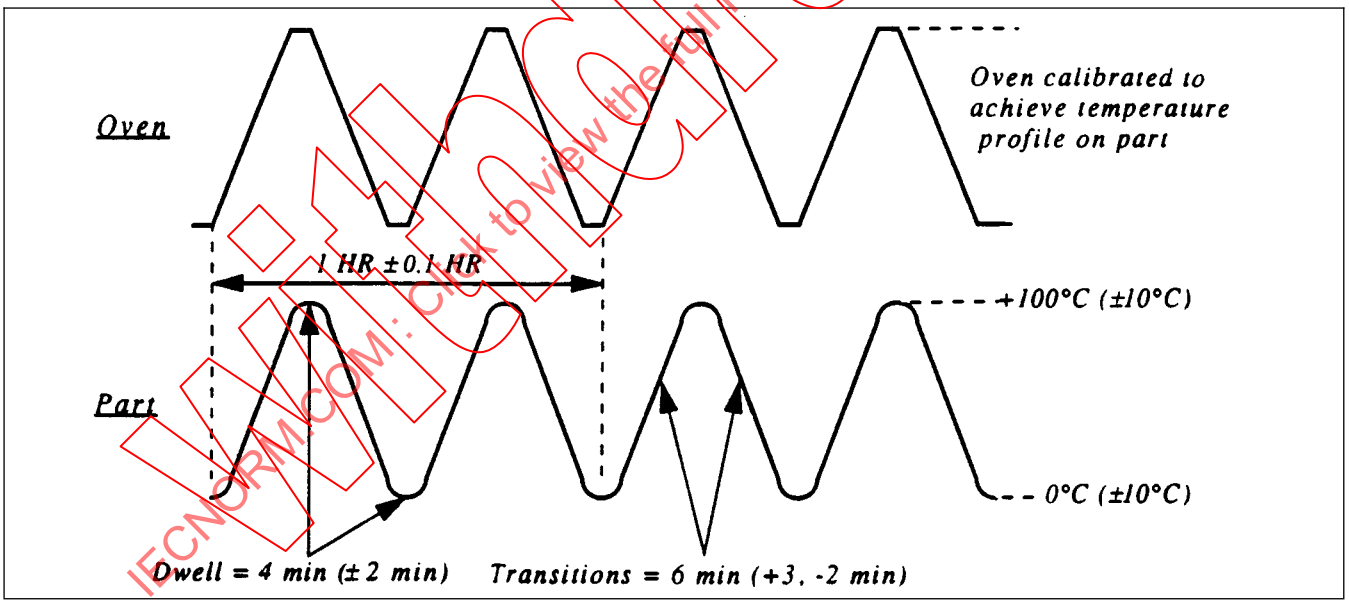


Figure 9−9¬ Thermal Cycling Test Profile

These test conditions are considered the most effective for accelerating fatigue wearout. They reflect a reasonable range of thermal cycling without introducing invalid failure mechanisms. Occasional temperature excursions such as those encountered in shipment are not significant contributors to solder bump fatigue.

Testing for shipment conditions may be conducted using pre-conditioning thermal shock tests (figure 9-10). These stress tests may be performed on the same samples prior to

sand to tens of thousands. This depends on the DNP, the substrate material, underfill, etc. Test duration should be long enough to generate sufficient failures (at least 50%) such that a cumulative percent failure (extrapolated) can be determined with a high level of confidence.

Other, non-solder bump fatigue stress testing involves corrosion testing in non-hermetic packages run for 1000-3000 hours at 85°C and 81% Relative Humidity (RH) with an electrical bias comparable to the product application. For

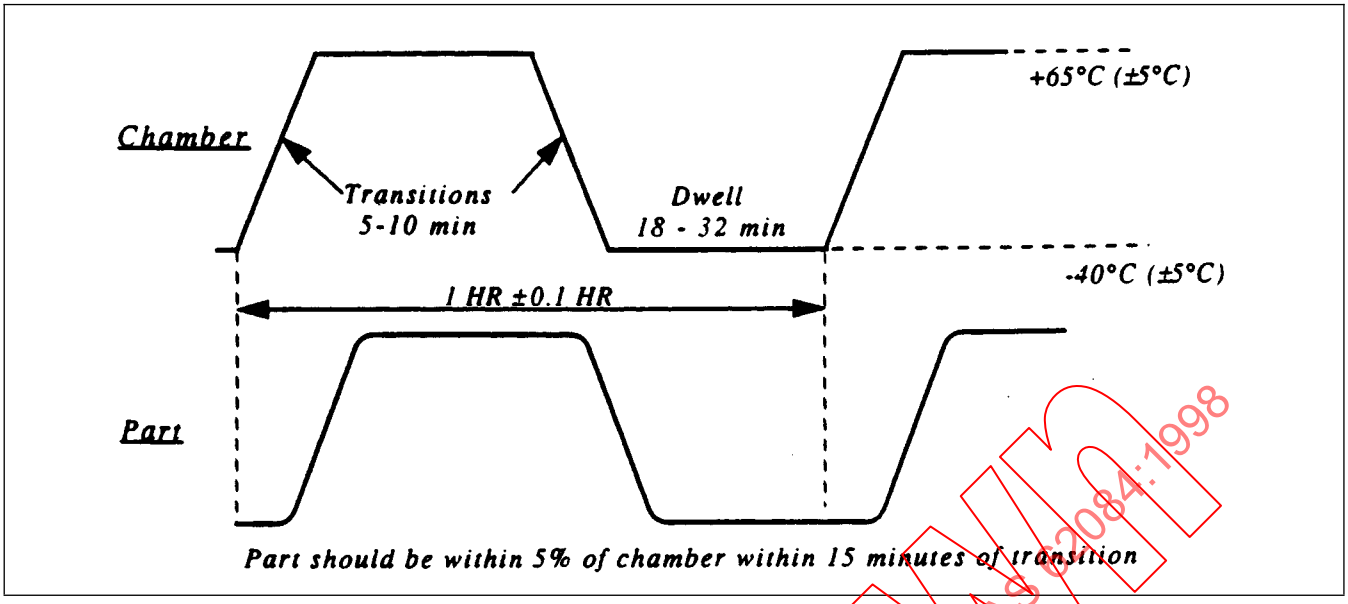


Figure 9−10¬ Pre-Conditioning Thermal Profile

metal migration, a higher electrical bias is used to accelerate the transport mechanism. To prevent condensation 81% RH is used during thermal cycling.

9.3.1 Wear-Out Mechanisms

Flip chip interconnects have many potential failure modes, but there are primarily five degradation or wear out mechanisms which affect metal interconnects. These are: creep, corrosion, electromigration, solid state diffusion, fatigue. These mechanisms can all lead to degradation and eventual failure of flip chip interconnects. They are very dependent on materials used, processing defects and reduced geometries, assembly stresses, and environmental conditions.

Creep is caused by stresses or displacements applied to the interconnect in one direction. Properly selected materials can minimize this mechanism for most environments.
Corrosion could become a concern in some applications. The flip chip assemblies are used in a number of nonhermetic applications. Reliable operation without corrosion affects will depend on process control in terms of well sealed passivation sites, cleaning of the flip chip attach sites, and controlling environmental exposure to avoid contaminants from the application environment. In general, this wear out mechanism is best controlled through design and process control.

Electromigration mechanisms are affected by current density. If the designer keeps current density within the constraints of the materials being used, and there is strict process control on defects, electromigration should not be a problem for most applications.

9.3.2 Reliability Factors

Fatigue is the primary mechanism for wear out of flip chip assemblies. Fatigue is accumulated damages caused by cyclic thermal stressing by the environment or application. This cycling behavior causes plastic deformation within the flip chip interconnects, which initiates a crack that will eventually grow to open or substantially degrade the interconnect. Most system applications will have temperature changes which will eventually fatigue the interconnects.

It is important that the user of flip chip technologies have a fundamental understanding of the intrinsic material properties and on the design/process driven dimensions for their application. The application environment factors involved in fatigue are temperature, temperature cycle range, and hold times of the temperature cycling. The flip chip assembly will affect the reliability in terms of presence of underfill, the symmetry of interconnect pattern, misalignments and variations of bump geometries and device size.

Under thermal cycling, the failures will first occur at locations farthest from the thermal expansion center point. It is possible to develop detailed models, as most are based on some form of the Coffin-Manson equation for low cycle fatigue. A model for each flip chip approach could be established, if the geometry, material and application conditions are known. There is little published data on flip chip reliability. It is extremely important that a user establish this level of knowledge to apply these technologies reliability.

Chip underfill can substantially enhance fatigue life. When the underfill is applied correctly, it reduced the flip chip joint strain level by constraining the expansion of the flip chip interconnect to be used in a wider range of environments and lager device sizes can be accommodated.

Underfill material must be carefully selected, so that it adheres to the assembly surface, but does not adversely stress the flip chip interconnect joints. The material must

Table 9–5¬ Representative Realistic Worse Case Use Environments for Surface Mounted Electronics and Recommended Accelerated Testing for Surface Mount Attachments by Most Common Use Categories

Use Category	Worst-Case Use Environment¬					Accelerated Testing					
	Tmin °C	Tmax °C	ΔT(1) °C¬	tpt hrs	Cycles/ year	Years of Service	Accept. Failure Risk %	Tmin °C	Tmax °C	ΔT(2) °C¬	$t_D$min
1. Consumer	0¬	+60¬	35¬	12¬	365¬	1–3¬	–1¬	+25¬	+100¬	75¬	15
2. Computers¬	+15¬	+60¬	20¬	2¬	1460¬	–5¬	–0.1¬	+25¬	+100¬	75¬	15
3. Telecom¬	–40¬	+85¬	35¬	12¬	365¬	7–20¬	–0.01¬	+25¬	+100¬	75¬	15
4. Commercial Aircraft	–55¬	+95¬	20¬	12¬	365¬	–20¬	–0.001¬	0¬	+100¬	100¬	15
5. Industrial Automotive Passenger Compartment	–55¬	+95¬	20 &40 &60 &80	12 12 12 12	185 100 60 20	–10¬	–0.1¬	0¬	+100¬	100 & COLD(3)	15
6. Military Ground & Ship	–55¬	+95¬	40 & 60¬	12 12	100 265	–10¬	–0.1¬	0¬	+100¬	100 & COLD(3)	15
7. Space leo geo	–269¬	+95¬	3 to 100	1 12	8760 365	5–30¬	–0.001¬	0¬	+100¬	100 & COLD(3)	15
8. Military Avionics	–55¬	+95¬	40 60 80 &20	2 2 2 1	365 365 365 365	–10¬	–0.01¬	0¬	+100¬	100 & COLD(3)	15
9. Automotive Underhood	–55¬	+125¬	60 &100 &140	1 1 2	100 300 40	–5¬	–0.1¬	0¬	+100 & COLD(3)	100 & LARGE	15 ΔT(4)

& = in addition

1) ΔT represents the maximum temperature swing, but does not include power dissipation effects; for power dissipation calculate Δ; power dissipation can make pure temperature cycling accelerated testing significantly inaccurate. It should be noted that the cyclic temperature range, ΔT is not the difference between the possible minimum, T_{MIN}, and maximum T_{MAX}, operational temperature extremes; Δ is typically significantly less.

2) All accelerated test cycles shall have temperature ramps <20°C/minute and dwell times at temperature extremes shall be 15 minutes measured on the test board. This will give ~24 test cycles/day.

3) The failure/damage mechanism for solder changes at lower temperatures; for assemblies seeing significant cold environment operations, additional "COLD" cycling from perhaps –40 to 0°C with dwell times long enough for temperature equilibration and a for a number of cycles equal to the "COLD" °C operational cycles in actual use is recommended.

4) The failure/damage mechanism for solder is different for large cyclic temperature swings traversing the stress-to-strain –20 to +20°C transition region; for assemblies seeing such cycles in operation, additional appropriate "LARGE ΔT" testing with cycles similar in nature and number to actual use is recommended.

have properties which allow easy application to avoid process defects, and it also must not contain or trap contaminants which could initiate corrosion related problems.

If an underfill fails, it will most likely lose adhesion to the device and/or substrate. This would lead to failure due to fatigue, creep, or it could also increase chances of corrosion and other failure mechanisms.

9.3.3 Event Related Failures

The assembled flip chip assembly can experience unexpected thermal and/or mechanical transients, which are isolated events related to maintenance or just abnormal system operating conditions.

These events could, if they are severe enough, cause catastrophic failure, or they can initiate failure sites which will then fail at a later date. These types of failures are best accommodated by selecting the most robust technology that can be used. There are some additional failure mechanisms which need to be considered when using flip chip technology.

Flip chip interconnects are usually high in lead content for ductility. This can lead to trace amounts of radioactive elements which emit alpha particles which can cause soft errors in semiconductor devices. It is important to factor the lead and underfill, much lower level particle emissions for proximity to sensitive device features to minimize this affect.

Most semiconductor devices have ESD protection placed near the pads and most available devices are not configured for flip chip interconnection. Rerouting layers will be

applied to the surface of the device to convert these devices to a flip chip array configuration. This will result in some metal runs in close proximity to device structures that are not protected.

9.4 DESIGN FOR RELIABILITY (DfR)

The reliability of electronic assemblies requires a definitive design effort that has to be carried out concurrently with the other design functions during the developmental phase of the product. There exists a misconception in the industry, that quality manufacturing is all that is required to assure the reliability of an electronic assembly.

While of course, consistent high quality manufacturing—and all that this implies in terms of Design for Manufacturability (DfM), Design for Assembly (DfA), Design for Testability (DfT), etc., is a necessary prerequisite to assure the reliability of the product, only a Design for Reliability (DfR) can assure that the design, manufactured to good quality, will be reliable in its intended application.

Thus, adherence to quality standards is necessary but not sufficient. For example, solder joint quality in the U.S. is generally measured against criteria in both IPC-A-620, Acceptability of Electronic Assemblies with Surface Mount Technologies, for overall workmanship and ANSI/J-STD-001, Requirements for Soldered Electrical and Electronic Assemblies. However, meeting these criteria does not assure reliable solder connections, only quality solder connections. To clarify the difference between the two requires an explanation and a definition of reliability. Reliability is defined in IPC-SM-785, Guidelines for Accelerated Reliability Testing of Surface Mount Solder Attachments, by:

Reliability is the ability of a product to function under given conditions and for a specified period of time without exceeding acceptable failure levels.

In the short term, reliability is threatened by infant mortality failures due to insufficient product quality; these infant mortalities caused by defects can be eliminated prior to shipping by the use of appropriate screening procedures. Long term failures are the result of premature wear-out damage caused by inadequate designs of the assembly.

It is for this reason that IPC-D-279, Design Guidelines for Reliable Surface Mount Technology Printed Board Assemblies, is being developed.

9.4.1 Damage Mechanisms and Failure of Solder Attachments

The reliability of electronic assemblies depends on the reliability of their individual elements and the reliability of the mechanical thermal, and electrical interfaces (or attachments) between these elements. One of these interface types, surface mount solder attachment, is unique since the solder joints not only provide the electrical interconnections, but are also the sole mechanical attachment of the electronic components to the PWB and often serve critical heat transfer functions as well.

A solder joint in isolation is neither reliable nor unreliable; it becomes so only in the context of the electronic components that are connected via the solder joints to the PWB. The characteristics of these three elements - component, substrate, and solder joint - together with the use conditions, the design life, and the acceptable failure probability for the electronic assembly determine the reliability of the surface mount solder attachment.

9.4.1.1 Solder Joints and Attachment Types

Solder joints are anything but a homogeneous structure. A solder joint consists of a number of quite different materials, many of which are only superficially characterized. A solder joint consists of:

(1) the base metal at the PWB
(2) one or more intermetallic compounds (IMC)—solid solutions—of a solder constituent—typically tin (Sn)—with the PWB bag metal
(3) a layer from which the solder constituent forming the PWB-side IMC(s) has been depleted
(4) the solder grain structure, consisting of at least two phases containing different proportions of the solder constituents as well as any deliberate or inadvertent contaminations
(5) a layer from which the solder constituent forming the component-side IMC(s) has been depleted
(6) one or more IMC layers of a solder constituent with the component base metal, and
(7) the base metal at the component.

The grain structure of solder is inherently unstable. The grains will grow in size over time as the grain structure reduces the internal energy of a fine-grained structure. This grain growth process is enhanced by elevated temperatures as well as strain energy input during cyclic loading. The grain growth process is thus an indication of the accumulating fatigue damage. At the grain boundaries contaminants like lead oxides are concentrated; as the grains grow these contaminants are further concentrated at the grain boundaries, weakening these boundaries. After the consumption of ~25% of the fatigue life micro-voids can be found at the grain boundary intersections; these micro-voids grow into micro-cracks after ~40% of the fatigue life; these micro-cracks grow and coalesce into macro-cracks leading to total fracture as is schematically shown in figure 9-11.

Surface mount solder attachments exist in a wide variety of designs. The major categories are leadless and leaded solder attachments. Among the leadless solder joints a differentiation has to be made between those without fillets, e.g.,

Flip-Chip C4 (Controlled Collapse Chip Connection) solder joints, BGAs with C5 (Controlled Collapse Chip Carrier Connection) solder attachments, BGAs with high-temperature solder (e.g., 10Sn/90Pb) balls, and CGAs with high-temperature solder columns; and solder joints with fillets, e.g., chip components, Metal Electrode Face components (MELFs), and castellated leadless chip carriers. The leaded solder attachments differ primarily in terms of their compliancy and can be roughly categorized into components with super-compliant leads {K_D<~9 N/mm (~50 lb/in)}, compliant leads (~9 N/mm<K_D<~90 N/ram), and non-compliant leads {(K_D>~90 N/mm (~500 lb/in)}.

The different surface mount solder attachment types can have significantly different failure modes. Solder joints with essentially uniform load distributions, e.g., Flip-Chip, BGA, CGA, show behavior as illustrated in Figure 1. Solder joints with non-uniform load distributions, e.g., those on chips components, MELFs, leadless chip carriers, and all leaded solder joints, show localized damage concentrations with the damage shown in Figure 9-9 preceding an advancing macro-crack.

The solder joints frequently connect materials of highly disparate properties, causing global thermal expansion mismatches [Refs. A-7: 1-4], and are made of a material, solder, that itself has often properties significantly different than the bonding structure materials, causing local thermal expansion mismatches [Refs. A-7: 3, 5].

The severity of these thermal expansion mismatches, and thus the severity of the reliability threat, depends on the design parameters of the assembly and the operational use environment. In Table 9-5 guidelines as to the possible use environments for nine of the more common electronic applications are illustrated [Refs. A-7: 6, 7]. However, it needs to be emphasized, that the information in Table 9-5 should serve only as a general guideline; for some use categories the description of the expected use environment can be rather more complex [Ref. A-7: 7].

9.4.1.2 Global Expansion Mismatch

The global expansion mismatches result from differential thermal expansions of an electronic component or connector and the PWB to which it is attached via the surface mount solder joints. These thermal expansion differences result from differences in the CTEs and thermal gradients as the result of thermal energy being dissipated within active components.

Global CTE-mismatches typically range from $\Delta\alpha$~2 ppm/°C (1 ppm=lx10^6) for CTE-tailored high reliability assemblies to ~14 ppm/°C for ceramic components on FR-4 PWBs. CTE-mismatches of $\Delta\alpha$ <2 ppm/°C are not achievable in reality as a consequence of the variability of the CTE values of the materials involved on both components and PWBs.

Global thermal expansion mismatches typically are the largest, since all three parameters determining the thermal expansion mismatch—the CTE-mismatch, $\Delta\alpha$, the temperature swing, ΔT, and the acting distance, L_D—are large.

This global expansion mismatch will cyclically stress, and thus fatigue, the solder joints. The cyclically cumulative fatigue damage will ultimately cause the failure of one of the solder joints, typically a corner joint, of the component causing functional electrical failure that is initially intermittent.

9.4.1.3 Local Expansion Mismatch

The local expansion mismatch results from differential thermal expansions of the solder and the base material of the electronic component or PWB to which it is soldered. These thermal expansion differences result from differences in the CTE of the solder and those of the base materials together with thermal excursions [Refs. A-7: 3, 5].

Local CTE-mismatches typically range from $\Delta\alpha$~7 ppm/°C with copper to ~18 ppm/°C with ceramic and ~20 ppm/°C

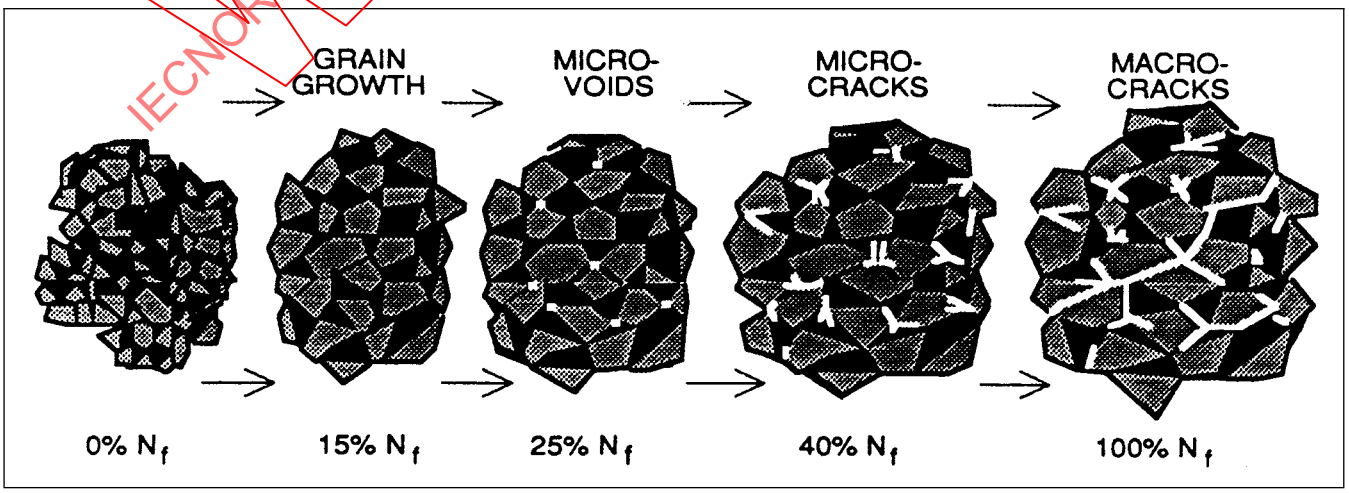


Figure 9−11¬ Depiction of the Effects of the Accumulating Fatigue Damage in Solder Joint Structure

with Alloy 42 and Kovar™. Local thermal expansion mismatches typically are smaller than the global expansion mismatches, since the acting distance, the maximum wetted area dimension, is much smaller—in the order of hundreds of ~µm (tens of mils).

9.4.1.4 Internal Expansion Mismatch

An internal CTE-mismatch of ~6 ppm/°C results from the different CTEs of the Sn-rich and Pb-rich phases of the solder. Internal thermal expansion mismatches typically are the smallest, since the acting distance, the size of the grain structure, is much smaller than either the wetted length or the component dimension—in the order of less than 25 µm (<1 mil) [Ref. A-7: 9].

9.4.1.5 Solder Attachment Failure

The failure of the solder attachment of a component to the substrate to which it is surface mounted is commonly defined as the first complete fracture of any of the solder joints of which the component solder attachment consists.

Given that the loading of the solder joints is typically in shear, rather than in tension, the mechanical failure of a solder joint is not necessarily the same as the electrical failure. Electrically, the mechanical failure of a solder joint results, at least initially, in the occasional occurrence of a short-duration (<1 µs) high-impedance event during either a mechanical or thermal disturbance. From a practical point of view, the solder joint failure is defined as the first observation of such an event.

For some applications this failure definition might be inadequate. For high-speed signals with sharp rise times signal deterioration prior to the complete mechanical failure of a solder joint might require a more stringent failure definition. Similarly, for applications which subject the electronic assemblies to significant mechanical vibration and/or shock loading, a failure definition that considers the mechanical weakening of the solder joints as the result of the accumulating fatigue damage might be necessary.

9.4.2 Reliability Prediction Modeling

9.4.2.1 Creep-Fatigue Modeling

It has been experimentally shown [Refs. 2,4, 12, 13] that the fatigue life of surface mount solder joints can be described by a power law similar to the Coffin-Manson low-cycle fatigue equation [Ref. A-7: 14] developed for more typical engineering metals. For practical reasons and as the direct consequence of the time-dependent stress-relaxation/creep behavior of the solder at typical use environments (see table 9-5), the specialized case of the Coffin-Manson equation requires reversion to the more general strain-energy relationship of Morrow [Ref. A-7: 13]; it also requires that the cyclic strain energy be based on the total possible thermal expansion mismatch and that the exponent is a function of temperature and time to provide a measure of the completeness of the stress-relaxation process. The Engelmaier-Wild solder creep-fatigue equation [Refs. A-7: 1-4, 7, 10], subject to some caveats listed later, relates the cyclic visco-plastic strain energy, represented by the cyclic fatigue damage term, ΔD, to the median cyclic fatigue life for both isothermal-mechanical and thermal cycling [Ref. A-7: 14]

$$N_f(50\%) = \frac{1}{2}\left[\frac{2\epsilon'_f}{\Delta D}\right]^{\frac{1}{c}}$$

[Eq. #1]

where

ϵ'_f = fatigue ductility coefficient, =0.325 for eutectic and 60/40 Sn/Pb solder (for other solders the value of E'_f is expected to be somewhat different).

Solder, uniquely among the commonly used engineering metals, readily creeps and stress-relaxes at normal use temperatures; the rate of creep and stress-relaxation is highly temperature- and stress-level-dependent. Thus, the cyclic fatigue damage term, ΔD, for practical reasons, has to be based on the total potential damage at complete creep/stress relaxation of the solder. For cyclic conditions that do not allow sufficient time for complete stress relaxation ΔD is larger than the actual fatigue damage. The temperature- and time-dependent exponent, c, compensates for the incomplete stress relaxation and is given by

$$c = -0.442 - 6x10^{-4}\,\overline{T}_{S_j} + 1.74x10^{-2}\ln\left(1 + \frac{360}{t_D}\right)$$

[Eq. #2]

where

T_{S_j} = mean cyclic solder joint temperature

t_D = half-cycle dwell time in minutes.

The half-cycle dwell time relates to the cyclic frequency and the shape of the cycles and represents the time available for the stress-relaxation/creep to take place.

Equations #1 and #2 come from a generic understanding of the response of SM solder joints to cyclically accumulating fatigue damage resulting from shear displacements due to the global thermal expansion mismatches between component and substrate. These shear displacements cause time-independent yielding strains and time-, temperature-, and stress-dependent creep/stress relaxation strains. These strains, on a cyclic basis, form a visco-plastic strain energy hysteresis loop which characterizes the solder joint response to thermal cycling and whose area, given as the damage term ΔD, is indicative of the cyclically accumulating fatigue damage. Hysteresis loops in the shear stress-strain plane have been experimentally obtained [Refs. A-7: 13, 17-19].

9.4.2.2 Statistical Failure Distribution and Failure Probability

While the physical parameters define the median cyclic fatigue life from physics-of-failure considerations, solder attachment failures for a group of identical components will follow a distribution like all fatigue results, which typically is best described by a Weibull statistical distribution [Ref. A-7: 18]. Given the statistical distribution, the fatigue life at any given failure probability can be predicted as long as the slope of the Weibull distribution is known. Thus, the fatigue life of surface mount solder attachments at a given acceptable failure probability, x, is —assuming a two-parameter (2P) Weibull statistical distribution—given by

$$N_f(x\%) = N_f(50\%)\left[\frac{1n(1-0.01x)}{1n(0.5)}\right]^{\frac{1}{\beta}} \quad \text{[Eq. \#3]}$$

where

β = Weibull shape parameter or slope of the Weibull probability plot; typically β≈3 for fatigue tests, from low-acceleration tests of stiff leadless solder attachments β≈4 and ≈2 for compliant leaded attachments.

Experimentally, β can be found to be quite variable with more severely accelerated reliability tests resulting in tighter failure distributions and thus giving larger values for β. Values of β in the range of 1.8 to 9.0 have been observed.

There is some, unfortunately as yet inadequate, evidence that for lower failure probabilities a three-parameter (3P) Weibull distribution, postulating a failure-free period prior to first failure [Refs. A-7: 19, 20], may be applicable. From physics-of-failure and damage mechanism considerations, a failure threshold as provided by a 3P-Weibull distribution makes sense, since the fatigue damage in the solder joints has to accumulate to crack initiation and complete crack propagation. While the 2P-Weibull distribution may be overly conservative for designs to very small acceptable failure probabilities (~x<0.1%), a too liberal choice of the failure-free period is definitely non-conservative. This area requires more work.

Also, when designing to low failure probabilities, the variability in the quality of the solder joints may no longer be negligible; also solder joints with latent defects that made it into the field will have in impact on the actual failure experience of a product in the field.

9.4.2.3 Damage Modeling

The assessment of the cyclically cumulating fatigue damage is not a straight-forward task. While Eq. #1 is widely used, the question of how to best quantify the cyclic fatigue damage is still hotly debated. The choices are primarily between more complex finite-element analyses (FEA), which can give more detailed information and can include second-order effects, but require a large number of not fully-supported assumptions [Ref. A-7: 21]; and closed-form empirically-based relationships of the first-order design parameters, which cannot include second-order effects and have use limitations due to their simple nature, but allow, due to their simple form, a direct assessment of the impact of the primary design parameters as well as design trade-offs.

The following cyclic fatigue damage terms are of the simplified closed-form type and should be utilized with the application caveats that follow [Refs. A-7: 1-4, 7, 10, 14, 22].

The cyclic fatigue damage term for leadless SM solder attachments, for which the stresses in the solder joints exceed the solder yield strength and cause plastic yielding of the solder, is

$$\Delta D(\text{leadless}) = \left[\frac{FL_D\,\Delta(\alpha\Delta T)}{h}\right] \quad \text{[Eq. \#4]}$$

Equation #4 contains the design parameters that have a first-order influence on the reliability of SM solder attachments. They are

F¬ = empirical ‘‘non-ideal’’ factor indicative of deviations of real solder joints from idealizing assumptions and accounting for secondary and frequently intractable effects such as cyclic warpage, cyclic transients, non-ideal solder joint geometry, different solder crack propagation distances, brittle IMCs, Pb-rich boundary layers, and solder/bonded-material expansion differences, as well as inaccuracies and uncertainties in the parameters in Eqs. #1 through #4; 1.5>F> l.0 for ball/column-like leadless solder joints (C4, C5, BGAs, CGAs), 1.2>F>0.7 for leadless solder joints with fillets (castellated chip carriers and chip components), F≈1 for solder attachments utilizing compliant leads;

h¬ = solder joint height;

$2L_D$ = maximum distance between component solder joints measured from component solder joint pad centers, L_D is sometimes referred to as the distance from the neutral point (DNP);

T_C, T_S = steady-state operating temperature for component, substrate (T_C>T_S for power dissipation in component) during high temperature dwell;

$T_{C,0}$, $T_{S,O}$ = steady-state operating temperature for component, substrate during low temperature dwell, for non-operational (power off) half-cycles $T_{C,O}$

$T_{S,O}$;¬ = (1/4)(T_C+T_S+$T_{C,O}$ +$T_{S,O}$), mean cyclic solder

joint temperature;

α_C, α_S = CTEs for component, substrate;

ΔD¬ = potential cyclic fatigue damage at complete stress relaxation;

ΔT_C = $T_C - T_{C,O}$, cyclic temperature swing for component;

ΔT_S = $T_S - T_{S,O}$, cycling temperature swing for substrate (at component);

$\Delta(\alpha\Delta T)$¬ = $|\alpha_S \Delta T_S - \alpha_C \Delta T_{C|}$, absolute cyclic expansion mismatch, accounting for the effects of power dissipation within the component as well as temperature variations external to the component;

$\Delta\alpha$ = $|\alpha_C - \alpha_S|$, absolute difference in CTEs of component and substrate, CTE-mismalch, because of the inherent variability in material properties $\Delta\alpha < 2x10^{-6}$ should not be used in calculating reliability.

9.4.2.4 Caveat 1 — Solder Joint Quality

The solder joint fatigue behavior and the resulting reliability prediction equations. Eqs. #1 through #2, were determined from thermal cycling results of solder joints that failed a result of fracture of the solder, albeit sometimes close to the IMC layers. For solder joints for which layered structures are interposed between the base material and the solder joints, these equations could be optimistic upper bounds if the interposed layered structures become the 'weakest link' in the surface mount solder attachments. Such layered structures could be: metallization layers that have weak bonds to the underlying base material, or are weak themselves, or dissolve essentially completely in the solder; oxide or contamination layers preventing a proper metallurgical bond of the solder to the underlying metal; brittle IMC layers too thick due to too many or improperly long high temperature processing steps.

Solder joints which have solder joint heights (gaps) of h<50 to 75µm also require special attention. For solder joints that thin, the gap is essentially filled with intermetallic compounds and those solder metals that do not go into solution with the base metals to form the IMCs. Therefore Eqs. #1 and #2 do not apply because these gaps are not filled with solder [Ref. A-7: 20]. These materials do not creep as readily, if at all, at the prevailing temperatures and are typically more brittle, but much stronger than solder. Thus, fatigue lives are longer than would be predicted from Eqs. #1 and #2 unless overstress conditions occur.

On the other hand, the fatigue lives of solder attachments can be underestimated by Eqs. #1 through #4 if the component is underfilled with a load-bearing substance, e.g., epoxy [Ref. A-7: 23]. Components that are glued-down to the substrate result in higher solder joint fatigue reliability, since the solder joints are loaded in compression when the adhesive contracts on cooling from the solder reflow temperatures. Covercoats can either increase or decrease solder joint fatigue lives depending on the properties of the covercoat and when and how it is applied. Parylene™ has been found to increase the solder joint fatigue live by about a factor of three.

In general, caution might be indicated in all instances were the predicted life is less than 1000 cycles, because the severe loading conditions producing such short lives are likely to produce different damage mechanisms or/and failure modes.

9.4.2.5 Caveat 2 — Large Temperature Excursions

Solder joints experiencing large temperature swings which extend significantly both below and above the temperature region bounded by −20°C to +20°C, in which the change from stress- to strain-driven solder response takes place, do not follow the damage mechanism described in Eqs. #1 and #2 [Ref. A-7: 24]. The damage' mechanism is different than for more typical use conditions and is likely dependent on a combination of creep-fatigue, causing early micro-crack formation, and stress concentrations at these micro-cracks causing faster crack propagation during the high stress cold temperature excursions, as well as recrystallisation considerations.

9.4.2.6 Caveat 3 — High-Frequency/Low-Temperatures

For high-frequency applications, f>0.5 Hz or t_D<l s, e.g., vibration, and/or low temperature applications, T_C< 0°C, for which the stress relaxation and creep in the solder joint is not the dominant mechanism, the direct application of the Coffin-Manson [Ref. 14] fatigue relationship might be more appropriate. This relationship is

$$N_f(50\%) = \frac{1}{2}\left[\frac{2\epsilon_{f'}}{\Delta\gamma_p}\right]^{\frac{1}{c}}$$

[Eq. #5]

where $\Delta\gamma_p$ is the cyclic plastic strain range and $c \approx -0.6$.

It has to be noted, that the determination of $\Delta\gamma_p$ depends on the expansion mismatch displacements and the separation of the plastic from the elastic strains.

For loading conditions of this character, it is possible that high-cycle fatigue behavior may be observed.

9.4.2.7 CAVEAT 4 — Local Expansion Mismatch

For applications for which the global thermal expansion mismatch is very small, e.g. ceramic-on-ceramic or silicon-on-silicon (flip-chip solder joints), the local thermal expansion mismatch becomes the primary cause of fatigue damage. Equation #4 does not address the local thermal expansion mismatch. This reliability problem needs to be assessed using an interfacial stress analysis [Ref. A-7: 25] and appropriate accelerated testing.

Suhir [Ref. A-7: 25] has shown that the interfacial stresses resulting from the local expansion mismatch follow

$$\tau \propto L\,(\alpha_{Solder} - \alpha_{Basr})(T_{max} - T_{min})$$

[Eq. #6]

where L is the wetted length of the solder joint. In addition, besides substantial shear stresses at the interface between the solder joint and the base material to which it is wetted, even larger peeling stresses occur. Both of these stresses are proportional to the parameters given in Eq. #6.

However, since in most applications, the local expansion mismatch results in contributory
damage to the more important damage caused by the global expansion mismatch, this
contra-indication can be ignored without suffering catastrophic consequences.
From the available experimental data, the damage term, to be used in Eq. #1, for the local expansion mismatch alone is

$$\Delta D(local) = \left[\frac{L\Delta\alpha\Delta T}{L_0}\right]$$

[Eq. #7]

where the parameters are the same as in Eq. #6 and L_0=0.1 mm, a scaling wetted length. The local expansion mismatch is than treated as an additional loading condition (see 9.4.2.8 & 9.4.2.9).

9.4.2.8 Multiple Cyclic Load Histories

The loading histories over the life of a product frequently includes many different use environments and loading conditions [Refs. A-7: 26, 27]. Multiple cyclic load histories (e.g., ''Cold'' temperature fatigue cycles combined with higher temperature creep/fatigue cycles (see table 9-5) combined with vibration and local expansion mismatches) all make their contributions to the cumulative fatigue damage in solder joints. Under the assumption that these damage contributions are linearly cumulative—this assumption underlies Eqs. #1 and #2 as well—and that the simultaneous occurrence or the sequencing order of these load histories makes no significant difference, the Palmgren-Miner's rule [Ref. A-7: 28] can be applied.

Frequently the initial reliability objective is stated as an allowable net cumulative damage ratio (CDR). The CDR is calculated as the sum of the ratios of the number of occurring load cycles to the fatigue life at each loading condition and is

$$CDR = \sum_{j=1}^{j} \frac{N_j}{N_{fj}} < 1$$

[Eq. #8]

where

N_j = actually applied number of cycles at a specific cyclic load level j,

N_{fd} = fatigue life at the same specific cyclic load level j alone.

The fatigue life is frequently not completely specified and is normally taken to be the mean cyclic fatigue life. Equation #8 can be used with the allowable CDR significantly less than unity to provide margins of safety, or more accurately, margins of ignorance.

Because the failure of solder joints results from wearout due to fatigue, the failure rate is continuously increasing. This is in stark contrast to the reliability design philosophy of MIL-HDBK-217 [Ref. A-7: 29] which presumes a constant failure rate. These increasing failure rates are properly represented by an appropriate statistical failure distribution.

Thus, to assure low failure risks, the fatigue life should be specified at the acceptable cumulative failure probability at the end of the design life as per Eq. #3. Thus, Eq. #8 is more appropriately written as

$$CDR(x\%) = \sum_{j=1}^{j} \frac{N_j}{N_{fj}(x\%)} = 1$$

[Eq. #9]

where

CDR(x%) = cumulative damage ratio resulting in a cumulative failure probability of x%,

$N_{fj}(x\%)$¬ = fatigue life at the cyclic load level j and a failure probability of x% .

This approach works very well for the design of the solder attachment for a single component. However, it is inadequate for a reliability analysis of a the whole assembly.

9.4.2.9 System Reliability Evaluation

Equations #1 through #9 address the reliability of the SM solder attachment of individual components. Systems consist of a variety of different components most of which occur in multiple quantities. Further, as shown in table 9-5, many use environments cannot and should not be represented by a single thermal cyclic environment, and accumulating fatigue damage from other sources, such as cyclic thermal environments as described in Caveats 2 to 4 as well as vibration, needs to be included also.

For a multiplicity of components, i, in the system, the effect of the various components on the system reliability can be determined from

$$F_{\Sigma}(N) = 1 - \exp\left\{1n(1 - 0.01x)\sum_{i=1}^{i} n_i\left[\sum_{j=1}^{j} \frac{N_{ij}}{N_{f,i,j}(x\%)}\right]^{\beta_i}\right\}$$

[Eq. #10]

where

$F_{\Sigma}(N)$¬ = system cumulative failure probability after N total cycles,

n_i = number of components of type i,

$N_{f,i,j}$ (x%) = actually number of cycles applied to component i at a specific cyclic load level j, Nf, $_{i,j}$ (x%)

β_i = Weibull slope for SM solder attachment of component i.

9.4.3 DfR-Process

For Flip Chip and Chip Scale technologies the biggest reliability concern is the large expansion mismatch between the chip silicon and the polymeric substrate. This either means relatively small chips or the use of organic underfill materials which relieve the solder joints from most of the thermal expansion mismatch loads. The underfill material does however make repairs difficult if not impossible.

Appropriate DfR-measures to improve reliability can take one of two forms, which are best employed in combination for improved reliability margins. These measures are:

1) CTE-tailoring to reduce the global expansion mismatch;
2) Increasing attachment compliancy, e.g., by increasing the solder joint height, to accommodate the global expansion mismatch;
3) Underfilling the gap between the component and substrate;

Further, a DfR procedure aiming at high-reliability should also include

4) Choosing base materials that have not too large a local CTE-mismatch with solder, or
5) In case item (4) cannot be done, reduce the continuous wetted length to reduce interfacial stresses.

CTE-tailoring involves choosing the materials or material combinations of the MLB and/or the components to achieve an optimum ΔCTE. An optimum ΔCTE for active components dissipating power is ~1-3 ppm/°C (depending on the power dissipated) with the MLB having the larger CTE, and 0 ppm/ °C for passive components. Of course, since an assembly has a multitude of components, full CTE-optimization cannot be achieved for all components—it needs to be for the components with the largest threat to reliability. For military applications with the requirement of hermetic—and thus ceramic—components, CTE-tailoring has meant the CTE-constraining of the MLBs with such materials as Kevlar™ and graphite fibers, or copper-Invar-copper and copper-molybdenum-copper planes. Such solutions are too expensive for most commercial applications for which glass-epoxy or glass-polyimide are the materials of choice for the MLBs. Thus, CTE-tailoring has to take the form of avoiding larger size components that are either ceramic (CGAs, MCMs), plastic with Alloy 42 leadframes (TSOPs, SOTs), or plastic with rigid bonded silicon die (PBGAs).

Increasing attachment compliancy for leadless solder attachments means increasing the solder joint height (C4, C5, shimming, gluing [Refs. A-7: 30, 31], 10Sn/90Pb balls, 10Sn/90Pb columns) or switching to a leaded attachment technology. For leaded attachments increasing lead compliancy can mean changing component suppliers to those having lead geometries promoting higher lead compliancy or switching to fine-pitch technology.

The DfR-process needs to emphasize a physics-of-failure perspective without neglecting the statistical distribution of failures. The process might involve the following steps:

A. Identify Reliability Requirements—expected design life and acceptable cumulative failure probability at the end of this design life;
B. Identify Loading Conditions—use environments (e.g., IPC-SM-785) and thermal gradients due to power dissipation, which may vary and produce large numbers of mini-cycles (Energy Star);
C Identify/Select Assembly Architecture—part and substrate selections, material properties (e.g., CTE), and attachment geometry;
D. Assess Reliability—determine reliability potential of the designed assembly and compare to the reliability requirements using the approach shown here, a 'Figure of Merit'-approach [Ref. A-7: 32], or some other suitable technique; this process may be iterative;
E. Balance Performance, Cost and Reliability Requirements.

9.4.4 Validation and Qualification Tests

The lessons learned over the past 15 years with surface mount technology (SMT) and fine-pitch attachments should be heeded and applied. However, some of the emerging advanced technologies fall outside the previous experience with SMT attachments. It is therefore important, that appropriate design validation and qualification tests be carried out to extend and, if necessary, alter and augment, the existing understanding.

The validation and qualification tests should follow the guidelines given in IPC-SM-785, Guidelines for Accelerated Reliability Testing of Surface Mount Solder Attachments.

However, for large components with significant heat dissipation and small global CTE-mismatches, temperature cycling tests are inadequate to provide the required information; full functional cycling—including external temperature and internal power cycling—is necessary.

9.4.5 Screening Proceedures

9.4.5.1 Solder Joint Defects

The solder joint defects of greatest reliability concern are those involving inadequate wetting for whatever reason. Well wetted solder joints, regardless of their geometric variations within the standards provided by IPC-A-620, Acceptability of Electronic Assemblies with Surface Mount Technologies and ANSI/J-STD-001, Requirements for Soldered Electrical and Electronic Assemblies, and somewhat beyond, will not pose a reliability threat due to inadequate quality.

Those solder joints have adequate strength even for severe mechanical loading conditions as well as no diminished thermal cyclic fatigue reliability. Only with severe offsets beyond Class 2 requirements is the reliability diminished. However, solder joints not properly wetted, can prematurely fail both as the result of mechanical and thermal cyclic loading [Refs. A-7: 1, 22]. Voids in the solder joints are generally regarded as not constituting a reliability threat [Ref. A-7: 28]. Possible exceptions are large voids reducing the solder joint cross-section enough to reduce a required thermal heat transfer function, and voids in high-frequency applications where the voids can cause signal deterioration.

9.4.5.2 Screening Recommendations

Effective screening procedures need to be capable of causing the failure of latent solder joint defects, i.e., weak inadequately wetted solder joints, without causing significant damage to high quality solder joints.

The best recommendation is random vibration (6-10 grams for 10-20 minutes), preferably at low temperature, e.g., -40 C. This loading does not damage good solder joints, but overstresses weakly bonded ones [Ref. A-7: 1].

Thermal shock can also be successfully used, however some damage to good solder joints can be expected, particularly for larger components.

10 STANDARDIZATION

Many standards are needed to accomplish flip chip and chip scale implementation strategies. Some of these standards relate to process details, others represent an overview of the technology. Information is also needed on the services that are available and these services need definition.

The requirements for underfill technology need to be available and well understood. Differences between the configurations that require and those that do not require, such as when silicon is mounted on silicon, need to be adequately described.

The differences between the peripheral and array package I/O relationship are important. Some of the chip scale technologies have addressed that issue by providing a ruggedized mounting structure that takes peripheral I/Cs and converts them into array type packages.

The ability to do failure analysis and methodologies for process and process control definition are also a major requirement in some of the standards. It has been stated that the industry needs a reliability database or a model that helps to clarify the issues on what standards exist and what standards are needed.

There are many questions to be answered for those people entering the technology. These questions address such items as:

- Inventory control (Lot identification)
- Flip chip and chip scale availability (bumped dice)
- The issue of known-good die (Who owns it?)
- Connection techniques (polymeric, solders, elastometers, etc.)
- Availability of CAD tools (for ASICs, for MCMs, for flex)
- Thermal extraction methodologies need definition
- Quality conformance and inspection techniques (die, bump, patterns)
- Process verification characteristics
- Process robustness descriptions
- Definition of critical assembly parameters
- Wafer dicing, wafer bonding, wafer burn-in
- Wafer availability vs. individual chip

In addition to the above list, users are interested in performance vs. cost relationships. Although these are not generally defined in a standard, the selection process should be clear and well understood so that those entering the technology have a clear understanding of the high cost for entry. The low volume user should consider techniques using chip scale packages, where handling the chip, as a package, may be more advantageous than the conventional flip chip high-volume approaches.

10.1 Standards for Development

Standards for development are categorized into four major sections. These sections deal with the following:

- Flip chip development and performance standards (10.2)
- Mounting substrate design and performance standards (10.3)
- Flip chip/substrate assembly design and performance standards (10.4)
- Material performance standards for flip chip manufacture and assembly (10.5)

The following sections describe those standards that need to be created. In each instance, a scope and purpose have been developed for consideration by the standard developers. At the time of the release of this publication, some

standard developers have already started work. Individuals are encouraged to contact the publishers of ANSI/J-STD-012, and volunteer to participate in the standardization effort by identifying those standard committees in which they wish to participate.

10.2 Flip Chip Development and Performance Standards

The following sections define standards that are proposed to be developed for flip chip and chip scale technologies. Each section deals with particular characteristics; a table including a proposed scope and purpose for each standard is proposed in each section.

10.2.1 Flip Chip IC/Component Design

This standard is intended to deal with chip finish requirements such as passivation and metalization. Pin die, marking conventions are described as well as the design and the rules for flip chip or chip scale technology. (See figure 10-1)

STD. No. 101	Semiconductor Design Standard for Flip Chip Applications
SCOPE:¬	This standard addresses semiconductor chip design. It is intended for applications utilizing standard substrates, assembly and test methods.
PURPOSE:¬	The purpose is to provide design standards that are commensurate with established manufacturing practices. Process details include wafer/chip processing, substrates, testing, and qualification methods. Also addressed are the electrical, thermal, and mechanical chip design parameters and methodologies as well as the reliability associated with these items.

Figure 10–1¬ Semiconductor Design

10.2.2 Mechanical Outline Standards

This standard is intended to deal with the grid, pitch, tolerances, and bump specifications related to the package outline for flip chip and chip scale technologies. Included are the bump pitch descriptions, total bump characteristics, as well as alignment and orientation. (See figure 10-2)

STD. No. 102	Mechanical Outline Standard for Flip Chip or Chip Scale Configurations
SCOPE:¬	This standard establishes mechanical requirements for devices supplied in flip chip or chip scale package formats, including die surface, die terminals and bump interconnections.
PURPOSE:¬	The purpose of this standard is to establish a family of mechanical outlines and footprints to minimize the number of variations. Pitch, bump size/locations, coplanarity, and associated tolerances will be included in this standard.

Figure 10–2¬ Mechanical Outlines

10.2.3 Performance Requirements for Bumps

This standard is intended to deal with performance characteristic related to bump materials and the techniques for reducing alpha radiation that may impact chip operating conditions. (See figure 10-3)

STD. No. 103	Performance Standard for Flip Chip/Chip Scale Bumps
SCOPE:¬	This standard establishes the performance requirements for flip chip and chip scale devices. All flip chip devices shall meet all the parameters detailed in this document.
PURPOSE:¬	The intent of this specification is to allow the manufacturer of flip chip or chip scale devices the flexibility to implement the best commercial practices, to the maximum extent possible, while providing a product that meets the electrical requirements provided in this specification.

Figure 10–3¬ Bump Performance

10.2.4 Physical Flip Chip Testing Requirements

This section is intended to cover bump integrity, vibration, shake, rattle and solderability testing. Tolerances, voids, and purity in flip chips are discussed as well as the chip finish including the passivation, metallization, etc.

Flip chip reliability requirements are defined as well as the issue of who own the known good die. (See figure 10-4)

STD. No. 104	Test Methods for Flip Chip or Chip Scale Performance
SCOPE:¬	This standard describes the tests and test methods required to ensure the quality and reliability of flip chip and chip scale products.
PURPOSE:¬	The purpose of this document is to establish standardized test methods and acceptance criteria, to facilitate test data correlation between vendors and users. Included are tests for tensile and shear strength, composition and bonding conditions.

Figure 10–4¬ Flip Chip and Chip Scale Test Methods

10.2.5 Trays for Flip Chip (Shipping and Delivery)

This standard is intended to cover die and substrate handling. It includes details on shipping trays for bumped dice or chip scale packages. (See figure 10-5)

STD. No. 105¬	Flip Chip/Chip Scale Carrier Tray Standard
SCOPE:¬	This standard defines the properties of handling trays used for shipping and pick & place equipment used for flip chip die or chip scale package placement. The intent is to protect the chips from physical, and ESD damage and from contaminates.
PURPOSE:¬	The purpose is to provide a standard for a system of trays that will allow pick and place equipment designers a uniform set of carriers from which to pick die for test, burn-in and assembly. Trays will need to accommodate chips having varying sizes and bump configurations yet have uniform outlines. This specification covers the environments that the trays must withstand, including storage, and type of environmental protection for the flip chip devices provided by the carrier.

Figure 10–5¬ Flip Chip/Chip Scale Carrier Trays

10.2.6 Configuration Management

This standard is intended to deal with inventory control, considering such items as lot identification and known good die. The die and substrate handling are discussed as well as wafer dicing, wafer bonding, and wafer burn-in. (See figure 10-6)

STD. No. 106	Bare Dice as Flip Chip or Chip Scale Configuration Management Standard
SCOPE:¬	This standard establishes the requirements for identification, tracking and inventory control for flip chip devices or chip scale packages (passive and active) intended to be mounted on a substrate or carrier such that visual examination of the attachment joint is not possible after any level of packaging. Included are instances where additional passivation, metalization, bumps, etc. hinder the identification of the device and/or its mask version, lot number, date code, etc.
PURPOSE:¬	The purpose of this specification is to permit determination of vendor, part number, revision, lot numbers, date codes, programmable contents, etc. of devices used in a ''flipped'' orientation. This document provides methods and systems for marking, inventory and device orientation.

Figure 10−6¬ Flip Chip/Chip Scale Configuration Management

10.3 Standard on Mounting of Substrate Design and Performance

Standards on mounting of substrate design and design performance for the mounting and interconnecting structure take several forms. The design standard include requirements of the mounting and interconnecting structure as well as the assembly. Qualification standards deal with organic and inorganic mounting structures.

10.3.1 Design Standard for Flip Chip or Chip Scale Package Mounting

This standard is intended to cover the details for organic and inorganic mounting structures intended for flip chip and chip scale mounting. (See figure 10-7)

STD. No. 107	Design Standard for Flip Chip and Chip Scale Mounting Structures
SCOPE:¬	This standard establishes requirements and other considerations for the design of organic and inorganic mounting and interconnecting structures to be used primarily for mounting passive and active devices in bumped dice or chip scale package formats.
PURPOSE:¬	The purpose of this standard is to establish principles and guidelines that shall be used to produce detailed designs to accommodate the mounting of passive and active devices in a face-down configuration. The standard covers layers, dielectric separation, via formation, etc., and metallized mounting site features that are unique to flip chip or chip scale application. Included are the variations and trade-offs for appropriate selection and tolerances of the MIS performance requirements.

Figure 10−7¬ Mounting Structures Design

10.3.2 Qualification and Performance of Organic Mounting Structures Intended for Flip Chip Mounting

This standard is intended to describe requirements for organic mounting and interconnection structures. (See figure 10-8)

STD. No. 108	Qualification and Performance Standard for Flip Chip Organic Mounting Structures
SCOPE:¬	This standard covers organic mounting structures intended primarily for the attachment of semiconductor die in face-down configurations. The structures may be single-sided, double-sided or multilayered with or without blind/buried vias.
PURPOSE:¬	The purpose of this specification is to provide requirements for qualification and performance of mounting structures made of organic materials with metallization whose primary purpose is to interconnect semiconductor dice mounted on the structure surfaces. Details of visual, dimensional, physical, electrical, environmental and construction integrity are described as well as the methodologies for their assessment.

Figure 10−8¬ Organic Mounting Structure Performance

10.3.3 Qualification and Performance of Inorganic Mounting Structures Intended for Flip Chip Mounting

This standard is intended to describe requirements for inorganic mounting and interconnection structures. (See figure 10-9)

10.3.4 Qualification, Quality Conformance, and In-process Test Methods used for Organic/Inorganic Flip Chip Mounting Structures

This standard is intended to describe the test methods to qualify the process and the product, total assembly related to the packaging issues. (See figure 10-10)

10.4 Flip Chip/Substrate Assembly Design and Performance Standards

This sections covers the details for those standards necessary to detail the flip chip and chip scale relationship to the