

Small Pitch Micro-Bumping and Experimental Investigation for Under Filling 3D Stacking

Antonio La Manna, K. J. Rebibis, J. De Vos, L. Bogaerts, C. Gerets and E. Beyne

IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

Tel: +32 16 28 199, Fax: +32 16 281 097; Antonio.lamanna@imec.be

ABSTRACT

Although 3D-IC integration is largely acknowledged as most promising technology to push further the scaling concept of Moore's law, a few important uncertainties are delaying high volume production. Among these uncertainties, the reliability of 3D-IC stacked components is the most important. A key element in 3D stacking reliability is the choice of the correct UF to integrate in the assembly processing: die stacking and final packaging. In this paper we report the recent results coming from the UF screening and 3D stacking activities. These activities are performed at Imec in the frame of the 3D Imec Industrial Affiliation Program (3D-IIAP).

After describing the test vehicle used in this work, we introduce some of the requirements in term of process improvement and stacking accuracy to enable 3D stacking in case of small pitch micro-bumps. We then present recent results coming from the UF screening activity and report on the electrical characterization of the 3D stacks. This characterization is done by measuring the electrical resistance of the daisy chains connections between top/bottom dies bumps. The stacks are done with thermo-compression bonding and using die-to-die approach. The stacks are lately exposed to thermo-cycling processing followed by further characterization.

Key words: underfill, 3D stacking, μ bumps.

INTRODUCTION

By 3D-IC Integration it is possible to put more transistors on the same footprint without the need to shrink transistor sizes. Anyway, as for any new technology, there are many challenges and issues that need to be addressed before moving to high volume manufacturing. A key element for 3D-IC integration is the process of stacking the IC components on top of each other.

To realize 3D stacking the bump pitches and die-substrate gap are scaled down to as low as possible, this results in increasing assembly complexity and making more evident the limits of standard processes like capillary UF (Under Fill). In view of this increased complexity, alternative under filling process like No Flow UF (NUF) or Non Conductive Paste (NCP) and Wafer Level UF (WLUF) are investigated [1,4].

In this paper we focus on die-to-die assembly and report most recent results obtained by screening several UFs for 3D stacking.

The UF materials are initially screened by considering processing criteria like thickness uniformity, voiding/delamination, and curing time.

In a second phase, selected materials are used to generate 3D stacks using a specific test vehicle. In this phase the screening criteria are electrical measurements (to check 3D connections), SAM (Scanning acoustic Microscope, to check voids and delamination), X-SEM (Cross-section, Scanning Electron Microscope, to check material entrapment and stacking misalignment). The work presented is a follow up of the activities already anticipated at IMAPS 2011 conference [1].

TEST VEHICLE DESCRIPTION

The test vehicle used to generate 3D stacks consists in two different dies. These dies are stacked on top of each other using die-to-die thermo-compression bonding. The two dies are named: **PTCM** (Packaging Test Chip version **M**) and **PTCN** (Packaging Test Chip version **N**).

During the stacking process, the PTCM (5,1 mm x 5,1mm) is typically used as top die, while PTCN (8,1mm x 8,1mm) is used as bottom die.

Both dies are processed using 130nm CMOS technology with 2 metal layers BEOL (Back End of Line). The test vehicle has two variations based on μ bumps configuration: periphery array (40 μ m pitch) and full area array (50 μ m pitch).

The same test vehicle may also allow 20 μ m and 15 μ m bump pitches plating. These latter configurations are not considered for 3D stacking characterization but just for bump development.

The stack PTCM+PTCN (refer to Fig. 1) is electrically characterized by measuring the resistance of the Daisy Chains (DC) structures. These structures are accessible on the bottom die (larger than top die).

The number of chains is different in case of periphery array (40 μ m pitch) and full area array (50 μ m pitch).

In case of periphery array the total number of bumps is 480, while in case of full area array the number of bumps is 8250.

The target bumps metallurgies are Cu (bottom die) and Cu(Ni)Sn (top die).

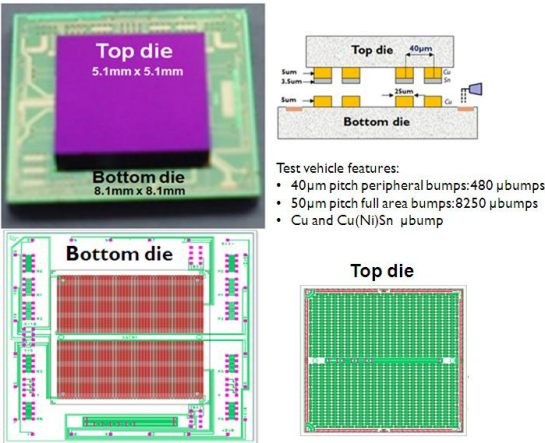


Fig. 1: Lay-out and characteristics of used test vehicle

SMALL PITCH MICRO-BUMPS

The need for more integration is directly linked to the reduction of the IOs pitch. In the case of 3D stacking, the IOs are mainly associated to the interconnection bumps, therefore increasing the number of IOs require a scaling down of bumps dimension and pitch. This scaling adds at the same time two major requirements: A) processing improvement for reliable micro-bumps and B) major stacking accuracy (X-Y and Z axis).

A. Processing improvement for reliable micro-bumps

By scaling the bump pitch from hundreds of microns to a few tens and below is not straightforward. Several parameters need to be taken into account to allow a feasible and reliable bumping process. In this paragraph we will report some of the challenges faced during micro-bumping process development. These challenges are mainly related to the higher aspect ratio of the µbump structures, reduced bump diameter and spacing between landing pads; plating uniformity, and seed etch optimization to enable minimal bump undercut.

A few parameters that will be described in this part are: selection and treatment of the plating resist, plating process optimization and seed etch processing. Considering that in our micro-bump scheme, we target bump heights of around 5µm and 10µm, we have selected a plating resist with a nominal thickness of 14µm. This resist offers good performance in case of larger bump pitch (40µm and 50µm) but present a few challenges when reducing bump pitch. In fact, the reduction of bump pitch required the reduction of bump spacing and bump diameter. This translates in high aspect ratios for the resist opening and bumps spacing. In case of 20µm bump pitch (refer to Fig. 2), the selected resist needs to enable an aspect ratio ~2 for both opening and spacing.

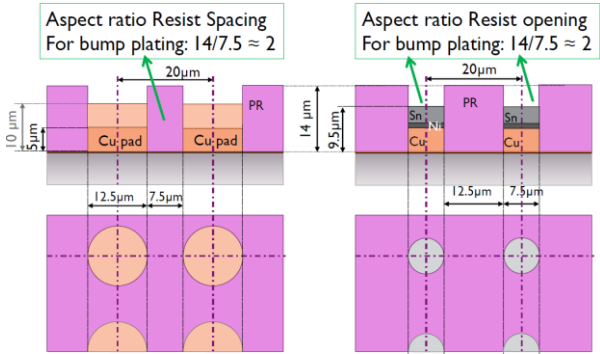


Fig. 2: Aspect ratio requirements for plating resist in case of 20 µm bump pitch.

To achieve these requirements, new process steps and optimization are necessary. To enable the micro-bumping we have introduced new resist pre-treatment steps to improve the adhesion to metal seed, resist wetting during plating and control seed oxidation. The process optimization has been obtained by improving resist adhesion and plating. The results have been validated by Scanning Electron Microscope (SEM) and shear test.

The SEM images in case of initial and optimized processing are illustrated in figure 3 and figure 4.

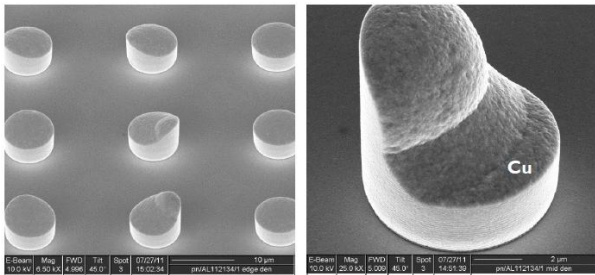


Fig. 3: Micro-bumps plated with initial processing.

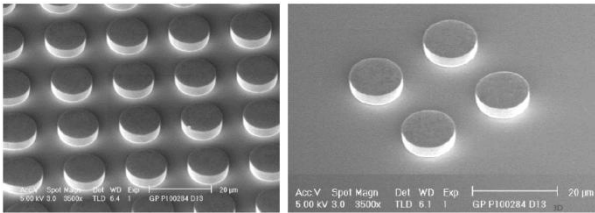


Fig. 4: Micro-bumps plated after optimized processing.

The table 1 illustrates the shear test values measured for a few conditions.

Condition	avg shear force (mg/µm²)
Condition 1	36.3
Condition 2	27.5
Condition 3	40.2
Condition 4	38.7

Table 1: Different resist pre-treatment conditions and bump shear strength after plating in case of 40µm bump pitch and 25µm bump diameter.

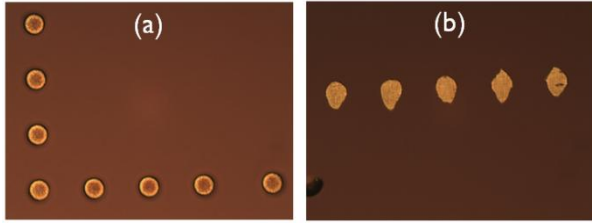


Fig. 5: Micro-bumps before (a) and after (b) shear test in case of 40μm bump pitch and 25μm bump diameter.

The target shear force in case of Cu bumps is 7.5mgf/μm² (shear test done at 46% of Cu height). The results in table 1 show that all conditions are fulfilling the spec.

Similar conditions have been also tried in case of smaller bump pitch (20μm) and diameter (7.5μm). In this case the observed failure mechanism was bumps delamination instead of bumps shearing. The root cause for this failure is mainly due to the bump undercut that has higher impact in case of smaller bumps (refer to figure 6).

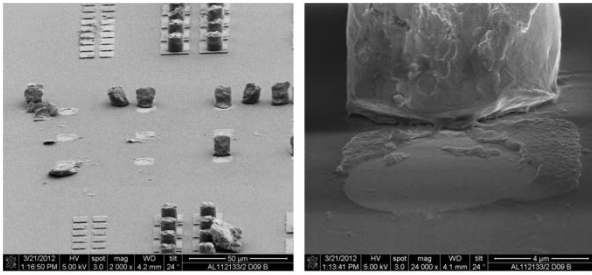


Fig. 6: Micro-bumps after shear test in case of 20μm bump pitch and 7.5μm bump diameter: bump delamination is the main failure mechanism in this case.

B. Stacking accuracy for small pitch micro-bumps

3D stacking in case of small bump pitch, not only requires good bumping processing but also higher staking accuracy. The requirement for higher accuracy could be enabled by investing in high accuracy stacking equipments or/and by modifying the bump design to accommodated lower accuracy [5].

The first option has a clear cost implication and therefore is not the preferred option. To enable better alignment by design, the key parameters to be considered are (refer to Fig. 7):

- Solder diameter: $\varnothing_b$
- Landing pad diameter: $\varnothing_p$
- Pad spacing: S
- μbump pitch: $P = S + \varnothing_p$

The maximum acceptable misalignment (X-Y axis) is determined by the minimum of $(\varnothing_p + \varnothing_b)/2$ in case of OPEN and $(\varnothing_p - \varnothing_b + 2S)/2$ in case of SHORT.

Anyway, to achieve good electrical connections, a minimal contact area between the bumps is also required. The figure 8 shows possible scenarios for acceptable contact areas.

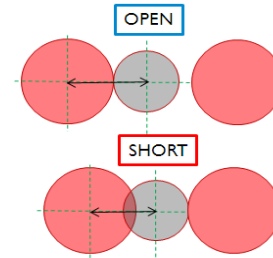
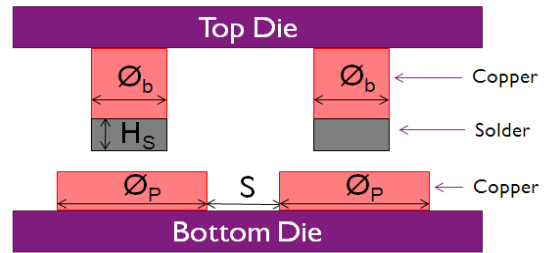


Fig. 7: Bump dimension parameters

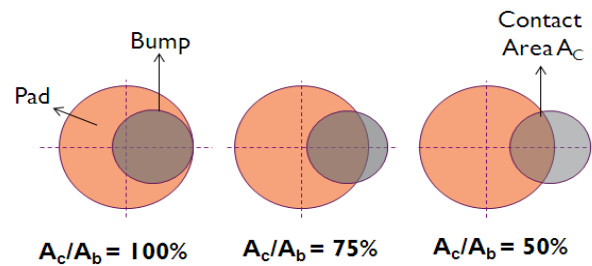


Fig. 8: Options for acceptable contact area (A_c =Contact area, A_b =bump area)

By geometrical considerations, a bump scheme that offers better margin for alignment error is where the top die bumps diameter is smaller than bottom pads.

As examples, for $A_c/A_b=50\%$ (~contact area is 50% of bump area) and a bump pitch of 20μm: the max alignment error is 4μm if bump and pad have same size (10μm diameter). The alignment error can increase to 6μm in case of smaller top bump diameter (7.5μm) versus a landing pad of 10μm diameter (figure 9 and figure 10).

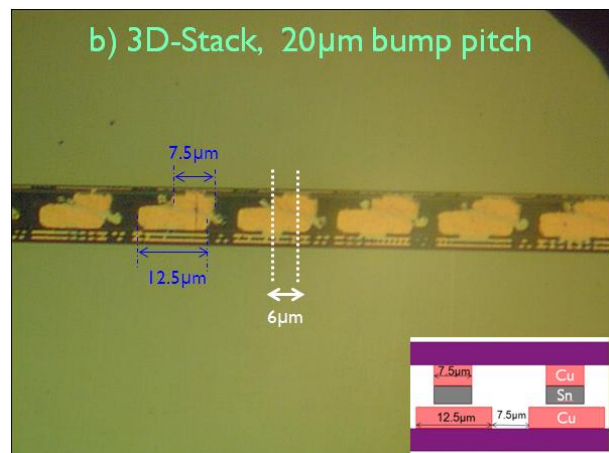


Fig. 9: Images of 3D-stacks in case of 6μm misalignment for 20μm bump pitch.

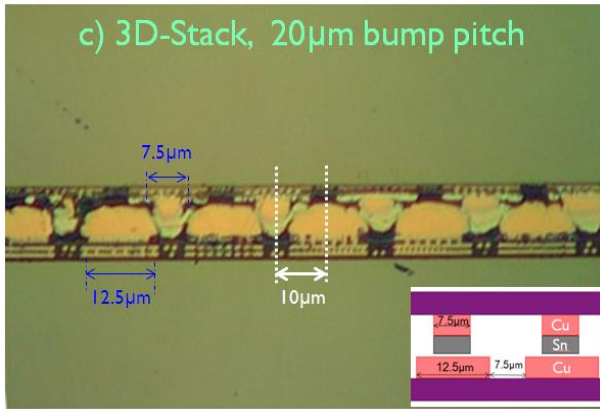


Fig. 10: Images of 3D-stacks in case of 10µm misalignment and 20µm bump pitch.

Another aspect to take into account for 3D stacking is the bump topography and the bump height variation. These parameters add new requirements for Z-axis tolerance that are not mentioned in this work as already discussed in another paper (reference [5]).

UF SCREENING AND EXPERIMENTAL RESULTS

In a previous paper [1] we have already reported the main differences between the standard underfill approach (capillary UF) dispensed after stacking operation); and newer approaches based on pre applied underfill: dispensed locally (NUF/NCP) and at wafer-level (WLUF). For 3D stacking, the capillary UF presents clear limits in case of narrow gap between dies and bumps; it also requires a dedicated dispensing area. These limitations are crucial for high density 3D integration and therefore our main focus is on pre applied UF. In case of pre applied UF, the material is dispensed on the landing die before stacking (refer to Fig.11).

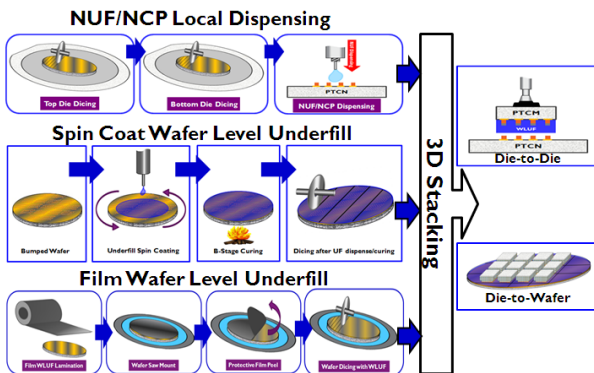


Fig. 11: Process flows in case of pre applied UF.

In this case, the stacking operation has the double function of bonding top/landing dies and curing the UF. This approach well fits the standard assembly operations and can allow a partial UF curing (typically ~90% of material is cured), later completed by the following assembly steps like post molding cure. Use of pre-applied UF presents anyway a few concerns like: material transparency to see alignment marks during assembly operation, dispense timing (done for each

single die), UF/filler entrapment between bumps, voids and other defects related to material flowing.

The main objective of our screening activity is to identify UF materials that can be used for 3D integration. The scope is to assess the key material properties and modify them in order to allow a reliable underfill process for 3D stacks.

In this continuous work, the materials are evaluated in three subsequent phases: *pre-screening* and material selection; *requirements assessment*; *assembly* [1].

In the pre-screening phase, we look at materials that could full-fill (at least partially) the following requirements:

- Able to reach a material thickness below 30µm, with 10µm as main target.
- Thickness uniformity of +/-10%.
- Able to fill gap for small bump pitch ($\leq 40\mu\text{m}$)
- High transparency to allow die alignment based on control marks
- Good tackiness at T ambient
- Low curing temperature
- Process capability up to 250°C

After this initial screening, mainly based on supplier datasheets and tests on blanket wafers, the selected materials are moved to the following phases: requirements assessment and assembly.

A. Requirements assessment

In the requirements assessment phase, the materials are processed by using the dedicated test vehicle (refer to a previous paragraph) in order to confirm the properties and requirements claimed in the pre-screening phase.

The major requirements for this phase are: minimum achievable thickness and uniformity; material transparency and absence of defects. In figure 12 we report the measured thickness for 2 WLUF materials. The materials are 2 different films UF. The target thickness is 15µm and the variation measured by an optical profile-meter is of 2µm.

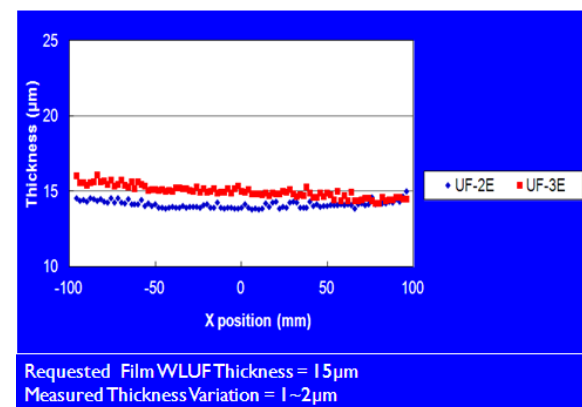


Fig. 12: Thickness measurements for two film WLUF materials: UF-2E and UF-3E.

In figure 13 we report the results on 2 other WLUF materials with different filler sizes and filler load. From these results it appears clear that filler size is a key element to improve material transparency: material

with similar filler load (~50%) is more transparent in case of small filler size (0.1µm vs. 0.5µm).

In figure 14 we report some defect observed after lamination of film WLUF. In this case the material is delaminating from the wafer, mainly in the area close to the micro-bumps.

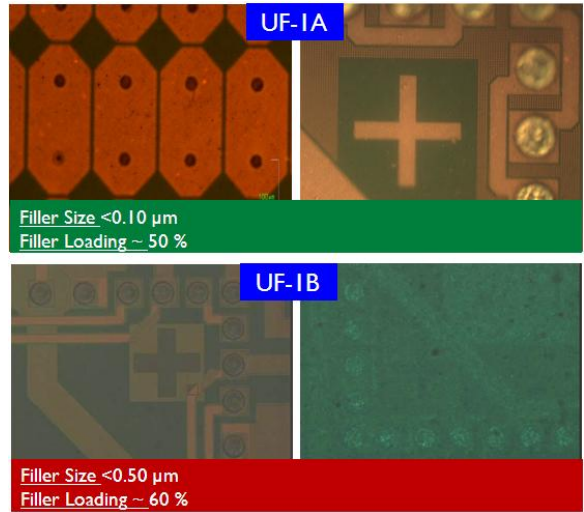


Fig. 13: WLUF transparency for two materials: UF-1A (50% filler load, 0.1µm filler size) and UF-1B (60% filler load, 0.5µm filler size).

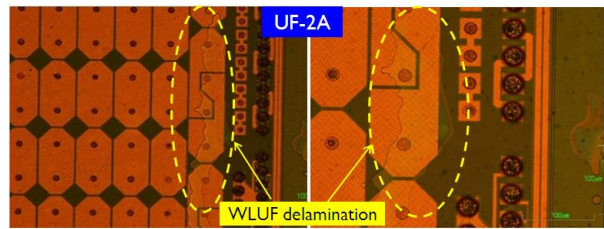


Fig. 14: Delamination observed in case of film WLUF.

B. Assembly

After completion of the previous phase, the materials have been moved to the assembly phase.

In this phase, for WLUF we check processes like dicing and material tackiness after ‘soft’ bonding; while for NUF/NCP we check material aging after dispensing. For both kinds of materials the main selection criteria is the electrical characterization of 3D stacks followed by SAM and X-SEM.

Furthermore, for materials that have showed good electrical results we have also introduced an initial reliability test. This reduced test consists in exposing the stacks to thermo-cycling (TC) steps followed by SAM inspections to detect eventual delamination.

To check the dicing process, the test wafers are coated with WLUF (spin coating and film materials), wafers are then diced and inspected for defects (material smearing, delamination, particles trapped, material degradation). Based on the results either dicing processing or WLUF material may require optimization.

In figure 15 we report some defects observed for one of the materials.



Fig. 15: Delamination observed after dicing the test vehicle with WLUF.

The material tackiness (WLUF) is evaluated by ‘soft’ bonding: dies with UF are placed on a landing wafer heated at 80°C with lowest pressure. Dies are required to stay in place (no movement) after wafer population (~1hour) and after wafer transferring from a tool to another one (2 meters distance). The eventual movement is characterized by optical microscope inspection. In case of local dispensable materials (NUF/NCP) we check the material staging: the material is dispensed on a landing wafer with a selected pattern and checked in time (refer to Fig. 16). The materials that have a minimal pattern modification are considered for die-to-wafer (D2W) assembly.

Material Type	Initial Dispense	After 20 minutes	After 24 Hours
UF-1C			
UF-6C			
UF-7C			
UF-8C			

Fig. 16: NUF/NCP materials dispensed and checked in time to assess material stability.

Electrical characterization and thermo-cycling for NUF/NCP

After have checked UF aging as indicated above, the selected materials are used to generate 3D stacks. The stacks are later electrically characterized to assess the stacking yield.

The UF is dispensed on the landing die and the stacks are done by using die-to-die approach. The stacking yield is defined by the number of functional daisy chain connections. The stacks are also checked with SAM and X-SEM to identify failure causes.

The results in term of electrical yield are reported in table 2. Among the selected materials, a few of them present an acceptable yield (>50%).

From these results, 2 materials (UF-5A, UF-1C) are identified as best performing. Both materials presented a relative high yield (83% for UF-5A and 93% for UF-1C), but after X-SEM inspection (Fig. 17) it appears clear that UF-5A present an higher filler entrapment between bumps, this is mostly due to the higher filler load (60% vs. 40% of UF-1C).

UF Type	UF name	Filler load	Electrical yield	SAM inspection	n. of stacks
NUF/NCP	UF-6A	0%	30%	NO voids NO delamination	18 stacks done
	UF-4A	60%	57%	NO voids NO delamination	66 stacks done
	UF-5A	60%	83%	NO voids NO delamination	110 stacks done
	UF-1C	40%	93%	NO voids NO delamination	82 stacks done
	UF-4C	0%	20%	NO voids NO delamination	18 stacks done

Table 2: Results of the electrical characterization for NUF/NCP materials. Number of stacks and filler load for each material are indicated.

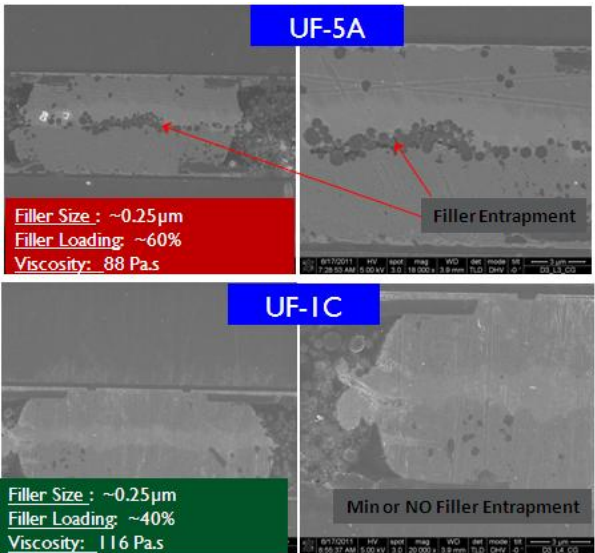


Fig. 17: X-SEM images of 2 highest yielding NUF/NCP materials.

As initial study of materials reliability, the stacks that presented the electrical yield above 50% have been processed by thermo-cycling (-40°C to 150°C). However, after TC the Cu based testing pads are highly oxidized and no electrical measurement is possible, therefore the response taken in this phase is SAM inspection to check for voids and/or delamination. The inspections were performed after stacking (T=0), after 500 cycles and after 1000 cycles: no voids and no delamination were observed (Fig. 18). Based on these results we could conclude that no major failure may be associated to the UF materials, it is anyway not possible to assess any yield loss due to TC. Anyway, we do believe that in case of materials presenting high filler entrapped between bumps, the TC should induce delamination and cracks between bumps. This assessment will be possible after the optimization of the test vehicle in order to allow post TC electrical test.

Thermo cycle: - 40°C to 150°C

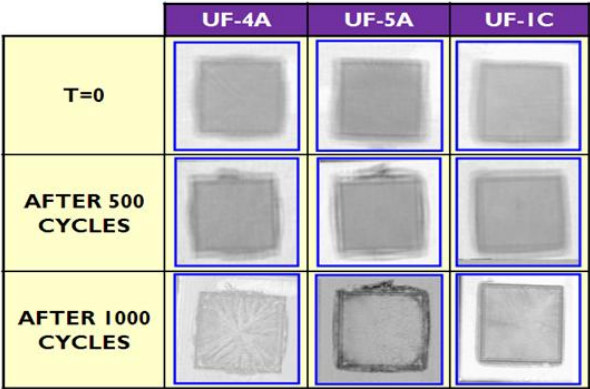


Fig. 18: SAM images after thermal cycling (-40°C to 150°C) of best electrical yielding NUF/NCP materials: NO voids and NO delamination observed.

Electrical characterization and thermo-cycling for WLUF

In case of WLUF, the material is coated on the top dies wafer. The wafer is then diced and inspected for defects. The 3D stacks are done also in this case using die-to-die approach. The bonding profiles have been optimized starting from supplier recommendations and taking into account the material viscosity curve (refer to Fig. 19 for an example). The results in term of electrical yield are reported in table 3. From these results, 3 materials have a very positive yield (>80%) and one of them has a yield above 94% (UF-1A). To be noted that material UF-1A was also the best performing in previous screening experiments refer to [1].

UF Type	UF name	Filler load	Electrical yield	SAM inspection	n. of stacks
WLUF	UF-1A	50%	94%	NO voids NO delamination	118
	UF-2A	60%	0%	Delamination seen	18
	UF-3A	35%	0%	Delamination seen	10
	UF-1B	50%	0%	Delamination seen	24
	UF-2C	40%	35%	NO voids NO delamination	66
	UF-2E	50%	85%	NO voids NO delamination	60
	UF-3E	50%	85%	NO voids NO delamination	60

Table 3: Electrical characterization for WLUF materials. Number of stacks and filler load for each material are indicated.

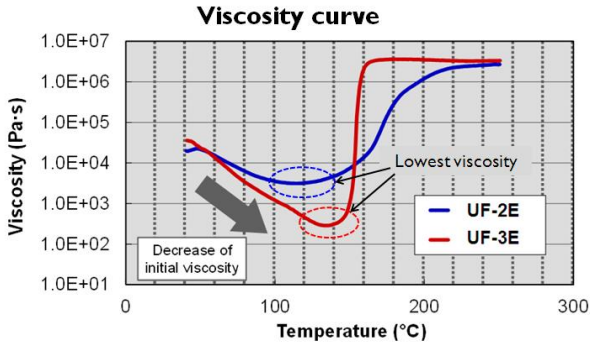


Fig. 19: Viscosity curve for 2 selected WLUFs.

After electrical test, a few stacks have been cross-sectioned and analyzed to check failure causes. In figure 20 we report the cases of UF-1A and UF-1B. In both cases we observe filler entrapment, but the case of UF-1B is so severe to avoid completely the bumps connection.

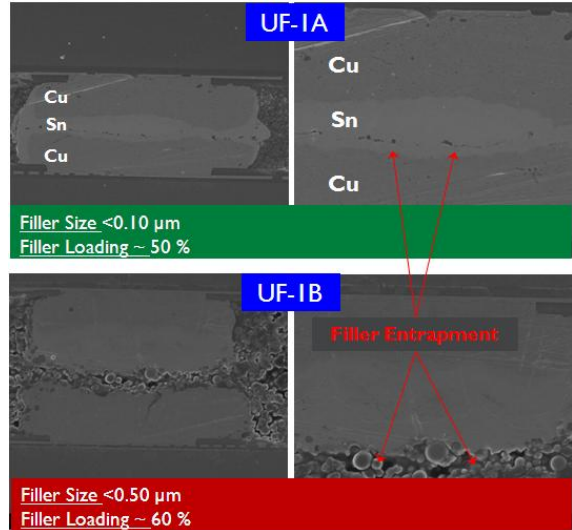


Fig. 20: X-SEM images for 2 WLUF. For material UF-1B the filler entrapped between bumps is the main reason for missing bumps connection.

Also in case of WLUF, the stacks have been processed by TC (-40°C to 150°C) in order to have an initial reliability test. As for the case of NUF/NCP, after TC the testing pads are highly oxidized and no electrical measurement is possible, therefore the response taken also in this phase is the SAM inspection. The inspections were performed after stacking (T=0), after 500 cycles and after 1000 cycles: no voids and no delamination were observed (Fig. 21).

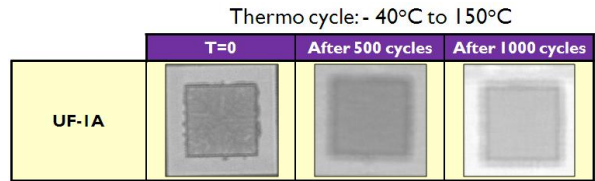


Fig. 21: SAM images after thermo-cycling in case of stacks done using UF-1A. Similar results observed also for the other yielding WLUFs.

Conclusions

We have provided a short introduction about requirements for 3D staking in case of small bump pitch. We indicated a few options to enable minimal misalignment error during stacking and presented some results in term of µbumping processes optimization. We have finally reported on 3D stacking characterization in case of pre-applied UF. We have reported cases with electrical yield above 90% in case of NUF/NCP and WLUF. We have also reported on initial reliability test based on 3D stacks thermo-cycling and SAM inspection. All the results reported

are encouraging and represent a push further to the adoption of pre-applied UF for reliable 3D stacking.

Acknowledgments

This work has been strongly supported by IMEC process support, Assembly Integration, Reliability and Modelling teams. We acknowledge the support of the partner companies in the imec 3D System Integration program.

References:

- [1] Antonio La Manna, K. J. Rebibis, C. Gerets, E. Beyne, *Use of Wafer Applied Underfill for 3D Stacking*. IMAPS 2011, Long Beach US.
- [2] A. Jourdain et al., *Integration of TSVs, wafer thinning and backside passivation on full 300mm CMOS wafers for 3D applications*, Proc. IEEE 61st Electronic Components and Technology Conference (ECTC), June 2011
- [3] T. Buisson et al., *Integration challenges of Cu Pillars with extreme wafer thinning for 3D Stacking and Packaging*, EPTC 2011, Singapore
- [4] Agarwal R, Zhang W, Limaye P and RuythoorenW2009 *Proc.Electronic Components and Technology Conf.* pp 345–9
- [5] J. De Vos et al., *Stacking Aspects in View of Scaling.*, IMAPS 2012, San Diego US