Overview and Outlook of 3D IC Packaging, 3D Si Integration, and 3D IC Integration

John H. Lau
ASM Pacific Technology
16-22 Kung Yip Street, Kwai Chung, N.T., Hong Kong
852-2619-2757, john.lau@asmpt.com

ABSTRACT

3D integration consists of 3D IC packaging, 3D Si integration, and 3D IC integration. They are different and in general the TSV (through-silicon via) separates 3D IC packaging from 3D Si/IC integrations since the latter two use TSV but 3D IC packaging does not. 3D Si integration and 3D IC integration are different. 3D IC integration stacks up the thin chips with TSV and microbump, while 3D Si integration stacks up thin wafers with TSV alone (i.e., bumpless). TSV is the heart of 3D Si/IC integrations and is the focus of this investigation. Also, the state-of-the-art, challenge, and trend of 3D integration will be presented and examined. Furthermore, supply chain readiness for high-volume manufacturing of TSVs is discussed.

I. INTRODUCTION

The Electronics Industry has been the largest industry since 1996 and may well reach 1.6 trillion dollars (i.e., $1.5x10^{12}) by the end of 2015 [1, 2, 3]. The most important invention of the Electronics Industry is, arguably the transistor (1947), which earned John Bardeen, Walter Brattain, and William Shockley the 1956 Nobel Prize in Physics. The invention of the integrated circuit (IC) by Jack Kilby in 1958 (which earned him the 2000 Nobel Prize in Physics), and six months later by Robert Noyce (who didn't share the Nobel Prize with Jack Kilby because he passed away in 1990) excited the generations of IC integrations.

The proposal of doubling the number of transistors on an IC chip (for minimum costs and innovations) every 24 months by Gordon Moore in 1965 (also called Moore's law) [4] has been the most powerful driver for the development of the microelectronics industry in the past ~50 years. This law emphasizes lithography scaling and integration (on a 2D surface) of all functions on a single chip, perhaps through system-on-chip (SoC). On the other hand, the integration of all these functions can be achieved through 3D integrations such as 3D IC packaging [1], 3D IC integration [1-3], [5-70], [107-125] and 3D Si integration [1-3], [69-125] as shown in Figure 1 [1-3].

![Figure 1 3D integration technologies vs. maturity](image-url)
TSV was invented more than 50 years ago [1, 126] by the 1956 Nobel Laureate in Physics, William Shockley. (Yes, the same Shockley who co-invented the transistor, which is generally considered the greatest invention in Semiconductor industry.) He filed the patent “Semiconductive wafer and method of making the same” on October 23, 1958, and as granted the U.S. patent (3,044,909) on July 17, 1962. One of the key claims is shown in Figure 2, which gets the Semiconductor world so excited today. Basically, the “deep pits” (which are called TSVs today) on the wafer allow the signals from its top side to its bottom side and vice versa.

TSV is the heart of 3D Si integration and 3D IC Integration [1, 127]. It provides the opportunity for the shortest chip-to-chip interconnects and the smallest pad size and pitch of interconnects. Compared with other interconnection technologies, such as wire bonding, the advantages of TSV include: (a) better electrical performance, (b) lower power consumption, (c) wider data width and thus bandwidth, (d) higher density, (e) smaller form factor, and (f) lighter weight, [1-3] and [5-125].

TSV is a disruptive technology. As with all disruptive technologies, the questions to ask are: “What is it displacing?” and “What’s the cost?” Unfortunately, TSV is trying to displace the wire-bonding technology, which is a most mature, high-yield, and low-cost technology [128]. However, just like solder-bumped flip-chip technology [129, 130], because of their unique advantages, TSVs will be here to stay and for a very long time for high-performance and high-density applications.

TSV has been in volume production for MEMS (micro-electro-mechanical systems) [131, 132] and CMOS (complementary metal-oxide-semiconductor) image sensor [133, 134]. However, they are out of the scope of this study, which is focused on memory, logic, processor, and SoC.

3D (IC and Si) integration is a very old idea [69, 70] which consists of two or more layers of active electronic components that are integrated vertically through TSV (it used to be called vertical interconnection) into a single circuit. It was triggered by the advance of the silicon-on-insulator (SOI) technology first reported by Gat and his colleagues more than 35 years ago [135], when semiconductor people thought Moore’s law could be hitting the wall by the 1990s. Of course, the fact showed otherwise.

3D IC integration is to stack up the thin chips with TSVs and microbumps. While 3D Si integration is to stack up the thin wafers/chips with TSVs alone, i.e., bumpless. The advantages of 3D Si integration over 3D IC integration are: (1) better electrical performance, (2) less power consumption, (3) lower profile, (4) less weight, and (5) higher throughput.

The most powerful proponent on 3D IC/Si integration is the 1965 Nobel Physics laureate, Richard
Feynman. Almost thirty (30) years ago, during his lecture, *Computing Machines in the Future* in 1985, said he: “Another direction of improvement (of computing power) is to make physical machines three dimensional instead of all on a surface of a chip. That can be done in stages instead of all at once – you can have several layers and then add many more layers as time goes on.” In this study, the overview, challenge, and outlook of 3D IC integration, and 3D Si integration will be presented and discussed. 3D IC packaging will be briefly mentioned first.

II. OVERVIEW AND OUTLOOKS OF 3D IC PACKAGING

Figure 1 shows that chip stacking by wire bonding and package-on-package (PoP) are now mature for high volume manufacturing (HVM). Chip-to-chip interconnects and 3D fan-out embedded wafer-level packaging are potential candidates for manufacturing.

(a) Chip Stacking by Wire Bonding

The first paper on stacking of memory chips in 3D by die-attach material and Au wire bonding was published by nCHIP [136] more than 20 years ago. Since then, memory chip (especially the NAND Flash) stacking by Au wire bonding has been in high volume production for, e.g., the smartphones and tablets. Because of the surge in Au prices and research and development progress in Cu wire-bonding technology, many companies have been looking for low-cost solutions, and the shift from Au to Cu wire bonding as shown in Figure 3 is genuinely picking up.

![Figure 3 AMKOR's 3D IC packaging (stacked chips by Cu wirebonding)](image)

(b) Package-on-Package (PoP)

PoP comes from many different forms. Figure 4 shows a wirebond package on top of a flip chip package. It can be seen that the top package consists of two chips cross stacked and wire bonded on a package substrate and then over molded. The bottom package consists of a solder bumped flip chip on another package substrate with underfill. All these package substrates are with solder balls. Again, PoP is in high volume production for, e.g., the smartphones and tablets [6].

![Figure 4 3D wirebonding package on flip chip package. TOP: Schematic, BOTTOM: Photo image](image)
(c) Chip-to-Chip Interconnects
Figure 5 shows the schematic of a 3D chip-to-chip interconnect [137] developed by IME. It consists of the mother chip which is face-to-face connected to a daughter chip. The backside of the mother chip can be attached to a heat spreader (and a heat sink if necessary). The whole module is attached (through the flip chip mother die) to a rigid or flexible substrate. It is a very cost-effective 3D IC package without using the TSVs [137]. In 2012, SONY’s PlayStation (CXD53135GG) attached Samsung’s 1Gb wide I/O SDRAM (synchronous dynamic random access memory) face-to-face to the processor and then wire bonded to the next level interconnects.

![Figure 5 3D IC packaging (chip-to-chip interconnects) [13]](image)

Figure 5 3D IC packaging (chip-to-chip interconnects) [13]

Figure 6 shows Amkor’s double POSSUM™ multi-stacked die configurations [138] without the use of TSVs. It can be seen that the grandma die is supporting the mother die, which is supporting the 3 daughter dies. The grandma die is solder bumped flip chip on a package substrate, which is then attached on to a PCB (printed circuit board). The packages shown in Figures 5 and 6 are not in manufacturing yet, however they are the potential candidates for medium-range performance applications.

![Figure 6 3D IC packaging (Amkor’s multiple chip-to-chip interconnects) [138]](image)

Figure 6 3D IC packaging (Amkor’s multiple chip-to-chip interconnects) [138]

(d) 3D Fan-Out Embedded Wafer-Level Package
Figure 7 shows the cross section SEM (scanning electron microscopy) images of a 3D fan-out embedded wafer-level package [139] developed by StatsChipPAC. It consists of a bottom package which is a fan-out embedded wafer-level ball grid array package (eWLB) and a top package which is a memory package. It can be seen that (a) the eWLP contains the logic, baseband, or application processor, (b) the eWLB is only 450μm thick, (c) the top package is 520μm thick and is housing the memory chips with wire bonding, and (d) the interconnection is from the PCB, solder balls, RDLs (redistribution layers), to processor, and solder balls, RDL, to the memory chips. This package is a potential candidate for mobile
and wearable products.

![Figure 7 3D IC packaging (STATSChipPAC’s package-on-package) [139]](image)

(e) Outlook of 3D IC Packaging

The outlook of 3D IC packaging is great! Stacked dies with wirebonding and PoP are in HVM for commercial products such as smartphones and tablets. Chip-to-chip interconnects are going to follow SONY and get into manufacturing soon. 3D fan-out embedded wafer-level package is also a strong candidate for production. All these 3D packaging technologies have been keeping 3D IC/Si integration technologies away from HVM.

III. OVERVIEW, CHALLENGES, AND OUTLOOK OF 3D Si INTEGRATION

Basically, wafer-to-wafer (W2W) is the only way to perform the bonding operation for 3D Si integration and yield is a big issue (e.g., some bad chips are forced to bond on the good chips). In addition, the absence of (or an infinitesimal) gap between wafers and thermal management could be a problem. Furthermore, the requirements of the bonding conditions, such as the surface cleanness, surface flatness, and the class of clean room for 3D Si integration are very high.

![Figure 8 3D Si integration: (L) IBM/RPI’s Cu-to-Cu bonding [71]. (R) NIMS/AIST/Toshiba/University of Tokyo’s Cu-to-Cu bonding [74]](image)

There are at least two different W2W bonding methods for 3D Si integration, namely, Cu-to-Cu bonding and oxide-to-oxide bonding, as shown in Figures 8 and 9, respectively. In general, for Cu-to-Cu bonding, the TSVs have to be fabricated before bonding. On the other hand, for oxide-to-oxide bonding, the TSVs are fabricated after bonding. Figure 8a shows a high-quality bonding interface by IBM and RPI [71 - 73]. Before bonding, the Cu interconnects (pads) are fabricated with the standard back-end-of-line damascene process, followed by the oxide CMP (chemical-mechanical polishing) process (oxide touch-up) to recess the oxide level to 40 nm lower than the Cu surface. The bonding temperature is ramped up to 400°C. Figure 8b shows a cross-section of the interface between the bumpless Cu-to-Cu electrodes (pads) given by the NIMS/AIST/Toshiba/University of Tokyo [74-80].
Figure 9 3D Si integration: (a) MIT’s oxide-to-oxide bonding [81-87], (b) Leti/Freescale/ STMicroelectronics’ oxide-to-oxide bonding [88-89]

Figure 9a shows a cross section of MIT’s oxide-to-oxide bonding structure of three-layer 3D (ring oscillator) bonded at 275°C [81-87]. It can be seen that: (1) the layers are bonded and interconnected with W-plugs, (2) the conventional inter level connections are in the bottom two layers, and (3) the 3D vias are located in the isolation (field) region between transistors. Figure 9b shows Leti/Freescale/STMicroelectronics’ dielectric-to-dielectric bonding structure of two device layers bonded at ~400°C [88-90]. It can be seen that: (a) first, a metal level is formed on a 200 mm bulk wafer and SOI wafer; next, these wafers are bonded face-to-face, and then the bulk silicon of the SOI wafer is removed down to the BOX (buried oxide) layer; (b) the ~1.5μm interstrata vias (ISVs) are formed, which make contact from upper strata to lower strata; (c) a metal layer is formed at the top of the back side of the SOI wafer, and (d) this ISV makes contact with both the top and bottom metal layers.

In order to use the 3D Si integration technology to HVM products, many research and development efforts have to be performed. Besides thermal management, vias formation, thin-wafer handling, more research and development efforts should also be placed on areas such as: cost reduction, design and process parameter optimization, bonding environment, W2W bonding alignment, wafer distortion, wafer bow (warpage), inspection and testing, contact performance, contact integrity, contact reliability, and manufacturing yield issues. In addition, packaging the 3D Si integration module systematically and reliably to the next level of interconnect pose another great challenge.

Besides technology issues just mentioned, the EDA (electronic design automation), which is the soul of 3D Si integration [127] is far from ready. Urgently, the industry needs to build standard and infrastructure and form an ecosystem for 3D Si integration. Then, the EDA can write the design, simulation, analysis & verification, manufacturing preparation, and test software with the following guidelines: (1) design automation from high level description to layout generation/optimization, (2) verification all dedicated and tuned to 3D integration, (3) addressing the 3rd dimension not like a packaging bumping, (4) addressing the true 3rd dimension, with partitioning, floor planning, automatic placing and routing, (5) full extraction with the 3rd dimension, full 3D DRC (design rule checks), 3D LVS (layout vs. schematic) with all tiers together in a same database, and (6) the 3D integrations have then to be seen as a whole system distributed in several tiers, and not just a stack of predefined chips.

In the next 10 years, the industry will be hard-pressed for HVM with the 3D Si integration technology, except for very niche applications. However, it should be noted and emphasized that 3D Si integration is
the right way to go and compete with Moore’s law. The industry should strive to make this happen!

IV. OVERVIEW, CHALLENGES, AND OUTLOOK OF 3D IC INTEGRATION

Unlike 3D Si integration, 3D IC Integration stacks up thin IC chips in the 3rd dimension with TSVs and microbumps (<25µm) [1] to achieve performance, low power consumption, wide bandwidth, and small form factor. The ones which are in and going into low volume production are: memory stacking with TSVs, hybrid memory cube (HMC) or wide I/O (input/output) DRAM, wide I/O DRAM 2, high bandwidth memory (HBM), and 2.5D IC integration (passive interposer).

(a) Memory Stacking with TSVs

Samsung mass-produced (August 2014) industry’s first TSV-based 64GB DDR4 (double data rate type 4) DRAM module which consists of 36 DDR4 DRAM chips, each of which consists of four 4-gigabit (Gb) DDR4 DRAM dies. The module performs twice as fast as a module that uses wire bonding packaging, while consuming approximately half the power. The module is for server application.

(b) Hybrid Memory Cube (HMC)

Figure 10 shows the very first sample [140] shipped by Micron/IBM at the end of September 2013. It is a hybrid memory cube (HMC) which consists of 4 DRAMs each with 2000+ TSVs stacking on top of a logic controller with TSVs. The HMC is then attached to an organic package substrate. The TSV DRAM cube is fabricated by Micron and the TSV controller is fabricated by IBM. The microbumps are Cu pillars (20µm-tall) with solder caps.

At the 2014 International Supercomputing Conference it was announced that Intel “Knights Landing” processor unit would debut in 2015. It will support for up to 384GB of on board DDR4 (double data rate type 4) RAM and 16GB of Micron HMC stacked DRAM on-package, providing up to 500GB/sec of memory bandwidth (Figure 11). Micron reports that having such HMC in the CPU (central processor unit)
package is expected to deliver 5X the sustained memory bandwidth vs. GDDR5 (graphics double data rate type 5) with one-third the energy per bit in half the footprint.

Figure 12 Schematic of JEDEX’s Wide I/O 2. Patent of quadrants of microbumps

(c) Wide I/O DRAM and Wide I/O 2
JEDEC standard (JESD229) [141], Wide I/O Single Data Rate (Wide I/O SDR), was published in December 2011 and JEDEC standard (JESD229-2) [141], Wide I/O 2 (WideIO2), was published in August 2014. They are meant for a stack of DRAMs with TSVs on a logic controller with TSVs; very similar to the HMC. The microbumps are divided into 4 quadrants with signal assignments mirrored both horizontally and vertically as shown in Figure 12, where the bump pitch (40μm) of the area array is also shown. The dimensions of each quadrant are 2880μm x 200μm. There will be a space between quadrants in the x-direction (1000μm) and in the y-direction (120μm) [141].

Figure 13 Schematic of JEDEX/Hynix’s High Bandwidth Memory (HBM)

(d) High Bandwidth Memory (HBM)
Figure 13 shows schematically a high bandwidth memory system developed by Hynix/AMD, which is based on JEDEC standard (JESD235) [142], High Bandwidth Memory (HBM) DRAM, published in December 2013. It is meant for graphics applications supporting bandwidth from 128GB/s to 256GB/s. A TSV/RDL interposer is used to support/connect mainly the lateral communication (HBM interface) between the HBM DRAM memory cube with TSVs and the SoC such as graphic processor unit (GPU) or central processor unit (CPU) without TSVs. The optional base chip is used for buffering and signal re-routing of the HBM DRAM cube.
(e) Passive Interposer (2.5D IC Integration)

A 2.5D IC integration is a TSV/RDL interposer system which consists of a piece of device-less silicon with TSVs, RDLs, and IC chips without TSVs. This piece of device-less TSV silicon (also called a passive interposer) is used to support the high-performance, high-density, fine-pitch chips and has RDLs (mainly) for lateral communication between the chips as shown schematically in Figure 14. Figure 15 shows a sample designed and fabricated by Altera/TSMC [41, 42]. It can be seen that even with more than 12 buildup layers (6-2-6) on the package substrate, it is still not enough to support the four sliced 28nm FPGA (field-programmable gate array) chips. In addition, a passive TSV interposer with 200,000+ microbumps on 45μm-pitch and four RDLs (three Cu damascene layers and one aluminum layer) at a minimum of 0.4μm-pitch is needed. This type of structure (Figures 14 and 15) is called by TSMC as chip on (interposer) wafer on (package) substrate (CoWoS) and has been in small production for Xilinx since the early of 2013.

![Figure 14 TSV/RDL passive interposer supporting chips on package substrate](image1)

**Figure 14 TSV/RDL passive interposer supporting chips on package substrate**

**Figure 15 Altera/TSMC's chips on wafer on substrate (CoWoS) [42]**

![Figure 16 TSV fabrication process flow](image2)

**Figure 16 TSV fabrication process flow**
(e1) Fabrication of TSVs
The fabrication process of TSVs for interposer is shown in Figure 16. The process starts with a SiN$_x$/SiO$_x$ insulation layer by either thermal oxidation or PECVD (plasma enhanced chemical vapor deposition) as shown in Figure 16. After photoresist and TSV lithography, the TSV is etched into the Si substrate by Bosch-type DRIE (deep reactive ion etch) [45] to form a high aspect ratio (10.5) via structure. The etched TSV structure is then processed with a SiO$_x$ liner by SACVD (subatmosphere chemical vapor deposition), a Ta barrier layer, and a Cu seed layer by PVD (physical vapor deposition) [17]. Cu electroplating is used to fill the TSV structure. The final blind TSV has a top opening of approximately 10μm in diameter and a depth of about 105μm, which gives an aspect ratio of 10.5. In such a high aspect ratio via structure, a bottom-up plating mechanism is applied to ensure a seamless TSV with a reasonably low Cu thickness in the field.

![Figure 17 SEM images of TSV cross sections [5]](image)

The SEM cross-sectional images are shown in Figure 17. It can be seen that the diameter of the TSV is slightly decreased at the bottom, which is expected from the etching process. The Cu thickness at the field is <5μm. The post plating anneal is at 400°C for 30 min. To complete the TSV process, excess Cu in the field is removed by CMP [13].

(e2) Fabrication of RDLs
There are at least two ways to fabricate RDLs [1, 2, 3, 5]. The first method is by using polymers, such as polyimide (PI) PWDC 1000 (Dow Corning), benzocyclobutene (BCB) cyclotene 4024-40 (Dow Chemical), polybenzo-bisoxazole (PBO) HD-8930 (HD Micro Systems), and the fluorinated aromatic AL-X 2010 (Asahi Glass Corporation) to make the passivation layer and electroplating (such as Cu) to make the metal layers. This method has been used by the OSAT (outsourced semiconductor assembly and test) to fabricate RDLs (without using semiconductor equipment) for wafer-level (fan-in) chip scale package [134], embedded wafer-level (fan-out) ball grid array package [143 - 147], and (fan-out) redistribution chip package [148, 149]. The second method is the Cu damascene method, which is primarily modified from the conventional semiconductor back-end-of-line to make the Cu metal RDLs such as those shown in Figure 15. In general, much thinner structures (both dielectric layers and Cu RDLs), finer pitches, smaller line-widths, and spacing can be obtained with the dual Cu damascene method, which will be briefly stated in the following.

If starting with the wafer from Figure 16, the fabrication process of RDLs with a dual Cu damascene technique is primarily based on the semiconductor back-end-of-line process. The details are shown in Figure 18 and listed in the following [5]: (1) SiO$_x$ layer by PECVD, (2) apply photoresist and mask, then use photolithography techniques (align and expose) to open vias on the SiO$_x$, (3) RIE (reactive ion etch) of SiO$_x$, (4) strip off portion of the photoresist, (5) repeat step 3, (6) strip off the photoresist, (7) sputter Ti and Cu and electroplate Cu over the entire wafer, (8) CMP the Cu and Ti/Cu and RDL1 is completed,
and (9) repeat step 1 through step 8 to complete RDL2 and any additional layers.

Figure 18 Process flow for fabricating RDLs by dual Cu damascene

SEM images of the RDL cross sections fabricated by the Cu damascene technique are shown in Figure 19. The minimum RDL line width is 3 μm. The thickness of RDL1 and RDL2 is 2.6 μm and of RDL3 is 1.3 μm. The passivation thickness between RDLs is 1 μm.

Figure 19 SEM images of cross sections of RDLs fabricated by the Cu damascene method [5]

(c3) Backside Processing and Assembly
The process flow of backside and assembly [5] is shown in Figure 20. It can be seen that after the fabrication of TSV, RDLs, passivation, and UBM (under bump metallurgy), the topside of the interposer

Figure 20 Conventional process flow for chip on interposer wafer on package substrate [8]
wafer is temporary bonded to a carrier by adhesive. The next step is backgrinding the interposer wafer, Si etching, low temperature passivation, and Cu revealing. Next, backside RDL (optional), UBM, and C4 (controlled collapse chip connection) wafer bumping are carried out. After that, the next step is to temporary bond another carrier wafer to the backside (with solder bumps) and de-bond the first carrier wafer. This step is followed by chip-on-wafer bonding and underfilling. After the whole (chip-on) interposer wafer is completed, the next step is to de-bond the second carrier wafer and transfer the thin interposer wafer with attached chips to a dicing tape for singulation. The individual TSV/RDL interposer with chips is attached to the package substrate by natural reflow and then underfilled.

(e4) Cu Revealing
Figure 21 shows more details on Cu revealing. Right after the temporary bonding of the support carrier, backgrinding the wafer to a few microns to the tip of TSVs, Si dry etching (by RIE) to a few microns below the tip of TSVs, and low-temperature passivating the SiN/SiO₂ are performed. Then, CMP for SiN/SiO₂ buffing and barrier and Cu seed layers polishing are carried out. Cu revealing is completed and shown in Figure 22 [5]. These processes also apply to device TSV wafers.

(f) Outlook of 2.5D/3D IC Integration
TSVs straight through the same memory chips, e.g., DRAMs to enlarge the memory capacity, increase the bandwidth, lower the power consumption, lower the latency (enhance the electrical performance), and reduce the form factor is the right thing to do and will be the major application of 3D IC integration. Besides Intel, HMC samples with capacities of 4GB and 8GB have already shipped to server and chip companies for testing. The first HMC modules will be used on FPGAs. HMC is targeted for high performance computing, cloud computing, in-memory database, networking, energy, wireless communications, transportation, security, and high-end servers and the potential customers are Intel,
Altera, Fujitsu, Cray, Cisco, Huawei, Xilinx, HP, etc.

TSV/RDL interposer (2.5D IC integration) is in small volume manufacturing by Xilinx/TSMC for the sliced FPGAs, which cannot be supported by the package substrate even with 12 build-up layers. Thus, interposer is for very high performance, high density, high I/O, and fine-pitch applications such as networking, communications, high-end servers, etc.

V. HVM SUPPLY CHAINS FOR TSVS AND MEOL
(a) Supply Chains before the TSV Era
Before the TSV era, the technology supply chains are very well defined and understood. Descriptions of the various entities comprising the supply chain before the TSV era are presented below.

(a1) FEOL (front-end-of-line)
This is the first portion of IC fabrication where the individual devices such as transistors or resistors are patterned. This process is from a bare wafer to (but not including) the deposition of metal layers. FEOL is usually performed in semiconductor fabrication plants (fabs).

(a2) BEOL (back-end-of-line)
This is the fabrication in which active devices are interconnected with wiring on the wafer. This process starts from the first layer of metal to bonding pads with passivation. It also includes insulators and metal contacts and is called MOL (middle-of-the-line). The term “MOL” is seldom used and embedded in the BEOL. Again, BEOL is usually performed in the fabs.

(a3) OSAT (outsourced semiconductor assembly and test)
This term is also called packaging, assembly and test. The process starts when the passivated wafer is received from the fab and then goes through circuit probing, bumping, thinning, dicing, wiring bonding, flip-chip, molding, ball mounting, final testing, and etc.

(b) Supply Chains for the TSV Era – Who Makes the TSV?
The following steps in the TSV fabrication process impact the various considerations that must be addressed:

(b1) TSVs Fabricated by the Via-First Process
The TSVs are fabricated before the FEOL. This can only be done by the fab. However, even in the fab, this seldom happens because the devices (e.g., transistors) are much more important than the TSVs.

(b2) TSVs Fabricated by the Via-Middle Process
The TSVs are fabricated right after the FEOL (e.g., transistors) and MOL (e.g., metal contacts), and before the BEOL (e.g., metal layers). In this case, the MOL is no longer embedded in the BEOL because the TSV fabrication process is between them. Owing to logistics and equipment compatibilities, usually the TSV by the via-middle process is done by the fab.

(b3) TSVs Fabricated by the Via-Last (From the Frontside) Process
The TSVs are fabricated (from the frontside of the wafer) after the FEOL, MOL, and BEOL. As of today, there is not a single creditable paper published with this process.

(b4) TSVs Fabricated by the Via-Last (From the Backside) Process
The TSVs are fabricated (from the backside of the wafer) after the FEOL, MOL, and BEOL process flows. The CMOS image sensor is an example. Strictly speaking, CMOS image sensors are not examples of 3D IC integration. For CMOS device sample wafers, the only creditable publication is given by LETI, et al., [150]. However, because of technical issues, such as hitting the various embedded alignment
targets in the x-, y- and z-directions (to enable the alignment between the metal layers on the topside of the wafer and the positioning of TSVs formed from the backside), TSVs fabricated by the via-last (from the backside) process should be avoided until these issues are resolved.

Based on the above discussions, it seems that for active device wafers being used for 3D IC integration applications, TSVs are better fabricated using the via-middle process. Also, the TSVs should be fabricated by the fabs, where all the equipment and expertise already exist and the cost to fabricate the TSVs is less than 5% of the cost in fabricating the (∆32nm) device wafers!

(b5) How About the Passive TSV Interposers?
When the industry defined the TSV processes for 3D IC integration, there were no passive interposers yet. Also, since there is no active device in the passive interposers, thus they don’t fit into any of the preceding!

(b6) Who Wants to Fabricate the TSV for Passive Interposers?
Both the fab and OSAT want to do it! It depends on the layout, design, and fabrication capabilities, especially the line width and spacing of the RDLs and the diameter of the TSVs. Usually, for a few microns of line width and spacing of the RDL and ≥5μm of TSV diameter it can be done by the OSAT. Otherwise, it should be done by the fabs.

(c) Supply Chains for the TSV Era - Who Does the MEOL?
For the thicknesses of memory-chip stacking and DRAMs in HMC and HBM, and interposers under consideration, all the TSVs fabricated are blind vias. The blind TSV wafer is followed by temporary bonding, backgrinding, TSV revealing, thin wafer handling, debonding, cleaning, solder bumping, etc. which, taken together, are called MEOL (middle-end-of-line). Except for the vertically integrated companies (e.g., TSMC and Samsung), it is better for the MEOL process flow to be performed by the OSAT.

(d) Outlook of HVM Supply Chains for TSVs and MEOL
For device wafers, the TSVs should be fabricated by the via-middle process and manufactured by the fabs. For deviceless wafers, it depends on the line width/spacing of the RDLs and the diameter of the TSVs. As to the MEOL, for both device and deviceless wafers, it should be done by the OSAT.

VI. SUMMARY AND RECOMMENDATIONS
The overview, challenge, and outlook of 3D IC packaging, 3D IC integration, and 3D Si integration have been presented and discussed. The fabrication processes of TSVs and RDLs have also been mentioned. Furthermore, the supply chains of TSVs and 3D IC integration at HVM have been examined. Some important results and recommendations are summarized in the following.

- The driving forces for consumer products such as smartphones, tablets, and wearables are cost, cost, and cost. The cost-effective 3D IC packaging such as the stacked dies by wirebonding, PoP, chip-to-chip interconnect, and 3D fan-out embedded wafer-level packaging are just the right technologies for these products.

- The driving forces for high performance computing, cloud computing, networking, wireless communications, high-end servers are performance and reliability. The considerable high-cost 2.5D/3D IC integrations such as memory chip stacking with TSVs, HMC, HBM, wide I/O 2, and passive interposer are the right technologies for these products.

- For device wafers, the TSVs should be fabricated by the via-middle process and manufactured by the fabs. For interposer wafers, if the diameter of the TSVs is ≥5μm and the line
width/spacing of the RDLs is \( \geq 3\mu m \) then it can be manufactured by the fabs and OSAT, otherwise it should be done by the fabs. However, since most interposers are for very high performance, high density, and fine-pitch applications, thus the TSV diameters and RDLs line width and spacing are most likely falling into the fabs’ territory.

➤ As to the MEOL, assembly, and test of both the TSV device and deviceless wafers they should be performed by the OSAT except the vertical integrated companies such as TSMC and Samsung. There are many important tasks in MEOL, assembly, and test, thus the ODAT should strive to make themselves ready for a robust and high-yield manufacturing process.

VII. ACKNOWLEDGMENTS
The author would like to thank his colleagues at Institute of Microelectronics, Hong Kong University Science and Technology, Industrial Technology Research Institute, and ASM Pacific Technology for their stimulating and fruitful discussions.

XIII. REFERENCES


141. JEDEC Standard, JESD229, Wide I/O Single Data Rate (Wide I/O SDR), December 2011. Also, JEDEC Standard, JESD229-2, Wide I/O 2 (WideI/O2), August 2014.


