A Short History of Flipped Chips

By Ken Gilleo, ET-Trends; gilleo@ieee.org

THE EARLY DAYS

In the early 1960's, the semiconductor industry was trying out various schemes for connecting the new Integrated Circuits (IC) to printed circuit boards. Although there were many variations, the three basic approaches were wire bonding, chip carriers with beam leads and direct chip connections. And 4-decades later, we use the same three approaches but with many more variations. Back then, the chip makers did all the packaging and each company had one favorite approach. While wire bonding worked, most companies didn't like the *one-lead-at-a-time* and the slow manual process. Many companies attempted to make integrated leadframes with cantilevered beams that could be bonded to the IC all at once. These "spider" circuits later became Tape Automated Bonding (TAB) where the metal spider was replaced by a flex circuit. Today, flex-based chip carriers with cantilevered beam connections are popular in Tessera's µCSP and IBM's TBGA (Tape Ball Grid Array). Figure 1 shows the early spider circuit.

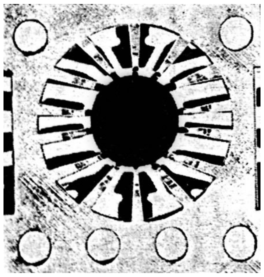

Figure 1 – Spider Circuit

IBM was a strong innovator even back in the 1960's as they are today. While the rest of the electronics industry was wiring up chips, IBM took a direct approach. Nothing could beat a direct chip connection (DCA), that was certain, but reduction to practice was non-trivial (a favorite IBM expression). IBM explored a number of methods for DCA and while soldering could work, results were too variable. Bridging could occur and the gap between the chip and substrate varied. If only the height could be controlled, DCA would be a great alternative to wired chips.

One early package, the SLT (Solid Logic Technology), solved the problem and several others. The interconnect choice was the copper micro-ball not solder. The copper spheres were bonded to the pads of the solid state device (a transistor) using high-melting solder. Alpha Metals (now Alpha-Fry Technologies) manufactured the copper spheres until the 1980's. The pre-cut copper segments were dropped in a fluidized bed of inert mineral powder and melted by an induction coil. As the segments passed though the hot mineral it melted and formed into spheres. [1]

The result was the world's first Ball Grid Array (BGA), the first DCA and apparently the first Surface Mount Technology (SMT) device. And since the package was chip size, the SLT was also the first CSP. While some may argue that flip chips are not real CSPs, this package did not use underfill and could probably be reworked since the copper balls were attached with high-melting alloy. Ironically, a “new” innovation in CSP industry is to use copper micro-spheres to achieve the same results obtained in 1961 [2]. Figure 2 shows the SLT, a still a great idea [3]. Note that the package development was so early, that it was applied to transistors. The product went into production in 1964 in the IBM System/360. Underfill, actually sealant, was used to exclude moisture from the bump interconnect to reduce concern about electromigration. Silicone was common in the beginning but an amide-imide, called ATP-10, was also used that had high enough modulus to improve thermocycle performance. So underfill is also an invention of the early 1960’s.

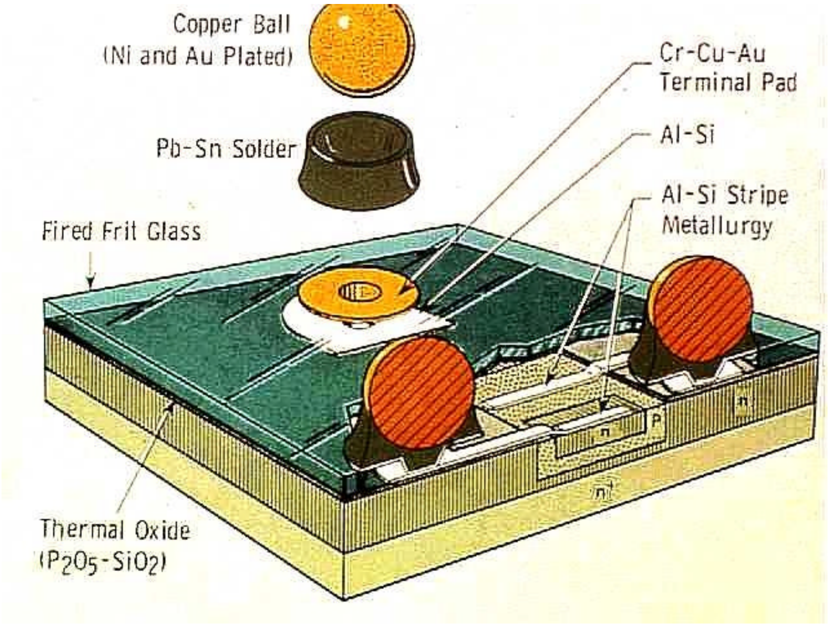

Figure 2 – SLT (Solid Logic Technology)– IBM (Paul Totta)

C4 (C^4) for Controlled Collapse Chip Connection, was the second flip chip technology and came the year after the SLT launch of 1964.

SECOND GENERATION FLIP CHIP

At IBM and a few years later at Delco, ceramic substrate became the standard substrate for flip chips. This was logical since no other practical circuit material had the desired low coefficient of thermal expansion (CTE). A CTE mismatch between the silicon device and substrate would cause high stress on the bumps during temperature cycling that could result from power on/off cycles. Still, the ceramic substrates were not perfect matches. Alumina range from 6.5 to 7 ppm/oC although glass-ceramic can get down to about 3 ppm/oC compared to 2.3 ppm/oC for silicon according to Marie Cole of IBM.

The Underfill Effect

How could low cost organic substrate, like FR4, be used? The solution had been found years ago in one of the rare serendipity events that we like to call a real invention. Retired IBMers tell this story and the following is based on hearsay. A variety of bump metals were being investigated including indium solder alloys. These were especially intriguing since joints could be formed without reflow by just using pressure. The indium alloys and some others could degrade in the presence of moisture especially under biased current. Although the IBM modules were sealed from the environment, Research decided investigate moisture-resistant sealants to place around the bonded flip chips. Materials ranging from silicones to epoxies were tested. It became apparent that low viscosity materials could flow under the tiny chips. Too make a long story short, testing unexpected revealed that some of the underfilled flip chips could be thermocycled longer. Investigation of the underfilling effect lead to the conclusion that properly formulated adhesives could extend lifetimes. Thermal cycle lifetimes ranged from thousands to tens of thousands of cycles before failure because of the small mismatch between chip and ceramic carrier. While silicones had very little affect on joint fatigue, the amide-imide gave up to a 300% cycle time improvement. But epoxies produced an astounding 10-fold improvement. Hitachi, in 1988, provided underfills to IBM that consistently improved thermal cycling and sealed out moisture [2]. However, the “underfill effect”, although well understood, was not really needed for these small FCs on low expansion ceramic.

Later, larger chips and experimentation with organic substrate made the underfill material more important. Epoxies were chosen as the preferred polymer because of availability, ease of formulation and last but not least, safety. IBM had been using epoxies for solder masks and other PWB applications and was comfortable with them. In fact, when the group testing underfills asked for coloration to make inspection easier, one formulator borrowed from the circuit materials group. Wanting to avoid the tedious process of qualifying a new “chemical” for use within the company, the formulator asked for the green colorant for masks. Finding that a mixture of yellow and blue colorant was used, he chose blue – and this is why IBM underfills are blue.

Later, the IBM New York based labs and manufacturing facilities sought to outsource their underfill. They were already using materials from a small coatings formulator near by called Dexter-Hysol. The company began producing the underfill for IBM on a toll basis. Viscosity was high, flow was slow and curing time was about 6 hours. By today’s standards, these first underfills would be commercially viable.

In the late 1980’s and early 1990’s, Motorola, an IBM flip chip licensee, worked diligently on getting second generation FC into production. If IBM was a “ceramic circuit” company, Motorola was a low-cost, high volume “polymer circuit” company. Motorola quickly released that slow flow. Long cure was not the path to productivity and initially sought shorter cure. In the early 1990’s, Alpha Metal’s polymer group, Advanced Products Division, decided to tackle the “underfill bottleneck” problem. Their first product drop the cure time down to 30 minutes and the flow was much faster. The Universal Instruments consortium tested the first fast flow/fast cure material. As fate would have it, the test chip, supplied by Texas Instruments, had a non-symmetrical layout with multiple rows around the perimeter, but no bumps in the middle. The theory back then, held that bumps slowed down underfill and fewer bumps was better. When the fast flow material was tested, a large void formed under the middle of the chip. The first theory was that the new underfill was outgassing, but this was not the case. A new innovative test set up was designed where the flip chip bumps were carefully coined flat and the chip was clamped against a glass plate that was figured over a video camera. Real time video disclosed what was really happening. The underfill was flowing quickly along the bumps and very slowly down the middle. In hindsight, this is a normal result of surface chemistry. Surface tension wetting or capillary action pulls along the underfill (hence the name – capillary underfills). Unfortunately, the conclusion was that "this is a weird underfill and you Alpha guys need to fix it? Figure 3 shows the flow problem.

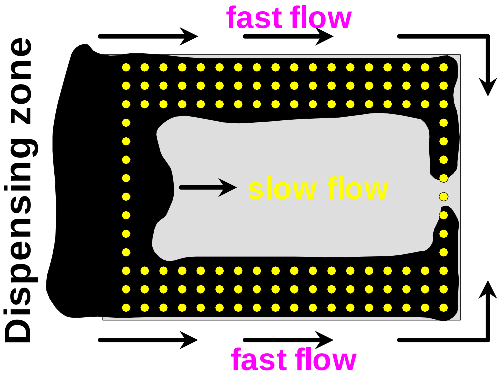

Figure 3 – Uneven Underfill Flow

The Great Underfill Race

The Jersey City based Alpha group, after checking into the basic principles of surface chemistry, decided that they were pursuing a correct strategy with fast flow. Underfills should have low viscosity, low surface tension and fast curing properties. Being new and being first in unconventional underfills had not been the right position. The consortium gave the newcomers inadvertent bad press with a, "Will you look at this weird underfill in the video". Then came the StarTac crash. The new Motorola StarTac cellular phone, the lightest and most sophisticated phone of the early 1990's was using a flip chip. But the classical underfill material, probably influenced by the old IBM formula, was causing die cracking. A legion of engineers worked long hours to save the project and the product In fact, the problem was so severe, that manufacturing resorted to manual wire bonding to get product out during the interim. The entire future of 2nd generation flip chip seemed to hinge on this high visibility product. How long could a company tout leading edge flip chip while quietly wire bonding? To make matters worse, marketing firm Prismark, did a study on the new StarTac and surprisingly found that their reverse-engineered phone had NO FLIP CHIPS.

Motorola became desperate and let it be know that anyone who could solve the problem would get some quick business. The Alpha group had nothing to lose by taking a wild swing since they had no customers. They already had met the desired characteristics of fast flow with fast cure and had passed thermocycling. Being an inexperienced underfill formulator was an advantage. But Motorola was allowing only a week and one shot for new entrants. The Alpha group applied well-know coating technology tricks and broke rules held sacred by the handful of underfill self-anointed gurus. Modulus was lowered but the glass transition temperature dropped far below the accepted limited of 125°C. Conventional wisdom held that if the Tg was less than maximum cycle temperature, the typically high CTE above Tg (called α_2) would cause damage. The new materials were submitted offering a range of modulus values and – they all passed. Motorola quickly qualified the new Alpha underfill and the StarTac with FC was back on the market.

While there were few contestants, the underfill race was on. Unconventional Alpha, with tricks from the epoxy coatings industry, continued to lead by introducing a 15-minute cure and finally a 5-minute system in 1995. Abelstik (National Starch), with newly acquired Amicon, also had a 5-minute cure with good properties. A year or two later, others including the market leader, had figured out the quick tricks. During the entire racing, flow was improved and by the late 1990's, the most of the products had reached a plateau. However, event in a different field would change strategies for underfilling.

NUF

Let's head back to Jersey City and the early 1990's and look into the flux research lab. Flux R&D was aimed at coming up with a better no clean flux. A polymer chemist, acquired with one of the many Alpha/Cookson acquisitions, was convinced that epoxy resin and acidic hardener was a viable path. The idea was to form a neutral residue and just leave it in place. Work led to flux formulations with common epoxy resins and organic acid hardeners. Results were encouraging and a family of epoxy-based flux materials and pastes was launched. Now back to flip chip problems.

One of the irksome problems with flip chip underfills was lack of compatibility with most fluxes. One could clean flux residue but this was difficult with the cleaning equipment of that day. One idea that begged to be tested was the use of epoxy-based flux with epoxy-based underfill. And, of course, it worked. In fact, the two epoxy systems polymerized together during underfill cure. But there was another obvious idea begging to get out to lead a life of its own. Why not use one material for flux and underfill? And this is just what happened. Alpha Metals, Kester and others formulated flux-underfills and this class continues to show promise as "No Flow" underfill or NUF, although pre-applied flux-underfill liquid is more accurate.

Georgia Tech added the flux-underfill process to their flip chip program and that group has done substantial work to quantify the process. So who really invented so called "no flow" underfills? Alpha Metals (now Alpha-Fry Technologies) holds some basic patents for epoxy-based flux. Kester probably introduced the first commercial products. And Georgia Tech elucidated the process and highlighted the concept. But the credit for the invention appears to belong to Motorola who received a patent the combined flux/underfill adhesive concept. Although this 1990 (filing date) invention may not be optimum from a formulation viewpoint, the idea is there.

WUF

But there's more to the underfill story. Back to Jersey City again, and yes, this is the last time. The Staystik^{TM} gang was formulating, testing and selling die attach adhesives, but not the ordinary silver-epoxy. Their materials were based on thermoplastic resins that could be remelted and REWORKED. The Staystik products were sold as films and pastes in solution. The group wanted to get on the underfill bandwagon but a solvent-born material was totally impractical. Customers were just starting to ask for reworkable materials. Solvent evaporation would cause voids and significant shrinkage. However, some customers were buying die attach paste, spin coating on wafers and drying. Others were stencil printing to get a higher yield. The result was a smooth, tightly bonded dry die attach film that could be bonded to a carrier by heat activation. Better yet, the coated wafer could be sawn. One more idea was struggling to get out. Why not substitute silica for silver filler and coat the front of the wafer? And here was the basic wafer-level underfill concept.

Research showed that the wafer-level concept was viable, but a lot of materials and process work would be required. Basic systems were tested ranging from single-layer flux-underfills to double-layer configurations that kept flux and underfill separate. The flux layer development tapped into the experience of the epoxy-flux group. A change in resins provided a solvent-applied solid flux. The underfill layer was adapted material from the thermoplastic Staystik die attach family. Functionally, both underfills and die attach adhesives must strongly bond the chip to substrate and the mechanical properties can be remarkably similar. After basic prototypes were developed and a large number of patents were filed and allowed, Alpha decided to wait and see since it was not really clear that wafer-level underfill would be accepted by the industry and that there would be a return on the substantial investment. The consortium made up Motorola, National Semiconductor, Loctite and others, continues to slowly develop a wafer-level underfill product and process, but will they run into the mine filed of patents by AFT and others.

MUF

But there is still one more piece to the underfill history. This time, we go to Alpharetta, GA where the former BP-Amoco Plaskon group developed molding compounds. The division was acquired by Alpha and moved a mile down the road to a major Alpha facility. One of the most intriguing projects was molded underfill (MUF). The group realized that high lead count Ball Grid Arrays (BGAs) would likely use flip chip and that many would be over-molded. Work with Hestia on molds lead to the development of tooling that could provide void-free underfilling. Epoxy Molding Compounds (EMC) were developed with the right fillers and melt profiles that gave good underfill performance. So with 4 classes of underfill vying for market space, lets leave the races and now look at the history of bumping.

BUMPING

The first bumps used discrete spheres as was shown with the IBM SLT. While IBM had mastered the assembly process, they sought a mass mumping method for simplicity and costs-savings.

Why Bumps?

First, it is very difficult to directly attach to thin aluminum, or even copper IC pads. Solder to aluminum is very difficult at best, but thin metals are "leached" or dissolved by solder. To, the new copper IC's don't really solve the pad connect issue.

A second reason for bumps is to provide a standoff that can produce a gap between chip and substrate. If the interconnect length was close to zero, extreme stress concentrations would result. Also, a bump provides a structure used by several connection deposition processes.

A layer of one or several metals is commonly created over the IC pad. The structure, called under bump metallization (UBM) creates a barrier between the fragile IC pad and the bump. The final finish must either be solderable or compatible with another joining material like conductive adhesives. Many new and old UBMs exist and most are listed in Table 1.

UBM TABLE 1

Need to make

Assuming that the proper UBM has been applied, let's look at bump choices.

Discrete Bumps: We can start with preformed spheres just like BGAs. Recall that the first FC used tiny copper balls that were soldered to the UBM-coated pads of the SLT package. This method is still viable but not too common. Both non-fusible copper balls and fusible solder alloys could be used.

Vacuum Deposition: Early on, IBM used vacuum deposition and this method is still used but declining. Equipment is expensive and maintenance is high. However, a variety of alloys can be produced by sequential deposition of two or more metals.

Plating: Electroplating has been known for decades in both the PWB and IC industry. IBM set up electrolytic bump plating several years ago to replace vacuum deposition. While electro-bumping is not the lowest cost process, it produces extremely good control and will likely be the preferred method for high I/O fine pitch.

Electroless nickel, a very old process in automotive and other industries, was applied to FC about a decade ago. While Delco and others developed processes, Technical University of Berlin (TUB) has done the most work and published extensively. PacTech was started a few years ago as the commercial spin-off of Fraenhufer Institute. PacTech and others offer services and technology know how for those wanting to do their own bumping. The resulting nickel bump is rather shallow since no mask is used to contain plating. Also, the nickel cannot be directly bonded and a joining material must be added. Several companies, including Motorola, have widely reported on forming solder bumps over the nickel. The nickel with solder paste bumping process appears to be one of the more popular for low to medium density.

Conductive adhesives can also be used to form bumps and Epoxy Technology (Epotek) invented and developed methods over the 1990's. They spun off Polymer Flip Chip (PFC) to handle customer bumping. Several companies have adopted polymer bumps. Conductive adhesive paste is used to join the bump to substrate. The bumping and assembly processes are low temperature and there are no α-particle emitters.

Last but not least, mechanical stud bumping is now covered. The most common version involves using a gold ball bonder to form a connection to the pad and then breaking the wire to leave a bump and tail. The process goes back at least to the late 1980's and was practiced in the IBM labs for prototyping. Several Japanese companies have improved the process by adding coining and shaping steps.

Equipment is now available that can form 10 or more gold stud bumps per second. While cost inevitably increases per die as IO count goes up, this is a practical cost-effective bumping method for many die like memory. It is also ideal for single chips that are the desirable bare die format for many companies.

CONCLUSIONS

The history of FC probably began in 1961 with the invention of IBM's SLT. The technology has evolved on many fronts, especially bumping and underfill. While FC is not for everyone, the technology will continue to grow, especially for CPUs with their high lead count.

REFERENCES

[1] Angelo Gulino, Alpha-Fry Technologies, personal communication, Jan. 8, 2001.

[2] Totta, P., IBM, personal communication, Jan. 8, 2001.

[3] Totta, P.; Sopher, R., "SLT Device Metallurgy and Its Monolithic Extensions", IBM J. Res. Development, 5: pp. 226 – 238, May 1969.

More Underfill Reading

Gilleo, K., "Flip or Flop?", pp. 40 – 42, 44 – 48, Circuits Assembly, Feb. 1997.

Gilleo, K., Cinque, T. and Silva, A., “Flip Chip 1, 2, 3: Bump, Bond and Fill”, pp. 32 – 34, Circuits Assembly, June 1996.

Gilleo, K. and Blumel, D., “The Great Underfill Race”, pp. 701 – 706, Proc. International Symposium on Microelectronics, IMAPS, San Diego, CA, Nov. 1 – 4, 1998.

Diep-Quang, H., “Epoxy Based, VOC-Free Soldering Flux”, US Patent 5,904,782, May 18, 1999.

Preveti, M., “No Flow Underfill Reliability is Here – Finally!”, pp. P-MT1/1-1 – 1-4, Technical Proc. APEX, Long Beach, CA, Mar 12-16, 2000.

Johnson, C. D. and Baldwin, D. F., “Pre-Applied Underfills for Low Cost Flip Chip Processing”, pp. 73 – 76, International Symposium on Advanced Packaging Materials, Chateau Elan, GA, March 14 – 17, 1999.

Gilleo, K., “Direct Chip Interconnect Using Polymer Bonding”, Electronic Connection Conference, (IEEE Proc. CHMT 3-39 8933552), Houston, TX, May 1989.

Gilleo, K. and Blumel, D., “New Generation Underfills Power the 2nd Flip Chip Revolution”, pp. 147 – 154, Proc. Pan Pacific Microelectronics Symposium, Mauna Lani, HI, Feb. 10 – 13, 1998.

Gilleo, K., and Blumel, D., “Transforming Flip Chip into CSP with Reworkable Wafer-Level Underfill”, pp. 159 – 165, Proc. Fourth Annual Pan Pacific Microelectronics Symposium, Kaua’l, HI, Feb. 2 – 5, 1999.

Gilleo, K., “Chip Scale or Flip Scale: The Wrong Question?”, pp. 30, 31 - 34, Circuits Assembly, Feb. 1998.

Ito, S., et al, “A Novel Flip Chip Technology Using Non-Conductive Resin Sheet”, 48th ECTC, Seattle, WA, May 1998, pp. 1047-1051.

Pascarella, N. and Baldwin, D., “Compression Flow Modeling of Underfill Encapsulants for Low Cost Flip Chip Assembly”, 48th ECTC, Seattle, WA, May 1998, pp. 463-470.

Baldwin, D. and Pascarella, N., “Manufacturability of Underfill Processing for Low Cost Flip Chip”, ASME International Congress and Expo., Dallas, TX, November 1997.

Han, S., Wang, K., and Cho, S., “Experimental and Analytical Study on the Flow of Encapsulant During Underfill Encapsulation of Flip-Chips”, Proc. 46th Electronic Components and Technology Conference, Orlando, FL, May 1996, pp. 327-334.

Pascarella, N. and Baldwin, D., “Advanced Encapsulation Processing for Low Cost Electronics Assembly – A Cost Analysis”, 3rd International Symposium and Exhibition

On Advanced Packaging Materials, Processes, Properties, And Interfaces, Braselton, GA, March 1997, pp. 50-53.

Pascarella, N. and Baldwin, D., “Advanced Encapsulation Processing for Low Cost Electronics Assembly – A Cost Analysis”, Advances in Electronic Packaging – ASME INTERpack ’97, Vol. 19-1, June 1997, pp. 359-363.

Lasky, R., Morvan, Y. and Gilleo, K., ”Advances in Materials and Processing in Flip Chip Processing”, Proceedings of InterPACK ’99, EEP-Vol. 26-2, pp. 1527 – 1526, Maui, HI, June 13 – 19, 1999.

Gilleo, K., Cinque, T. and Silva, A., “Flip Chip 1, 2, 3: Bump, Bond and Fill”, pp. 32 – 34, Circuits Assembly, June 1996.