TSV package architectures and trade-offs

By Paul Silvestri, Rama Alapati, Mike Kelly [Amkor Technology, Inc.]

The proliferation of connected devices is driving system requirements for smaller form factor, higher data transfer rates, improved signal integrity, higher memory bandwidth, and/or increased thermal performance. The problem is technology scaling limitations present a cost barrier to traditional semiconductor die shrinks and device integration using 2D packaging techniques. Through-silicon via (TSV) packaging technology provides a solution and enables homogenous and heterogeneous integration of logic and memory co-located closely together in a small form factor assembly. Concurrently, high-density signal routing and very high bump counts are driving line and space widths with very fine-pitch interconnects using copper pillar technology on interposers.

Choices for TSV

Combined with the appropriate wafer bumping technology and advanced packaging, TSV interconnects enable very high-density I/O so application-specific integrated circuits (ASIC) and memory die can be located near each other on a TSV silicon interposer. This integration scheme is called 2.5D TSV. Additionally, silicon layers can be stacked tier-to-tier on top of each other, which reduces the physical area allocated for each component. This tier-to-tier stacking is called 3D TSV technology. In the dynamic random access memory (DRAM) space, 3D TSV has been deployed in both high-bandwidth memory (HBM) and 3DS-DDR 4. By combining the two TSV technologies into a flip-chip ball grid array (FCBGA) assembly, systems with high-bandwidth and low-power per bit transferred and high memory density are realized.

A 2.5D assembly process can be achieved through either wafer-level or substrate-level integration. The decision to use one method or the other requires a comparison of the process flow. In general, architects should trade off test coverage vs. repair capability at the ASIC design level. Because automated test equipment (ATE) platforms are much more flexible for substrate technology rather than wafer technology, assembly integration that starts at the substrate level (chip-on-substrate, or CoS) will readily enable flexible test solutions early in the assembly flow.

In contrast, an assembly methodology that begins with a wafer (chip-on-wafer, or CoW) will require full population of top die and completion of the wafer processing plus assembly to a printed circuit board (PCB) before suitable testing can begin. This "dies first" strategy of placing the top die (logic + logic or logic + memory) to the interposer while in wafer form (CoW) can become a very high-cost process if redundancy methods and post-assembly die repair schemes are not deployed in the architecture to manage yield loss. Assembly methods such as chip-on-substrate and hybrid forms of CoW + CoS (partial chip-on-wafer flows) are very attractive solutions where die yield, cost, and test strategy are crucial to a device's economic viability.

TSV assembly technologies

To address the changing market requirements, multiple 2.5D TSV technologies have been developed and commercialized. At Amkor, a 2.5D TSV CoS process has been in production for several years and the CoW process is being qualified in 2016. The process flow details of each of these approaches are shown in **Figure 1** and **Figure 2**. The TSV process flow starts with middle-end-of-line (MEOL) processing to expose the TSVs and metallize the front and back of the wafers to form the interconnects.

Figure 1, **2** and **3** show three major assembly techniques. The choice of the appropriate TSV technology to use depends on the size of the interposer and number of component dies attached to it. Designers need to understand the nuances between these

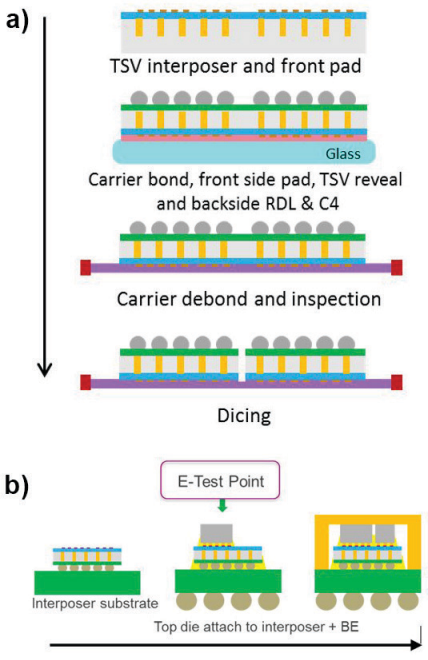

Figure 1: a) The MEOL process flow; and b) assembly process flow with interim test for CoS packaging.

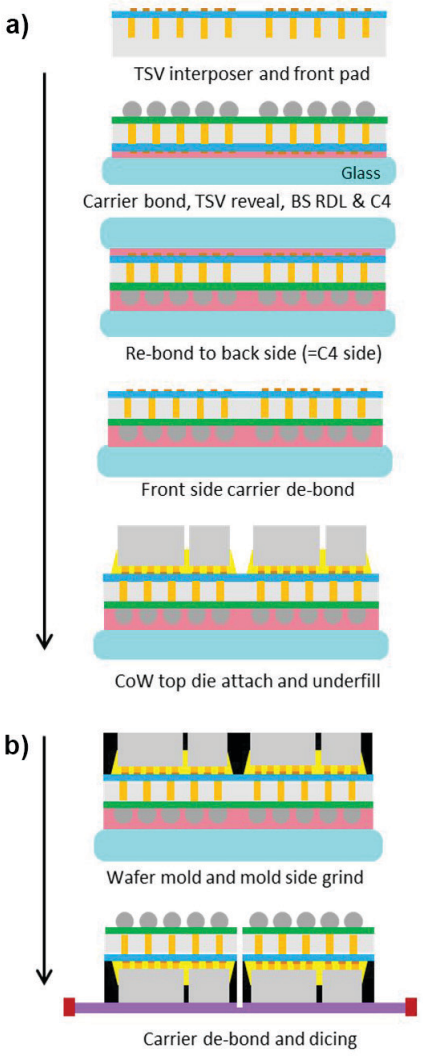

Figure 2: a) MEOL process flow and die attach; and b) wafer finish option of molded wafer format for CoW packaging.

approaches including the advantages and disadvantages of each. In most cases, TSV processing whether it is CoS, CoW or even 3D TSV for memory has essentially the same basic approach on the wafer side: the foundry processes the TSVs that do not go all the way through. The outsourced semiconductor and test supplier (OSAT) receives the wafer with the TSVs and then processes both the front and back side to expose the TSVs and bump the wafers. **Figure 1a** and **Figure 2a** show the MEOL processing. The CoW approach adds a carrier flip (**Figure 2b**) to present the interposer front-side for top die assembly.

Interposers for 2.5D are designed to enable side-by-side die interconnects. As a result, interposer die sizes are large, sometimes larger than the reticle size. As such, 2.5D TSVs are deeper in the silicon, generally 100µm deep. During the MEOL process, more bulk silicon will remain, which increases the die resistance to warpage.

In contrast, 3D TSV integration is generally integrated in DRAM with small die area. The 3D TSVs are integrated at 50µm deep, making the die more sensitive to warpage. HBM is built using CoW molded style assembly technology, which does limit the test coverage available because the memory cube can only be tested while in wafer form. Because HBM test coverage is reduced, it is desirable for 2.5D assembly integration to consider assembly methods that can enable test before HBM cubes are committed to the assembly.

With a CoW molded style assembly, all interposer die sites must be populated before wafer molding. The problem with wafer-level testing is it can only be performed at wafer test sites that are typically limited to how fast the vectors can be run, and parallelism, and even temperature can be problematic. In contrast, on a substrate using a handler and some form of automation, much higher volume testing can occur with much better test vectors, more I/O on the test board, and more. This provides much more efficient testing at the substrate level. For many customers, this is the right approach today.

The flow that is the most flexible and low cost is CoS because the location in assembly where the top die has to be committed is late in the flow. **Figure 1b** shows the E test point where acceptable or unacceptable product is identified. A multiple die design frequently has two, three, four, or five dice that sit on top of a very large interposer. The die can be very expensive if it uses 7 or even 14 or 16nm technology, or even the HBM device itself. If there is an issue with the assembly, committing the die in a CoW process without testing them can potentially result in the loss of many other die. This is where the avoidable expense occurs.

CoS allows an OSAT supplier to perform interim tests prior to committing the other top die that could be a logic layer or memory. With the current cost of HBM, there is a lot of interest in placing die on a tested substrate. However, the thinned interposer will assume the substrate warpage characteristics once assembly to the PCB is completed, making assembly of the top die to the interposer difficult if warpage is not managed. CoS assembly can be a suitable solution for interposer designs up to the reticle size. Because of the coupling of the interposer and PCB, the PCB design and materials selection must be carefully selected.

In other cases, CoW can offer a desirable solution. Again, there are advantages and disadvantages. In the CoW approach shown in **Figures 2a** and **2b**, the difference occurs when the wafer is flipped (**Figure 2b** bottom) to expose the front side of the interposer to place the die. Because the wafer warpage is managed by the carrier, the front side of the interposer wafer remains flat and subsequent die placement is not sensitive to warpage. This enables very large interposer designs with fewer constraints on PCB selection. During wafer assembly, all top die must be populated before wafer molding starts.

CoW process flexibility is shown in **Figure 2b** with a mold and **Figure 3a** with a non-mold process flow. **Figure 3b** shows a side-by-side comparison of the module assemblies. Typically, the industry is pursuing the molded process flow. In the over-mold flow, everything is locked in place after molding. After assembly and wafer molding, the module is mechanically diced from the interposer wafer and placed on the PCB. Test can then be performed at that point to check functionality and perform any post-package repair. A defect that is not repairable or recoverable through design redundancy (redundant TSVs) or other external means (such as remapping I/O) is lost and a very large and expensive top die (ASIC and HBM) is scrapped.

The mold compound's mechanical properties dictate the use of less mold material to reduce stresses on the interposer die and the adjacent top die. This is accomplished by keeping the die-to-die spacing very tight (<100µm) and minimizing all unused interposer die area that would be exposed to the mold compound. The resulting design rule requires any top die (ASIC and/or memory) to be similar in size such that the interposer floor plan to minimize the interposer size. CoW molded is a good platform for logic + logic integration where the logic dies are the same size. Additionally, the wafer dicing step must use mechanical dicing methodology in order to accurately dice through the mold and silicon interposer. This requires a sufficiently large dicing lane to allow for singulation. FPGAs are an excellent example of CoW-molded integration.

In the CoW no-mold path, an interim test flow (**Figure 3c**, middle) is supported. Die size alignment is not design rule-limited because mold compound is not used in this platform. Furthermore, stealth dicing of the interposer is possible, which could reduce the interposer dicing street width. Because the wafer will not be over-molded, every die site on the interposer does not need to be populated. This feature allows for interim test options, either at the wafer level or at substrate level. In this process, the partially populated wafer is put onto a substrate so it can be tested similar to a CoS and then the remaining dice are added later. This hybrid chip-on-wafer process is an attractive flow for heterogeneous 2.5D top die integration that does not limit the top die size synchronization. It is a combination of CoW and CoS.

With the variations discussed above, there are interesting tradeoffs. CoS provides the

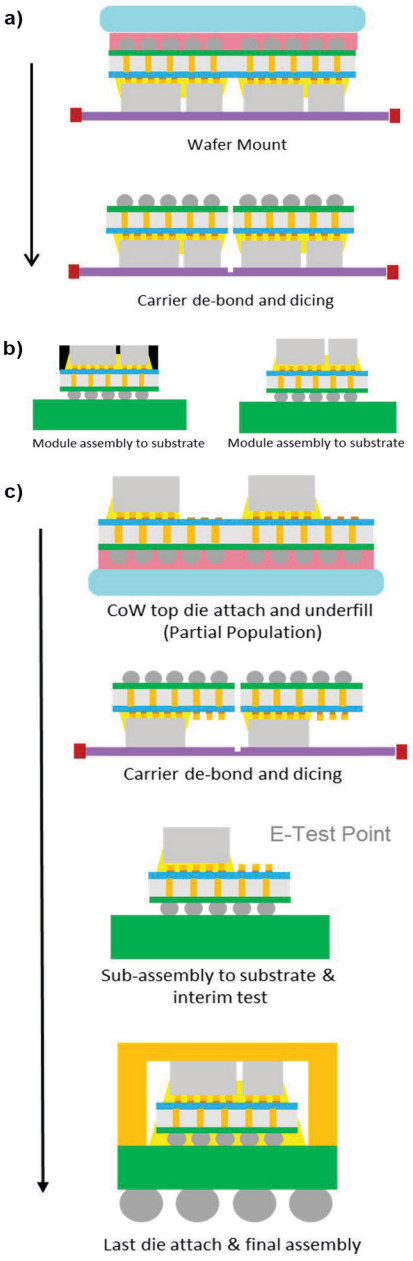

Figure 3: a) Wafer finish option of no-mold wafer for CoW; b) CoW module assembly – molded and no-mold module; c) CoW + CoS hybrid flow with eTest.

TSV Assembly Process	Interim Test	Interposer Size Support	Die to Die Spacing	Interposer Whitespace	Interposer Dicing	Application
Chip on Substrate (CoS)	Yes	< Reticle Size	Largest for Interim Test; Variable OK Otherwise	Top Die can be any size	Stealth	Logic + Logic; Logic + Memory
Chip on Wafer (CoW) Molded	No	> Reticle Size	Minimum required	Top Die should be balanced	Mechanical	Logic + Logic
Chip on Wafer (CoW) Non Molded	No	> Reticle Size	Variable	Top Die can be any size	Stealth	Logic + Logic; Logic + Memory
Chip on Wafer (CoW) Hybrid	Yes	> Reticle Size	Largest for Interim Test; Variable OK Otherwise	Top Die can be any size	Stealth	Logic + Logic; Logic + Memory

Table 1: TSV processes and attributes comparison.

most flexibility. The CoW molded flow has the least flexibility. However, there are two process flows between these extremes. One is the CoW no mold. While it is not as aggressive, it can be modified to perform the hybrid process that combines the best of CoS and CoW.

Table 1 shows a quick high-level summary of several high-level design considerations. There are more criteria depending on the application and specific customer constraints, such as improved system performance, or high performance at relatively low energy dissipation as measured by bandwidth per unit of energy expended, thermal performance, or others.

TSV solutions

With TSVs commonly accepted as a solution for advanced IC packaging and several TVS processes available, today's design decisions focus on identifying the right implementation for specific applications. To make the correct choice, there are several design considerations and tradeoffs. Today's 2.5D TSV solutions enable new power and form factor efficient systems. As a result, many leading original equipment manufacturers (OEM) of semiconductors have taken advantage of their flexibility and ability to achieve performance and cost goals. While 2.5D technology has matured, significant effort is underway to scale the technology to enable faster and more efficient systems. At the same time, much of the effort for 2.5D is applicable to 3D packages that will provide even further performance and system advantages. With these and other process variations, different customers have the flexibility to choose the process that meets their specific needs.

Biographies

Paul Silvestri received his BS degree in Electrical Engineering from Florida Atlantic U. and is Director, TSV Product Development, at Amkor Technology, Inc.; email Paul.Silvestri@amkor.com

Rama Alapati received his BS degree in Chemical Technology from Osmania U., Hyderabad, India, and a Master's in Chemical Engineering from the U. of Kansas, Lawrence, and is VP, Strategic Development, at Amkor Technology, Inc.

Mike Kelly received his Master's degrees in Mechanical and Chemical Engineering and is Sr. Director, Advanced Package & Technology Integration, at Amkor Technology, Inc.