Could copper pillar change the ecosystem?

The future of bumping is copper, and that could change everything. Though originally developed at the IDMs, players from foundries to OSATs are now developing a variety of differing copper pillar and micro-bumping solutions for a range of customers, looking re-designing the substrate, weighing extending reflow vs scaling thermal compression bonding, and developing probe and test systems.

"For copper pillar bumping, the industry will have to solve similar challenges as it did for copper wire bonding—the materials are different, the substrates are different, and the designs are different. Moving beyond solder will require a whole new ecosystem, and will require working closely with our customers and with our materials and equipment partners," says Bill Chen, ASE Fellow and Senior Technical Advisor.

"Copper is a node change for bumping," concurs David McCann, VP of Packaging R&D at GLOBALFOUNDRIES, pointing out that this node should be extendible more than the usual 7 years, all the way down through 10μm pitch connections and below. "We see copper as a core competency for several generations," he adds, noting that the company is investing significantly in people and projects to develop copper expertise, counting on the same core knowledge and tools to extend to beyond copper pillar and micro-bumping down to copper-to-copper bumpless bonding below 14μm. GLOBALFOUNDRIES is developing Cu pillar technology in its Dresden facility and will be adding a bump facility, focusing on Cu pillar production at its Fab 8 facility in New York. The New York facility will enable interconnect and package R&D co-located with its front-end development teams for advanced nodes. Development in Dresden focuses on 100-110μm pitch bumps for 28nm, and 40μm/80μm pitch for 20nm.

STATS ChipPAC is currently doing staggered 40μm/80μm pitch Cu columns, with 20-30μm bumps on 80μm pitch centers for 28nm mobile processors, reports Raj Pendse, VP and Chief Marketing Officer. Similar technology should be extendible to 20nm processors, he suggests, as the relatively low I/O density of the power and ground in the center of the processor die means there will likely be enough room to just add more rows of the same 80μm pitch columns.

Though copper bumping has been around for years, higher density devices mean growing demand for higher density interconnect than the ~130-140μm pitch that can be done with solder bumps. As mobile application or baseband processors increasingly require the high density of copper pillars for small size at 28nm and below, and memory soon starts to go to wide I/O and DDR4 stacked with processors needing higher pin counts and reduced parasitics, manufacturers are focusing on scaling copper technology to tighter pitches, higher volumes, and of course lower costs.

"For copper pillar bumping, the industry will have to solve similar challenges as it did for copper wire bonding—the materials are different, the substrates are different, and the designs are different. Moving beyond solder will require a whole new ecosystem, and will require working closely with our customers and with our materials and equipment partners," says Bill Chen, ASE Fellow and Senior Technical Advisor.

"Copper is a node change for bumping," concurs David McCann, VP of Packaging R&D at GLOBALFOUNDRIES, pointing out that this node should be extendible more than the usual 7 years, all the way down through 10μm pitch connections and below. "We see copper as a core competency for several generations," he adds, noting that the company is investing significantly in people and projects to develop copper expertise, counting on the same core knowledge and tools to extend to beyond copper pillar and micro-bumping down to copper-to-copper bumpless bonding below 14μm. GLOBALFOUNDRIES is developing Cu pillar technology in its Dresden facility and will be adding a bump facility, focusing on Cu pillar production at its Fab 8 facility in New York. The New York facility will enable interconnect and package R&D co-located with its front-end development teams for advanced nodes. Development in Dresden focuses on 100-110μm pitch bumps for 28nm, and 40μm/80μm pitch for 20nm.

STATS ChipPAC is currently doing staggered 40μm/80μm pitch Cu columns, with 20-30μm bumps on 80μm pitch centers for 28nm mobile processors, reports Raj Pendse, VP and Chief Marketing Officer. Similar technology should be extendible to 20nm processors, he suggests, as the relatively low I/O density of the power and ground in the center of the processor die means there will likely be enough room to just add more rows of the same 80μm pitch columns.
Copper pillar bumping means a new era for the OSATs, says Chen, noting that the earlier solder bumping technologies were developed at the IDMs like IBM and Intel, and the OSATs' job was simply to do just what the IDMs had done. But this time the OSATs are developing their own versions of the technology with their customers. “This generation of copper pillar that started with TI follows a general pattern of laying down an under-bump metallization layer, then plating over it and using some kind of photo resist, but after that there is plenty of room for people to do things differently,” he notes.

Foundries are also developing their own solutions with their customers. “Some foundries want to control the entire supply chain but we want to enable our customers’ supply chain choices. We also need the fast feedback of probe results to the fab drive early yields,” says McCann. “But we don’t want to do production packaging. We want to drive our logic sales, and we want to enable the OSATs’ business.” While the leading edge work at future nodes will likely start at the foundry, high volume readiness will be verified at the OSAT, and bump and probe will move to the OSATs for mature nodes and devices. The need for co-development of BEOL stacks, interconnect and assembly will necessitate close collaboration on assembly. At future nodes, however, interconnect and packaging processes that happen at the wafer level will require more fab-like processes and investment in traditional CVD, PVD, and CMP fab tools. “Both foundry and OSAT will likely support middle-of-the-line processing, and there may be a merging of business models in the future,” he suggests.

Pendse concurs that OSATs will take a larger role as the technology matures, as dividing responsibility for yield between the fab and OSAT when transferring bumped wafers will become easier as bumping reaches consistently high yields, and can be more easily handled by agreements on expected performance.

“At 20-14nm, it’s not clear what kind of pitch will be required,” says Dr Raj Pendse, STATS ChipPAC.

Foundries may also revisit if the lower-margin bumping business makes sense for them as the technology matures. And the chip stacking technology may eventually move away from the front-end fabs’ area of expertise, if it moves away from using silicon wafers for re-distribution or interposer substrates and instead finds ways to use the lower cost LCD infrastructure for fan-out WLP, or to use panel or glass 2.5D interposers.

Bringing down costs by re-thinking substrate design

Plenty of technical issues remain, however, to move copper pillar and micro-bumping to high volume production at smaller geometries and lower costs. Finding ways to reduce the costs of the denser and denser substrates is one key path.

Bumping interconnect technology roadmap for FC BGA

(Flip Chip report, Yole Développement, March 2013)
Moving from bond pads to the finer traces of the bond-on-lead approach has helped bring down costs initially by allowing more efficient routing, to reduce the number of layers required to reduce the significant cost of the substrate. Pendse notes that this may eliminate the need for the solder mask for solder confinement, and the associated tight tolerances on solder mask alignment. The use of copper column bumps, with the greater stand-off heights achievable, also makes it easier to use mold instead of capillary underfill.

Re-thinking substrate design will continue to be the main path to bring down costs, argues Chen.”The idea of a pillar gives us the opportunity to significantly re-think what flip chip means, and allows us to re-think everything we’re doing to design solutions for each application. It gives the design a lot of freedom.” He notes that by working closely with chip designers the copper pillar substrate can be simplified from expensive build up board down to lower cost four layer or all subtractive substrates.

Or perhaps the industry may need to re-think the entire approach of routing and attaching ever denser die I/Os on ever denser multilayer substrates. Pendse suggests that at some point this no longer makes sense, since new generations of die quickly come down in cost, but the substrate and packaging costs remain high even in maturity. "At 20μm/40μm you start to pay a really high premium [for the high density substrate]," he notes. "We need to move to a new technology to get around this. We think it will be fan-out wafer level packaging, which can give a 2X jump in density." He suggests moving to more scalable chip-like solutions for the routing could be a cheaper alternative at some point.

Instead of attaching the die to a ready-made substrate, the packaging house would redistribute the dies across a 300mm silicon wafer, then build thin film redistribution layers of metal and dielectric on top to carry the signal from the die pad to the BGA pad, an approach that could potentially make 10μm traces at lower cost. "For these sorts of I/O densities we can’t afford to be pushing the leading edge of substrate technology. We should move to a different technology that’s more in the sweet spot for the required geometries," he argues. Though FOWLP is in high volume production now using wafer-like carriers, it could be scaled several fold by changing the carrier format. "We’re now using 300mm wafers as the carriers, but we don’t need to use the silicon infrastructure for 10μm lines," he notes. "We could use large area substrates like 2G LCD panels and tools from the LCD industry."

Reflow or thermal compression bonding?

Finding ways to extend the use of mass reflow attachment to tighter pitch copper bumps could also help hold down costs, perhaps putting off the need to convert to more expensive thermal compression bonding. Bond-on-lead interconnection allows reflow on current products, but finer pitches may eventually push the industry to thermal compression bonding, as reflow placement and accuracy may not be sufficient. "At 20-14nm, it’s not clear what kind of pitch will be required," says Pendse, suggesting that 14nm devices might need 30μm/60μm pitch bumping with 15μm lines and spaces for escape routing in 2-3 years, which could be still doable with reflow, or might need thermal compression bonding. But thermal compression bonding will likely be needed for 2.5D/3D applications at <40μm pitch.

McCann suggests pitches below 100μm will likely need thermal compression bonding with pre-dispensed underfill, and work is also progressing in the industry to bring down the cost of that technology as well. But he notes that mass reflow technology is also improving, with narrow substrate lead traces instead of pads allowing narrower and more flexible interconnect, reducing the strain on the solder joints and stress on the silicon from thermal mismatch, while substrates with lower coefficients of expansion and more consistent metal balancing will also help make mass reflow more reliable at smaller pitches.

Much of the industry is using thermal compression mainly because TI used it initially, suggests Chen, noting that both reflow and compression bonding processes will likely be used by different players for different products. "Reflow will need to be optimized, but thermal compression bonding has still had very limited use in high volume production," he notes. "Reflow will be extended further than you might think. We believe it will be the lower cost solution."

Probe and test, wet etching, re-work issues also

Just how to test these finer structures also remains an issue. "The ability to test is as important as the ability to bump," says McCann. "100μm pitch is the current limit of vertical probe." Potential probe solutions could be MEMS cantilever deflecting beams, released at one end for soft contact, which exist but...
remain expensive. Another solution might be for vertical contact on a pad next to the Cu pillar. Some are using an area-array, membrane-like probe solution.

McCann also notes that the industry may need to go to dry etch at finer geometries. At 20μm copper pillar, undercutting becomes an issue by wet etch. Etching to remove the blanket UBM attacks the edge of the bump and undercuts it, and with smaller pillars the undercut becomes an increasing percentage of the smaller cross section, so GLOBALFOUNDRIES is developing a dry etch process.

Chen is also concerned about re-work. “You can re-work solder, but there’s no solution for copper pillar. It’s possible, but people just haven’t looked at it yet.” He also warns to not count out competition from older technologies. “Copper wire bonding will also be competition,” he notes. “We used to think 1 mil would be its limit, but now it’s looking at 0.5mil.” He notes that copper to copper bumpless bonding remains in the research labs, with nothing demonstrated yet for high volume production.

Paula Doe for Yole Développement