IEEE International Test Conference (ITC) 2014:

Direct Probing on Large-Array Fine-Pitch Micro-Bumps of a Wide-I/O Logic-Memory Interface

Erik Jan Marinissen[1] Bart De Wachter[1] Ken Smith[2]
Jörg Kiesewetter[3] Mottaqiallah Taouil[4] Said Hamdioui[4]

[1] IMEC vzw
Kapeldreef 75
B-3001 Leuven
Belgium
erik.jan.marinissen@imec.be
bart.dewachter@imec.be

[2] Cascade Microtech, Inc.
9100 SW Gemini Drive
Beaverton, OR 97008
United States of America
ken.smith@cmicro.com

[3] Cascade Microtech GmbH
Süss Straße 1
Thiendorf 01561
Germany
joerg.kiesewetter@cmicro.com

[4] Technische Universiteit Delft
Dept. of Computer Engineering
Mekelweg 4, 2628CD Delft
The Netherlands
m.taouil@tudelft.nl
s.hamdioui@tudelft.nl

On-line: http://dx.doi.org/10.1109/TEST.2014.7035314

Direct Probing on Large-Array Fine-Pitch Micro-Bumps of a Wide-I/O Logic-Memory Interface

Erik Jan Marinissen[1*] Bart De Wachter[1*] Ken Smith[2] Jörg Kiesewetter[3*] Mottaqiallah Taouil[4] Said Hamdioui[4]

[1] IMEC vzw
Kapeldreef 75
B-3001 Leuven
Belgium
erik.jan.marinissen@imec.be
bart.dewachter@imec.be

[2] Cascade Microtech, Inc.
9100 SW Gemini Drive
Beaverton, OR 97008
United States of America
ken.smith@cmicro.com

[3] Cascade Microtech GmbH
Süss Straße 1
Thiendorf 01561
Germany
joerg.kiesewetter@cmicro.com

[4] Technische Universiteit Delft
Dept. of Computer Engineering
Mekelweg 4, 2628CD Delft
The Netherlands
m.taouil@tudelft.nl
s.hamdioui@tudelft.nl

Abstract

In order to obtain acceptable compound stack yields for 2.5D- and 3D-SICs, there is a need to test the constituting dies before stacking. The non-bottom dies of these stacks have their functional access exclusively through large arrays of fine-pitch micro-bumps, which are too dense for conventional probe technology. A common approach to obtain pre-bond test access is to equip these dies with dedicated pre-bond probe pads, which comes with drawbacks such as increased silicon area, test application time, and reduced interconnect performance. In order to avoid the many drawbacks of dedicated pre-bond probe pads, we advocate the usage of advanced probe technology that allows to directly probe on these micro-bumps. This paper reports on the technical and economical feasibility of this approach.

1 Introduction

There is a lot of excitement around and expectations from 2.5D- and 3D-stacked integrated circuits [1]. In 2.5D-SICs, multiple active dies are placed side-by-side on top of and interconnected by a passive interposer die. In 3D-SICs, multiple active dies are stacked vertically. Both 2.5D- and 3D-SICs are enabled by the capability to manufacture through-silicon vias (TSVs) that provide an electrical connection between the front- and back-side of a silicon substrate [2–4]. In 2.5D-SICs TSVs connect the stacked active dies through the silicon interposer to the package substrate. In 3D-SICs TSVs provide vertical interconnections between the various stacked dies. Both types of SICs serve their particular market segments and are here to stay; 2.5D-SICs provide better chip cooling options and hence typically target high-performance computing and networking applications, whereas 3D-SICs with their small footprint are better suited for mobile applications.

In order to obtain acceptable compound stack yields, there is a need to perform *pre-bond testing* of the various dies before stacking [5, 6]. For non-bottom dies in the stack, the typical functional interface is through an array of fine-pitch micro-bumps. These micro-bumps are too small and too dense for conventional probe technology. Consequently, the current industrial approach to enable test access for pre-bond testing is to provide non-bottom dies with dedicated pre-bond probe pads [5, 7–9]. Although these dedicated probe pads achieve the job, they come at the expense of extra design effort, extra silicon area, possibly extra processing steps, extra test application time, extra load on the micro-bump I/Os during post-bond functional stack operation, and still leave the micro-bumps themselves untested.

In this work, we set out to directly probe on large-array fine-pitch micro-bumps. We are capable to do this at wafer level with a probe card in a single-site set-up. This enables a test flow in which the die's internal circuitry (logic, DRAM) is tested through dedicated pre-bond probe pads, possibly in a (massive) multi-site arrangement, and in which the micro-bumps and underlying TSVs are separately tested in a single-site set-up. It also enables an alternative test flow, in which the entire pre-bond test is performed single-site by probing directly on the micro-bumps; this will circumvent the need for dedicated pre-bond probe pads with all its associated drawbacks and costs.

Direct probing on fine-pitch micro-bumps requires advanced probe technology: fine-pitch low-force probe cards and accurate probe stations. Prior work in this domain has been reported by others [10–15] and by us [16–18], but, to the best of our knowledge, this is the first paper that reports on pre-bond contact resistance, probe marks on both top and landing micro-bumps, and impact on stack interconnect yield. In this paper, we are using the JEDEC Wide-I/O Mobile DRAM interface (JESD-229) [19–21] as a typical target for today's 2.5D- and 3D-SIC micro-bump arrays. We have designed and manufactured test wafers with this micro-bump interface and report on our experiences probing and subsequent stacking of that interface. We have used the 3D-COSTAR test flow cost modeling tool [22–25] to analyze the cost-effectiveness of our approach, in comparison to performing pre-bond testing through dedicated pre-bond probe pads.

The remainder of this paper is organized as follows. Section 2

* Part of the work of Erik Jan Marinissen and Jörg Kiesewetter has been performed in the project ESiP (winner of the ENIAC Innovation Award 2013), which was funded by the ENIAC Joint Undertaking (http://www.eniac.eu), and, for Jörg Kiesewetter, by the public authorities of Saxony. Part of the work of Erik Jan Marinissen, Bart De Wachter, and Jörg Kiesewetter is performed in the project SEA4KET (http://www.sea4ket.eu), sub-project 3DIMS; this project receives funding from the European Union's Seventh Programme for research, technological development, and demonstration under grant agreement No. IST-611332.

978-1-4799-4722-5/14/$31.00

discusses the importance of pre-bond testing. Section 3 describes the micro-bump probe targets. Section 4 details the selected probe technology, while Section 5 describes the test vehicle. Experiment results are given in Section 6. Our cost modeling case study is described in Section 7. Section 8 concludes this paper.

2 Pre-Bond Testing

The post-bond compound stack yield y_{stack} of a stack consisting of n dies cannot be greater than the product of the individual die yields y_d (for $1 \leq d \leq n$) and the interconnect yields y_i (for $1 \leq i \leq n-1$), where y_i is the yield of the interconnects between adjacent Dies i and $i+1$:

$$y_{\text{stack}} \leq \prod_{d=1}^{n} y_d \cdot \prod_{i=1}^{n-1} y_i \qquad (2.1)$$

Figure 1 plots the post-bond compound stack yield y_{stack} for varying die yields y_d for various values of n and y_i. The graph demonstrates that the compound stack yield decreases drastically if y_d decreases. Consequently, it is important to test dies before stacking (the so-called pre-bond test) and only stack dies passing that pre-bond test in a die-to-die or die-to-wafer scheme.

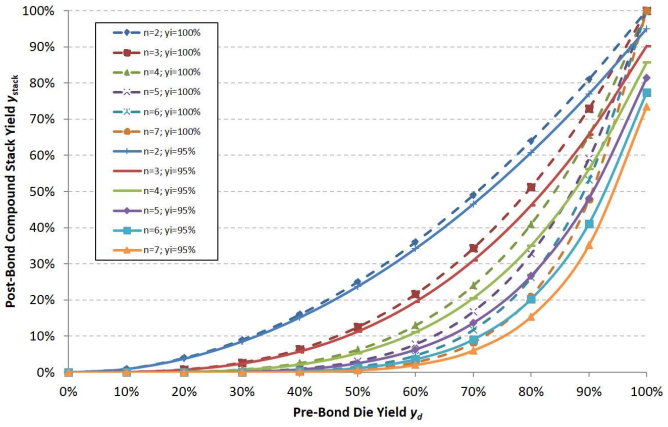


Figure 1: Post-bond compound stack yield y_{stack} as function of pre-bond die yield y_d for various stack heights n and various interconnect yields y_i.

Compared to skipping it, pre-bond testing obviously requires additional costs and the better the pre-bond test, the higher those costs will be. However, this investment typically pays off, as the alternative is that bad dies get detected only after stacking, at which point they are filtered out of the production flow together with the good dies to which they are now attached. Figure 2 shows an example of a total stack cost price calculation made with 3D-COSTAR [22–25]. We assumed a three-die stack, in which Die 1 was fully tested before stacking, but for which we varied the pre-bond test coverage and associated pre-bond test costs for Dies 2 and 3. The graph shows that more pre-bond testing (at assumed linearly increasing pre-bond test cost) actually decreases the overall stack cost price.

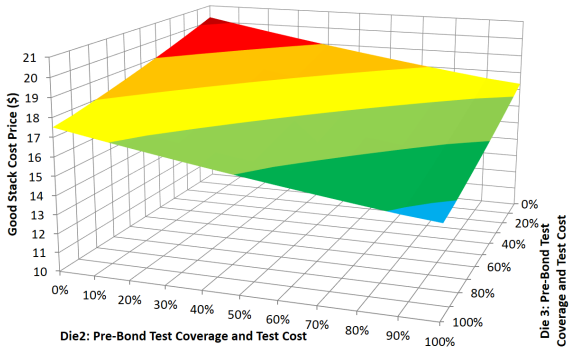


Figure 2: Good-stack cost price for a three-die stack as function of pre-bond test coverage and associated test cost for Dies 2 and 3.

Test access for pre-bond testing is through probing. Probing the bottom die of a stack is relatively easy, as the natural interface to the package substrate is implemented with large C4 bumps or copper pillars; a typical diameter is 50μm at 120μm pitch, which is no problem for today's probe technology. However, this is not true for the non-bottom dies; see Figure 3. All their functional connections (for power, ground, control, clocks, digital, analog, etc.) go through large arrays of fine-pitch micro-bumps. Typical micro-bumps have a diameter of $\sim$20μm at 40μm pitch and come in arrays of several hundreds to thousands of micro-bumps. Cantilever probe cards can achieve these small pitches, but cannot handle such large arrays. Vertical probe cards can be made in arbitrary array configurations, but are limited to pitches around 60μm.

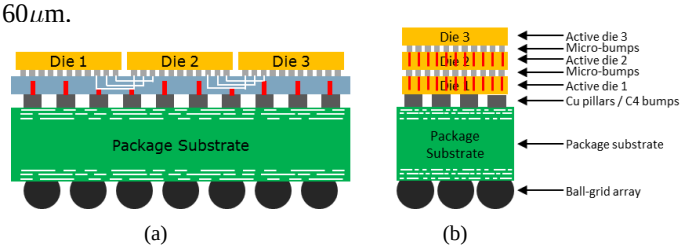


Figure 3: Cross-sections of typical (a) 2.5D- and (b) 3D-SICs containing three active dies.

Today's solution in the industry is to equip non-bottom dies with dedicated pre-bond probe pads, with sufficiently large size and pitch to accommodate today's probe technology [5, 7–9]. This solution requires extra design effort and possibly extra processing steps. Moreover, it causes a trade-off between extra silicon area and extra test time. The probe pads are larger than the micro-bumps; that is their whole purpose. Hence, typically one cannot afford as many probe pads as there are micro-bumps, as they would simply consume too much silicon area. As a result, the same pre-bond stimulus/response data needs to be pumped in and out of the die-under-test through a narrower interface and consequently the die's pre-bond test time smears out over more clock cycles, increasing the pre-bond test application cost. Furthermore, after performing a pre-bond test through dedicated probe pads, one can still not be certain of the correct operation of the functional interface through the micro-bumps. Finally, the dedicated pre-bond probe pads cause an extra capacitive load on the micro-bump I/Os during post-bond functional stack operation.

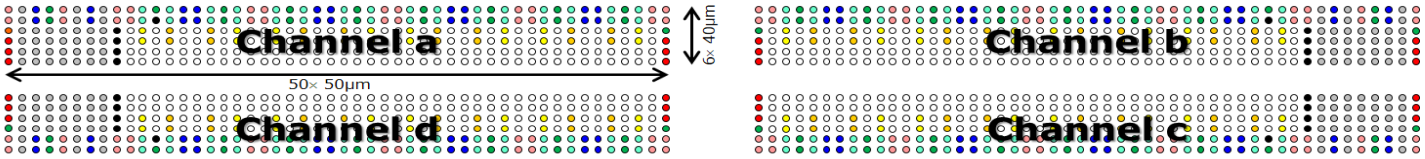
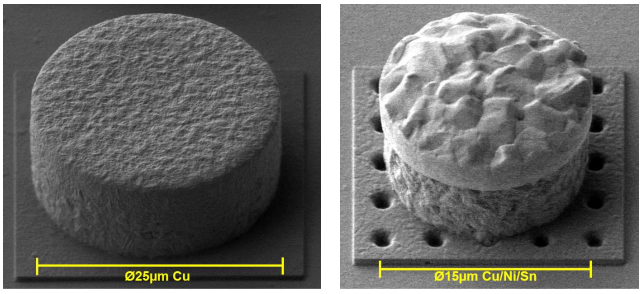


Figure 4: Standardized micro-bump lay-out according to the JEDEC Wide-I/O Mobile DRAM specification [19].

3 Micro-Bumps

Micro-bumps come in different metallurgies, forms, and shapes. IMEC's 40μm-pitch micro-bumps reference process utilizes copper (Cu) landing bumps of 9μm height and 25μm diameter and copper-nickel-tin (Cu/Ni/Sn) top bumps of 9μm height and 15μm diameter [26]. Two such micro-bumps are depicted in Figure 5. The micro-bumps have a cylindrical shape. As can be seen, the Cu micro-bumps have a rather smooth surface. As no reflow was applied (yet) on the Cu/Ni/Sn micro-bumps, their surface is significantly more rough. During stacking, the two micro-bumps form an intermetallic bond under thermo-compression.



(a) Cu micro-bump, $\varnothing$25μm (b) Cu/Ni/Sn micro-bump, $\varnothing$15μm

Figure 5: Typical micro-bumps at IMEC: (a) copper landing bump of 25μm diameter and (b) copper-nickel-tin top bump of 15μm diameter.

Micro-bumps typically come in large arrays. For this work, we took as target the representative micro-bump array of the JEDEC Wide-I/O Mobile DRAM standard [19–21]. This first standard for stackable Wide-I/O DRAMs, published as JESD-229 in December 2011, defines the functional and mechanical aspects of the Wide-I/O logic-memory interface. The interface consists of four DRAM channels (named a, b, c, and d), each consisting of an array of 6 rows $\times$ 50 columns = 300 micro-bumps with a horizontal pitch of 50μm and a vertical pitch of 40μm. The pad locations are symmetric between the four channels and also the spacing between the four channels is defined. The total interface occupies 0.52mm $\times$ 5.25mm. Figure 4 shows the lay-out of the 1,200 JEDEC Wide-I/O micro-bumps.

Direct probing on large arrays of fine-pitch micro-bumps has to meet the following criteria.

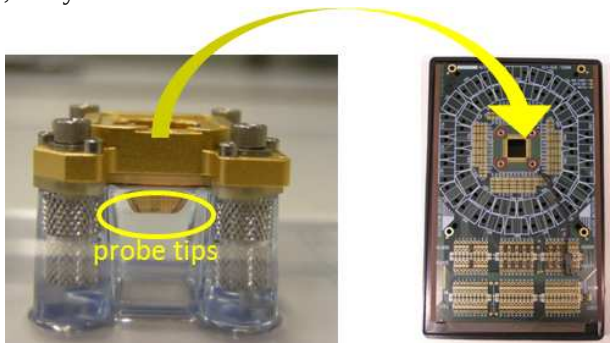
- *Good electrical contact with low contact resistance*, to allow for pre-bond testing of the die-under-test. We used as specification a contact resistance $<5\Omega$.
- *Probe marks with a limited profile*, to not impair downstream bonding or negatively impact the yield of that bonding process. We used as specification a probe mark profile $<$500nm.
- *Affordable test cost*, i.e., the cost of the required probe technology should not be excessive. We address this issue in Section 7.

4 Probe Technology

4.1 Probe Cards

Conventional probe cards are insufficient to probe on large-array fine-pitch micro-bumps, such as specified by JEDEC's Wide-I/O Mobile DRAM interface. Traditional cantilever probe cards do not come in the required array size, and vertical probe cards do not come in the required fine pitch. Hence, we needed to turn to advanced MEMS-type probe cards.

We have used the second generation of Cascade Microtech's Pyramid Probe® technology, named Rocking Beam Interposer (RBI), which is currently in its development phase. As depicted in Figure 6, this technology has a modular set-up comprising two components: the probe *core* and the probe *card*. The probe core contains the IC-design specific probe tips, whereas the probe card fits to the probe station-specific probe card holder. The probe core's rectangular frame has a screw in each of its four corners, with which it is screwed right on top of the hole in the middle of the probe card, such that the core's probe tips stick out under the probe card, ready to touch the wafer.



(a) Probe core (b) Probe card

Figure 6: Modular Pyramid Probe set-up consisting of (a) probe core and (b) probe card.

The Pyramid Probe RBI is a vertical, non-see-through probe technology. Two thin-film membranes are spanned across a 'plunger' with adjustable spring. The first membrane contains the routing layer from probe card I/Os to probe tips and vice-versa, whereas

the second (= outer) membrane contains the RBI probe tips; see Figure 7 [18]. The MEMS-type probe tips have a square probe surface of $6\times6\mu m^2$ and are placed on a rocking beam that connects to the upper routing-layer membrane through a copper post. Figure 8(a) shows a top view of an array of probe tips, while Figure 8(b) shows a cross-section of a single probe tip.

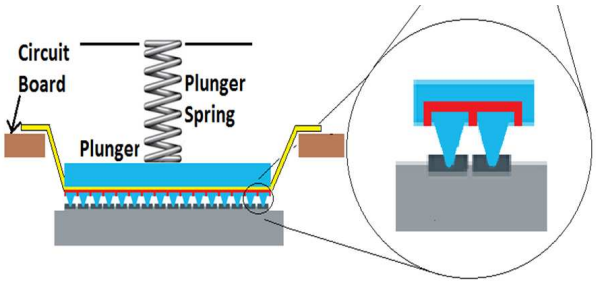
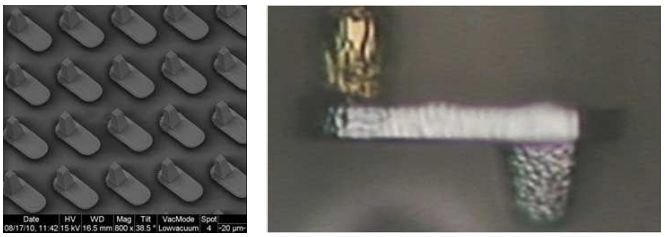


Figure 7: Conceptual cross-section view of the probe core [18].



(a) Top view (b) Cross section view

Figure 8: Pyramid Probe® RBI tips.

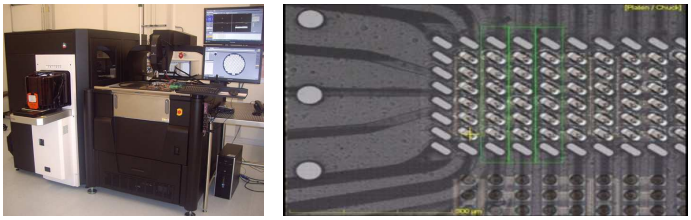
The Pyramid Probe® RBI technology has three main benefits: (1) it can be manufactured in large arrays at fine pitches, down to $20\mu m$; (2) the probe tips exercise a low probe force of up to 1 gram force per tip at a user-defined chuck over-travel, thereby inflicting minimal probe mark damage; and (3) the separate tip-layer coupon allows easy repair of inadvertently damaged probe tips.

4.2 Probe Station

The selection of a probe station had to fulfill three main criteria.

- The probe station needs to be able to work in a clean-room environment, as the stacking operations which follow after pre-bond testing are also performed in a clean-room environment and hence wafers/dies cannot be contaminated.
- The probe station has to be able to work with non-see-through vertical probe cards, which implies overlay *probe-to-pad alignment* (PTPA) in the presence of upward-looking (to the probe tips) and downward-looking (to the micro-bumps on the wafer) cameras (see Figure 9(b)).
- The probe station has to have a x, y, and θ touch-down accuracy and stepping accuracy sufficient to work with the small diameters and pitches of our micro-bump arrays.

We selected the Cascade Microtech CM300 probe station for our task. This is a brand-new prober platform with features for measurement accuracy and unattended testing inherited from the Cascade Microtech Elite300 and the Süss MicroTec MicroAlign probers respectively. Installed in IMEC's 300mm clean-room is the world's first demonstrator prototype of this probe station along with an auto-loader and material handling unit (see Figure 9(a)).



(a) (b)

Figure 9: CM300 probe station in IMEC's 300mm clean-room (a) and the prober's tip training software in action on the Wide-I/O probe core (b).

5 Test Vehicle: Vesuvius-2.5D

The IMEC-designed test vehicle for our Wide-I/O direct probing experiments is named Vesuvius-2.5D. It consists of two active Vesuvius test chips, stacked face-down side-by-side on a passive Interposer die. Figure 10 shows a picture of a single Vesuvius-2.5D die stack.

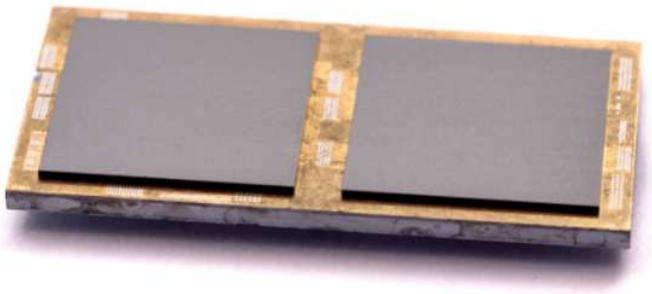


Figure 10: Vesuvius-2.5D die stack photo.

Figure 11 shows the various dimensions of the Vesuvius-2.5D die stack, both in top and cross-section view. The two Vesuvius dies atop measure $8.1\times8.1mm^2$ in a custom technology consisting of 65nm CMOS and five metal layers manufactured by GLOBALFOUNDRIES, and Cu/Ni/Sn micro-bumps (as described in Section 3) manufactured by IMEC. The bottom Interposer die measures $10\times20mm^2$ in an experimental silicon interposer technology containing four metal layers, $10\times100\mu m$ 'via-middle' TSVs and Cu micro-bumps (as described in Section 3), developed and manufactured by IMEC [27].

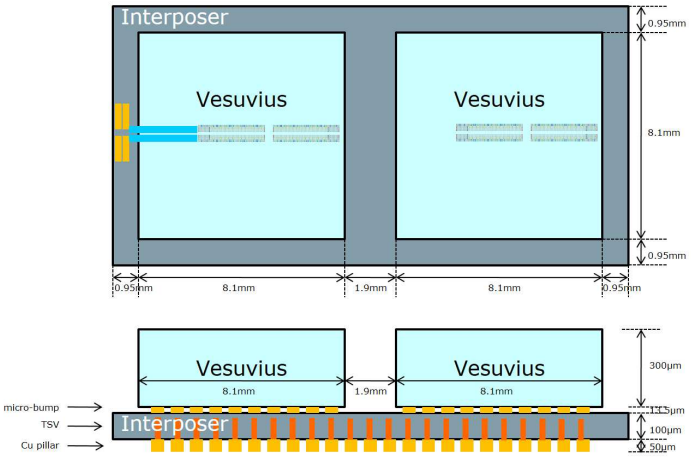


Figure 11: Vesuvius-2.5D test vehicle top view and cross-section.

Each Vesuvius die contains many test structures [28], including one full JEDEC-compliant four-channel Wide-I/O micro-bump interface [19]. Each Wide-I/O channel of 300 micro-bumps is divided in ten equal groups of 30 micro-bumps each, which after stacking form an up-down daisy-chain between Vesuvius and Interposer dies; see Figure 12. Hence, there are in total 40 daisy-chains for the four Wide-I/O channels. As depicted in Figure 11, the daisy-chains in the left-hand Vesuvius die are routed through the Interposer die to regularly-sized ($80\times60\mu m^2$) probe pads on the Interposer front-side to the left of the left-most Vesuvius die.

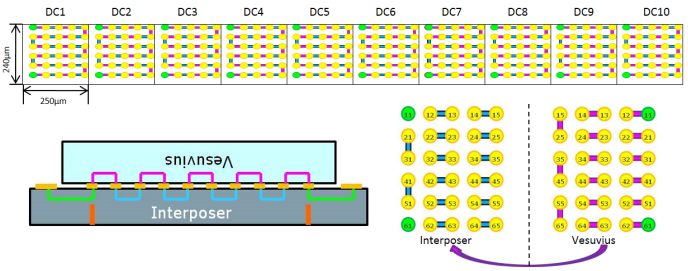
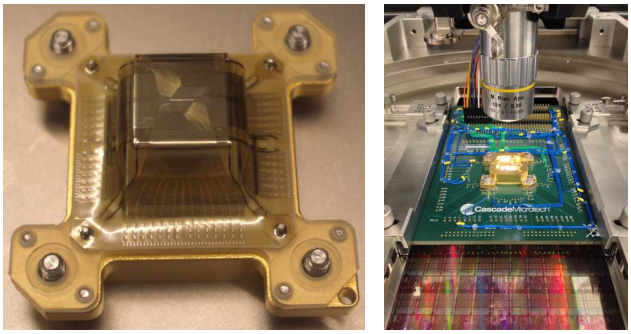


Figure 12: Vesuvius-Interposer daisy-chain consisting of 30 adjacent micro-bumps and metal interconnects.

The Pyramid RBI probe core (shown in Figure 13(a)) designed for our test vehicle is a so-called MSI core for exactly one Wide-I/O channel. The reason to probe on a single Wide-I/O channel (and not all four) is rooted in experimental considerations; this set-up allows us to evaluate the impact of probe marks on stack interconnect yield for all four cases: probed on top and bottom, probed only on either top or bottom, or not probed at all. The probe core routes all 300 probe tips out to the corresponding core-to-card I/Os; therefore the same probe core can be used on both Interposer and Vesuvius dies. We have two 4.5-inch rectangular engineering-type probe cards for our CM300 probe station: one which completes the ten daisy-chains when probing on an Interposer die (depicted in Figure 13(b)) and another one which completes the ten daisy-chains when probing on a Vesuvius die. Concatenating 30 micro-bumps in a daisy-chain limits the resolution of the probe-to-bump contact resistance that can be measured, but this design decision was due to a limitation in the number of available tester channels.



(a) (b)

Figure 13: Probe core for single-channel Wide-I/O interface (a) probe tips face-up and (b) attached to a probe card, in action in our probe station.

We have defined three test phases for Vesuvius-2.5D probing.

- In *Test Phase 1* we use the Pyramid RBI probe core with the dedicated Interposer probe card to probe on Wide-I/O channels a and b of the pre-bond Interposer dies in two subsequent touch-downs. We check the landing of the probe tips on the 25μm-diameter Cu micro-bumps, both by means of visual and scanning electron microscope (SEM) inspection of the probe marks, as well as by electrical continuity of the probe card-to-wafer daisy-chains.

- In *Test Phase 2* we use the Pyramid RBI probe core with the dedicated Vesuvius probe card to probe on Wide-I/O channels b and c of the pre-bond Vesuvius dies in two subsequent touch-downs. We verify the landing of the probe tips on the 15μm-diameter Cu/Ni/Sn micro-bumps, both by means of visual and SEM inspection of the probe marks, as well as by electrical continuity of the probe card-to-wafer daisy-chains.

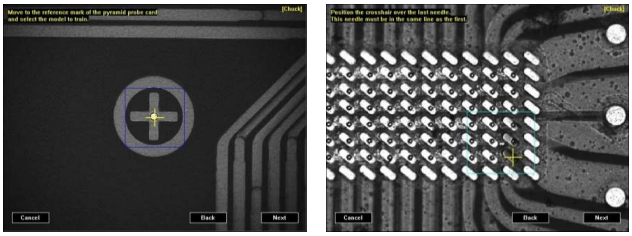
- In *Test Phase 3*, after stacking, we assess the impact of the micro-bump probing on the interconnect yield. We use a conventional cantilever probe card to probe on the regularly-sized post-bond probe pads on the front-side of the Interposer and measure the electrical continuity of the various micro-bump daisy-chains. In this, we can compare the four channels:

  - Channel a: only Interposer probed
  - Channel b: Vesuvius and Interposer both probed
  - Channel c: only Vesuvius probed
  - Channel d: no micro-bumps probed.

6 Experiment Results

6.1 Initial Hurdles

Initially, the PTPA software of the CM300 probe station was not optimally suited for automatically recognizing the Pyramid Probe RBI probe tips. The software was made to handle conventional cantilever and vertical probe needles, but turned out not to work reliably and repeatedly for the small and very different RBI probe tips. We developed a dedicated probe tip recognition routine for the RBI tips, which consists of three steps. The pattern recognition uses (1) the large cross-hair fiducials included on the four corners of the probe core's membrane (see Figure 14(a)), (2) dummy bumps on the probe membrane, and (3) two probe tips, typically located on opposite extremes of the probe tip array (see Figure 14(b)). With the deployment of the new software routine, the automatic probe tip recognition works without problems.



(a) Step 1 (b) Steps 2+3

Figure 14: Dedicated probe tip recognition routine for RBI probe tips.

During probing operation, probe tips pick up dirt, which increases the contact resistance and ultimately might obstruct electrical contact completely. Hence, probe tips need to be cleaned at regular intervals. This is also true for RBI probe tips. The recommended cleaning medium for RBI tips is a tungsten-carbide (WC) substrate, on which the to-be-cleaned tips need to touch down with regular over-travel at 15 fresh locations. For holding a cleaning substrate, the CM300 probe station is equipped with various vacuum-providing auxiliary chucks, positioned just outside the main wafer chuck (see Figure 15). The initial software version of the CM300 prober did not support usage of these auxiliary chucks with non-see-through probe cards like the Pyramid Probe RBI. This obstacle for RBI probe tip cleaning was quickly resolved in a new software version, and now the probe station can perform automatic probe tip cleaning after a user-defined number of touch-downs.

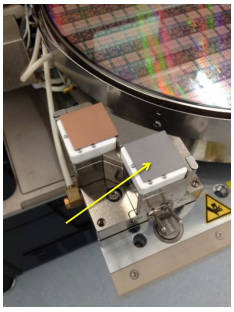
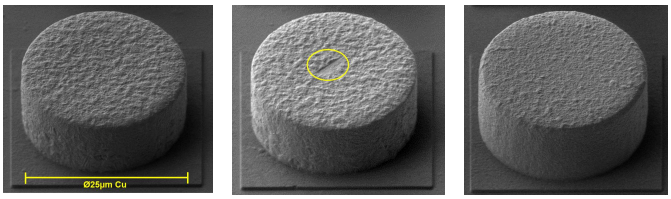


Figure 15: Tungsten-carbide (WC) tip cleaning substrate on CM300's auxiliary chuck.

6.2 Probe Marks

For given probe tip material and shape, the resulting probe marks depend on the chuck over-travel and the micro-bump metallurgies. All our experiments were performed at 150μm over-travel, which corresponds to 1gf/tip for the global plunger spring in the RBI probe cores we used.

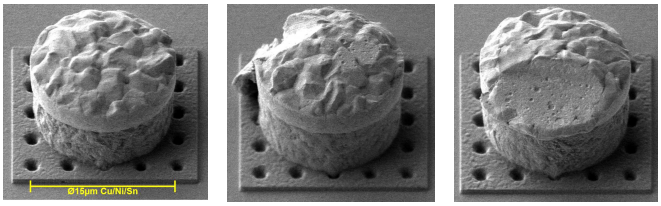
Test Phase 1 – Figure 16 shows SEM pictures of 25μm-diameter Cu landing micro-bumps before and after probing. Figure 16(a) shows such a Cu micro-bump *before* probing; the micro-bump is cylindrical in shape with a very smooth surface. Figure 16(b) shows a similar micro-bump *after* probing. On all such Cu micro-bumps, the probe marks are very uniform: a diagonal line of approximately $6\times1\mu m^2$, caused by the heel of the diagonally placed probe tip which itself measures $6\times6\mu m^2$. The probe mark is very shallow, on the order of the surface roughness of the Cu micro-bump. We do not expect any negative impact of the probe mark on the interconnect yield. Figure 16(c) shows a probed Cu micro-bump equipped with a 10nm-thick nickel-boron (NiB) cap. This NiB cap is meant to prevent the Cu micro-bump from oxidizing and thus improve the stack interconnect yield. The NiB cap is quite hard and consequently hardly any probe mark can be seen, although proper electrical contact was made [29].



(a) Before probing (b) After (c) After (NiB cap)

Figure 16: Probe marks on 25μm-diameter Cu landing micro-bumps.

Test Phase 2 – Figure 17 shows SEM pictures of non-reflowed 15μm-diameter Cu/Ni/Sn top micro-bumps before and after probing. Figure 17(a) shows such a Cu/Ni/Sn micro-bump *before* probing and reflow; the micro-bump is cylindrical in shape with a rather rough Sn surface. Figures 17(b) and 17(c) show two similar

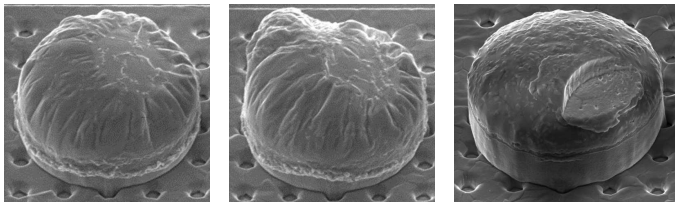


(a) Before probing (b) After – 1 (c) After – 2

Figure 17: Probe marks on non-reflowed 15μm-diameter Cu/Ni/Sn top micro-bumps.

micro-bumps *after* probing. On the softer Sn material, the probe mark is significantly larger than on the much harder Cu micro-bumps.

Figure 18 shows SEM pictures of reflowed 15μm-diameter Cu/Ni/Sn top micro-bumps before and after probing. Figure 18(a) shows such a Cu/Ni/Sn micro-bump which was only reflowed, and *not* probed; the originally rough horizontal Sn surface (as seen in Figure 17(a)) has been transformed by the reflow process in a dome-shaped cap. Figure 18(b) shows a Cu/Ni/Sn micro-bump which was first probed and subsequently reflowed. The hope was that the reflow process would eliminate the probe mark, but as can be seen from the figure, this is not entirely the case. Finally, Figure 18(c) shows a Cu/Ni/Sn micro-bump which was first reflowed and subsequently probed; as expected, the probe mark is clearly visible in the otherwise nicely smooth dome-shaped Sn cap. The remaining probe marks in Figures 18(b) and 18(c) could potentially form a location for particle or filler entrapment and hence negatively affect the bond's reliability. The smallest risk for this to happen is in the scenario where micro-bump probing precedes the reflow operation [30].



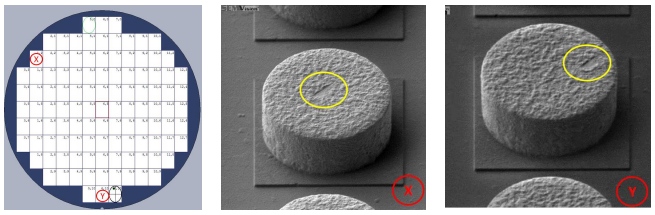
(a) Only reflowed (b) Probed; reflowed (c) Reflowed; probed

Figure 18: Probe marks on reflowed 15μm-diameter Cu/Ni/Sn top micro-bumps.

6.3 PTPA Accuracy

We want to minimize the probe mark damage to the micro-bump, in order not to negatively impact the downstream bonding yield; from that viewpoint, obtaining good electrical contact while leaving no visible probe mark is ideal. On the other hand, visible probe marks form reassuring evidence that a micro-bump was actually touched by a probe tip and allow us to determine the PTPA accuracy of the probe station.

To analyze the initial stepping accuracy of the CM300 demonstrator probe station, we stepped over all 111 dies in a 300mm Interposer wafer, starting top-left and zig-zagging down row-by-row until bottom-right, performing two touch-downs per Interposer die (on channels a and b). Analysis of the probe mark locations showed that all touch-downs were *on* the 25μm-diameter Cu micro-bumps, indicating that the stepping accuracy of the probe station was sufficient for Test Phase 1. There was little variation detected in the y-axis. However, the maximum variation in the x-axis was between Die X (-1μm, see Figure 19(b)) and Die Y (+6.2μm, see Figure 19(c)) (die locations indicated on the wafer map (Figure 19(a)). This was considered too inaccurate, as the probe mark was getting close to the micro-bump edge.



(a) Interposer wafer map (b) Probe mark X (c) Probe mark Y

Figure 19: Interposer wafer map (a) with Dies X (b) and Y (c) that showed the left-most resp. right-most probe mark location variation.

Cascade Microtech and IMEC have jointly taken several steps to improve the PTPA accuracy.

- *Probe card adapter*.
  The MSI-sized RBI probe cores have a vertical z height (*'draft'*) of 11.1mm measured from the top side of the probe card. This is higher than many other probe cards, due to which the probe tips ended up below the field-of-view of the probe station's side-view camera, which is meant to assist in the touch-down procedure. We had to lift the probe card in order to bring the probe tips back in sight of the side-view camera. Initially, this lifting was achieved with some dedicated shims. To further improve PTPA stability, this temporary workaround has been replaced with a new probe card adapter with larger z lift.

- *Thermal stability*.
  The ambient temperature in our clean-room is 22°C. However, during operation, the probe chamber of the probe station heats up to 30°C. This temperature increase can lead to a maximum radial wafer expansion of

$$\begin{aligned} r \cdot \Delta t \cdot CTE_{\mathrm{Si}} &= 150mm \cdot 8^{\circ}\mathrm{C} \cdot 2.6 \times 10^{-6}/^{\circ}\mathrm{C} \\ &= 3.12\mu m \end{aligned} \tag{6.1}$$

  where r is the wafer radius (in mm), Δt the temperature increase (in °C), and CTE_{Si} the coefficient of thermal expansion of silicon (in ppm/°C). Direct micro-bump probing requires thermal stability to avoid wafer expansion, and therefore we keep the chiller that controls the thermo-chuck on at 22°C.

- *Automatic ReAlign*.
  The CM300 software Velox has an option for automatic ReAlign after n touch-downs, with n a user-defined parameter. This software feature is mainly meant for temperature testing, where different thermal expansions of wafer and probe system might necessitate its usage. However, for fine-pitch micro-bump probing at stable ambient temperatures, the feature also provided great benefits. We have used it with good results after every 20 touch-downs. The ReAlign routine takes some time to execute (about 30 seconds in our case) and hence its usage slightly increases the overall test time. Therefore, the trade-off between touch-down accuracy and test time can be optimized.

Right now, the PTPA accuracy requirements are satisfied, as the electrical measurement results in the following section confirm.

6.4 Contact Resistance

Proper electrical contact of the RBI probe tips on the micro-bumps was analyzed by performing two-point resistance measurements of the daisy-chains through probe card and wafer, for all ten daisy-chains per touch-down. Due to a probe card fault, initially daisy-chains DC6 and DC9 were found to be consistently non-continuous. This was quickly diagnosed as a problem in the probe card wiring and fixed. From this moment onward, all daisy-chains were continuous, apart from confirmed bad dies. This demonstrated that the probe tips all make proper contact to the micro-bumps.

Figures 20, 21, and 22 show three representative wafer maps of two-point resistance measurements through a 30-long micro-bump-to-probe-tip daisy-chain of a DRAM channel on a micro-bumped wafer. The colors in the wafer map bin the daisy-chain resistance value R into three bins: (1) green: $R < 90\Omega$, (2) yellow: $90\Omega \leq R < 150\Omega$, (3) orange: $150\Omega \leq R < 300\Omega$, (4) red: $300\Omega \leq R$, (5) gray: $R = \infty$ ("Not-A-Number" = daisy-chain non-continuous).

Test Phase 1 – Figure 20 shows the wafer-map for DC2 of Channel a on an Interposer wafer with 25μm-diameter Cu micro-bumps. There are 111 dies on this 300mm wafer. Most daisy-chains are continuous. The non-continuous daisy-chains were confirmed (through other tests on the same dies) to be caused by wafer manufacturing issues on these particular dies. For this particular wafer, the median daisy-chain resistance was 118Ω, resulting in 3.9Ω per micro-bump.

DC 2 — Channel A — <90Ω <150Ω <300Ω >300Ω =open

Y / X	0	1	2	3	4	5	6	7	8	9	10	11	12
0	#N/A	#N/A	#N/A	#N/A	#N/A	79	68	101	#N/A	#N/A	#N/A	#N/A	#N/A
-1	#N/A	#N/A	100	73	91	80	63	75	89	81	86	#N/A	#N/A
-2	#N/A	111	100	67	168	94	80	84	78	76	68	80	#N/A
-3	74	75	64	88	79	72	64	91	90	74	75	72	84
-4	82	68	86	77	96	78	98	84	99	104	126	97	102
-5	132	75	105	95	100	85	77	95	96	101	77	117	99
-6	100	79	127	99	102	89	85	97	102	116	115	83	111
-7	129	102	NaN	105	99	105	120	NaN	91	91	88	101	113
-8	#N/A	136	89	NaN	NaN	NaN	NaN	NaN	150	NaN	NaN	NaN	#N/A
-9	#N/A	#N/A	172	151	99	99	146	140	115	174	131	#N/A	#N/A
-10	#N/A	#N/A	#N/A	#N/A	#N/A	125	75	114	#N/A	#N/A	#N/A	#N/A	#N/A

Figure 20: Wafer map with two-point resistance measurement values for DC 2 of Channel a on an Interposer wafer with Cu micro-bumps.

DC 2 — Channel A — <90Ω <150Ω <300Ω >300Ω =open

Y / X	0	1	2	3	4	5	6	7	8	9	10	11	12
0	#N/A	#N/A	#N/A	#N/A	#N/A	108	149	138	#N/A	#N/A	#N/A	#N/A	#N/A
-1	#N/A	#N/A	172	155	129	185	113	170	144	148	182	#N/A	#N/A
-2	#N/A	154	136	149	205	125	135	158	144	130	130	143	#N/A
-3	136	129	115	137	172	134	106	137	176	133	147	141	164
-4	124	126	117	144	126	129	144	174	143	173	155	206	166
-5	177	149	170	190	161	117	117	111	135	105	145	136	222
-6	204	166	195	152	166	151	134	143	144	164	194	175	211
-7	191	189	187	203	162	172	161	173	206	206	184	167	221
-8	#N/A	205	143	NaN	158	167	179	198	230	174	147	177	#N/A
-9	#N/A	#N/A	235	179	153	215	179	222	184	211	184	#N/A	#N/A
-10	#N/A	#N/A	#N/A	#N/A	#N/A	216	159	212	#N/A	#N/A	#N/A	#N/A	#N/A

Figure 21: Wafer map with two-point resistance measurement values for DC 2 of Channel a on an Interposer wafer with Cu micro-bumps with a NiB cap.

Figure 21 shows the wafer-map for DC2 of Channel a on an Interposer wafer with 25μm-diameter Cu micro-bumps with NiB cap. All daisy-chains (apart from one) are continuous. The NiB cap clearly increases the daisy-chain resistance. For this particular wafer, the median daisy-chain resistance was 170Ω, resulting in 5.7Ω per micro-bump.

Test Phase 2 – Figure 22 shows the wafer-map for DC3 of Channel b on a Vesuvius wafer with 15μm-diameter Cu/Ni/Sn micro-bumps. There are 255 dies on this 300mm wafer; due to a technical error, the testing was aborted half-way the last-but-one row at the bottom of the wafer. Most daisy-chains are continuous. The non-continuous daisy-chains were confirmed (through other tests on the same dies) to be caused by wafer manufacturing issues on these particular dies. For this particular wafer, the median daisy-chain resistance was 98Ω, resulting in 3.3Ω per micro-bump.

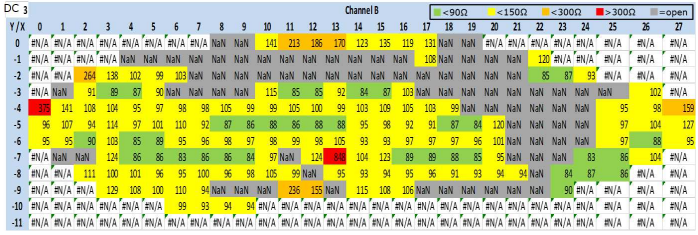


Figure 22: Wafer map with two-point resistance measurement values for DC 3 of Channel b on a Vesuvius wafer with non-reflowed Cu/Ni/Sn micro-bumps.

6.5 Probe Impact on Stack Interconnect Yield

In Test Phase 3, we verified the impact of the probe marks on stack interconnect yield. Table 1 lists the interconnect yields for 320 daisy-chains, 80 of each type. The table shows *no* significant impact. The differences between channels a–d are all explained by the variation in sheet resistance of the Interposer wires between Wide-I/O micro-bumps and post-bond probe pads, due to variations in lay-out locations of the micro-bumps.

Wide-I/O Channel	***a***	***b***	***c***	***d***
Vesuvius probed	no	yes	yes	no
Interposer probed	yes	yes	no	no
Interconnect yield	100%	100%	100%	100%
Daisychain resistance R	32.0Ω	42.4Ω	45.0Ω	33.1Ω
Std. deviation R	9.8Ω	5.8Ω	9.2Ω	8.2Ω

Table 1: Interconnect yield in Test Phase 3.

7 Cost Modeling Case Study

TU Delft and IMEC have developed a software tool named 3D-COSTAR to analyze product quality and test cost trade-offs in the many possible 3D test flows [22–25]. The tool uses as inputs lump-sum cost numbers for (1) design, (2) manufacturing, (3) test, (4) packaging, and (5) logistics. It models many different stacking approaches: simple linear 3D stacks, 2.5D stacks, complex multi-tower stacks, D2D/D2W/W2W stacking, etc. It assumes that no

manufacturing process is perfect and takes into account yields (in %) of die processing, interconnect layers, stacking, and packaging, as well as test coverage (in %) and test escape rates (in ppm). Furthermore, it attributes all costs made along the way to the end-of-line passing products.

We have used 3D-COSTAR to analyze the cost-effectiveness of our direct micro-bump probing approach, as alternative to performing pre-bond testing through dedicated pre-bond probe pads. In this cost modeling case study, we compare two scenarios.

1. *Probing through dedicated pre-bond probe pads.*
   These probe pads are by definition larger than the fine-pitch micro-bumps in order to allow probing on them with conventional probe technology. Consequently, they present a trade-off between the number of probe pads and corresponding silicon area on one hand, and the test input/output bandwidth provided and corresponding test time on the other hand (assuming a constant test data volume that needs to be pumped in and out of the device-under-test).

2. *Direct probing on micro-bumps.*
   This will require advanced (and hence expensive) probe cards, and hypothetically the micro-bump probe marks might decrease the interconnect yield after stacking.

Table 2 lists some of 3D-COSTAR's key cost model parameters. Note that there are many more parameters, which are not shown. The stack set-up and die sizes are inspired by the Vesuvius-2.5D test vehicle as described in Section 5: two active dies of 8.1×8.1=65.61mm^2 stacked on side-by-side on top of a passive interposer die of 10×20=200mm^2. We assume single-site testing on 300mm wafers with 3mm edge clearance. Unlike what was the case on our actual test wafers, we assume that the entire wafers are populated with only Vesuvius and Interposer dies respectively. The Interposer technology is assumed to be relatively cheap and mature; its defect density is fixed at 0.1 defects/cm^2. On the other hand, the active dies are assumed to be in an advanced technology node and hence have relatively expensive wafers; their defect density is varied from 0.0 to 1.0 defects/cm^2.

Parameter	**Interposer**	**Scenario 1** **Die 1+2**	**Scenario 2** **Die 1+2**
Pre-bond test contacts	n.a.	120	1200
300mm wafer cost	$ 700	$ 3000	$ 3000
Die area	200mm^2	66.61mm^2	65.61mm^2
Gross die / wafer	302	953	968
Defect density	0.1/cm^2	0.0–1.0/cm^2	0.0–1.0/cm^2
Die yield	84.52%	100–65.48%	100–65.76%
Pre-bond fault coverage	n.a.	99%	99%
Pre-bond test time	n.a.	100s	10s
Pre-bond probe card cost / die	n.a.	$ 0.00	$ 0.50
Pre-bond test cost	n.a.	$ 5.00	$ 1.00
Stack interconnect yield	100%	99%	98%
Final fault coverage	100%	99%	99%
Final test time	1s	10s	10s
Final test cost	$ 0.05	$ 0.50	$ 0.50

Table 2: Some key cost model parameters for the two test scenarios.

The case study concentrates on the pre-bond test of the active dies, and hence we modeled a test flow in which there is no pre-bond test for the Interposer die. We assume each of these active dies has a JEDEC Wide-I/O compliant micro-bump interface of 1,200 micro-bumps [19]. In Scenario 1, we are providing extra dedicated pre-bond probe pads. As we do not want to implement as many as 1,200 extra probe pads, we are assuming that we provide 120 extra probe pads only; we optimistically assume that this leads to only a $10\times$ increase in pre-bond test application time and hence test cost. In Scenario 2, we probe directly on the 1,200 micro-bumps. For this we need an expensive advanced probe card. We assume pessimistically its lifetime to be 100k touch-downs and its cost to be $50k. Assuming a single touch-down per die, this advanced probe card alone adds $0.50 costs to each die tested, on top of the assumed $0.05/s test cost. In addition, we pessimistically assume that its probe marks on the micro-bumps deteriorate the interconnect yield after stacking from 99% down to 98%. We have underlined the main differences between the two scenarios under comparison.

Figure 23 shows the test cost and the total stack cost per good stack for varying defect density of the two active dies for both scenarios. In all cases, direct probing on micro-bumps is cheaper than testing through dedicated pre-bond probe pads. The main differentiator for Scenario 1 is the $10\times$ increase in test time and hence test application costs, which makes pre-bond test a significant cost contribution in the overall stack cost price. Die yield works as a multiplier, as for example at 50% yield, two dies need to be manufactured and tested to find one good one, whose cost price should carry the cost of both dies. The increased area for dedicated pads, the expensive cost of the advanced probe card, and the (pessimistic) yield loss are all minor contributors in the overall cost calculation.

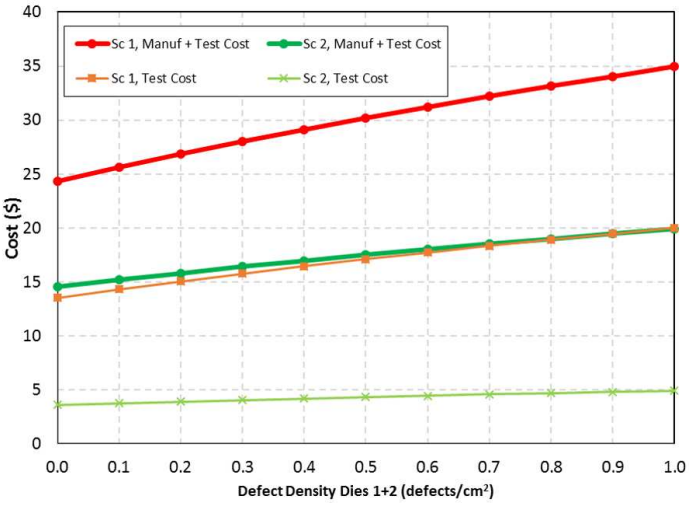


Figure 23: 3D-COSTAR test cost and total cost results.

Figure 24 shows the number of test escapes for both scenarios. In all cases, direct probing on micro-bumps results in a slightly lower number of test escapes except for the case where the defect density

value equals zero; here, the active dies have a 100% yield for both scenarios. Note that the interconnect yield, which is different for both scenarios, has no impact on the test escapes as the interconnect fault coverage is assumed to be 100% during final test.

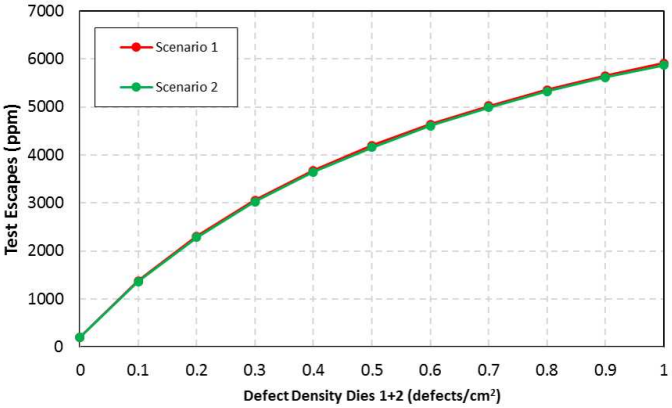


Figure 24: 3D-COSTAR test escape results.

8 Conclusion

In this paper we discussed direct probing of large-array fine-pitch micro-bumps in the context of 2.5D- and 3D-SICs. We have successfully conducted wafer-level direct probe experiments on single-channels of the JEDEC Wide-I/O Mobile DRAM interface, consisting of 6×50 arrays of 25μm-diameter Cu micro-bumps and 15μm-diameter Cu/Ni/Sn micro-bumps at 40/50μm pitches. Our experiments have shown the technical feasibility of the direct probing approach, with probe tips making proper electrical contact to the micro-bumps (i.e., contact resistance $<5\Omega$), causing only limited probe marks (i.e., probe mark profile $<$500nm, for Cu/Ni/Sn micro-bumps obtained with a post-probing reflow operation), and no measureable impact on stack interconnect yield. Our cost modeling indicates economical feasibility for single-site testing. The next step is to prepare this technology for volume production.

Acknowledgments

The authors thank many colleagues. At IMEC: Filip Beirnaert, Gerald Beyer, Eric Beyne, Kristof Croes, Miroslav Cupak, Inge De Preter, Jaber Derakhshandeh, Mikael Detalle, Joeri De Vos, Luc Dupas, Luke England, Antonio La Manna, Mireille Matterne, Julien Ryckaert, Michele Stucchi, George Vakanas, Marc Van Dievel, Geert Van der Plas, Joris Van Laer, and Dimitrios Velenis. At Cascade Microtech: Debbora Ahlgren, Juliane Busch, Claus Dietrich, Jens Fiedler, Gavin Fisher, Steve Harris, Ulf Hackius, Geert-Jan Hendricks, Torsten Kern, Clint VanderGiessen, and Eric Wilms.

References

[1] Erik Jan Marinissen. Challenges and Emerging Solutions in Testing TSV-Based 2.5D- and 3D-Stacked ICs. In *Proceedings Design, Automation, and Test in Europe (DATE)*, pages 1277–1282, March 2012. doi:10.1109/DATE.2012.6176689.

[2] Robert S. Patti. Three-Dimensional Integrated Circuits and the Future of System-on-Chip Designs. *Proceedings of the IEEE*, 94(6):1214–1224, June 2006. doi:10.1109/JPROC.2006.873612.

[3] Eric Beyne and Bart Swinnen. 3D System Integration Technologies. In *Proceedings of IEEE International Conference on Integrated Circuit Design and Technology (ICICDT)*, pages 1–3, June 2007. doi:10.1109/ICICDT.2007.4299568.

[4] Philip Garrou, Christopher Bower, and Peter Ramm, editors. *Handbook of 3D Integration – Technology and Applications of 3D Integrated Circuits*. Wiley-VCH, Weinheim, Germany, August 2008. ISBN 978-3-527-33265-6.

[5] Erik Jan Marinissen and Yervant Zorian. Testing 3D Chips Containing Through-Silicon Vias. In *Proceedings IEEE International Test Conference (ITC)*, November 2009. doi:10.1109/TEST.2009.5355573.

[6] Erik Jan Marinissen. Testing TSV-Based Three-Dimensional Stacked ICs. In *Proceedings Design, Automation, and Test in Europe (DATE)*, pages 1689–1694, March 2010. doi:10.1109/DATE.2010.5457087.

[7] Jung-Sik Kim et al. A 1.2V 12.8GB/s 2Gb Mobile Wide-I/O DRAM with 4×128 I/Os Using TSV-Based Stacking. In *Proceedings International Solid State Circuits Conference (ISSCC)*, pages 496–498, February 2011. doi:10.1109/ISSCC.2011.5746413.

[8] Christian Freund. Wide-IO DRAM – ST-Ericsson's First Mobile Processor Using TSV 3D-IC Technology. In *CDNLive! EMEA*, May 2011.

[9] Jung-Sik Kim et al. A 1.2 V 12.8 GB/s 2 Gb Mobile Wide-I/O DRAM With 4×128 I/Os Using TSV Based Stacking. *IEEE Journal of Solid-State Circuits*, 47(1):107–116, January 2012. doi:10.1109/JSSC.2011.2164731.

[10] Ben Eldridge and Marc Loranger. Challenges and Solutions for Testing of TSV and Micro-Bump. In *Digest of IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits (3D-TEST)*, September 2011. Paper 7.2, `http://3dtest.tttc-events.org`.

[11] Matt Losey et al. A Low-Force MEMS Probe Solution for Fine-Pitch 3D-SIC Wafer Test. In *Digest of IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits (3D-TEST)*, September 2011. Paper 7.3, `http://3dtest.tttc-events.org`.

[12] Onnik Yaglioglu and Ben Eldridge. Direct Connection and Testing of TSV and Microbump Devices Using NanoPierce Contactor for 3D-IC Integration. In *Proceedings IEEE VLSI Test Symposium (VTS)*, pages 96–101, May 2012. doi:10.1109/VTS.2012.6231086.

[13] Joseph Foerstel and Amy Leong. 40μm Pitch Probing Evaluation. In *Digest of IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits (3D-TEST)*, November 2012. Paper 5.1, `http://3dtest.tttc-events.org`.

[14] Gunther Böhm et al. Very Small Pitch Micro Bump Array Probing. In *Proceedings IEEE South-West Test Workshop (SWTW)*, June 2013. `http://www.swtest.org/swtw_library/2013proc/swtw2013.html`.

[15] Onnik Yaglioglu and Ben Eldridge. Contact Testing of Copper Micro-Pillars with Very Low Damage for 3D IC Assembly. In *Proceedings IEEE International Conference on 3D System Integration (3DIC)*, pages 1–4, October 2013. doi:10.1109/3DIC.2013.6702361.

[16] Ken Smith et al. KGD Probing of TSVs at 40μm Array Pitch. In *Digest of IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits (3D-TEST)*, November 2010. Paper 4.1, `http://3dtest.tttc-events.org`.

[17] Erik Jan Marinissen et al. Wafer Probing on Fine-Pitch Micro-Bumps for 2.5D- and 3D-SICs. In *Proceedings IEEE South-West Test Workshop (SWTW)*, June 2011. `http://www.swtest.org/swtw_library/2011proc/swtw2011.html`.

[18] Ken Smith et al. Evaluation of TSV and Micro-Bump Probing for Wide I/O Testing. In *Proceedings IEEE International Test Conference (ITC)*, pages 1–10, September 2011. doi:10.1109/TEST.2011.6139180.

[19] JEDEC. *Wide I/O Single Data Rate (JEDEC Standard JESD229)*. JEDEC Solid State Technology Association, December 2011. `http://www.jedec.org`.

[20] Sergej Deutsch et al. DfT Architecture and ATPG for Interconnect Tests of JEDEC Wide-I/O Memory-on-Logic Die Stacks. In *Proceedings IEEE International Test Conference (ITC)*, pages 1–10, November 2012. doi:10.1109/TEST.2012.6401569.

[21] Sandeep K. Goel et al. Test and Debug Strategy for TSMC CoWoS Stacking Process Based Heterogeneous 3D IC: A Silicon Case Study. In *Proceedings IEEE International Test Conference (ITC)*, pages 1–10, September 2013. doi:10.1109/TEST.2013.6651893.

[22] Mottaqiallah Taouil et al. 3D-COSTAR: A Cost Model for 3D-SICs. In *Digest of IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits (3D-TEST)*, November 2012. Paper 10.2, `http://3dtest.tttc-events.org`.

[23] Mottaqiallah Taouil et al. 3D-COSTAR: A Cost Model for 3D-SICs. In *3-D Architectures for Semiconductor Integration and Packaging (3D-ASIP)*, December 2012.

[24] Mottaqiallah Taouil et al. Impact of Mid-Bond Testing in 3D Stacked ICs. In *Proceedings IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT)*, pages 178–183, October 2013. doi:10.1109/DFT.2013.6653603.

[25] Mottaqiallah Taouil et al. Using 3D-COSTAR for 2.5D Test Cost Optimization. In *Proceedings IEEE International Conference on 3D System Integration (3DIC)*, pages 1–8, October 2013. doi:10.1109/3DIC.2013.6702351.

[26] Joeri De Vos et al. Key Elements for Sub-50μm Pitch Micro Bump Processes. In *Proceedings IEEE Electronic Components and Technology Conference (ECTC)*, pages 1122–1126, May 2013. doi:10.1109/ECTC.2013.6575714.

[27] Mikael Detalle et al. Interposer Technology for High Bandwidth Interconnect Applications. In *Proceedings IEEE Electronic Components and Technology Conference (ECTC)*, pages 323–328, May 2013. doi:10.1109/ECTC.2013.6575590.

[28] Erik Jan Marinissen et al. Vesuvius-3D: A 3D-DfT Demonstrator. In *Proceedings IEEE International Test Conference (ITC)*, October 2014. Paper 20.2.

[29] Luke England et al. NiB Capping of Cu Landing Pads for Thermocompression Bonding. In *Workshop on Materials for Advanced Metallization*, March 2014.

[30] Jaber Derakhshandeh et al. Reflow Process Optimization for Micro-Bumps Applications in 3D Technology. In *Proceedings IEEE Electronics System-Integration Technology Conference (ESTC)*, September 2014.