3D Packages and Assembly Methodologies

Bernd K Appelt
ASE Group
1255 E Arques Ave.
Sunnyvale, CA 94085
bernd.appelt@aseus.com

Abstract

3D packaging is coming of age. Initially it was conceived to provide more memory in the same space. Memory die were stacked after thinning the die and bonding them on top of each other using conventional, thin die bond adhesive. The interconnect method remained wire bonding. While at first same die were stacked, the functionality was soon expanded by stacking different type of memory and or controllers in the same package. The explosive growth in mobile products and new applications while shrinking the form factor demanded a new packaging concept: the stacking of packages. The bottom package contains typically the application processor while top package holds memory, often stacked as well. Here, interconnects between packages are solder joints. As the demand for functionality and performance continues to grow relentlessly, bandwidth and electrical performance increase again demand more advanced packaging. It appears that silicon interposers, also called 2.5 D, will be the next type of packaging. The silicon interposer acts as a very high density substrate interconnecting die of different functionality. The interposer still requires an organic substrate as a CTE mismatch mitigator to connect to the motherboard. The ultimate step of the packaging technology evolution will be the real 3D package based on die to die interconnects based through silicon vias (TSV) providing the densest levels of interconnects between heterogeneous die and components.

Key words: 3D stacking, package on package, silicon interposer

Introduction

3D packaging was born out of the need to provide more memory capacity in the same special volume before the wafer technology was able to provide a solution. Memory providers therefore drove the development of wafer thinning in conjunction with die stacking. The first packages were therefore based on stacking two identical die with die bond adhesives and a spacer in an overmolded wire bond PBGA package. This approach quickly evolved into die stacks of higher die count as well as thinner packages with varying number of die per package (see Figure 1). The other evolution was to eliminate spacers by rotating die 90 degrees so that the top die was resting between the two sided peripheral wire bonds. Likewise stacking schemes included pyramidal arrangements so that smaller die were placed on top of the larger die within the wire bond free area. The latest evolution was based on the development of special die attach film adhesives which allowed bonding of the next die over the wire bonds of the bottom die. These film over wire (FOW) adhesives were enabled by decreasing the loop height of the wire bonds to below 40 µ as in Figure 2. FOW eliminates the need for spacers and one layer of adhesive which increases productivity and decreases stack height.

Soon die of different memory type and/or function were combined in the same package to provide more functionality in a package which may be viewed as the beginnings of heterogeneous packaging and system in package (SiP) technology. ITRS defines SiP as any functional package with at least one die and one component which precludes homogeneous memory stacks.

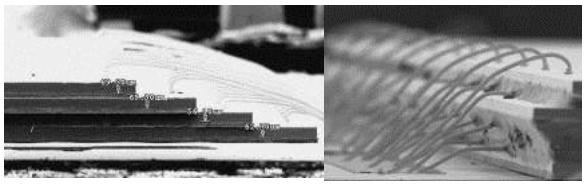


Fig. 1: Examples of stacked die packages: left – pyramidal stack; right – stack with film over wire and staggered die.

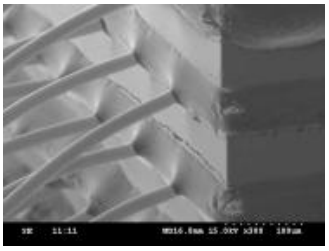
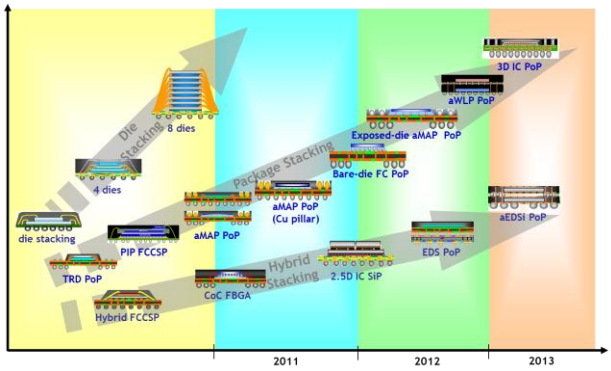


Fig. 2: Stacked die package with film over wire

The next significant step was the development of package stacking. In parallel two concepts were put forth: package on package (PoP) and package in package (PiP). The common feature was that one package contained a processor/controller and the other contained memory which often was stacked as well. Eventually the business model decided in favor of PoP even though PiP enabled a more compact package. Essentially the end customer, e.g. cell phone company, can decide what types of packages to stack at the time of board assembly because the board assembler will stack the two packages on the motherboard and finish the product by mass reflow. Therefore, at time of assembly it is possible to select memory configuration according to phone model type or according to memory (spot) pricing, etc. In the case of PiP, the point of control shifts to the PiP component supplier as the integrator. Of course the lead time for configuration changes becomes longer as well.

A typical roadmap for stacking technology is shown Figure 3. The first branch reflects stacked wire bond die as in memory packaging. The second branch depicts the evolution in PoP and the third addresses hybrid stacking with TSVs.



PoP Challenges and Development

The first generation bottom package had a top molded cap with peripheral ball pads for interconnection exposed on the top side (Figure 5). Ball size of the top package and mold cap thickness were specified to ensure that a small gap between mold cap and top package was formed. This gap was supposed to ensure that the packages did not touch during any time of the assembly process. The size of the balls after reflow collapse then define the achievable interconnect pitch.

Package warpage was traditionally specified as a property at room temperature. It was quickly discovered that for PoP assembly the warpage during reflow was even more important. The warpage behavior of top and bottom packages can be in opposing directions thereby leading to opens as illustrated in Figure 5. Stacking yield turned out to be the major detractor for this new packaging technology. Warpage in concave direction is also termed as 'smiling' and convex direction as 'crying'. Shadow Moire observations as a function of temperature was introduced to characterize warpage behavior during reflow and a new specification was developed to control the stacking yield.

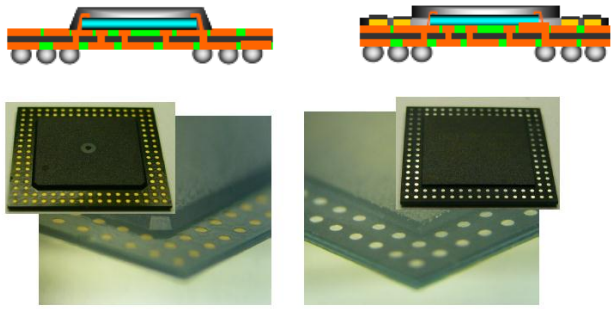


Fig. 4a: Bottom packages of first generation (left) and second generation (right) with molded in solder balls after solder ball reveal by grinding.

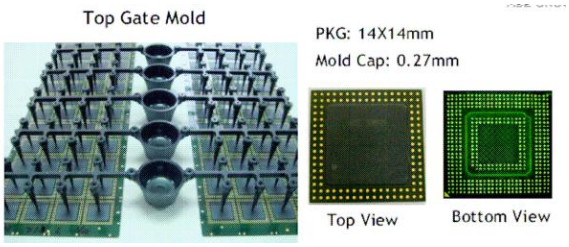


Fig. 4b: Top gate mold of 1st generation PoP and resulting bottom package.

Many solutions were attempted to improve the stacking yields. One approach was to form a cavity on the bottom substrate to lower the die top surface and loop height allowing a thinner mold cap and in turn providing a higher stand-off. The cavity was formed either directly in the laminate or by increasing the solder mask thickness in the ball pad area. These approaches proved to be of limited effectiveness and rather costly. Another approach was to provide solder balls on the bottom substrate as well which in turn required fixturing of the two packages so they would not misalign during the reflow process.

A third type of approach was to attach solder balls to the bottom package interconnect pads prior to molding and overmold die and solder balls simultaneously. This allowed the use of conventional side gate mold chases which are more universal and cost effective. The balls were then exposed by grinding off of mold compound and solder ball tops in the shape of a window frame. This provided now a higher stand-off for the top package even with the use of smaller solder balls and in turn facilitated a smaller interconnection pitch. While these packages die get qualified and move in to production, the market penetration was low.

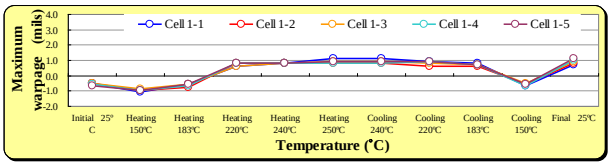


Fig. 5a: Bottom package warpage as a function of temperature in mils for 1st generation PoP.

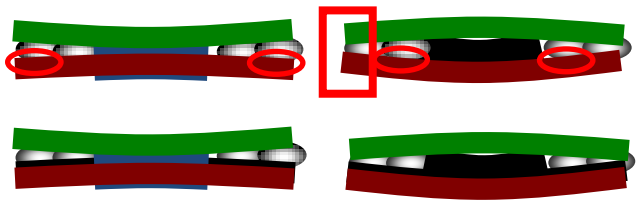


Fig. 5b: Schematic of smiling top package and crying bottom package leading to open interconnects. Top is for 1st generation PoP and bottom for improved 2nd generation package.

Lastly, the current PoP package was developed and is now in high volume manufacturing. It is based on the concept of overmolded solder balls which are accessed via laser vias. Vias can be controlled in registration and depth quite effectively. By joining two solder balls (top and bottom package) the individual ball diameters and therefore pitches can be reduced significantly. I/O density increases corresponding by decreasing the ball pitch from 0.65 mm to 0.5 mm and even down to 0.4 mm for the interconnects. The vias also provide a mechanical alignment during top package placement which the previous method did not have (Figure 6). The most common name for this package type now is through mold via (TMV).

As an aside, there was another version of the previous package. The die was overmolded without any solder balls present. Laser vias were drilled down to the ball pads and then solder balls were inserted and reflowed inside the vias. This process proved to be less effective than the above process but it did coin the name TMV.

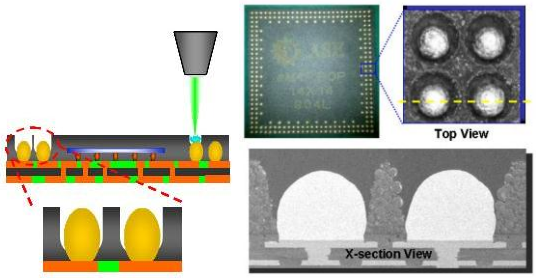


Fig. 6: Schematic of 3rd generation PoP with laser ablation of mold compound and revealed solder balls (left) and actual micrographs (right).

With this new configuration, warpage is controlled more easily (Figure 6) because of the full mold coverage of the bottom package. Care must be taken about the solder ball height uniformity to ensure there is uniform depth in the via to the top of the solder ball.

While early PoP packages were based on wire bond die assembly, the latest offerings are based on flip chip. When underfill is used to secure the solder bumps then the bleed out must be controlled to avoid wicking over the ball pads. This concern can be avoided if the is compatible with molded underfill i.e. die stand-off, pitch and die size must follow published design rules to facilitate sufficient mold flow. Filler size in the mold compound may have to be reduced to enable the mold flow.

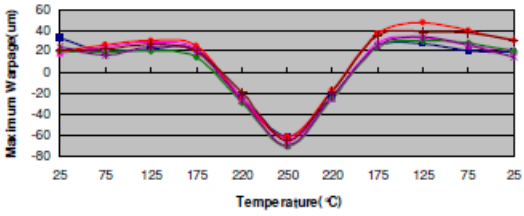


Fig. 6: Bottom package warpage as a function of temperature in micron.

In order to reduce the bottom package height, film assisted molding has been applied to the bottom package. This enables a bare back side flip chip and eliminates any mold compound over the die's backside. In addition to thinning the bottom package this way, it also reduces bottom interconnect ball size i.e. it facilitates reduced interconnect pitches.

Through Silicon Vias

As the demand for higher performance and smaller volume continues, TSV based substrates and die are emerging as the next proposal for a solution.

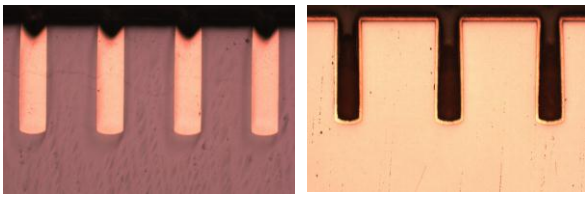


Fig. 7: Cross-sections of Cu filled TSVs (left) and conformally Cu plated TSVs (right)

The concept of TSVs (Figure 7) has been in development for many years, and now, the technology is available to form high aspect vias in silicon wafers. Further, the vias can be insulated and Cu plated to form a solid Cu plug in typical wafer fab formats of and 200 mm and 300 mm. It seems that wafer fabs are ideally suited for producing TSV products since the processes are compatible with current or older node technology. Wafer level packaging (WLP) equipment is also suitable for making TSVs with the addition of few extra processes. So the business model choices by the market will have to determine how the supply chain will develop.

As long as TSVs are used to stack die on each other, the fabs are probably destined to perform the via formation as they can use via first or via middle processes. The via last process seems to be no longer promoted. WLP suppliers can work with the via last process and therefore can also build silicon interposers or silicon TSV substrates. The interposer has TSVs as through connections and redistribution layers on either side (Figure 8). The advantage of the silicon interposer is that the coefficient of thermal expansion (CTE) essentially matches the CTE of the die which is a key reliability requirement for very fragile ELK die. Wafer processing or WLP processing is able to provide high density redistribution layers on either side of the interposer facilitating excellent die to die communication while providing sufficient fan out for external interconnection. Unless the interposers are small in x and y dimension, they will still require an organic substrate to connect to the motherboard in order to overcome the CTE and stress mismatch between interposer and board.

Except for memory stacking with TSVs other die stacks seem to be a few years away. There are many hurdles to overcome to stack TSV dies from different sources as there not any standards yet. Commercial design software, test tools and software, business model, etc. are all still in development. Therefore, the silicon interposer (SiI) seems to be the first product emerging that will have die attached via flip chip technology to the SiI using either micro bumps or micro Cu pillars. A first product announcement by Xilinx® introduces a SiI with four processor die flip chip bonded to the interposer and mounted on a traditional build-up substrate of 5+2+5 layers.

There are many other developments to fabricate interposers, and just as importantly, to assemble die to interposers. Likewise there is significant development in progress to stack TSV based die to each other and to interposers (Figure 9). The common element is that the TSVs which are formed as blind vias need to be exposed to form small Cu pillars or bumps which can then be prepared to assembly to the next interface. This requires the silicon to be back-etched to 'reveal' the vias. There are a many concerns still on handling and therefore wafer shipment of these very thin wafers. To efficiently manufacture vias, the wafers should be thin enough to keep the aspect ratio below ten. Typical via diameters today are larger than five micron and pitches larger than ten micron although smaller dimensions are being developed.

Fig. 8: Schematic of silicon interposer with TSVs (left) and actual cross-section (right)

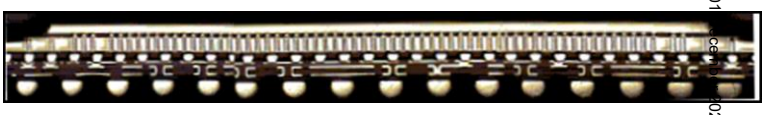


Fig. 9: Silicon interposer with 65 nm ASIC assembled on organic substrate.

The SiI therefore enables heterogeneous integration as depicted in the schematic of Figure 10. All the process and manufacturing tools for this integration now are available and the first volume products are expected to roll out in 2012.

This will be followed then possibly in 2013 with the first heterogeneous die stacked packages

combining application processors with stacked memories but without SiI but possibly a silicon substrate. In this case the silicon substrate may serve as 'conventional' substrate for the die stack to fan out and interconnect to a motherboard or organic substrate. If the silicon substrate is small enough to have a manageable DNP (distance to neutral point in terms of thermal expansion) relative to the motherboard then no organic substrate will be required to mitigate the stress.

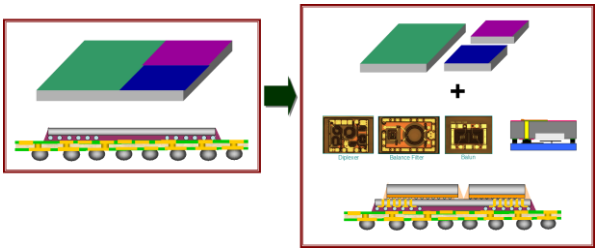


Fig. 10: Schematic of 3D heterogeneous integration.

Summary

3 D integration has advanced from homogenous wire bond die stacking t via package on package stacking to high density silicon interposer stacking. Wiring density and electrical performance have thus reached the highest level of performance to date. This will only be surpassed of direct die to die stacking with through silicon vias.