KNOWN GOOD DIE TECHNOLOGY

The term known good die (KGD) entered into the microelectronics lexicon in the early 1990s as a term to represent the quality and reliability of certain integrated circuit (IC) products prior to assembly. The term "known" refers to the *probability* that the IC product will function as designed in an application for a given time. The concept of KGD came into being to describe the quality and reliability requirements for ICs; applicable when more than one IC is to be assembled directly onto an interconnecting substrate. Thus KGD are supplied to the assembler without the intervening lead-frame, protective case, and large terminal contacts that traditionally formed the IC "package." KGD also implies that the IC supplier has employed quality and reliability testing and screening beyond the "normal" wafer and/or bare die testing generally done on ICs that are destined to be "packaged" in single-chip modules.

The IC package provides a number of benefits for both the IC fabricator and the system assembler. The IC package provides protection from the environment, especially humidity and contamination. Package sizes are relatively large scale, allowing ease of manufacturing and processing. The package provides a mechanism for thermal management, that is, a means to conduct heat generated by the IC away through the metal connections to the chip or to the backside (via thermal die attach). Electrical connection to the package leads is facilitated, and the package can be mounted temporarily in a socket for ease of testing or debugging. IC fabricators take advantage of these benefits for test and reliability screening, and avoid extensive bare die test and screens that may damage the device or produce false negative test results. The bare die test environment is also more difficult to control, making correlation of test results to the use environment more difficult for the IC manufacturer.

However, the benefits of the electronics package come at a cost. The electrical performance of a packaged IC is degraded by the additional capacitance, inductance, and lead length added to the IC bond pads. In addition to the performance limitations of the IC package itself, the larger footprint of the package implies that the next level of interconnect will also be suboptimal due to size and fanout of the inter-IC leads. The size, weight, and typical board space of IC packages is shown in Fig. 1. As the figure makes apparent, eliminating the first level of package, and directly attaching the chip to the interconnecting substrate holds a number of attractions, namely, higher input–output (I/O) densities, lower profile, shorter lead length, and lower weight.

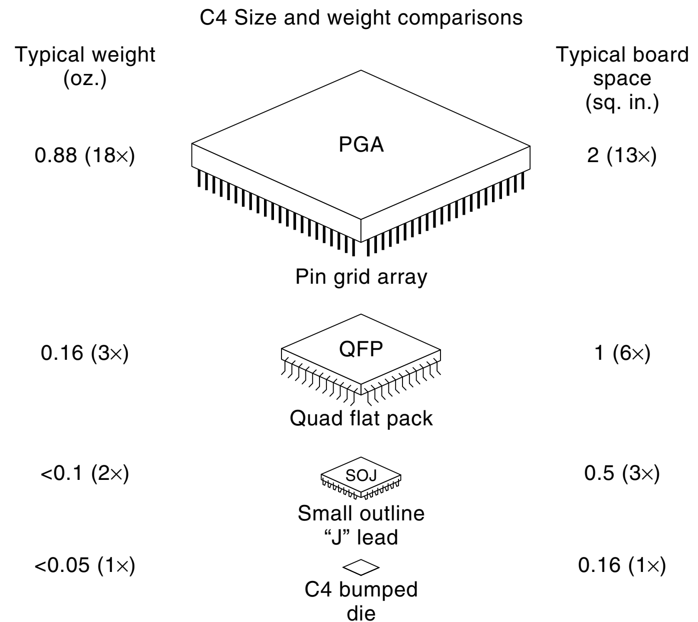

Figure 1. Comparison of various IC packaging technologies. Courtesy of Motorola/IBM.

The IC industry currently does an excellent job of routinely testing and burning-in traditionally packaged chips, although this can at times be a costly process. The resulting high-quality, high-reliability devices allow board-level products to be assembled with a high probability of first-pass acceptance. However, the probability of successfully assembling multiple bare die on a multichip substrate is decidedly poorer than that of the packaged ICs on a PC board because of the difficulties associated with conditioning the die appropriately. The next section will develop the issues for utilizing bare die in multichip systems.

BARE DIE TECHNOLOGIES

The process of wire-bonding bare chips directly onto electronic circuit boards is known as chip on board (COB) assembly (see Fig. 2). These COB technologies use die that have been prepared for wirebonding into conventional single-chip packaging, which is the bulk of IC production today. An advantage of COB assembly is that the die user may obtain die from many sources, and there is a mature industry making hybrid microcircuits using bare chips on ceramic substrates. The challenge for COB manufacturers is obtaining ICs that have been tested to higher levels of quality and reliability. These

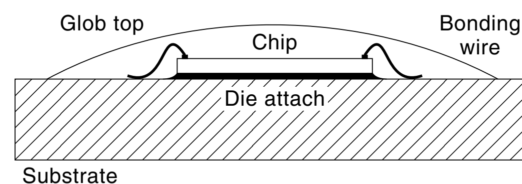

Figure 2. Cross-section of wirebonded die for chip-on-board application.

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering.* Copyright © 1999 John Wiley & Sons, Inc.

ICs command a premium price in the marketplace, and have been an obstacle to widespread adoption of COB technologies.

Tape automated bonding (TAB) is an IC chip on board technology that is often compared to wire bonding technology (1) (see Fig. 3).

TAB typically uses movie film-like tape material containing prepatterned leads with a rectangular cross-section instead of individually assembled round wires to distribute input/output (I/O) signals between the IC chip and its next-level interconnect. An opening, appropriately called a window, is formed near the center of the conductor array in the dielectric base film. The window permits the etched conductor leads to extend over the opening to create the essential beam-type interconnect array (2). The die is gang-bonded to the beams of copper traces on the tape. Each tape site contains one die and processing may occur on reels or slide carriers that incorporate several tape sites per slide. Advantages of TAB include extremely fine-pitch inner-lead bonding (ILB), the connection to the die, and the capability of the TAB tape to provide test access to the die. TAB requires a large investment in infrastructure and has not found wide use in mainstream products in the United States.

Another bare die technology is direct chip attach (DCA), using bumped die in a "flip-chip" configuration. Flip-chip ICs are postprocessed to deposit solder bumps over the bond pads on the die, and reflowed for mounting onto the substrate. The best known example of solder-bumped flip-chip technology was introduced by IBM in the 1960s and has been used extensively in IBM products. Known as C4 (controlled-collapse chip connection) technology, it utilizes a solder bump deposited on wettable terminals on the chip and a matching footprint of solder wettable terminals on the substrate (see Fig. 4). The solder bumped flip chip is aligned to the substrate, and all solder joints are made simultaneously by reflowing the solder (3). Generally viewed as the optimal IC packaging solution in terms of performance, lack of infrastructure, including test and burn-in of the bumped die, has kept the technology from being widely adopted.

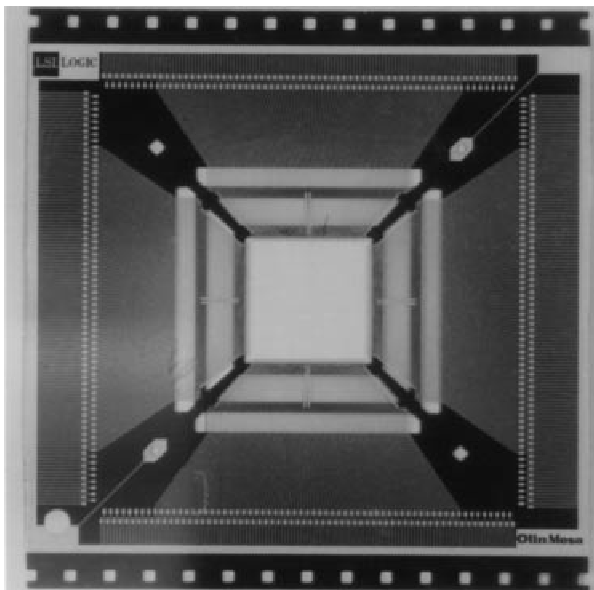

Figure 3. TAB (Tape Automated Bonding) frame. Window in the center of frame designed to accommodate the die, which is bonded to beam leads cantilevered over the edge of the polyimide tape.

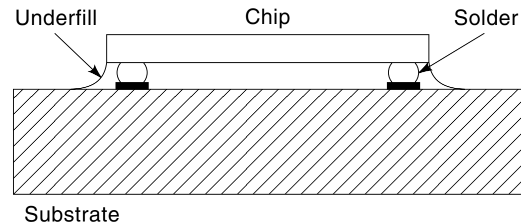

Figure 4. Cross section of direct chip attach application.

The need for fully conditioning a bare die to achieve quality and reliability levels required for multichip assembly is a critical enabler to the widespread acceptance of COB/DCA assembly for dramatic size and performance improvements.

IC QUALITY AND RELIABILITY

The probability that a multichip system will function correctly depends on, among other parameters, the incoming quality of the ICs that are being assembled according to the formula:

$$Y_b = 100(P_c)^n$$

Y_b is the predicted assembled module yield (assuming a perfect test is performed after assembly), P_c is the probability that the IC functions correctly, and n is the number of ICs. Figure 5 plots Y_b against n for several values of P_c.

Packaged ICs can approach 99.999% probability (considered to be known good) of performing correctly for a specified time in the final application (4). This probability index means that less than 10 parts out of one million will fail to perform their function correctly throughout a minimum guaranteed lifetime. The ability to fully test at-speed and over-temperature and to eliminate weak components with burn-in is not generally cost-effective for bare die. IC manufacturers prefer to do performance testing and burn-in after the die has been

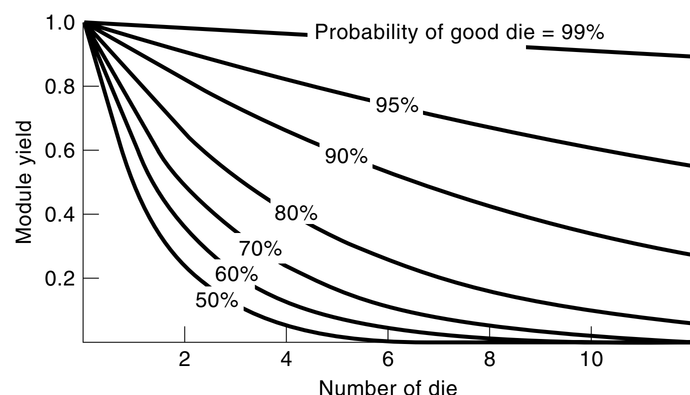

Figure 5. Plot of $Y_b = 100(P_c)^n$ for several probabilities of known good die. This plot makes the simplifying assumption that all die in a module have the same probability of being known good and that a test capable of detecting all defects is performed on the module after assembly.

assembled into a package—taking advantage of the package for ease of handling and test access. This significantly lowers the probability that an unpackaged IC will perform as specified over the expected service lifetime unless some extended test regime is available at the wafer or die level.

Figure 5 shows the effect of lower KGD probability for an assembly of bare die for an interconnecting substrate or multichip module (MCM). Even with a 90% probability of KGD—which is typical of wafer probe results for mature products—the resulting yield of the assembled module is unacceptable for systems with more than a few chips. In addition, the MCM final test coverage required to identify the low yielding MCMs must be extremely high to avoid escapes.

Figure 6 shows the MCM defect level or percentage of defective modules in a lot after final test as a function of the fault coverage of the test (4). These are modules that escaped detection because of the less than perfect final test. Notice that MCM defect level is a function of both final test fault coverage and the quality level and number of incoming die.

Figure 7 shows a simplified comparison of several incoming die quality levels for an MCM containing five ICs (each with identical incoming quality level) with MCM final test fault coverage. The incoming die quality levels that are less than 99% cause unacceptably high MCM failure rates, with attendant rework costs. But even worse, the high defect level indicates that faulty MCMs are not detected at final test, and thus are likely to cause system level failures later in the life cycle, perhaps in the field.

So, for most applications, KGD can be defined as a greater than 99% probability that the die in the wafer lot are True Good Die. This level of KGD is considered the absolute minimum probability that can be tolerated in an MCM assembly process. There are a variety of methods used to achieve this probability level. The KGD problem, therefore, can be summarized as follows:

- Unless incoming die quality level is at least 99%, the yield of all but the simplest MCMs will be unacceptable.

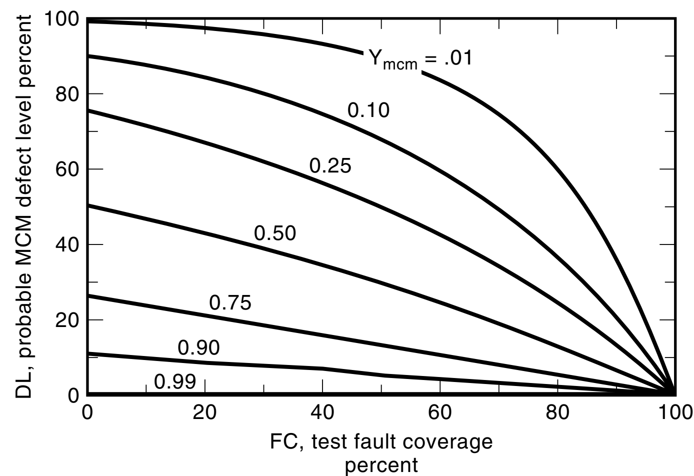

Figure 6. Defect level of outgoing MCMs as a function of MCM yield and final test fault coverage. Defect level (in percent) can be calculated as follows: $DL = (1 - Y^{(1-FC)}) \cdot 100$, where DL = defect level, Y = yield of the device under test, and FC is the fault coverage of the test program.

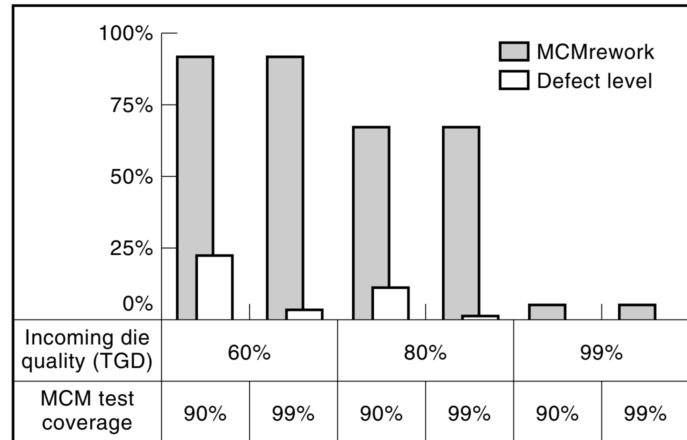

Figure 7. Example of how incoming die quality and MCM final test coverage affect rework and defect levels for a 5-chip MCM. This example assumes that all 5 die have the same incoming quality level. The defect level of the MCMs represent defective product that will be shipped to the customer for each scenario.

- Low incoming die quality will require an exceedingly high fault coverage at final test to detect defective modules. High fault coverage is very costly or may be impossible to achieve at MCM final test.

The following sections will highlight the issues of test and burn-in as they relate to KGD.

IC Test

The complexity of VLSI circuits prohibits exhaustive testing. For example, it takes 2^{64} clock cycles (minimum) to exhaustively test a 64 bit adder. If the clock rate is 50 MHz, it would take over 11,500 years! Obviously, test engineers have developed techniques which allow effective testing, with defect levels approaching zero, which take much less time, but no one claims to be able to exhaustively test VLSI circuits. So, as a matter of fact, all ICs, whether packaged or bare, have a finite probability of containing defects. This probability depends on the effectiveness of the testing done to the IC during manufacturing and assembly.

Fault coverage, a measure of test effectiveness, is usually quoted for a particular logic fault model, the single stuck-at fault. This simple fault model (5) assumes that any line in a circuit may have a fault which causes it to remain permanently either at logic 1 or at logic 0. Fault coverage then is a measure of how many lines in the circuit are checked for stuck-at faults with a particular test program. Single stuck-at fault coverage can sometimes be misleading in today's VLSI CMOS circuits. For example, the single stuck-at model does not directly take into account defects which only occur at the rated speed of the device nor bridging faults where two lines are inadvertently connected together by a resistive path, with neither line meeting the requirement of "stuck." So there is a finite probability that a device which passes a test program which is graded at 100% fault coverage for single stuck-at faults will fail to perform its function in a particular application, at speed, over temperature, or at power supply extrema.

Failure Modes. The mechanisms of semiconductor failure can be classified into three main areas (6):

1. Electrical stress (in-circuit) failures
2. Intrinsic failure mechanisms
3. Extrinsic failure mechanisms

The first category of failure mechanism is said to be event related and is directly associated with the equipment design in which the IC is assembled, or to handling damage due to electrostatic discharge (ESD). This failure mechanism is very important to MCM assemblers, as they have more control over this than the other mechanisms, but the other two are more germane to the discussion of KGD, as they are under the control of the IC manufacturer.

The second category of failure mechanism is inherent to the die itself, and is the result of defects introduced during fabrication of the wafers. These may be crystal defects, contamination, flaws in the deposition layers, or mishandling during processing. VLSI products with high wiring densities, such as logic and SRAMs, are driven primarily by failures in the metal interconnect and the insulator. These failures are predominately driven by dielectric breakdown and electromigration. DRAM products exhibit failure mechanisms associated with the movement of mobile charge due to an electric field. Sodium is the most common ion seen (7). These are the key mechanisms which determine the quality and reliability levels of bare die as shipped to customers.

The third category, the extrinsic failure mechanisms for IC failure, refers to the defects introduced during the "back-end" IC processing. Traditional packaging (bonding), test, burn-in, mark, pack and ship, all contribute failure mechanisms, usually to more traditional packaged parts. Failure mechanisms unique to KGD may be introduced at this point, as the KGD back-end processes usually are somewhat different from traditional packaged devices.

Operational Tests. Operational tests are those used to detect defects in the IC. There are a number of tests which may be applied to an IC in a manufacturing environment. For the purposes of this article, these tests are characterized as follows:

Functional Test. A functional test is performed to verify that a circuit performs its intended function. This is sometimes termed "truth-table verification" for digital circuits. The functional test performed at wafer probe test is usually done at reduced clock speed, due to the moderate performance nature of traditional needle-type epoxy ring probe cards, and is used primarily to determine whether circuit behavior is correct. The final test program usually incorporates a full at-speed functional test and relies on the IC package inserted into a socket for test access.

Dc Parametric Test. Dc parametric tests are those in which steady-state voltages and currents are applied to certain inputs of the device and corresponding voltages and currents are measured at the outputs. These tests are used to detect opens, shorts, input/output levels, noise margins, static and dynamic supply current levels, leakages, and so on. These tests may be performed at both wafer probe test and final test, with satisfactory results.

Ac Parametric Test. Ac parametric tests are those used to measure the frequency-dependent characteristics of the circuit. These characteristics are propagation delay, setup/hold times, duty cycles, clock period, signal timing, and so on. These tests are usually not done at wafer probe due to the poor electrical environment offered by a standard probe card. High-performance probe cards are emerging in the industry, which offer the capability of high-speed tests, so ac parametric testing (as well as at-speed functional testing) is becoming an option for wafer probe test, especially for bare die which do not need burn-in type stress tests. These bare die are then fully tested and may be adequate to meet the quality and reliability specifications for certain applications at lower cost than fully conditioned KGD. However, correlation of high-speed wafer probe test results to the actual performance in the application environment can be a significant problem for this type of testing.

I_{DD}Q Testing. This is a technique for detecting certain faults in CMOS circuits by monitoring the quiescent current of the device between clock edges. Once all transistors in a CMOS circuit have switched, no appreciable current should be flowing in the internal logic circuits. Any value of current detected above a threshold indicates a potential problem. There is some evidence that I_{DD}Q testing identifies certain reliability problems that burn-in would normally eliminate.

Built-In Self Test (BIST). Of course, a number of VLSI devices now incorporate built-in test, where test structures are built into each die to enhance its testability. These structures can be accessed from the I/O pins of the IC and provide a variety of levels, coverages, and effectivity. BIST can be particularly effective at wafer or bare-die burn-in. Boundary scan (IEEE Std. 1149.1), a form of BIST, is being designed into more and more advanced VLSI devices to support assembly level testing of advanced substrates. Boundary scan can also be effective as a "design for burn-in" feature, utilizing the boundary scan chain to provide the functional test patterns to the IC, and again to "read" the results out of the device (7).

Visual Inspection. Although not an electrical test, visual inspection is being used to increase the probability of KGD. Visual inspection criteria for bare die is an issue that is complicated by the fact that cosmetic effects cloud the quality and reliability issues, especially for current VLSI memory products which incorporate redundancy and laser repair. Differentiating between strictly cosmetic effects and defective die is not straightforward.

Screening. Screening ICs is the process of applying stress to the device and using operational tests to detect failures. Stresses applied may be electrical, thermal, or mechanical; the operational tests are then used to determine that the voltage and current levels and functional performance of the device are within specification.

Individual defective die can be detected by 100% screening of a batch. Process- or design-related problems can be detected by sample screening. For optimum cost benefit, screening tests must focus on the requirements of the customers and the possible failure mechanisms associated with the device and technology. An analysis of the failure mechanisms enables the development of good screening processes, and

tests are then based on the activation of the relevant defect mechanism (6). IC manufacturers generally work to eliminate defects from production processes. Screens are useful to reduce the number of defects that a customer may encounter while the process improvements are under way. Because of test access issues, screening is always more problematic at the wafer or bare die levels.

Burn-in stress screening is probably the most common screen for detection of infant mortality type defects due to manufacturing anomalies. This is usually a 100% screen used in production of leading-edge ICs to eliminate those devices containing random latent defects that will otherwise fail in the final application. Careful attention to the design of stresses for screening tests is necessary to ensure that defective devices are stressed to failure but that the useful life of the remaining devices is not adversely affected. Defects may be introduced through an improper screen.

Tests To Detect Faults. Table 1 lists some failure modes which are typical of VLSI devices today (6). The table also lists the appropriate operational tests to detect the faults that typify the failure mode and suggests a screen which will accelerate the defect.

Burn-In Screening

Burn-in is a reliability screening tool used in the microelectronics industry to reduce the risk of early field failures of IC devices (8) (see Fig. 8). It is an effective means for screening out defects contributing to infant mortality. This screening typically combines electrical stresses with elevated temperature over a period of time in order to activate the temperature- and voltage-dependent failure mechanisms for a particular device or process in a relatively short time (9).

Burn-in is a production technique for improving the reliability of IC devices as delivered to the customer. The entire population of devices is aged by the application of accelerated stress to make the weak devices fail (10). Burn-in typically is done with maximum power supply voltages applied. Random voltage patterns are applied to the inputs of the device to cause internal nodes to toggle. This assures that all nodes are stressed during the burn-in, which may last for a few hours up to a few days, depending on the level of maturity of the device and the level of reliability desired. Some burn-in systems are capable of applying deterministic vectors to the inputs and monitoring the outputs of the devices. It is then possible to provide a functional test during the burn-in cycle. This is becoming popular for detecting pattern sensitive failures in large memory arrays.

There is not a consensus today among IC manufacturers as to whether burn-in is a requirement for VLSI devices. There is a claim that current early failure rates are low, less than 5000 ppm (11). Certainly, for mature products, burn-in is not typically done for commercial or industrial use, although burn-in may be used for mature products to identify

Table 1. IC Defects with Relevant Detection Methods and Screens

Life Cycle Period	Failure Mode	Fault	Detection	Lifetime Region	Screen
Design	Parasitic elements	Latch-up	Functional test	Event related	Possible destructive test of sample
Fabrication	Ionic contamination	Degradation	ac, dc parametric test	Early life/wearout	Burn-in
	Gate oxide breakdown	Shorts/opens	Functional test	Early life	Burn-in with over voltage
	Corrosion	Opens/shorts	Visual inspection, functional test	Wearout	Temperature-humidity-bias tests
	Surface charge spreading	Degradation	ac, dc parametric test	Wearout/early life	Burn-in
	Piping	Parametric shifts	ac, dc parametric test	Wearout	Burn-in
	Dislocations	Threshold shift	ac, dc parametric test	Early life	Process control
	Slow trapping	Threshold shift	ac, dc parametric test	Wearout	High temperature anneal
Assembly	Microcracks	Opens	Functional test	Early life	Burn-in
	Electrostatic discharge	Shorts/opens	Functional test	Event related	None
		Degradation	ac, dc parametric test	Event related	
	Storage	Degradation	Functional test	Wearout	None
Use	Hot electrons	Degradation	ac, dc parametric test	Wearout	Low-temp life test
	Corrosion	Opens/shorts	Visual inspection, functional test	Wearout	Temperature-humidity-bias tests
	Electromigration	Opens/shorts	Functional test	Early life/event related/wearout	Burn-in
	Contact migration	Shorted junctions/ open contacts	Functional test	Wearout	Burn-in
	Mechanical stress relief migration	Shorts/opens	Functional test	Wearout	None
	Radiation	"Soft" errors, threshold shift, activation of parasitic elements	Functional test, ac/dc parametric tests	Event related	Testing with accelerator
	Electrical overstress	Shorts/opens	Functional test	Event related	None

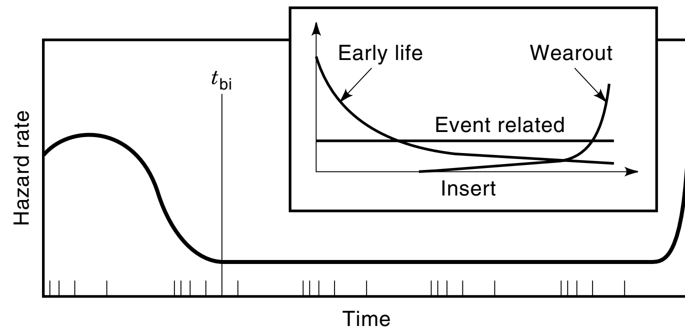

Figure 8. Traditional bathtub curve for IC lifetimes. The insert shows the three distinct models of device lifetime. Burn-in is designed to screen the early life failures from the population. Time t_0 is the time the device is manufactured. The failure rate shows an initial increase, then a steep decline during the early life, then it is constant during the long normal operating period, again exhibiting a sharp increase at the end of life. This bimodal distribution is caused by a population of weak devices which are a result of anomalies in the fabrication and manufacturing processes. This weak population can be accelerated to failure through the application of appropriate stresses for a relative short period as part of the manufacturing process. The resulting population of normal devices appears to have the failures in time beginning at t_{bi}.

"rogue" lots, that is, wafer lots which have a much higher than average failure rate due to processing anomalies (see Fig. 9) (12).

Systems for high-reliability use, especially nonrepairable systems, still require substantial IC burn-in to achieve acceptable field failure rates. Studies published by IBM indicate that burn-in effectiveness, defined as the ratio of hazard rate without burn-in to hazard rate with burn-in equivalent to t_{BI} hours. The data indicate that burn-in reduced the reliability failures of IC devices by close to 2 to 3 times through 3000 fielded h (13).

Jensen and Petersen (14) differentiate between devices that are congenitally weak, that is, the weakness was manufactured in (freak population), and devices which have unwittingly been subject to some stress subsequent to manufacture but prior to being put into service (infant mortality). This distinction has not been important to traditional packaged ICs, as devices receive burn-in after most postfabrication operations have been completed, and burn-in stress can be applied to accelerate both the freak population and the infant mortalities to fail. However, as burn-in strategies aimed at die level burn-in, either singulated bare die or die still in wafer form become available, more post-burn-in processing and handling will occur. The question of how this will affect infant mortalities experienced in the field is an open question, and will need to be analyzed as part of any KGD strategy.

Another consideration for burn-in of KGD is the thermal environment that the device sees. At preset, burn-in of bare die requires the use of a temporary carrier to provide the function of the IC package, namely, environmental protection, thermal management, and electrical access to chip power and I/O pads. However, it is not a given that a temporary carrier will provide the same characteristics as a permanent package. The thermal resistance of the path from the die junctions to the ambient ($R_{\theta JA}$) may be quite different, necessitating a different burn-in regime (time, temperature, and atmosphere), which could require a separate back-end process for KGD production.

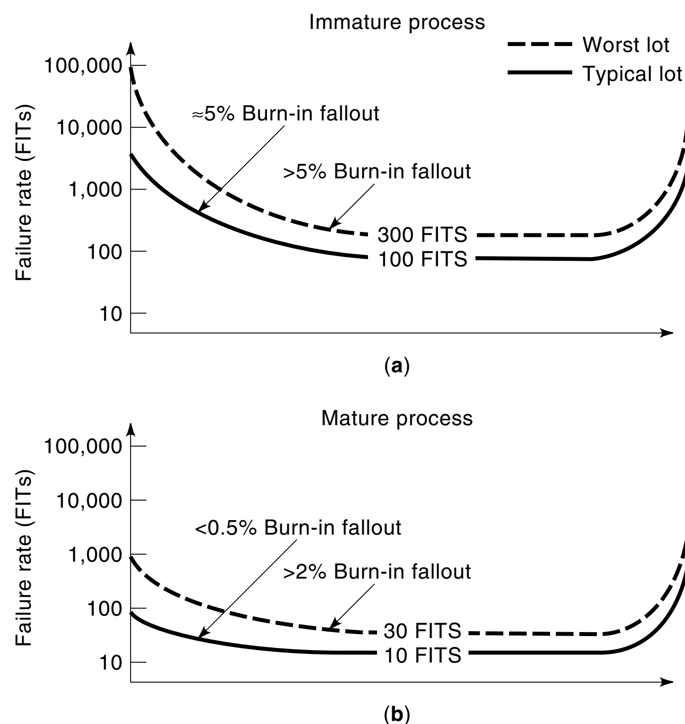

Figure 9. Curves for typical and worst-case (rogue) wafer lots for (a) immature IC processes; and (b) mature IC fabrication processes.

METHODS OF ASSURING KGD

Process Control

The move to 6-sigma processing has improved semiconductor yields and reliability over the past several decades.

The data in Fig. 10 are indicative of sound manufacturing and process-control improvements over the past few years. Vertically integrated companies that control the IC fabrication process as well as assembling modules can use improvement processes to raise known-good die probabilities. By correlating failures at the MCM or system level with testing done at the wafer level, defects can be migrated back to the front end and detected at wafer test (15). Early identification

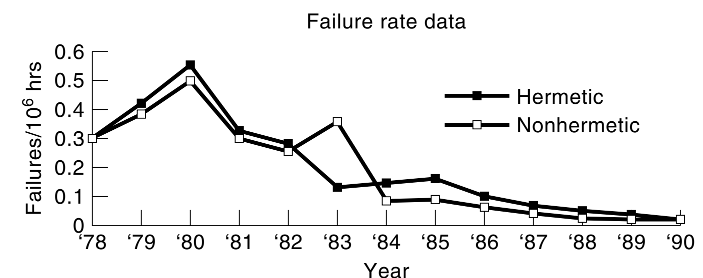

Figure 10. Historic IC failure rate data. The total devices tested is close to 400 million over 12 years. These data were collected from an equipment maker by the Reliability Analysis Center, Rome AFB.

Table 2. Yield Learning Leverages Typical of IC Product Life Cycles

	Y_{cum}	=	Y_{wp}	*	Y_{ft}	*	Y_{eol}
Entry yields	>10%		>70%		>20%		>90%
Mature yields	>80%		>95%		>90%		>99%

Y_{cum} is the cumulative yield of the device. Y_{wp} is the yield of the wafer probe operation, Y_{ft} is the final test yield and Y_{eol} is the packaging, burn-in, marking, etc. yield.

of defective devices is the most cost-effective approach to improving the probability of known good ICs [see Table 2 (7)].

Statistical Sampling

IC suppliers may provide statistical probabilities of known good die by packaging a sample of die in a wafer lot, performing exhaustive test and burn-in on the packaged sample, and applying the statistics of the sample to certify the entire lot. This technique is effective in detecting whether the batch (wafer lot) has been fabricated such that the statistics vary significantly from the "normal" population. To achieve KGD to an acceptable level (i.e., >99%) requires that test and burn-in statistics be obtained on a significant sample of devices. Figure 11 shows the probability of accepting a lot with a LTPD (lot tolerance percent defective) of 0.10, or a probability of .90 that any lot accepted will have at least 99% TGD. The various sample size/lot accept criteria are plotted. For an accept on 0, reject on 1 defect criteria, a sample size of 230 devices is needed to obtain the .90 probability of accepting a lot that has at least 99% TGD.

The above assumes that defects are evenly distributed throughout a large population and randomly sampled in each lot. It also assumes that a test is available to detect all possible defects. This is not generally the case. A test program that detects all possible defects on a VLSI device is not possible within the constraints of time, cost, and complexity. As discussed above, this is especially true of wafer test, where the probe card provides a very difficult electrical environment that is hostile to performance type testing (at-speed, over-temperature, full functionality, etc.).

In summary, statistical methods are adequate to predict lot to lot variations only. A fairly large sample of devices must be packaged and tested to demonstrate that the probability of each lot is acceptable.

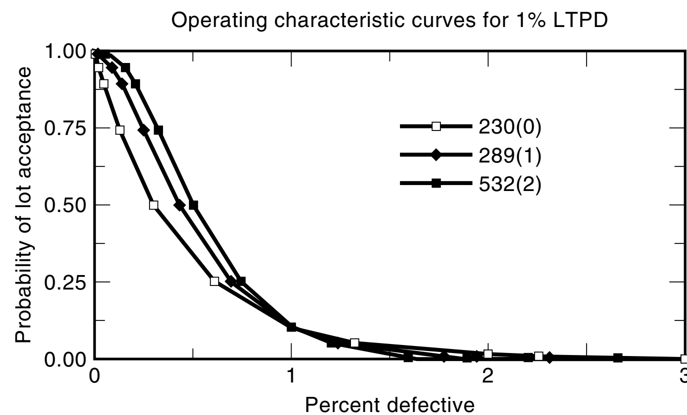

Figure 11. This plot shows the probability that a lot of ICs have a certain percentage of defective units based on packaging and fully testing a sample of devices from the lot. The legend shows the number of parts sampled and the pass criteria, that is, 230 parts sampled with accept on 0 rejects, 389 parts samples with accept on 1 or less rejects, 532 parts sampled with accept on 2 or less rejects. This particular set of curves is based on the criteria of less than 10% probability of accepting a lot of more than 1.0% defective devices.

Test/Screen Every Die

The most effective method of assuring KGD is to test every device to the quality and reliability levels demanded by the particular application by using test and screening methods as appropriate for the process and particular device.

KGD ASSURANCE TECHNOLOGY

The term Known Good Die implies that bare or unpackaged ICs will have the same quality or probability of failure as those same die assembled in "conventional" packages. Packaged die are exposed to a thorough suite of tests designed to ensure that specified functional performance will be maintained over a variety of supply voltage and temperature conditions; also, these packaged devices may be subjected to a burn-in screen to eliminate those devices subject to infant mortality. For simplicity, the suite of tests to which the packaged die has been subjected can be described using three categories:

1. Wafer level prepackage
2. Postpackaging at-speed functional testing
3. Burn-in, when required, to eliminate infant mortality

A basic assumption is that to achieve the quality and reliability levels required, known good die must undergo the same suite of tests as the equivalent packaged parts. There are several realities to be faced when considering full test and burn-in on an unpackaged IC.

Wafer Probe Test

Wafer probing, as it is generally practiced, is not an adequate test to deliver the quality levels demanded for high-performance, multichip packaging applications. Wafer test is designed to find only the obviously defective die, and is usually a static or dc functional test, performed at room temperature. It is organized so that the most likely faults are detected first. It generally detects somewhere between 75% and 95% of the defective devices, depending on the maturity of the IC fabrication process. Table 3 shows the typical final (packaged test)

Table 3. Typical Final Test Yield of Certain IC Devices

IC Product	Final Test Yield
8-bit MPU	95%
20,000 gate array	90%
4M DRAM	95%
16M DRAM	90%
64M DRAM	75%
4K GaAs SRAM	80%
32-bit MPU (386)	90%
32-bit MPU (P54C)	75%

Source. Integrated Circuit Engineering Corp.

yield for certain technologies. Table 3 also illustrates yield improvement with the maturity of the devices being tested. But, more importantly, for KGD considerations, it provides a measure of defective die that escape being detected at wafer probe under normal probing conditions. The defect level of bare die needs to be much better than that which is achieved with today's wafer probe testing.

Wafer Test Technologies

There are three technologies that provide the environment used today for test at the wafer level.

1. Needle (epoxy ring) probe cards
2. Buckling beam probe cards
3. Membrane probe cards

Needle (Epoxy Ring) Probe Cards. Although the vast majority of probe cards in use today are epoxy ring needle probes, the performance of needle probe card technology is inherently limited. The needles have an unreferenced electrical length of 25 mm or more. This unreferenced length results in several nanohenries of inductance on the V_{dd} and V_{ss} lines, as well as significant reflections and switching noise generators on the clock lines. As a result, needle probe cards can provide useful signal integrity bandwidth only up to 100 MHz or so. This is no longer sufficient for completely testing leading edge ICs that are intended for today's advanced packaging applications. Figure 12 shows the needle probe card cantilever beam action of an integrated circuit I/O pad.

Vertical (Cobra, Buckle Beam, SiCard) Probe Cards. Vertical probes were introduced by IBM in the form of buckling beam probes. They alleviate several of the problems associated with needle probes. They can be easily designed to probe area array pads on die, they can be designed with shorter uncontrolled impedance than needle probes, thus improving high-frequency performance. A variety of designs are appearing as solutions to area array, bumped die. The shape of the force-deflection curve of some vertical probes is advantageous in overcoming nonplanarity in the probes or pads/bumps of the

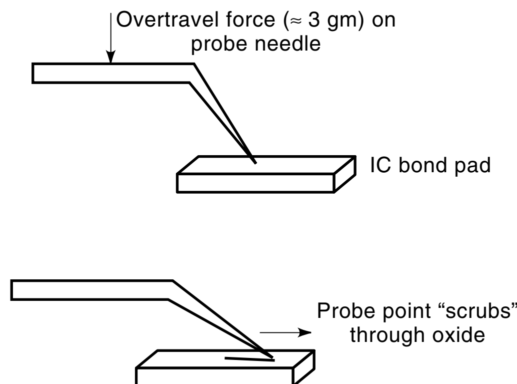

Figure 12. Needle probe forms a cantilever beam that moves horizontally across aluminum bond pad as the probe is forced downward after the tip contacts the surface of the pad.

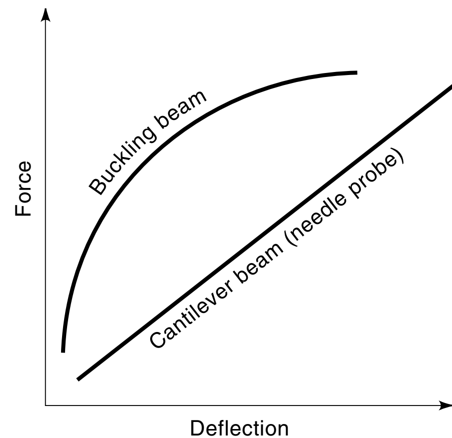

Figure 13. Force-deflection curves for buckling beam prove compared to the cantilever effect of a needle probe.

device under test. Figure 13 shows the force-deflection curve of a buckling beam compared to a cantilever beam.

Vertical probes can be very expensive (compared to needle probe cards), the lead times can be long (numbers of weeks), and they are difficult to rework. These issues, coupled with the fact that full test is not generally being done at wafer probe, are hindering the acceptance of vertical probe cards in the industry today.

Membrane Probe Cards. As with vertical probe cards, membrane probe technologies were introduced into the market several years ago, but have failed to make inroads into the market share of needle probe cards, despite several advantages. They have the potential to achieve higher speed than either needles or vertical probes, they are area array/multiple die capable, and are extremely rugged resulting in lower cost of ownership. However, they suffer from some of the same drawbacks as the vertical probe cards, high initial cost (NREs may be high, unit costs low), and long lead times to produce.

Again, since the majority of IC products are shipped as single-chip modules, the requirement for "heroic" efforts at wafer or die level probe test is not seen as cost effective by the IC manufacturers.

Test and Burn-In Carriers

The solutions for delivering the highest probabilities of producing KGD today is to assemble the bare die into temporary

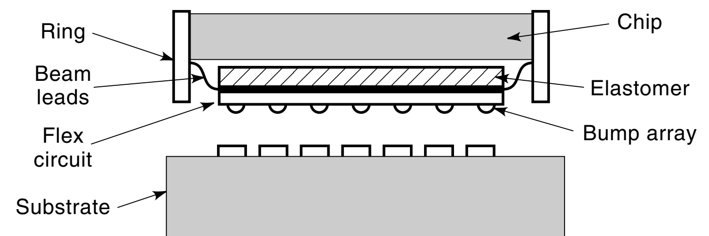

Figure 14. Example of a chip scale package which consists of flex circuit with a rerouted bump array, bonded to the I/O pads on the chip through beam leads. The rerouted bump array can be held in contact to pads on a substrate for test and burn-in by applying force on the backside of the package. Based on Tessera μBGA technology.

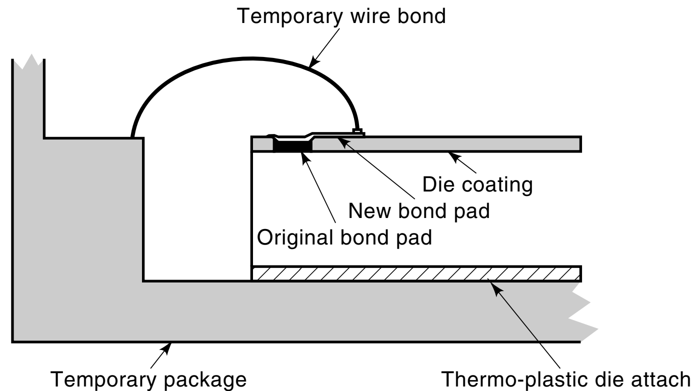

Figure 15. Semipermanent KGD carrier. The wafer is coated with dielectric material and the bond pads enlarged. Wires are temporarily bonded to the enlarged pad and the die attached to the temporary carrier. After test and burn-in, the wire bonds are “clipped” off and the die is removed from the temporary package.

“carriers.” These carriers serve the purpose of a single-chip package to allow the complete final test and burn-in infrastructure currently in place for packaged ICs to be used. Temporary electrical connection is made to the bond pads and the device is qualified through test and burn-in processes identical to the packaged part; so quality and reliability levels are achieved comparable to packaged devices. Automatic Test Equipment (ATE), component handlers, burn-in boards, burn-in ovens, and loaders can be used. Once the die is qualified, electrical connections to the bond pads are released and the die is taken from the carrier. The result is a fully tested, qualified IC device with specifications comparable to those of an equivalent packaged part. These technologies can be divided into categories based on the type of die-to-carrier connection: permanent, semipermanent or temporary (16).

Permanent Carrier Approaches. These approaches take advantage of the fact that some minimal packaging approaches enhance the ability to do test and burn-in. Ruggedized packages, wider pitch, and gold contacts can help solve many of the problems with handling and contacting bare die. Tape automated bonding (TAB) is an established technique for testing integrated circuits. The IC is bonded to a lead frame which fans out the electrical connections to peripheral pads on tape, which may be readily contacted. TAB carriers are available which allow the device to be socketed for test and burn-in.

Another minimal packaging approach, the chip scale package, is, like TAB, designed to be permanently assembled with the die into the next level interconnect. The chip scale package is designed to take up little more real estate on the board than the chip itself. In the chip scale package approach, the die bond pads are routed to a standard footprint, usually an area array, of bumps, which provide mechanical and electrical connections to a substrate. Thermal management can be provided through the backside of the die. A major advantage to the chip scale package, from a KGD point of view, is that the chip to substrate connection can remain fixed through die shrinks, upgrades, and so forth.

Semipermanent Carriers. Semipermanent carrier approaches consist of making a nonstandard (or rerouted) metallurgical connection (wirebond, C4, etc.) to a reusable standard package type; sending the assembly through test and burn-in; then breaking the connection and removing, inspecting, and shipping the die. Several semipermanent carrier technologies are available for preparing KGD, including temporary TAB connection, temporary wirebond, and reduced radius removal (R3) based on the IBM C4 flip-chip technology.

These technologies rely heavily on established processes and tools to condition die. In some cases, semipermanent carrier technology is targeted exclusively toward specific final assembly methods, such as TAB or R3 for solder-bumped die. Concerns with the cost-effectiveness and final assembly limitations of some semipermanent carrier methods have resulted in heightened interest in temporary carrier solutions that are applicable to all die, regardless of final assembly methods.

Temporary Contact KGD Carrier. Temporary contact carriers have a microprobe set for contacting pads on the IC under test. The microprobes are built onto an interconnect that routes the signals to a fanned out connector set to provide the electrical connection to a socket, in much the same manner as a package I/O pins contact a test socket. The temporary contact carriers have few limitations on characteristics of incoming die. The carrier (including interconnect and microprobe set) may be used many times, increasing cost effectiveness.

The die is held in alignment to the probe set with force to insure reliable electrical contact to the IC bond pads. The main technical challenges with the probes/contacts are compliance to nonplanar pads on the die, penetration of the native aluminum oxide present on the IC pads, and maintaining adequate contact without causing damage, which could preclude the next assembly operation.

Pad Penetration. A temporary carrier approach requires some form of “scrub” or penetration through the native oxide on the Al pad of the IC, typically on the order of 50 Å to 80 Å. Many KGD approaches currently being developed use some form of z-axis penetration, in which the microprobe is designed to be used with a piercing vertical on-axis force (see Fig. 16).

Compliance. Scrubbing or penetrating actions on the bond pad require that some form of compliant member be available

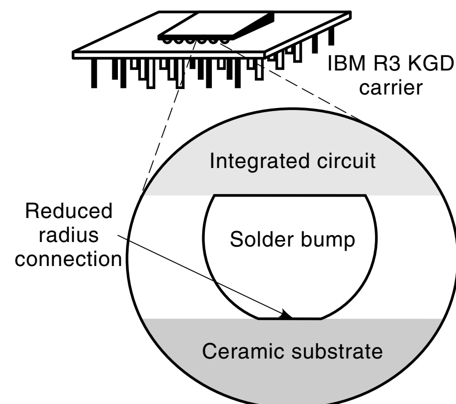

Figure 16. Semipermanent KGD carrier. The IBM R3 carrier uses a C4 process to join a bumped die to reduced radius pads on a ceramic substrate. After test and burn-in, the chip is sheared off the substrate with a lateral force. The bumps remain intact on the die.

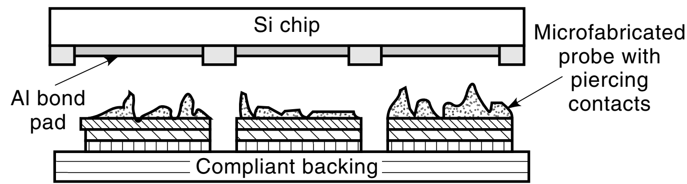

Figure 17. Microfabricated probe set makes reliable contact to the aluminum IC bond pad by piercing through the oxide layer on the surface of the pad.

to equalize the force on each contact. For traditional epoxy ring needle probe cards, the spring constant of the needle provides compliance (see Fig. 12). For temporary carriers, especially those that rely on z-axis penetration to make reliable, low-resistance contact, this problem is nontrivial. Since force is applied on-axis through the probe structure, the compliance must be built into the substrate on which the probe set is mounted—unless some form of compliant or deformable probe contact capable of z-axis penetration is available. This compliant substrate must be capable of providing two conflicting functions—the ability to transmit probe forces independently in the z-direction while maintaining x- and y-direction positional accuracy. This is especially difficult given that the pitch of the contact is sometimes less than 100 μm. In addition, the compliant members must be capable of maintaining these properties during burn-in temperature excursions typically to 150°C. Many temporary carrier approaches use a form of membrane as the probe set substrate. Thin films, laminates, organic, and inorganic membranes have all been proposed as KGD carrier interconnect substrates. These membranes may be backed by an elastomer to provide the compliance required for the probe set.

Pad Damage. Another major problem to be addressed by both semipermanent and temporary carrier methods is bond pad damage after the removal of the die from the carrier. A primary concern is determining how many retests a given approach is capable of before next level assembly is precluded. The generally accepted industry standard—Mil-Std-883, Method 2010, Section 3.1.1.1—states the criteria for rejection as 25% of the passivation underlying the bond pad being exposed. This serves as a reference to determine acceptable pad damage. A probe should be capable of ≥ 2 touches on the pad before unacceptable damage occurs. At no time should the probe touch down on the top or sides of the top passivation on the die.

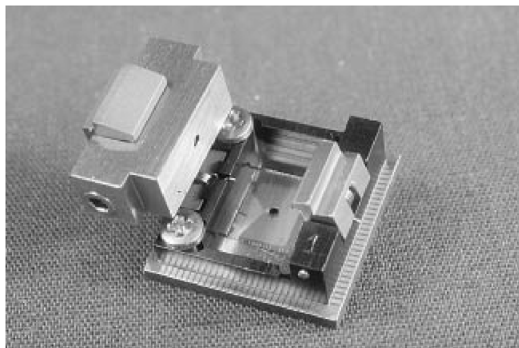

Figure 18. DiePak™ carrier showing hinged lid open to accept die.

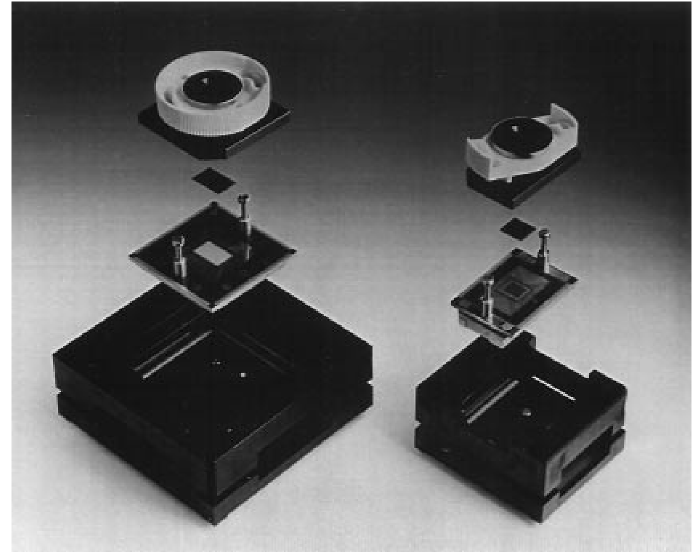

Figure 19. DieMate™ carrier showing assembly open to accept die. Photo courtesy of Texas Instruments.

Alignment. Bare die handling and alignment is the next critical issue. The die must be accurately aligned with the probe. If mechanical alignment is used, standard practice for wafer saw tolerances must be improved. If not, vision systems will be required. If mechanical die alignment is required, that is, using the sides of a die for alignment to a probe set, the following issues must be taken into account:

- *Saw Cut Dimensions and Placement Accuracy.* Saw cut width must be maintained within tolerance ($\approx$ 3 μm) over the life of the blade. The distance of the cut to a designated die feature must also be maintained within tolerance.
- *Saw Alignment Accuracy.* The placement of the saw cut centerline should be within the maximum tolerance of intended location.
- *Sawn Edge Camber.* Total deviation (or bevel) of sawn edge from a straight vertical cut should not be more that a small percentage ($\approx$5%) of the wafer thickness.

As IC pad pitches and sizes become smaller, improved saw drift control and placement accuracy may be required for die that must be mechanically aligned onto a probe set in a temporary KGD carrier. See Figs. 17–19.

Wafer-Level Burn-In

If an IC which is targeted for bare die application needs burn-in to meet current reliability targets, the industry accepted solution is to perform burn-in at the die level using one of the carrier technologies described above. There is a consensus that, while die level carrier technologies that take advantage of existing test and burn-in infrastructure are a requirement today, the long-term cost effective solution to KGD is to perform test and burn-in at the wafer level.

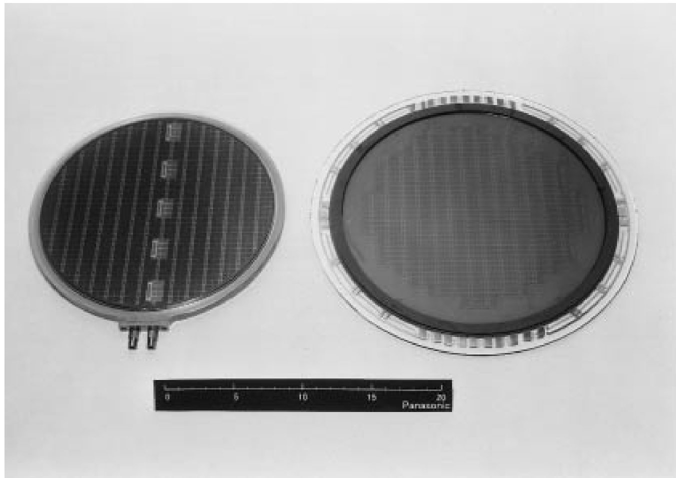

Figure 20. Wafer Level burn in probe card. A polyimide membrane with nickel bumps is fixed on a ring of ceramic having a thermal expansion coefficient similar to the silicon substrate. Localized pressure-sensitive conductive rubber establishes connections between the bumps provided on the polyimide membrane and the wiring substrate. The atmospheric pressure in the space between the bumps and the wafer is decreased in order to apply a uniform pressure in-between. The picture shows a 200 mm wafer (in wafer tray on right) and the associated probe card on the left. Photo courtesy of Matsushita.

In fact, wafer level test and burn-in promises to reduce costs for all die, regardless of whether the final package format is single-chip or multiple-chip modules. Recent studies indicate that burn-in failures in today's VLSI devices are wafer process related, not package related (17), so fully conditioned die at the wafer level can be packaged without degrading reliability. Full wafer probe cards are being developed and will be in trial usage before the year 2000. A particular wafer burn-in technique (18), based on voltage stress applied to special circuits on DRAM circuits resulting in an accelerated burn-in is being applied in industry today. For this method, the probe card contacts only special burn-in pads on the wafer on several die at a time while stress is applied. The probe then steps to the next several die and applies stress. The DRAM circuitry is designed to allow voltage stress to be applied to the array cells only. See Fig. 20.

For a traditional burn-in process using a full-wafer probe, the issues of thermal management, mechanical alignment through temperature and over time, routability, interface to ATE and burn-in electronics, and reliable contact to ever-shrinking die I/O are extremely challenging.

Though Wafer Level Burn In (WLBI) is considered a powerful method to address these issues simultaneously, the problems below have to be solved before establishing the WLBI technology.

1. Establishment of several thousand electric connections by bringing all of the bond pads on the wafer into contact with the corresponding bumps on the contactor simultaneously. Moreover, loads of tens to hundreds of kg must be applied simultaneously.
2. Since the burn-in has to be executed at a temperature range of 125° to 150°C, the coefficient of thermal expansion should be matched between the contactor and the wafer. Misregistration between the bond pads on the wafer and the bumps is possible by the difference in thermal expansion coefficients between the contactor and the wafer.
3. Application of uniform loads to all of the bumps provided on the wafer is essential.

SUMMARY

Known good die have been characterized as "in the eye of the beholder" in terms of quality and reliability. Thus, the emphasis should be on *known,* or the quality of the information available on the condition of the device that is of importance to the assembler. The issue of conditioning die is not a question of technology, but of costs. The current paradigm shift to array packaging, and ultimately, to flip-chip mounting, will bring KGD issues to the forefront for IC manufacturers.

GLOSSARY

Chip-on-board (COB). Bare chip wire-bonded to interconnecting substrate, usually with other chips, forming an MCM.

Controlled collapse chip connection (C4). Solder bumps placed on the IC terminals are joined to matching PC board terminals by reflowing the solder, making a mechanical and electrical connection of the chip to the board.

Defects. Voids, bridging, pinholes, and so on, occurring in a critical area of a device that cause electrical faults (opens, shorts, excessive current, etc.).

Defect level. Ratio of defective devices in a lot to total number of devices in the lot.

Direct chip attach (DCA). Refers to IC's that are flip-chip mounted on interconnecting substrate, usually with other chips, forming an MCM.

Known good die (KGD). Probability that die are defect-free (and will remain so through some predictable lifetime).

Multichip module (MCM). Group of highly functional ICs interconnected on a fine-line substrate.

Tape automated bonding (TAB). A packaging technology that bonds IC to fine-line conductor patterns on a flexible printed circuit tape.

True good die (TGD). Percentage of defect free die in a lot compared to total die remaining in lot.

Yield. Ratio of devices passing a test to total devices subjected to test.

BIBLIOGRAPHY

1. J. Lau (ed.), *Chip On Board Technologies for Multichip Modules,* New York: Van Nostrand Reinhold, 1994.
2. G. L. Ginsberg and D. P. Schnorr, *Multichip Modules and Related Technologies,* New York: McGraw-Hill, 1994.
3. J. Lau (ed.), *Flip Chip Technologies,* New York: McGraw-Hill, 1996.
4. J. K. Hagge and R. J. Wagner, High yield assembly of multichip modules through known-good IC's and effective test strategies, *Proc. IEEE,* **80**: 1965–1994, 1992.
5. E. R. Hnatek, *Integrated Circuit Quality and Reliability,* New York: Marcel Dekker, 1987.

6. E. A. Amerasekera and F. N. Najm, *Failure Mechanisms in Semiconductor Devices,* 2nd ed., New York: Wiley, 1997.
7. A. H. Landzberg (ed.), *Microelectronics Manufacturing Diagnostics Handbook,* New York: Van Nostrand Reinhold, 1993.
8. W. Kuo and Y. Kuo, Facing the headaches of early failures: A state-of-the-art review of burn-in decisions, *Proc. IEEE,* **71**: 1257–1266, 1983.
9. D. J. Klinger et al., eds., *AT&T Reliability Manual,* New York: Van Nostrand Reinhold, 1990.
10. M. H. Woods, MOS VLSI reliability and yield trends, *Proc. IEEE,* **74**: 1715–1719, 1986.
11. J. M. Flaherty, A burnin' issue—IC complexity *Test & Measurement World,* October 1993.
12. D. Maliniak, Known-good die poised to take off, *Electronic Design,* November 21, 1994.
13. H. H. Huston, M. Wood, and V. M. DePalma, *Burn-in effectiveness—Theory and Measurement.*
14. F. Jensen and N. E. Petersen, *Burn-in: An Engineering Approach to the Design and Analysis of Burn-in Procedures,* New York: Wiley, 1982.
15. J. Eastman et al., Defect migration of multi-chip modules using structural test, *Proc. ISHM-IEPS Int. MCM Conf.,* 1994, pp. 230–235.
16. B. Vasquez and S. Lindsey, The promise of known good die technologies, *Proc. ISHM-IEPS Int. MCM Conf.,* 1994.
17. S.-M. Tang, New burn-in methodology based on IC attributes, family IC burn-in data, and failure mechanism analysis, *1996 Proc. Annu. Reliability Maintainability Symp.,* 1996.
18. T. Furuyama et al., Wafer burn-in (WBI) technology applied for RAMs, *Proc. IEDM,* 1993.
19. Y. Nakata et al., A wafer-level burn-in technology using the contactor controlled thermal expansion, *Proc. 1997 Int. Conf. Multichip Modules,* 1997.

LARRY GILG
MCC