NEW GENERATION UNDERFILLS POWER THE 2ND FLIP CHIP REVOLUTION

Dr. Ken Gilleo
Cookson Electronics Group
Providence, Rhode Island

Abstract

Various forms of Flip Chip technology have been around since the 1960’s but within the framework of ceramic substrate. The 2nd generation of Flip Chips, primarily involving the porting over to more cost-effective organic substrate, is now a major force in the Packaging Revolution. The seemingly simple change from ceramic to rigid and flexible circuitry requires post-packing for reliability in the form of underfill encapsulant. Once a bottleneck, underfill has been transformed into a snap flow, snap cure (5 minute) system that delivers the efficient, high through-put required for consumer electronics and other high volume business sectors. This paper will first explore the technological history of Flip Chips before describing the many breakthroughs in underfill chemistry.

A short comparison of Flip Chip and Chip Scale Packaging systems will be made. It will be shown that Flip Chip On Board (FCOB) is not a true package while properly dimensioned Flip Chip In Package (FCIP) meets all of the requirements to be a CSP.

The technical discussion will include the 5 generations of underfill and the significance of each new development. Finally, an attempt will be made to predict what the 6th generation of underfill will bring. The future Flip Chip may well become a standard SMT with integral flux and underfill.

Keywords: BGA, bumping, bumps, CSP, DCA, FC, FCOB, FCIP, Flip Chip, packaging, underfill.

INTRODUCTION

Technical revolutions and their “Next Generation” products are sometimes difficult to quantify. We are unquestionably within the midst of a powerful Packaging Revolution, however. Our 1990’s revolution, inspired by faster-smaller-cheaper, has driven us to efficient Area Array format and dense minimal packages. The last clearly discernible packaging revolution took place throughout much of the 1980’s under the banner of Surface Mount Technology. The SMT revolution brought us productivity improvements and moderate density increases. SMT has been primarily about changing the assembly process although new packages were obviously required to move from feed-through to surface mount. The present packaging revolution builds on the SMT foundation, but it is mostly about changes in the design, fundamental package construction, footprint, materials and package manufacturing processes. The revolution is taking place in the packaging foundries as much as on the assembly floor. The 1990’s area array transition can also be differentiated from the 1980’s SMT event by the assembly equipment - we are utilizing the infrastructure. The very fact that we can use SMT lines for most of the new BGAs, CSPs and other micro-packages, means that the rate of change is much faster and successful changeover is virtually assured by working mostly within the infrastructure.

FIRST LEVEL CONNECTIONS

Fundamentally, there are only two chip connection methods - wires and joints. Yet, there are countless packaging concepts that use these two basic connections and their modifications. Flip Chips, BGAs, CSPs, SMTs and TAB use only wires, joints or combinations. We need not debate whether CSPs are better or worse than Flip Chips. That is the wrong question since excellent CSPs can be built with Flip Chip Technology, but more on this later. The most basic issue is the type of first level connection - the means of connecting to the chip. Is it wire or joint? Flip Chip or Direct Chip Attach (DCA), by definition, is the technology that encompasses joints as the means of connection to the IC.

During the early 1960’s, packaging technologists sought to develop the most efficient means of connecting solid state electronics to the system. Much of the early semiconductor industry worked on developing wire connection schemes with some success. Wire bonding, and various forms of Tape Automated Bonding (TAB) with its chemically-crafted wires, eventually emerged. Wire bonding gained popularity as an important IC connection process, but
not without the resolution of significant problems. However, not everyone was focused on wire technology.

IBM’s laboratories in upstate New York pursued the direct approach. Why not eliminate the chip pad wires and put any routing on the circuit or package? This should improve reliability, enable higher speed, increase density, add manufacturing simplicity and ultimately reduce cost. IBM, noted for very complex “thinking machines”, was taking an approach that was the epitome of simplicity - directly soldered ICs. Nothing could be simpler.

The earliest Direct Chip Attach concepts involved reflow soldering the chips directly to ceramic modules (packages) that could be tested and then assembled to sub-system or full-system circuitry. In today’s jargon, this 1960's technology was Flip Chip In Package (FCIP). Solder was applied to the module, chips were placed and reflow heating produced the connections. The board, rather than the chip was bumped. The concept of bumping the board has gained renewed interest and the technology has become available through Bell Labs. Figure 1 shows an early computer direct solder module.

IBM moved to another phase of Direct Chip Attach where the joint was added to the chip. This very early bumping process involved reflow soldering non-collapsible spheres to the transistor or IC pads. The copper micro-spheres produced a controlled stand-off and facilitated testing. Interestingly, Tessera and other Chip Scale Package (CSP) companies, are exploring the use of copper micro-spheres for their micro-BGAs some 30 years later. The ease of testing afforded by the use of non-collapsible spheres appears to be the driving force for their resurrected use today. Figure 2 shows the IBM copper sphere attachment process used in the 1960’s. Please note that this 3-decades old package is a micro-BGA utilizing both area array and SMT principles.

Although the copper micro-sphere invention proved reliable, IBM was driven toward a mass bumping process as Flip Chip use was dramatically increased. The method selected was multi-metal vacuum deposition of high lead-content tin solder. The use of precise area pads on chips and circuits allowed the reflowing solder to partially collapse during assembly to produce a predictable and well-controlled stand-off. Thus, the Controlled Collapse Chip Connection (C4 acronym) was derived which degenerated to C4. Vacuum deposition of bumps is still used today by IBM, Motorola and others, typically with alloy modifications. Motorola has reported on the E-3 (Evaporated Extended Eutectic) process that uses a multi-step, multiple alloy process [1]. A tin cap is deposited over the high-melting high lead bump to provide low temperature soldering while maintaining a stand-off.

2nd GENERATION FLIP CHIP

Direct Chip Attach (DCA) has served the computer and automotive industries for 3 decades under the C4 moniker, but the term Flip Chip is becoming more popular. The most important shift for this well-proven technology and fundamental IC interconnect system has been the move from ceramic to organic substrate. The transition of C4 to organic substrate can be described as 2nd generation Flip Chip. While the ceramic-based packaging and substrate industries have done a remarkable job of increasing performance and reducing cost, the excellent benefits derived from organic materials, make them the dominant systems for
PCBs and chip carriers today and in the future. Newer methods for producing micro-vias, ultra-fine lines and extremely dense multilayers are being implemented to further boost capabilities of polymer-based circuitry and interconnects. What’s more, organic circuitry and packaging still have considerable untapped capability that will be delivered in the near future as the Printed Circuit Revolution gains momentum.

2nd generation Flip Chip, the move to organic substrate, is well underway in a number of business sectors. Watches, pagers, cellular phones, disk drives, BGAs, CSPs and a host of other products are now in production with Flip Chips on FR4 and flexible circuitry. These products are already in the market place. The StarTAC® cellular phone from Motorola, the first wearable product in its class, was launched with 2nd generation Flip Chip. PC disk drives are faster and now boast higher capacity because of Flip Chips on the read-write flexible circuit.

**FLIP CHIPS vs. CSPs?**

Before moving onto some of the technical issues associated with 2nd generation Flip Chip, let’s briefly explore the over-popularized “CSP vs. FC” issue. Perhaps too many conferences and debates have asked us to decide which technology is better. This, of course, implies that Flip Chip and Chip Scale Packages are mutually exclusive. First we need to determine if a FC is actually a CSP. The 1st generation Flip Chips were Chip Scale Packages, but the 2nd generation FCs are not. Here is an explanation of this paradox.

The package can have many attributes, but most will agree that it must provide chip protection if this is required for reliability. The bare Flip Chip can require added protection depending on how well the pads are sealed after attachment and on the type of passivation. The C4 high lead solder bumped FC for ceramic were designed so that they did not require additional protection. Chips could be solder assembled and reworked if required.

However, 2nd generation FCs on organic are incomplete until underfill is added. The underfill serves at least two purposes: (1) performance enhancement and (2) chip/joint protection. Underfill is actually an encapsulant that is coated over the active area of the die and interconnect structure. Keep in mind that we have flipped the chip over so that the active surface is down. The underfill completes the assembly and can be considered a post-packaging step. Unlike the C4 chip on ceramic, it can not be eliminated in most cases for organic substrate.

We then need to determine if the underfill-protected FC is now a CSP? Most packaging experts concluded, in a recent Internet survey [2], that a package must be removable after assembly and preferably, reworkable. Today’s commercial underfills are not considered reworkable although IBM and others have announced technology in this area. Since the underfilled FC is not reworkable, the system is not a true package and therefore it is not a CSP. If underfill can be either eliminated or made readily reworkable, the Flip Chip is then a Chip Scale Package. But there is still no need to debate FC vs. CSP because it is relatively easy to transform a Flip Chip into a CSP.

**FCOB vs. FCIP - THE REAL ISSUE**

The DCA and other bare die processes can be used to apply the chip to a circuit board or to a chip carrier, commonly called an interposer. Wire bonded chips can be placed on a circuit to produce the familiar Chip-on-Board (COB) or on an interposer to build a BGA. Similarly, Flip Chip can be placed on the circuit board with the Direct Chip Attach method or onto an interposer. The Flip Chip placed onto an interposer is now being referred to as Flip Chip In Package (FCIP). Amkor, IBM, Motorola, TI and a number of Japanese companies are now developing and producing FCIP products. A Motorola product is described in the preceding paper of this conference. When the interposer is designed with a footprint meeting the 1.2 size limit, the FC package is a CSP. The FCIP product is reworkable and hence a true package. Underfilling is done by the packaging foundry not the board assembler. Each of these Flip Chip technologies have their advantages and limitations, but this will be the topic of another paper. Figure 3 shows a Flip Chip-CSP from the now famous Sony HandiCam, the most “over-exposed” camera in the market.

**FLIP CHIP THERMOMECHANICS**

The seemingly simple shift from ceramic to an organic circuit platform for the veteran Flip Chip would not
appear to require anything new. But let’s examine one fundamental characteristic of matter before commenting. All materials undergo dimensional changes under the influence of temperature change. Most materials expand when heated as increased thermal energy causes atomic or molecular distances to increase. The term used to quantify the dimensional-temperature relationship is the Coefficient of Thermal Expansion (CTE). It is commonly expressed as the change in length per unit length per degree (C or K). Since the thermal expansion of common materials is only a few millionths of an inch per linear inch of material when temperature is increased by a single degree, it is convenient to use dimensionless units; parts per million/°C. Americans seem to like dimensionless units perhaps to more gracefully avoid metrification. Table one lists CTE values for common materials.

<table>
<thead>
<tr>
<th>Material</th>
<th>CTE (ppm/deg-C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aluminum</td>
<td>23.0</td>
</tr>
<tr>
<td>copper</td>
<td>17.5</td>
</tr>
<tr>
<td>solder</td>
<td>25-26</td>
</tr>
<tr>
<td>FR4</td>
<td>18-25</td>
</tr>
<tr>
<td>ceramic</td>
<td>~6</td>
</tr>
<tr>
<td>silicon</td>
<td>2.3</td>
</tr>
<tr>
<td>underfills</td>
<td>19-35</td>
</tr>
</tbody>
</table>

Two conclusions should become apparent in looking at Table 1. First, silicon has a very low CTE value. Secondly, ceramic has an expansion rate that is fairly close to silicon’s while FR4 is nearly an order of magnitude higher. The switch from ceramic to organic is clearly a non-trivial transformation that will encounter problems. Thermocycling in such a mismatched environment will place strain on the joint as graphically represented in Figure 4.

Since there is only a small thermomechanical mismatch between silicon and ceramic, the amount of stress on the solder joints will be small. Unless a very large chip is involved, the deformation of the joint will be small and within safe limits especially with high lead solders. However, there will be an excessive mismatch for the chip on organic substrate. Heating and cooling will cause the PCB to expand and contract considerably compared to the silicon. The solder joints will be damaged and will eventually fracture. This thermomechanical strain problem will occur even for small die (~ 6 mm) on FR4. The transition from ceramic to organic circuit substrate thus has a major obstacle to overcome. Figure 5 shows the expected thermomechanical induced fracture of a eutectic Sn/Pb solder joint.

There are several solutions for solving this thermomechanical mismatch problem resulting from assembling flip chips to polymer-based substrate. The simplest and most cost-effective approach is to fill the gap between the chip and circuit with a composite that transfers the stress away from the fragile bump zone to a more strain-tolerant region. Figure 6 is a diagrammatic representation of the stress displacement produced by underfill. The stress is displaced to the organic circuit board that can accommodate the forces.
We should note that underfill movement is constrained by the very rigid and low expansion chip, but only in the X-Y plane. The underfill will expand as described by its own CTE value in the vertical or Z-direction. Underfill thermal expansion rates are thus asymmetric. This is why the underfill should have a CTE value that approximates that of the joint. Underfills are therefore designed to have CTE values close to solder’s. No attempt should be made to match the CTE of the chip as this would produce a vertical mismatch between underfill and joint.

**THE INVENTION OF UNDERFILL**

Underfill appears to be one of those rare “Eureka” inventions where the unexpected really did happen. Veteran Flip Chip developers from IBM claim that filling the gap between the chip and ceramic hybrid circuit was originally tried to protect indium solder joints from corrosion since these alloys are much more sensitive than common tin-lead solders. As the story goes, a number of polymer sealing systems were flowed under chips and hardened. The materials ranged from silicones to common epoxies. Along with temperature & humidity testing, the lab carried out long-duration thermocycling. A small, but positive trend was seen in some of the data for the “underfilled” parts. Harder resins appeared to give the most significant improvement.

IBM correctly deduced that the rigid polymers were limiting the small differential movement between the chip and substrate. The discovery was of dubious value for the ceramic systems that had been run as high as 60,000 cycles without failure (on small die). The discovery was a solution to a non-problem at this point in time. However, the underfill invention had some merit for large die and higher expansion (and lower cost) ceramic. The underfill was formulated by adding silica to reduce the CTE and better confine flow. Color was later added for visibility. One of the early formulators claims to have come up with blue, not so much for corporate loyalty, but for expediency. The blue is said to have been borrowed from the pigment mix used to make the green solder mask then captivity-produced by IBM.

The original IBM colored epoxy-fine “sand” mix is the 1st generation underfill. The material satisfied the needs for building mainframe computers with ceramic circuits. However, the Packaging Revolution, with its craving for low cost organic substrates, would place new demands on underfill. So when Flip Chip began making a serious move onto PCBs a few years ago, the high volume consumer-oriented products, such as pagers and cellular products, needed higher productivity. No longer was the “flows-like-molasses, cures-in-half-a-day” underfill sufficient. The 1st generation underfill was indeed a bottleneck that required attention.

**SPEED, THE INGREDIENT FOR 2ND GENERATION**

2nd generation underfill vanquished the slow flow problem. New entrants into the underfill arena introduced materials made with much lower viscosity resins and fillers of smooth, spherical silica in 1995. These new underfills gave a flow rate improvement of 10 to 12 times over the current materials as reported in the IBM-Universal Instruments Corporation consortium. The fast flow feature brought the productivity breakthrough needed for the new consumer products ready to go into production with Flip Chips. But faster flow caused some seemingly abnormal phenomena observed by researchers. Large Flip Chips ended up with voids in the underfill beneath the die that were initially attributed to outgassing of the new, low viscosity epoxies. Inexplicably, many large peripherally-bumped die ended up with unacceptable voids in the center [3]. Flow was faster along bump patterns, not slower, as predicted by conventional wisdom. Figure 7 shows the flow rate test using transparent quartz flip chips.

Eventually, the mystery was uncovered by the use of “glass” circuits and video cameras that showed fast-flow underfill moving most quickly along the bump rows thus trapping air. This is shown in Figure 8. Fastest flow always occurred where bump density was highest. Although increased flow rate along bump rows is a simple consequence of surface chemistry, it was viewed as a bad feature of this first 2nd generation underfill. Later, it would be found that all low viscosity,
highly wetting underfills behaved the same way. Today, the simple solution is to use the correct dispensing pattern. But, being first was not the best place to be in the first days of 2nd generation Flip Chip.

![Figure 8 - Air Entrapment Mechanism Diagram](image)

**3rd GENERATION UNDERFILL SPELLS RELIEF**

First and second-generation underfills were designed as stiff, high modulus composites. This was logical since the underfill was required to lock the chip and circuit board together to prevent differential movement. However, some assemblers were beginning to experience die cracking problems with Flip Chips on thin FR4. Newly launched 2nd generation Flip Chip was already in trouble before achieving orbit. In fact, a line producing cellular phones using relatively small Flip Chips had to be shut down until the problem could be solved. The entire FC program of a major electronics company was in jeopardy.

A successful answer to die cracking was to employ a principle called Intrinsic Strain Relief (ISR) - enter now the 3rd generation underfill. Underfills had been designed with very high modulus, but apparently too high for thin and easily bent PCBs. In-process handling of the thin organic boards populated with brittle silicon certainly exacerbated the problem. A lower modulus was clearly indicated, but the solution is not so simple as flexibilizing the underfill. Since the function of the underfill is to restrain lateral movement relative to chip and circuit, a flexible material might prevent die cracking but allow solder joints to fatigue and fracture. The first successful approach to the problem was to formulate an underfill with temperature-varied modulus. The modulus of most polymers generally decreases slowly as temperature is increased. The material would slowly as temperature is increased. The material would

<table>
<thead>
<tr>
<th>Material</th>
<th>Modulus@ 25°C</th>
<th>Modulus@ 125°C</th>
<th>Radius of Curvature</th>
<th>Apparent Tg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical</td>
<td>11-13 GPa</td>
<td>7-9 GPa</td>
<td>700-800 mm</td>
<td>170°C</td>
</tr>
<tr>
<td>ISR1</td>
<td>11</td>
<td>1</td>
<td>1250-1600 mm</td>
<td>128</td>
</tr>
<tr>
<td>ISR2</td>
<td>9.2</td>
<td>0.15</td>
<td>1500-2100 mm</td>
<td>119</td>
</tr>
<tr>
<td>ISR3</td>
<td>7.3</td>
<td>0.05</td>
<td>1750-4000 mm</td>
<td>100</td>
</tr>
</tbody>
</table>

The paradox is that the apparent glass transition temperature (Tg) is well below the cycling temperature. Conventional wisdom had dictated that Tg must be well above the test and operating temperatures. The reason for the high Tg was that most underfills undergo a 300% to 400% expansion rate increase above their Tg. The increased CTE value, called a₂, would cause the underfill to expand and strain the solder joints in the vertical plane. However, no such failure occurred during extended (>1000 cycles) thermocycling on actual parts. The success of the ISR underfill is thought to result from less dramatic a₂ increase combined with very low modulus above the Tg that does not exert destructive vertical forces on the die/circuit structure. Regardless of exactly how the variable modulus, low Tg underfill works, the product is field-proven with high reliability record. The excellent performance launched the 3rd generation ISR underfill into high volume production in 1996 the 2nd generation Flip Chip was again on track.

Fast cure speed was moved to the top of the underfill wish list once the serious issues had been addressed by 2nd and 3rd generation products. The speed boost from a 6 hour cure for 1st generation underfill down to about 30 minutes for some of the newest systems, was not good enough as Flip Chip sought to become an in-line process. The industry wanted snap cure. Snap cure, of course, has been around almost as long as die attach adhesives. However, the so-called snap cure epoxy hardeners are typically solids (like dicy/epoxy systems) or very viscous imidazole liquids that have been poor choices for fast-flow, low viscosity underfills. Furthermore, most high performance epoxy formulas relied on anhydride hardeners that have relatively slow reaction rates.

Faster cure was made in late 1996 that allowed the epoxy-anhydride reaction to be greatly accelerated. A 5 minute cure, often considered the
criterion for snap cure, was now possible in a fast-flow, high performance anhydride system. The 4th generation underfill, with the suggestive code name HEL (for “Hot” Epoxy Liquid), was nearly ready to go. Extensive testing showed that all cured-polymer properties, including adhesion and moisture resistance improved. However, the new “hot” catalyst also increased the modulus due to more efficient polymerization. The concern over die cracking again became a concern. Furthermore, faster cure often translates to higher stress. This has been seen in BGA encapsulants where a quick, single-stage cure results in greater stress-induced warp. So while the 4th generation, snap cure underfill was awaiting launch, the obvious decision was made to combine fast-flow, fast-cure and intrinsic strain relief into a single material. Figure 9 shows the Differential Scanning Calorimetry curve (DSC) for a snap cure underfill.

![Figure 9 - DSC Curve](image)

DSC Curve - Time of Cure @ 165C

One more new feature, recently added to state-of-the-art underfills is Color-Change-On-Cure. The idea is to add a polymerization end-point detector to signal when the underfill has been fully cured. The color change can serve as a safety check against incomplete cure due to any number of causes. A special class of indicator dyes has been discovered that can be used with epoxy-anhydride systems. The new Color Change underfills convert from a brilliant red to an amber color during the final stages of polymerization. The color change does not occur even if the underfill is hardened but not fully polymerized. The new color change system has been approved by a large FC assembler and is in the final stages of qualification. The indicator is effective in detecting inadequate oven temperatures and cure inhibition due to contaminants. Figure 10 shows the color change. The circle is cured underfill while the darker, features are freshly deposited uncured material.

![Figure 10 - Color Change Smile](image)

5th generation underfill has been qualified on both rigid and flex circuit substrate using a criterion of 1000 cycles @ -55 to 125°C. Today, it is possible to underfill a common-size chip (6 mm x 6 mm) in just seconds and cure in only minutes, all without concern over stress. Several assemblers of smaller Flip Chips no longer use the sealing step where underfill is dispensed around the die opposite to the dispensing edge. This means that the Flip Chip can be efficiently underfilled by depositing material at one edge of the chip and going directly into the curing oven. The underfill bottleneck has certainly been shattered by 5th generation underfills.

**GENERATION 6?**

What’s next? First, it is worth mentioning that many less dramatic, but very significant improvements have been occurring in parallel. Fillers have greatly improved in just the past 6 months. Very small and extremely smooth spherical silica has become commercially available that enables faster and smoother flowing underfills to be manufactured. Fillers under 3 microns in diameter are now available that are being used in underfills designed for use with the smallest gaps without clogging. New Flip Chip joining techniques, including thermocompression bonding and conductive adhesive assembly, can produce gaps of 15m and less. Ultra-clean synthetic silica is now available to make underfills for memory chips that are sensitive to α-particle emissions.

We should also mention attempts to combine underfill and flux into a single product. This type of underfill is commonly referred to as “pre-applied”. This concept has been under development and testing for at least 3 years, but with only marginal success [4]. The idea is to
apply underfill to the chip site, place the chip and form solder joints while curing the underfill, all within the reflow oven. One of the few commercialized products, however, requires a 1 hour post bake [5]. Even if the entire process could be completed on the reflow line, underfill-flux must still be dispensed with added equipment that may not be part of a standard SMT line. Figure 11 compares capillary flow vs. Pre-applied underfill processing.

We would also expect a reworkable underfill for the pre-applied system since in-circuit testing would only be possible after the simultaneous reflow soldering/underfill hardening step. Although reworkable underfills were developed several years ago [6,7,8], the task of combining these “depolymerizable” chemistries with a self-fluxing system is a major challenge.

The real 6th generation underfill would be a solid. The material would be applied at wafer level by the silicon foundry or the bumping service. The bumped wafer would be coated with flux-underfill on the bumped side by screen printing or spin coating. Both of these methods are used today for the application of thermoplastic die attach adhesives. As with the die attach adhesive, the flux-underfill would be dried. Such a system could be made with solid epoxies and hardener in solvent. Organic acids used in commercial liquid epoxy-based fluxes are solids that readily react with epoxy resins under heating. The flux-underfill can be brought to a solid phase by simple solvent evaporation just like the die attach adhesives. The “dry” flux-underfill coated Flip Chip could then be packaged in tape and reel and used like a standard SMD. The assembler could then run a Direct Chip Attach process on a standard SMT line at high speed. The IR reflow oven would melt the flux-underfill, reflow the solder bumps and polymerize the epoxy-hardener into the finished underfill. Figure 12 shows the proposed process. Is 6th generation underfill possible? Stay tuned.

REFERENCES


