Semiconductor Technology from A to Z

Advanced Packaging / 3D Integration

3D Stacking

Die-to-Die vs. Wafer-to-Wafer Bonding

While 2.5D integration arranges dies side by side, true 3D stacking places several dies directly on top of one another and connects them vertically via TSVs. Two fundamental manufacturing approaches are available:

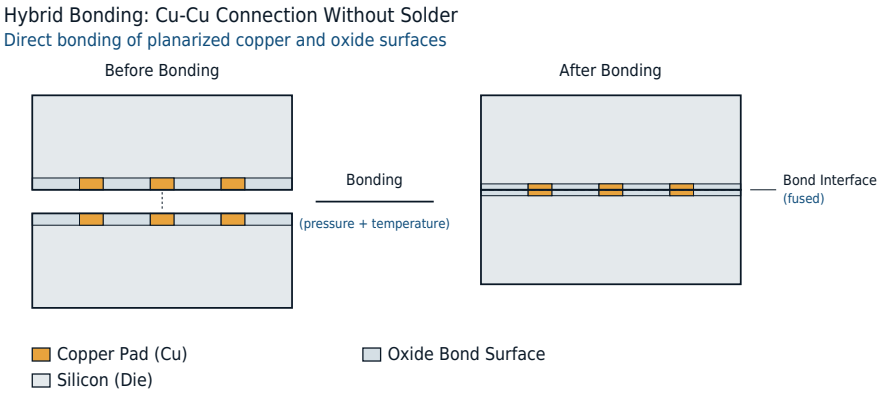
- Die-to-die (D2D): individual dies that have already been singulated and tested are stacked on top of each other. Advantage: before stacking, each die can be individually verified as functional (“known good die”), which significantly improves the overall yield of the stack.
- Wafer-to-wafer (W2W): two complete, unsingulated wafers are bonded together as a whole and only singulated afterward. Advantage: significantly higher throughput and placement accuracy; disadvantage: a defective die on either wafer causes the entire stacked device at that position to fail, since no pre-selection is possible.

A hybrid approach, die-to-wafer (D2W), combines the advantages of both methods: tested individual dies are stacked onto a complete wafer.

Hybrid Bonding: Direct Cu-Cu

The densest form of 3D interconnect currently available is hybrid bonding. Here, both surfaces to be joined have copper areas (for the electrical connection) and oxide areas (for mechanical bonding and isolation) planarized with high precision at the same time – using chemical-mechanical polishing, see the corresponding chapter – and are then joined directly, without any additional solder or bump material. Under pressure and moderate temperature, the oxide surfaces first fuse chemically; the copper then expands thermally and forms a direct metallic connection. This makes it possible to achieve interconnect pitches below one micrometer – an order of magnitude unreachable with classical solder or bump processes.

Hybrid bonding before and after the bonding process



Thermal and Mechanical Challenges

Stacked dies generate heat during operation that must be conducted away through several active layers – a die in the middle of a stack has a significantly worse thermal path to the heat sink than a single die sitting flat on a substrate. In addition, mechanical stress caused by differing thermal expansion coefficients of the stacked materials increases with every additional layer. Both factors limit, in practice, how many active, high-power logic dies can reasonably be stacked on top of each other, while passive or low-power memory layers (as in HBM) are far less critical.