# Chapter 11
# 3D IC Heterogeneous Integration by FOWLP


## 11.1 Introduction

Two 3D IC heterogeneous integrations by Fan-Out Wafer-Level Packaging (FOWLP) technology are presented in this chapter. The emphasis of the first such method is on the design, and of the other method, the emphasis is on the manufacturing process. The heterogeneous integration versus SoC (system-on-chip) will be briefly discussed. Some examples on the TSV (Through-Silicon Via)-less heterogeneous integration by FOWLP will also be presented. Since MCM (multichip module) is the *frontier* of heterogeneous integration and thus it will be briefly mentioned first.

## 11.2 Multichip Module (MCM)

MCM integrates different chips and discrete components side-by-side on a common substrate such as ceramic, silicon, or organic to form a system or subsystem for high-end networking, telecommunication, servers, and computer applications. Basically, there are three different kinds of MCM, namely, MCM-C, MCM-D, and MCM-L.

### *11.2.1 MCM-C*

MCM-C are multichip modules that use thick film technology such as fireable metals to form the conductive patterns, and are constructed entirely from ceramic or glass-ceramic materials, or possibly, other materials having a dielectric constant above five. In short, an MCM-C is constructed on ceramic (C) or glass-ceramic substrates [1].



J. H. Lau, *Fan-Out Wafer-Level Packaging*,
https://doi.org/10.1007/978-981-10-8884-1_11

### 11.2.2 MCM-D

MCM-D are multichip modules on which the multilayered signal conductors are formed by the deposition of thin-film metals on unreinforced dielectric materials with a dielectric constant below 5 over a support structure of silicon, ceramic, or metal. In short, MCM-D uses deposited (D) metals and unreinforced dielectrics on a variety of rigid bases [1].

### 11.2.3 MCM-L

MCM-L are multichip modules which use laminate structures and employ PCB (printed circuit board) technology to form predominantly copper conductors and vias. These structures may sometimes contain thermal expansion controlling metal layers. In short, MCM-L utilizes PCB technology of reinforced organic laminates (L) [1].

There was much research performed on MCMs during the 1990s. Unfortunately, at that time, due to the high cost of ceramic and silicon substrates and the limitation of line width and spacing of the laminate substrate, compounded with business models such as difficulty in getting the bare chips, the high-volume manufacturing (HVM) of MCMs never materialized, except some niche applications. Actually, since then, MCM has been a "dirty" word in semiconductor packaging.

## 11.3 System-in-Package (SiP)

### 11.3.1 Intention of SiP

SiP integrates different chips and discrete components, as well as 3D chip stacking of either packaged chips or bare chips (e.g., wide-bandwidth memory cubes and memory on logic with TSVs) side-by-side on a common (either silicon, ceramic, or organic) substrate to form a system or subsystem for smartphones, tablets, high-end networking, telecommunication, server, and computer applications. SiP technology performs horizontal as well as vertical integrations. Some people also called SiP vertical MCM or 3D MCM.

### 11.3.2 Actual Applications of SiP

Unfortunately, because of the high cost of TSV technology [2, 3] for smartphones and tablets, it never materialized. Most SiPs that went into HVM in the past

10 years are actually MCM-L for low-end applications such as smartphones, tablets, smart watches, medical, wearable electronics, gaming systems, consumer products, and internet of things (IoT)-related products [4] such as smart homes, smart energy, and smart industrial automation. Most actual applications of SiPs by OSATs (outsourced semiconductor assembly and test providers) integrate two or more dissimilar chips and some discrete components on a common laminated substrate.

### *11.3.3 Potential Applications of SiP*

The applications of SiP for the high-price, high-margin, and high-end products are, e.g., dual-lens camera modules. However, right now this SiP cannot be all done by the OSATs, but also involves optical design, testing, lenses, micromotors, flexible substrate, and system integration capabilities which still need to be strengthened.

## 11.4 System-on-Chip (SoC)

Moore's law [5] has been driving the system-on-chip (SoC) platform. Especially in the past 10 years, SoCs have been very popular for smartphones, tablets, and the like. SoCs integrate different-function ICs into a single chip for a system or subsystem. Two typical SoC examples are shown in the followings.

### *11.4.1 Apple Application Processor (A10)*

The application processor (AP) A10 is designed by Apple and manufactured by TSMC using its 16 nm process technology. It consists of a 6-core graphics processor unit (GPU), two dual-core central processing unit (CPUs), 2 blocks of static random access memories (SRAMs), etc. The chip area (11.6 mm $\times$ 10.8 mm) is 125 $mm^2$, Fig. 11.1a.

### *11.4.2 Apple Application Processor (A11)*

The application processor A11 is also designed by Apple and manufactured using TSMC's 10 nm process technology. The A11 consists of more functions, including a tri-core Apple-designed GPU, neural engine for face ID, etc. However, the chip area (89.23 $mm^2$) is about 30% smaller than that of the A10 because of Moore's law, i.e., the feature size is from 16 nm down to 10 nm, Fig. 11.1b.

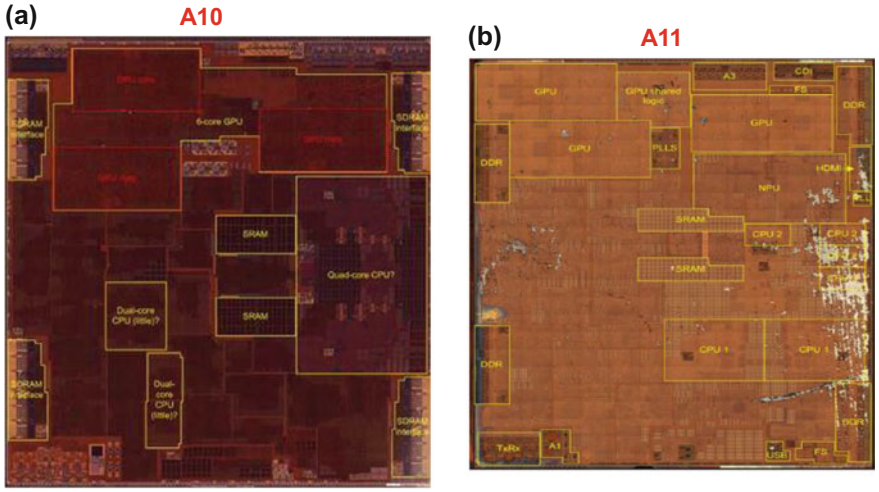



**Fig. 11.1** SoC platforms for the A10 and A11 APs

## 11.5 Heterogeneous Integration

Some of the early researches in heterogeneous integration have been provided by Georgia Institute of Technology [6–8], where they reported a differential Si CMOS (complementary metal–oxide semiconductor) receiver IC (operating at 1 Gbps) integrated with a large-area thin-film InGaAs/InP I-MSM (metal–semiconductor–metal) photodetector (Fig. 11.2). Today, most heterogeneous integrations focus on higher density, finer pitch, and more complex system.

### *11.5.1 Heterogeneous Integration Versus SoC*

Why is the heterogeneous integration of such great interest? One of the key reasons is because the end of Moore's law is fast approaching and it is more and more difficult and costly to reduce the feature size (to do the scaling) to make SoCs.

Heterogeneous integration contrasts with SoCs as follows. The heterogeneous integration uses packaging technology to integrate dissimilar chips (either side-by-side or stack) with different functions from different foundries, wafer sizes, and feature sizes (as shown in Fig. 11.3) into a system or subsystem on different (e.g., organic, silicon, or RDL) substrates, rather than integrating most of the functions into a single chip and going for a finer feature size. Heterogeneous integration and SiP are similar, except that heterogeneous integration is for finer pitch and higher density applications.

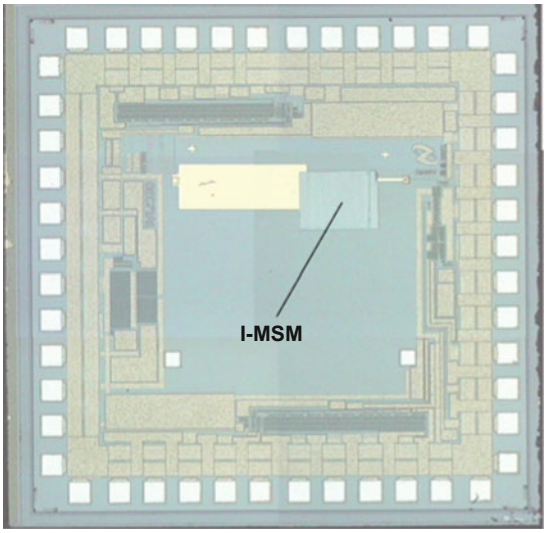



**Fig. 11.2** InGaAs/InP I-MSM integrated onto differential Si CMOS receiver IC

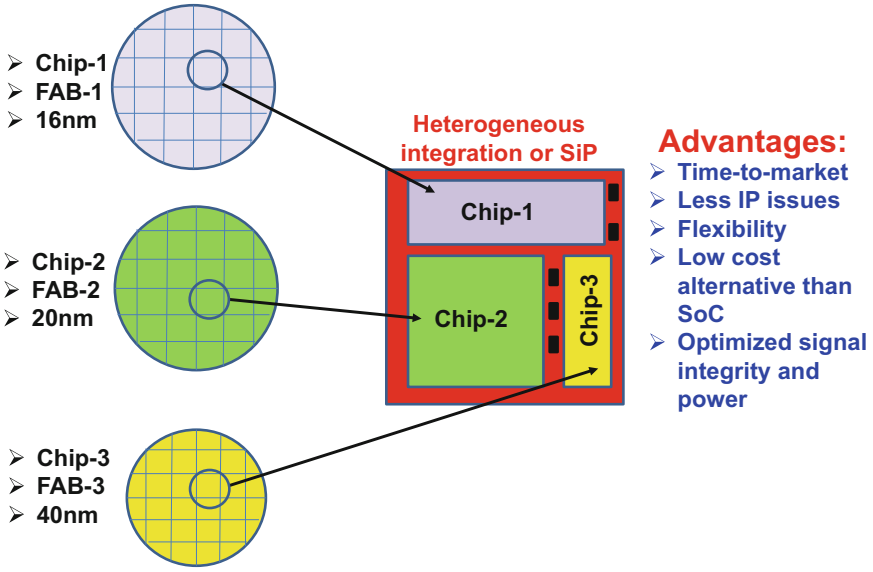



**Fig. 11.3** Heterogeneous integration or SiP

### 11.5.2 Advantages of Heterogeneous Integration

For the next few years, we will see more of a higher level of heterogeneous integration, whether it is for time-to-market, performance, form factor, power consumption, signal integrity, or cost. Heterogeneous integration is going to take some of the market shares away from SoCs on high-end applications such as high-end smartphones, tablets, wearables, networkings, telecommunications, and computing devices. How should these dissimilar chips talk to each other, however? The answer is redistribution layers (RDLs) [9]! How should those RDLs be made? One key method is by FOWLP technology.

## 11.6 Heterogeneous Integration on Organic Substrates

Today, the most common applications of heterogeneous integration are on organic substrates, or the so-called SiP. The assembly methods are usually SMT (surface mount technology) including flip chips with mass reflow as shown in Fig. 2.16a and wire bonding chips on board. In general, this is for low-end to middle-end applications.

### 11.6.1 Amkor's SiP for Automobiles

Amkor's SiP for automobiles focuses on autonomous driving, infotainment, and ADAS (advanced drive assist systems), and computer in a car. Figure 11.4a, b shows a couple of examples of Amkor's SiP for automobiles. It can be seen from Fig. 11.4a that the 42.5 mm × 42.5 mm infotainment organic substrate is supporting the processor and DDR (double data rate) memories. While from Fig. 11.4b, the 55 mm × 72 mm organic substrate is supporting the network switch, ASIC (application-specific integrated circuit), and memories.

### 11.6.2 Apple Watch II (SiP) Assembled by ASE

Through USI (Universal Scientific Industrial), ASE is a sole backend provider for Apple's custom-designed S2 SiP modules (Fig. 11.5) for use in the Apple Watch II. It can be seen from Fig. 11.5 that there are 42 chips and are on an organic substrate. Some of these chips are discrete passive components such as capacitors and resistors, ASIC, processors, controller, converter, DRAM (dynamic random access memory), NAND, Wi-Fi, NFC (near-field communication), GPS (global positioning system), sensors, etc.

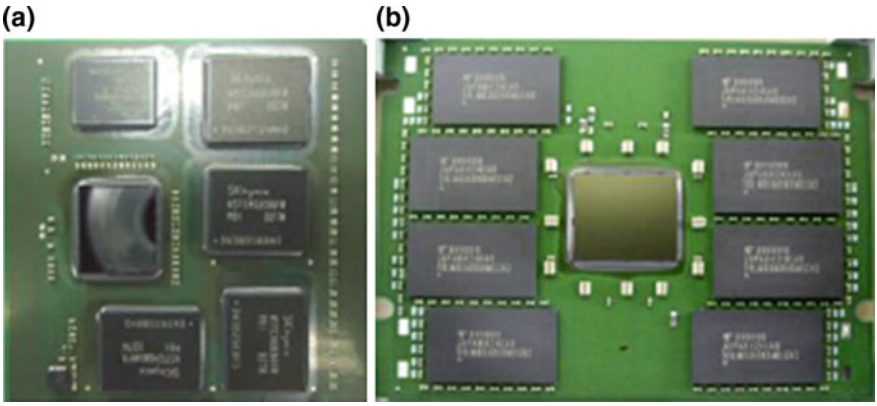



**Fig. 11.4** Amkor's SiP for automobiles. **a** 42.5 mm × 42.5 mm infotainment. **b** 55 mm × 72 mm organic substrate

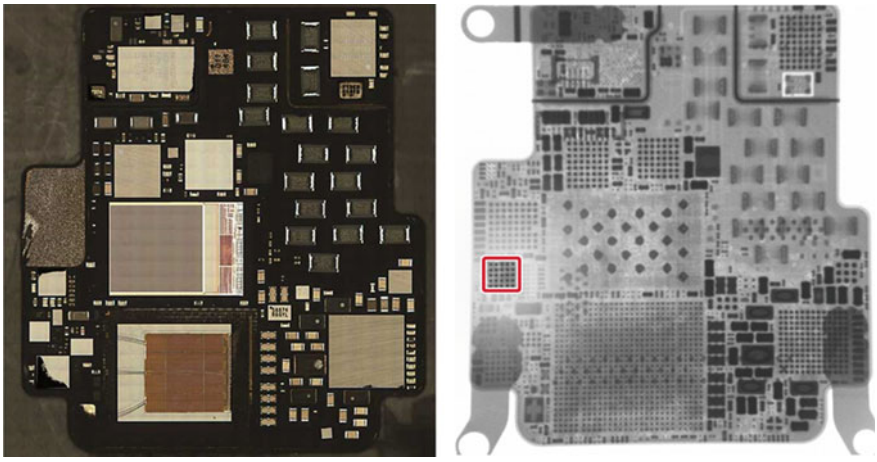


**Fig. 11.5** Apple's smart watch SII assembled by ASE

### *11.6.3 Cisco's ASIC and HBM on Organic Substrate*

Figure 11.6 shows a 3D system-in-package (SiP) designed and manufactured with a large organic interposer (substrate) with fine-pitch and fine-line interconnections by Cisco/eSilicon [10]. The organic interposer has a size of 38 mm × 30 mm × 0.4 mm. The linewidth, spacing, and thickness of the front-side and backside of the organic interposer are the same and are, respectively, 6, 6, and 10 μm. A high-performance ASIC die measured at 19.1 mm × 24 mm × 0.75 mm is attached on top of the organic interposer along with four HBM (high-bandwidth memory) DRAM die stacks.

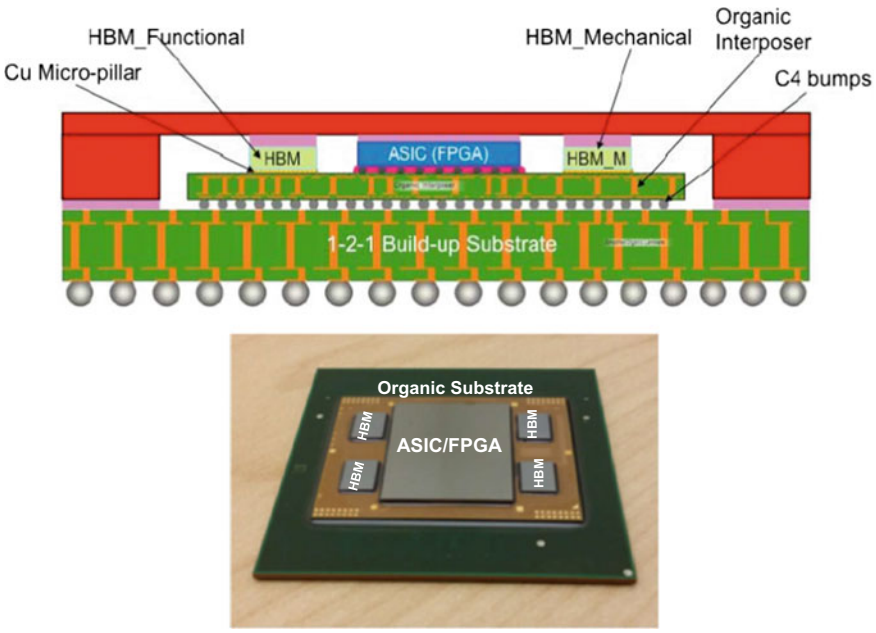



**Fig. 11.6** Cisco's networking system with organic interposer

The 3D HBM die stack with a size of 5.5 mm × 7.7 mm × 0.48 mm includes one base buffer die and four DRAM core dice, which are interconnected with TSVs and fine-pitch micro-pillars with solder caps. This is for the high-end application.

### *11.6.4 Intel's CPU and Micron's HMC on Organic Substrate*

Figure 11.7 shows Intel's Knights Landing CPU with Micron's HMC (hybrid memory cube), which have been shipping to Intel's favorite customers since the second-half of 2016. It can be seen that the 72-core processor is supported by 8 multichannel DRAMs (MCDRAM) based on Micron's HMC technology. Each HMC consists of 4 DRAMs and a logic controller (with TSVs), and each DRAM has >2000 TSVs with C2 bumps (Fig. 2.6). The CPU and the DRAM + logic controller stack are attached to an organic package substrate. Micron's current HMC assembly process is by using a low-force TCB (thermocompression bonding) with CUF (capillary underfill) as shown in Fig. 2.16b. This is for the high-end application.

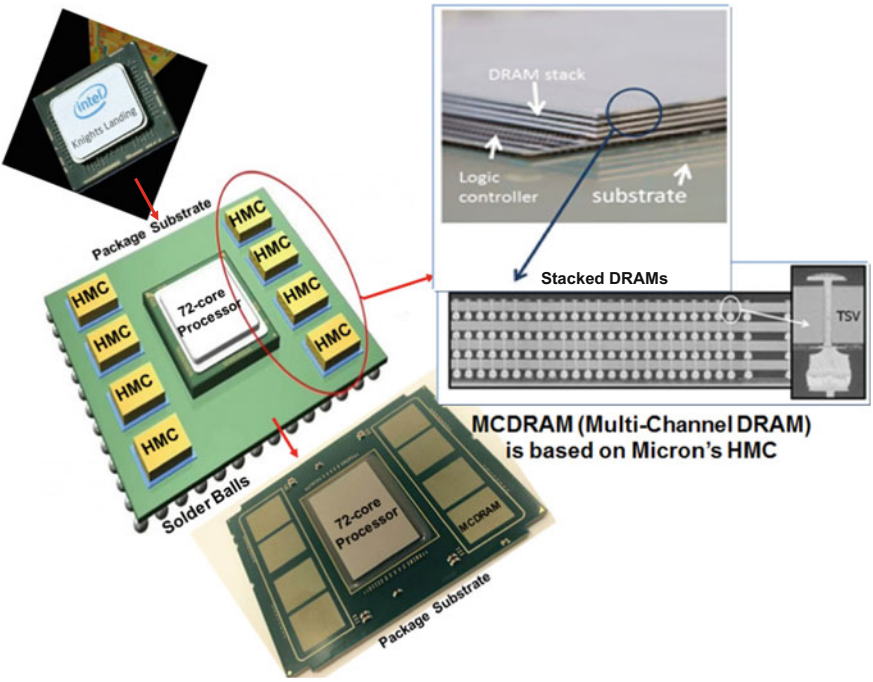



**Fig. 11.7** Intel's Knights Landing and micron's HMC on an organic substrate

## 11.7 Heterogeneous Integration on Silicon Substrates (SoW)

In general, heterogeneous integrations on silicon substrates are for multichips on silicon wafer or system-on-wafer (SoW). The assembly methods are usually flip chips-on-wafers (CoW) with TSVs (through-silicon vias) with mass reflow (Fig. 2.16a) or with thermocompression bonding (Fig. 2.16b, c) for very fine pitches. In general, this is for high-end applications.

### *11.7.1 Leti's SoW*

One of the early applications of SoW is given by Leti [11, 12] as shown in Fig. 11.8. It can be seen that a system of chips such as ASIC and memories, PMIC (power management IC) and MEMS (microelectromechanical systems) are on a silicon wafer with TSVs. After dicing, the individual unit becomes a system or subsystem and can be attached on an organic substrate or stand alone.

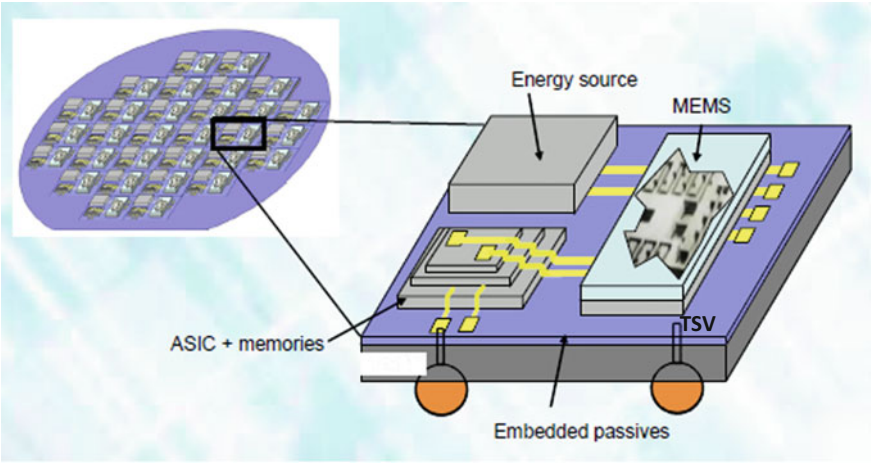



**Fig. 11.8** Leti's SoW

### *11.7.2 Xilinx/TSMC's CoWoS*

In the past few years, because of the very high-density, high I/Os, and ultrafine pitch requirements such as the sliced field-programmable gate array (FPGA), even a 12 build-up layers (6-2-6) organic package substrate is not enough to support the chips and a TSV-interposer is needed [13–22]. For example, Fig. 11.9 shows the Xilinx/TSMC's sliced FPBG chip-on-wafer-on-substrate (CoWoS) [15–17]. It can be seen that the TSV (10 μm diameter) interposer (100 μm deep) has four top RDLs: three Cu damascene layers and one aluminum layer. The 10,000+ of lateral interconnections between the sliced FPGA chips are connected mainly by the 0.4 μm pitch (minimum) RDLs of the interposer. The minimum thickness of the RDLs and passivation is <1 μm. Each FPGA has more than 50,000 microbumps (200,000+ microbumps on the interposer) at 45 μm pitch as shown in Fig. 11.9.

### *11.7.3 Analog Devices' MEMS on ASIC Wafer*

Figure 11.10 shows Analog Devices' MEMS on ASIC wafer. It can be seen that the MEMS chip is bonded on the ASIC wafer with TSVs. After dicing the wafer into individual units, then they can be attached to the PCB (printed circuit board) with solder bumps (balls).

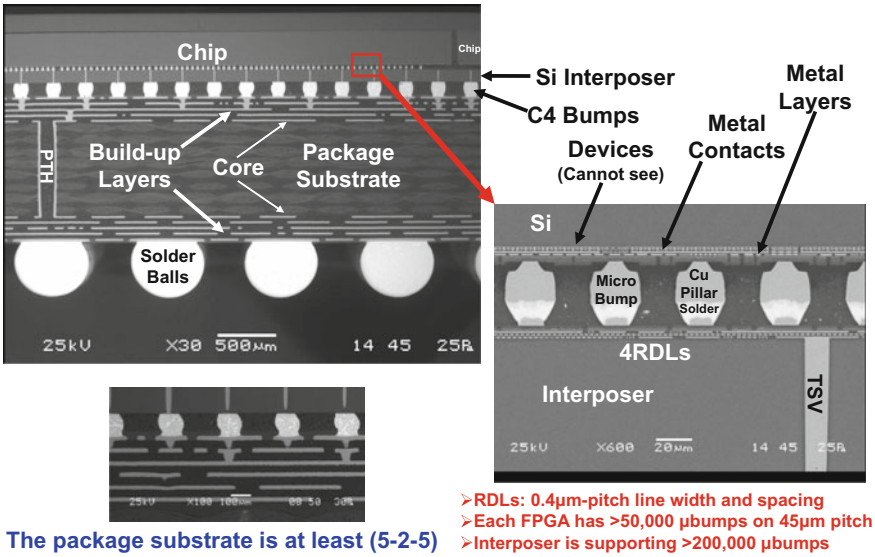



**Fig. 11.9** Xilinx/TSMC's CoWoS

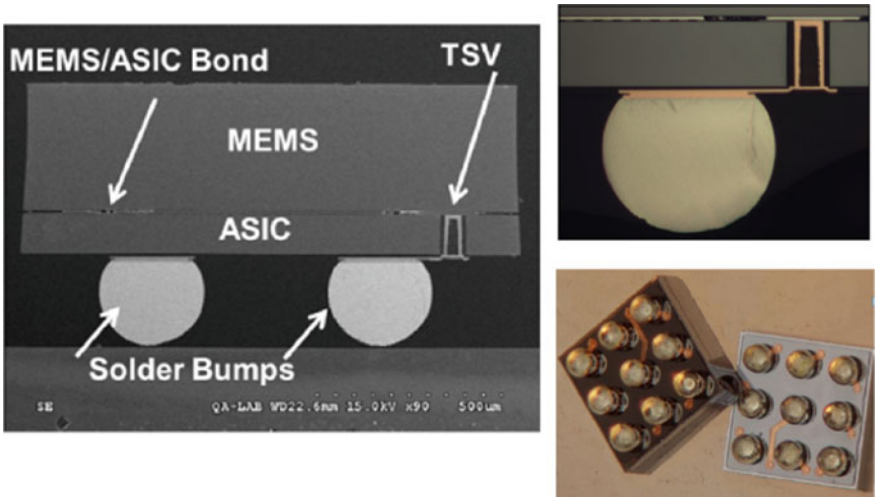



**Fig. 11.10** Analog devices' MEMS on ASIC wafer

### *11.7.4 AMD's GPU and Hynix's HBM on TSV-Interposer*

Figure 11.11 shows AMD's Radeon R9 Fury X graphic processor unit (GPU) shipped in the second-half of 2015. The GPU is built on TSMC's 28 nm process technology and is supported by four HBM cubes manufactured by Hynix. Each HBM consists of four DRAMs with C2 bumps and a logic base with TSVs straight through them. Each DRAM chip has >1000 TSVs. The GPU and HBM cubes are on top of a TSV interposer (28 mm × 35 mm), which is fabricated by

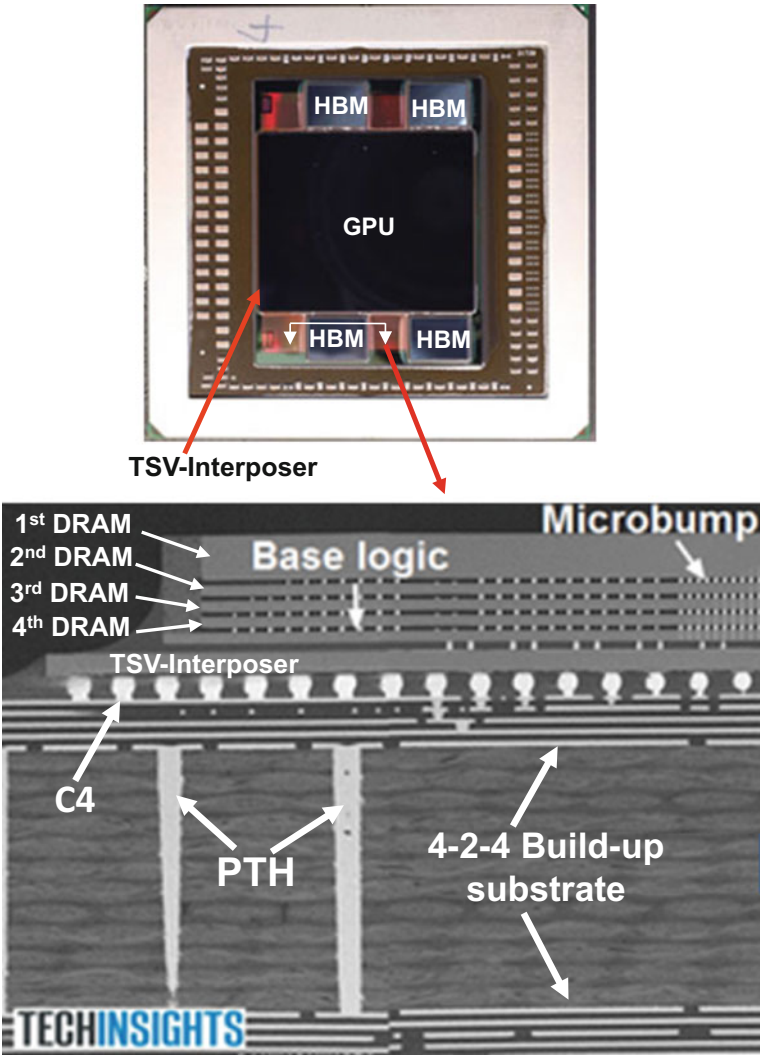



**Fig. 11.11** IMD's GPU and Hynix's HBM on Si interposer

UMC with a 64 nm process technology. The final assembly of the TSV interposer (with C4 bumps as shown in Fig. 2.4) on a 4-2-4 organic package substrate (fabricated by Ibiden) is by ASE.

### *11.7.5 Nvidia's GPU and Samsung's HBM2 on TSV-Interposer*

Figure 11.12 shows Nvidia's Pascal 100 GPU, which was shipped in the second-half of 2016. The GPU is built on TSMC's 16 nm process technology and is supported by four HBM2 (16 GB) fabricated by Samsung. Each HBM2 consists of four DRAMs with C2 bumps and a base logic die with TSVs straight through them. Each DRAM chip has >1000 TSVs. The GPU and HBM2s are on top of a TSV interposer (1200 $mm^2$), which is fabricated by TSMC with a 64 nm process technology. The TSV interposer is attached to a 5-2-5 organic package substrate with C4 bumps.

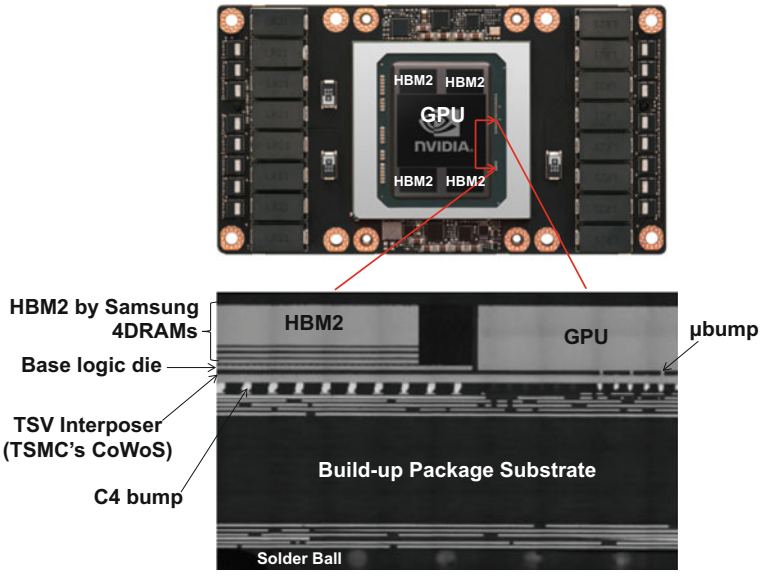



**Fig. 11.12** Nvidia's GPU and Samsung's HBM2 on Si interposer

### 11.7.6 UCLA's SoW

Figure 11.13 shows the complete fabricated Si-IF (silicon interconnect fabric) by UCLA [23]. It can be seen that the test Si-IF has 4 dielets of size (4 mm × 4 mm) with an interconnect pitch of 10 μm and with a total of 640,000 connections. The Si-IF is fabricated using conventional Si-based BEOL (back end of line) processing with up to four levels of conventional Cu damascene interconnects with wire pitches in the range of 1–10 μm and is terminated with Cu pillars of 2–5 μm height and diameter also using a damascene process. Au-capped Cu–Cu thermocompression direct bonding has been used.

## 11.8 Heterogeneous Integration on RDLs

Recently, in order to lower the package profile, enhance the performance, and lower the cost, the heterogeneous integration on RDLs have been very popular, especially with the FOWLP technology. In general, this is for middle-end to high-end applications.

### 11.8.1 Xilinx/SPIL's TSV-Less SLIT

In the past few years, through-silicon via (TSV)-less interposer [24] to support flip chips is a very hot topic in semiconductor packaging. In 2014, Xilinx/SPIL

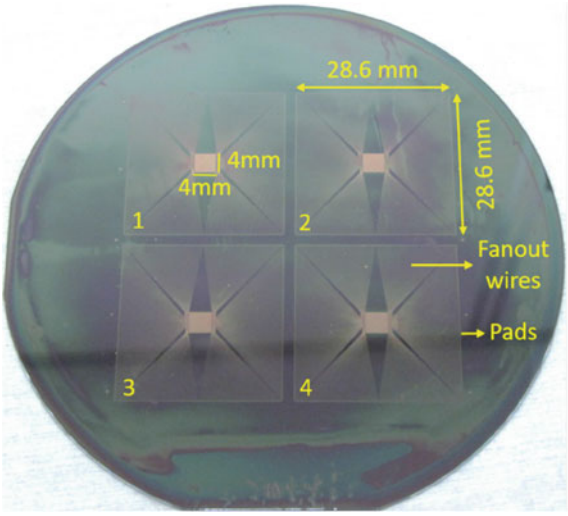



**Fig. 11.13** UCLA's SoW

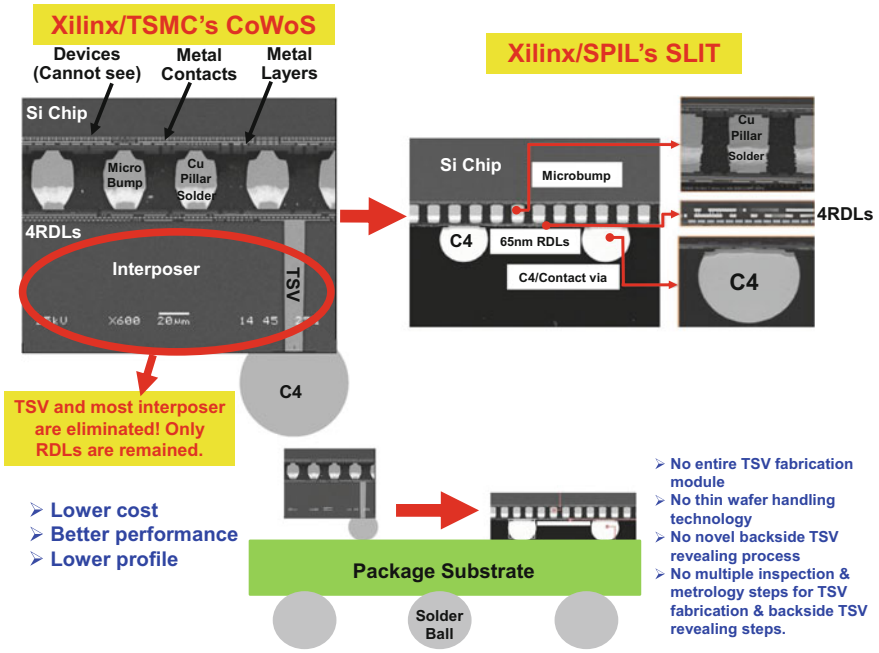



**Fig. 11.14** Xilinx/SPIL's SLIT

proposed a TSV-less interposer for sliced FPGA chips called silicon-less interconnect technology (SLIT) [25]. The upper right-hand corner of Fig. 11.14 shows the new packaging structure along with the old one, which is shown in the left-hand corner. It can be seen that the TSVs and most of the interposer are eliminated and only the four RDLs needed for performance, mainly, the lateral communication of the sliced FPGA chips, remain.

The SLIT process flow is shown in Fig. 11.15. It starts off by fabricating the RDLs—examples on a bare silicon wafer can be seen in [9] (Fig. 11.15a). That process is followed by chip-to-wafer bonding (i.e., bonding the FPGA chip to the silicon wafer with RDLs; Fig. 11.15b), and underfilling/curing (Fig. 11.15c). These processes are followed by overmolding the whole wafer with an epoxy mold compound (EMC) (Fig. 11.15d). It is followed by backgrinding the over mold to expose the backside of the chips and attaching an optional reinforcement wafer on the backside of the chips (Fig. 11.15d). Then, backgrind the silicon wafer (Fig. 11.15e). Next come passivation, photoresist, mask, patterning, etching, sputtering TiCu, photoresist, mask, and patterning (Fig. 11.15f). Finally, Cu-contact pad plating (Fig. 11.15g), photoresist stripping, TiCu etching, and controlled-collapse chip connection (C4) wafer bumping are done (Fig. 11.15h). The final assembly of the heterogeneous integration package on the substrate and then on PCB is shown in Fig. 11.16.

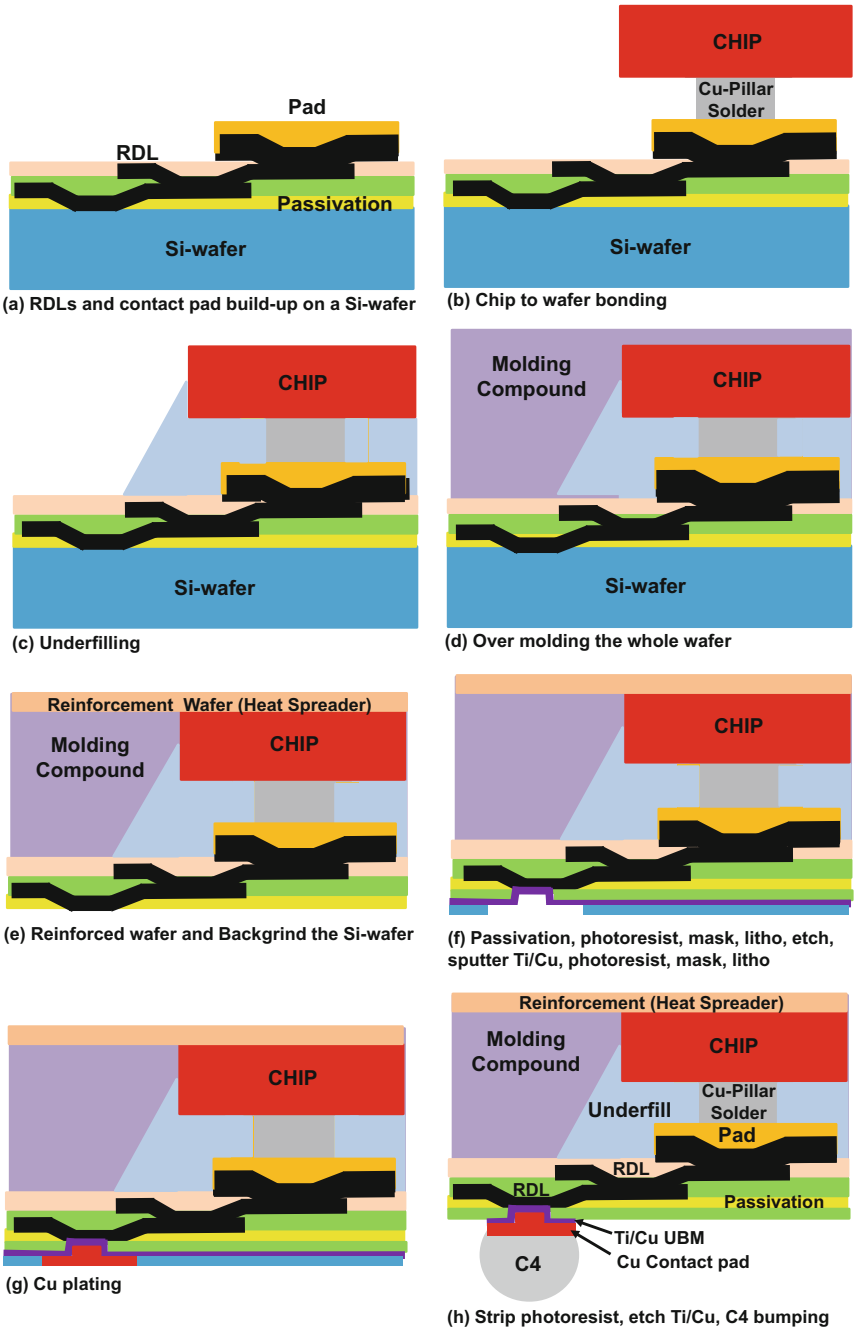



**Fig. 11.15** Process flow for implementing SLIT technology

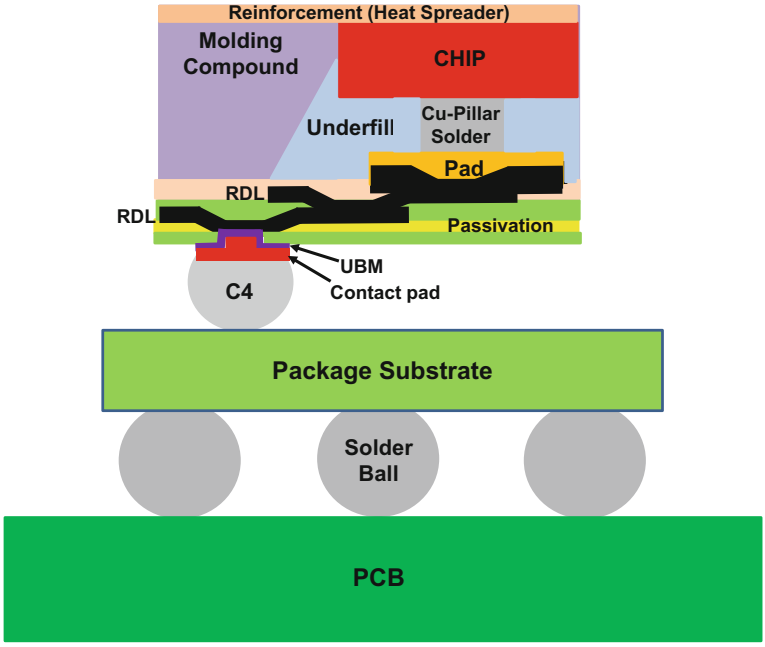



**Fig. 11.16** Final assembly of the Xilinx/SPIL's SLIT

Depending on the linewidth/spacing of the RDLs' conductive wiring, the fabrication method of the RDLs can be accomplished either by using a polymer for the dielectric layer and Cu plating of the conductive wiring (line width/spacing $\geq$5 μm), or by using plasma-enhanced chemical vapor deposition (PECVD) to make the $SiO_2$ dielectric layer and Cu damascene plus chemical mechanical polishing (CMP) to make the conductive wiring (linewidth/spacing <5 μm). In 2016, SPIL/Xilinx published a similar paper [26] with more characterization results including warpage data and called it non-TSV interposer (NTI).

### *11.8.2 Amkor's TSV-Less SLIM*

In 2015, Amkor announced a very similar technology to SLIT and is called silicon interposer-less integrated module (SLIM) [27].

### 11.8.3 Intel's TSV-Less EMIB (RDL) for FPGA and HBM

Intel proposed an embedded multi-die interconnect bridge (EMIB) [28] RDLs to replace the TSV interposer [29]. The lateral communication between the chips will be taken care of by the silicon embedded bridge and the power/ground and some signals will go through the organic package substrate as shown in Fig. 11.17. There are two major tasks in fabricating the organic package substrate with EMIB. One is to make the EMIB, and the other is to make the substrate with EMIB. To make the EMIB, one must first build the RDLs (including the contact pads) on a Si wafer. The way to make the RDLs depends on the line width/spacing of the conductive wiring of the RDLs. Finally, attach the non-RDL side of the Si wafer to a die-attach film, and then singulate the Si wafer.

To make the substrate with an EMIB, first place the singulated EMIB with the die-attached film on top of the Cu foil in the cavity of the substrate, Fig. 11.18a. It is followed by laminating a resin film on the whole organic package substrate. Then, drilling (on epoxy resin) and Cu plating to fill the holes (vias) to make connections to the contact pads of the EMIB. Continue Cu plating to make lateral connections of the substrate as shown in Fig. 11.18b. Then, it is followed by laminating another resin film on the whole substrate and drilling (on resin) and Cu

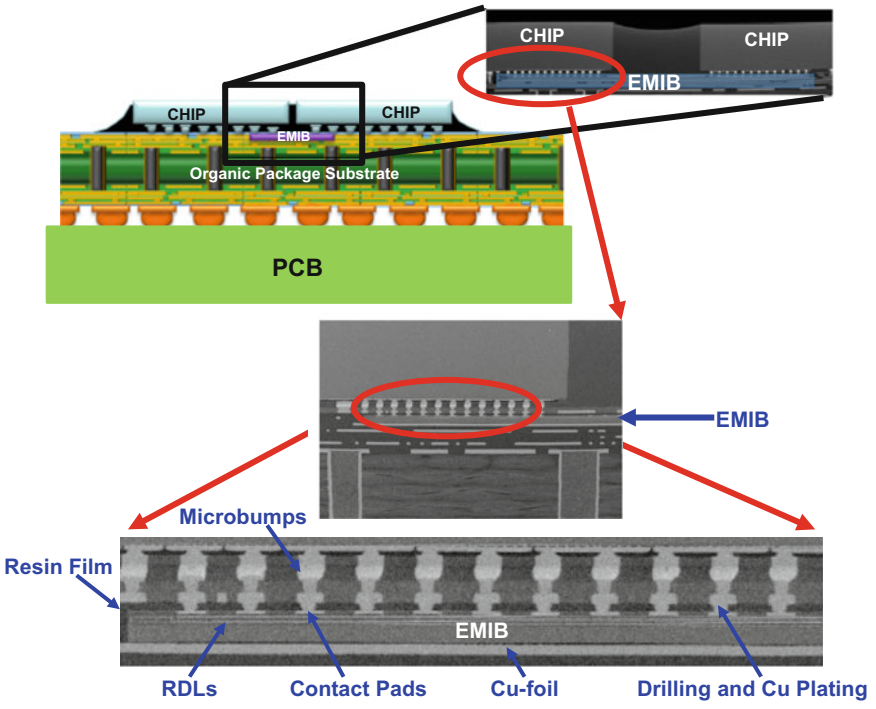



**Fig. 11.17** Intel's TSV-less interposer—EMIB

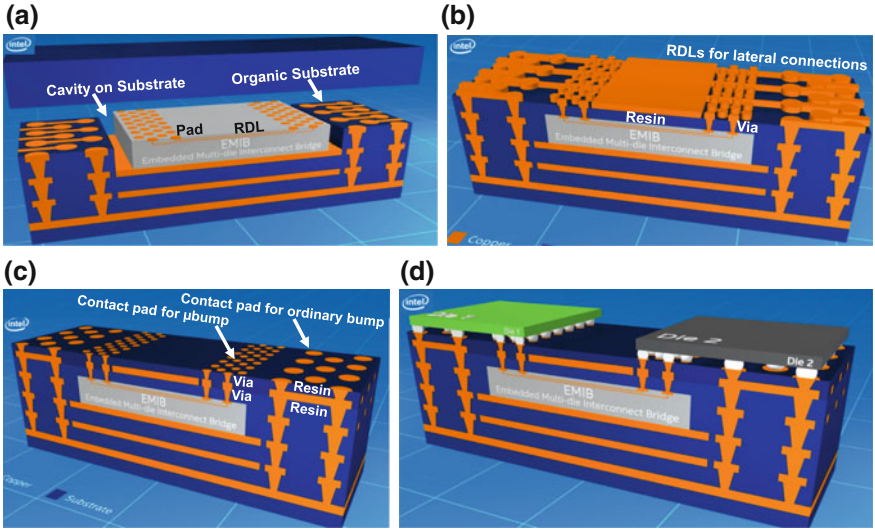



**Fig. 11.18** Assembly process of Intel's EMIB

plating to fill the holes and make contact pads, Fig. 11.18c. (Smaller pads on a finer pitch are for microbumps, while larger pads on a gross pitch are for ordinary bumps). The organic package substrate with an EMIB is ready for bonding of the chips as shown in Fig. 11.18d.

On November 9, 2015, Altera/Intel announced the industry's first heterogeneous integration devices that integrate stacked HBM from SK Hynix with high-performance Stratix® 10 FPGAs and SoCs as shown in Fig. 11.19. It can be seen that the TSV interposer is gone and replaced by Intel's EMIB.

It is interesting to note that in order to use the EMIB, the chips will have different kinds/sizes of bumps as shown in Fig. 11.19, i.e., C4 bumps and microbumps (Cu pillar + solder cap). Wafer bumping and flip chip assembly could be challenging.

### *11.8.4 EMIB (RDL) for Intel's CPU and AMD's GPU*

On November 6, 2017, Intel has formally revealed it has been working on a new series of processors that combine its high-performance ×86 cores CPUs with AMD GPUs (Radeon Graphics), as shown in Fig. 11.20, into the same processor package (heterogeneous integration) using Intel's own EMIB multi-die technology. If that wasn't enough, Intel also announced that it is bundling the design with the latest high-bandwidth memory, HBM, as shown schematically in Fig. 11.21.

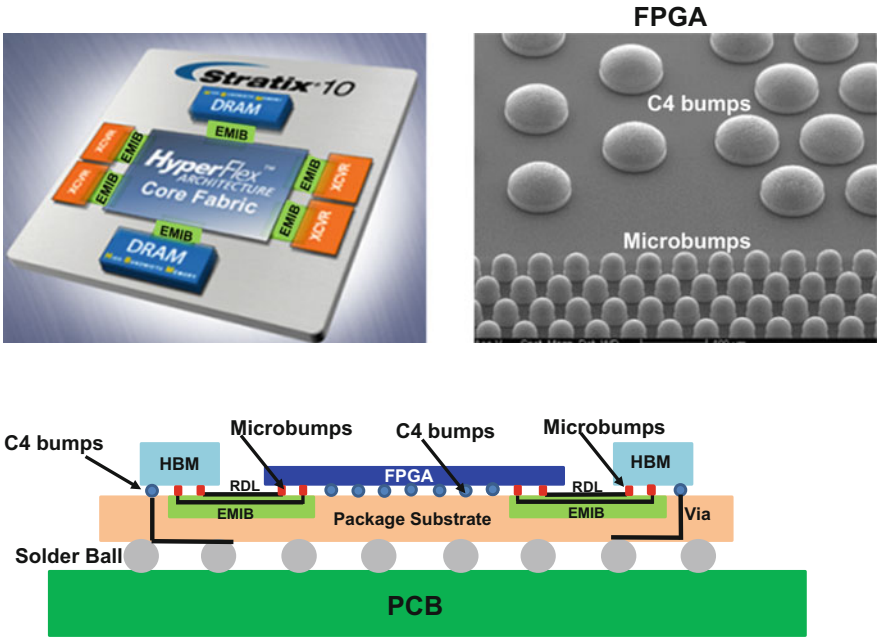
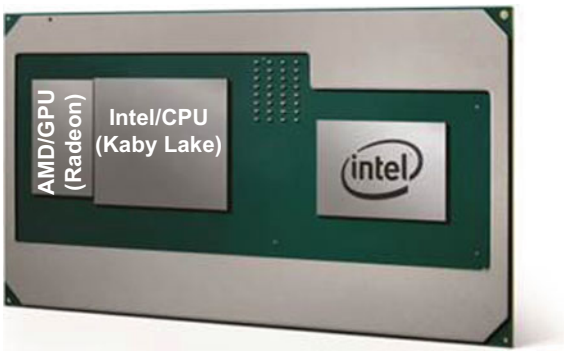



**Fig. 11.19** Intel's FPGA and HBM on EMIB

**Fig. 11.20** Intel's CPU and AMD's GPU on EMIB




## *11.8.5 STATS ChipPAC's FOFC-eWLB*

At ECTC2013, STATS chipPAC proposed [30, 31] using the fan-out flip chip (FOFC)-eWLB to make the RDLs for the chips to perform mostly lateral communications as shown in Fig. 11.22. It can be seen that the TSV interposer, wafer bumping, fluxing, chip-to-wafer bonding, cleaning, and underfill dispensing and curing are eliminated.

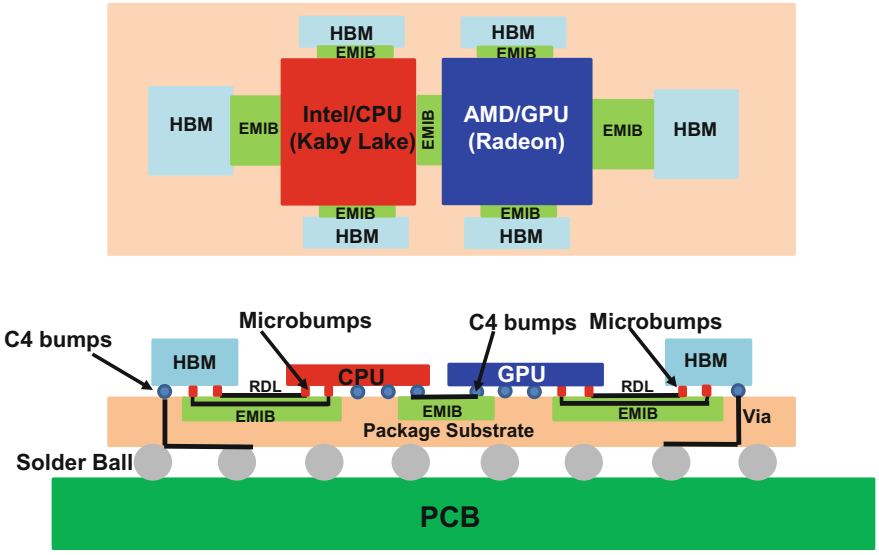



**Fig. 11.21** Schematic of Intel's CPU, AMD's GPU, and Hynix's HBM on EMIB

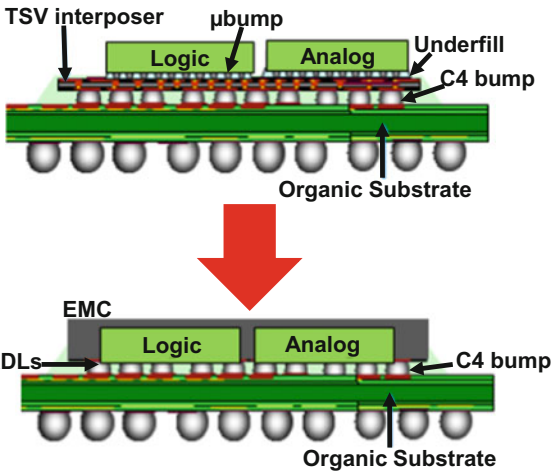



**Fig. 11.22** STATSChipPac's TSV-less FOFC-eWLB

## *11.8.6 ASE's FOCoS*

In 2016, ASE [32] proposed using the fan-out wafer-level packaging (FOWLP) technology (chip-first and die-down on a temporary wafer carrier and then over-molded by the compression method) to make the RDLs for the chips to perform mostly lateral communications as shown in Fig. 11.23; the technology is called

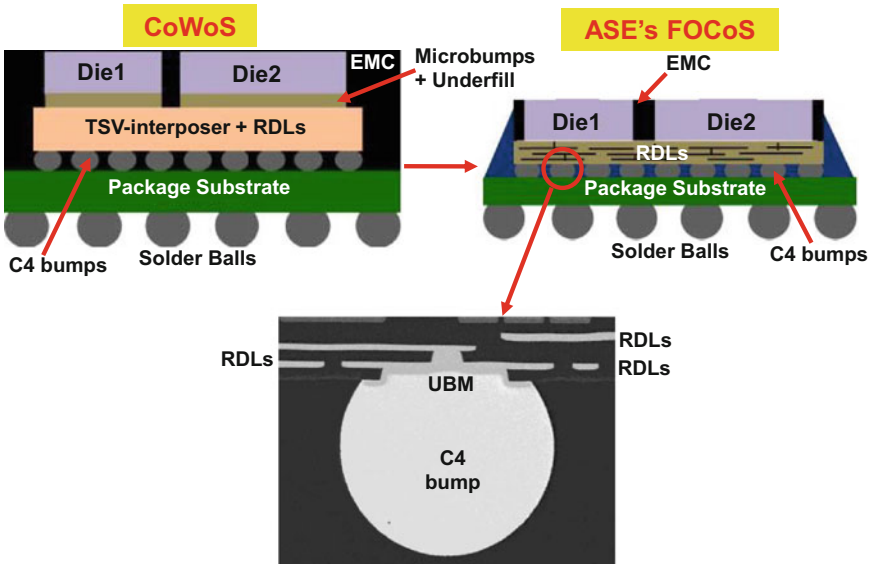



**Fig. 11.23** ASE's FOCoS

fan-out wafer-level chip-on-substrate (FOCoS). The TSV interposer, wafer bumping of the chips, fluxing, chip-to-wafer bonding, and cleaning, and underfill dispensing and curing are eliminated. The bottom RDL is connected to the package substrate using under bump metallurgy (UBM) and the C4 bump as shown in Fig. 11.23.

### *11.8.7 MediaTek's RDLs by FOWLP*

In 2016, MediaTek [33] proposed similar TSV-less interposer RDLs fabricated with FOWLP technology as shown in Figs. 11.24 and 11.25. Instead of the C4 bump, they used a microbump (Cu pillar + solder cap) to connect the bottom RDL to the 6-2-6 package substrate.

## 11.9 3D IC Heterogeneous Integration by FOWLP

A low-profile and low-cost 3D IC heterogeneous integration of the application processor chipset by FOWLP is presented in this section. The emphasis is placed on the design of the package.

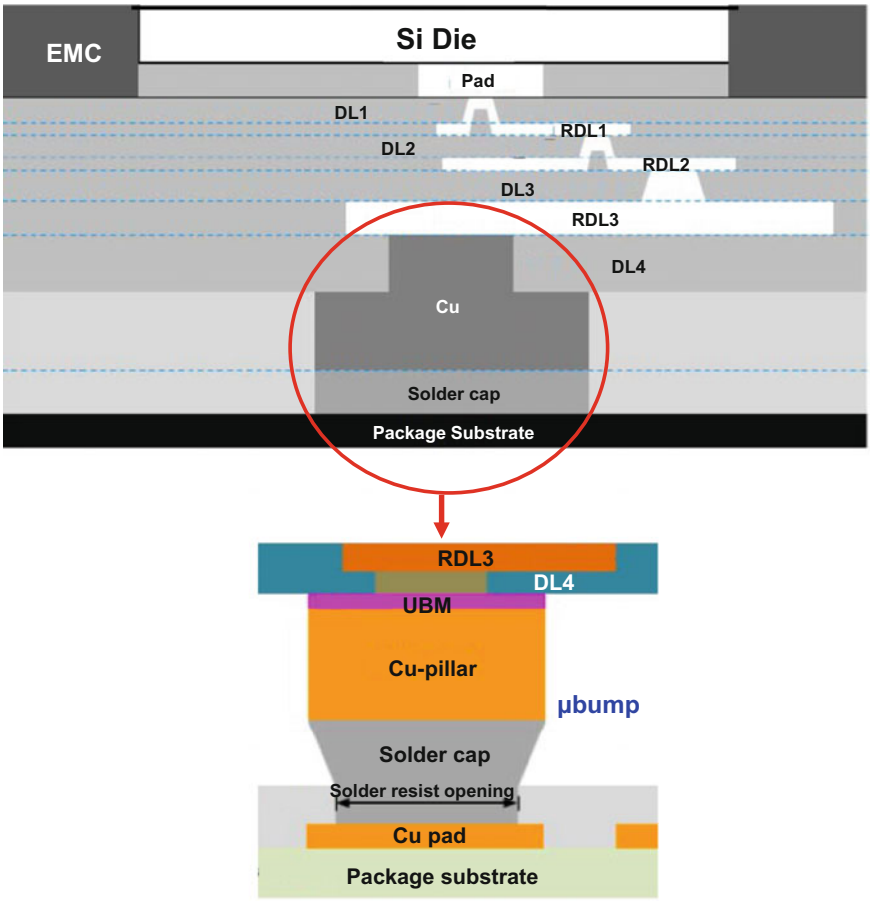



**Fig. 11.24** Schematic of MediaTek's RDLs by FOWLP

### *11.9.1 Application Processor with FOWLP*

The A10 and A11 application processors are packaged using TSMC's InFO (integrated fan-out) wafer-level packaging method [34–44]. The mobile dynamic random access memories (DRAMs) are wire bonded on a 3-layer core-less package substrate and the substrate is area-array solder balled on top of the application processor package—a package-on-package (PoP) format as shown schematically in Fig. 11.26. The interconnections between the application processor and the mobile DRAMs are mainly through the RDLs, through-InFO vias (TIVs), solder balls, and core-less substrate.

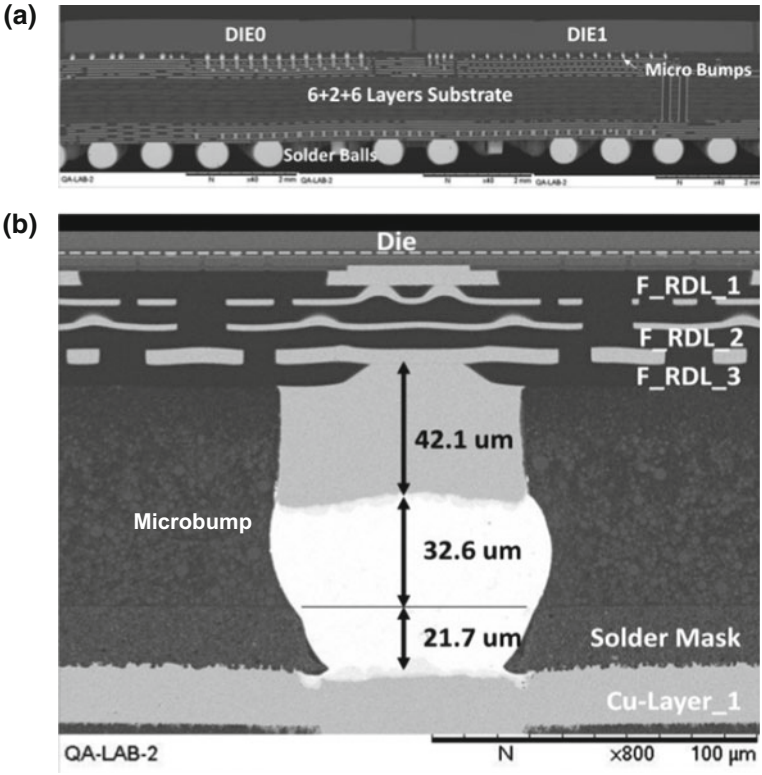



**Fig. 11.25** SEM images of MediaTek's RDLs by FOWLP

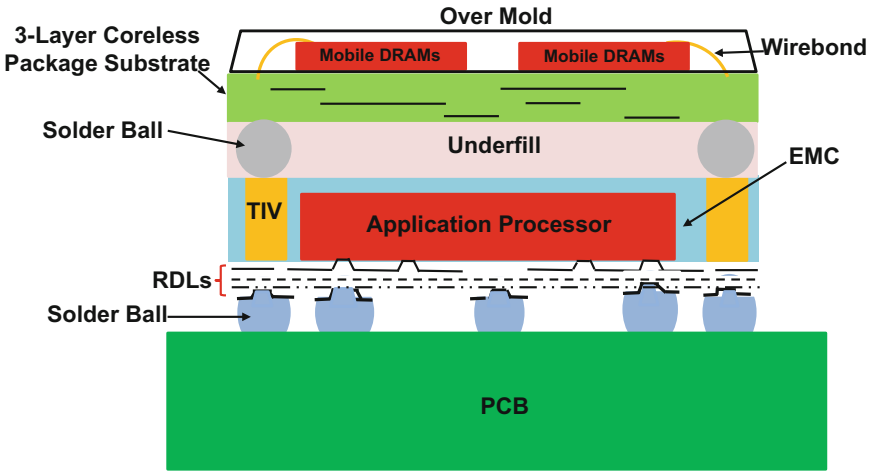



**Fig. 11.26** PoP for packaging the application processor and mobile memory

### 11.9.2 Application Processor by 3D IC Heterogeneous Integration with FOWLP

A new 3D IC heterogeneous integration by FOWLP, as shown in Fig. 11.27, is proposed in this chapter. It consists of the SoC, chips, and the mobile DRAMs. Their interconnections are mainly through the RDLs, which can be fabricated by the FOWLP method. Depending on the number of layers of the RDLs, usually the total thickness of a 3-layer RDL is about 40 μm. The DRAMs ($\leq 50$ μm thick) are cross-stacked with wire bonds and then encapsulated. The diameter of the solder ball is usually 200 μm.

Figure 11.28 shows a special case of Fig. 11.27 (when there is no other chip and the SoC is the application processor). Comparing the new design (Fig. 11.28) with that of Fig. 11.26 (the 3D IC heterogeneous integration vs. the PoP), it is obvious that: (1) the new design leads to a lower package profile; (2) the new design has less interconnects; (3) the new design is more reliable because of less interconnects; (4) the new design has better electrical performance; and (5) the new design leads to lower cost.

The manufacturing process of the proposed 3D IC heterogeneous integration is very simple. First, the device wafer has to be modified by sputtering an under bump metallurgy (UBM) and electroplating a Cu contact pad (for building the RDLs later), as shown in Fig. 11.29. This step is followed by spin coating a polymer on top of the device wafer and laminating a die-attach film (DAF) at the bottom of the device wafer. Meanwhile, a light-to-heat conversion (LTHC) layer is spin coated onto the temporary glass carrier wafer. Then the individual known-good die (KGD) (chip) from the device wafer is placed face-up on the LTHC carrier. This step is followed by epoxy molding compound (EMC) dispensing, compression molding, and finally, post mold cure (PMC). These steps are followed by backgrinding the EMC and polymer to expose the Cu contact pad for making the RDLs and for mounting the solder balls, as shown in Fig. 11.29. This is the conventional FOWLP method to package the application processor [34–44], as shown in Chap. 6.

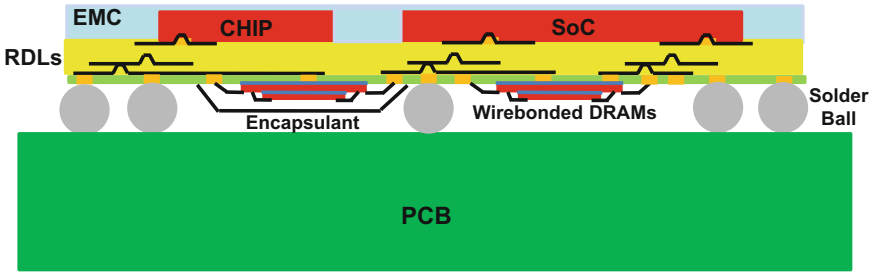



**Fig. 11.27** 3D IC heterogeneous integration by FOWLP

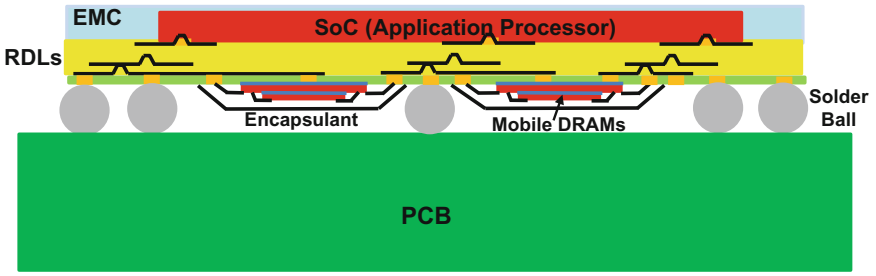



**Fig. 11.28** 3D IC heterogeneous integration to package the application processor chipset

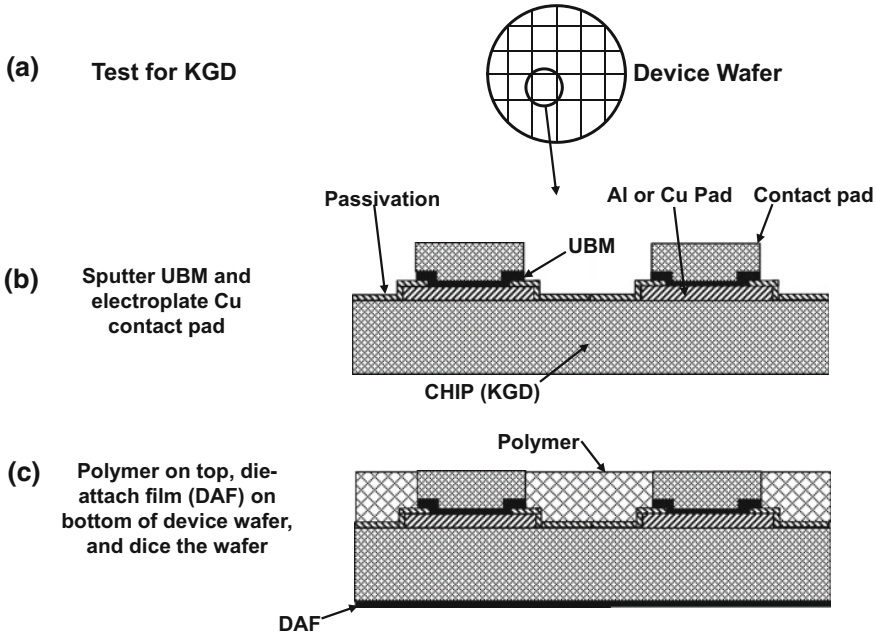



**Fig. 11.29** Manufacturing process for packaging the application processor chipset

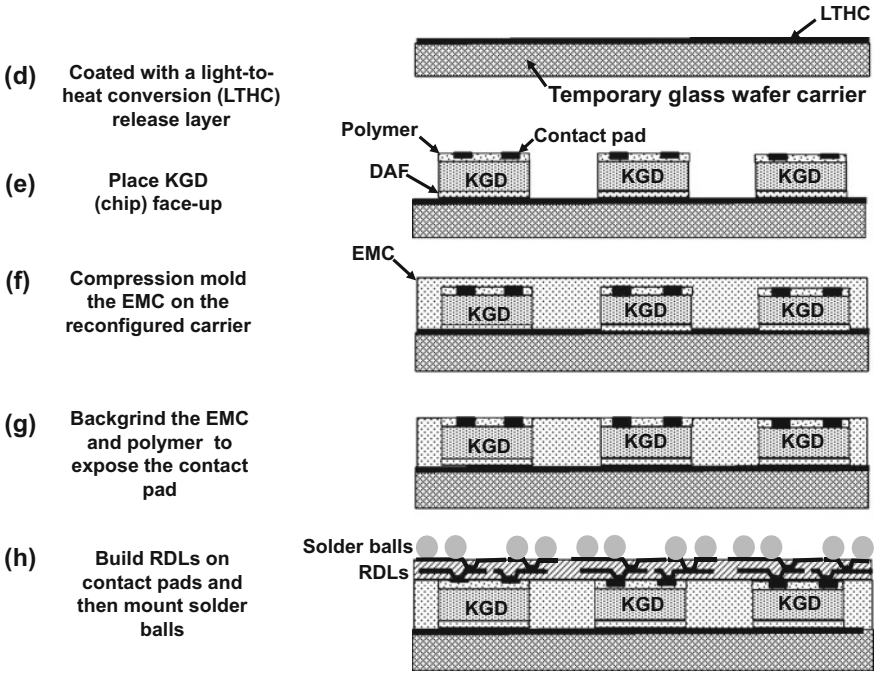



**Fig. 11.29** (continued)

There are two methods to attach the mobile DRAMs to the bottom of the application processor fan-out wafer-level package. The first method comprises the following steps: (1) removing the glass carrier by a laser (Fig. 11.30a); (2) dicing the reconstituted wafer into strips with individual packages (Fig. 11.30b); (3) wire bonding the memory chips to the bottom side of the individual package (Fig. 11.30c, d); (4) and then glob topping the wires and memory chips with an encapsulant (Fig. 11.30c, d).

The second method to attach the mobile DRAMs to the bottom of the application processor fan-out wafer-level package comprises the following steps: (1) wire bonding the memory chips to the bottom side of every package on the reconstituted wafer; (2) glob topping the wires and memory chips with an encapsulant; and (3) then dicing the reconstituted wafer into individual packages (Fig. 11.31).

Figure 11.32 is a special case of Fig. 11.27. This is when it is difficult and costly to reduce the feature size to make the SoC. Therefore, some of the functions (for example, the GPU) are not integrated into the SoC and the GPU chip is placed side-by-side with the SoC.

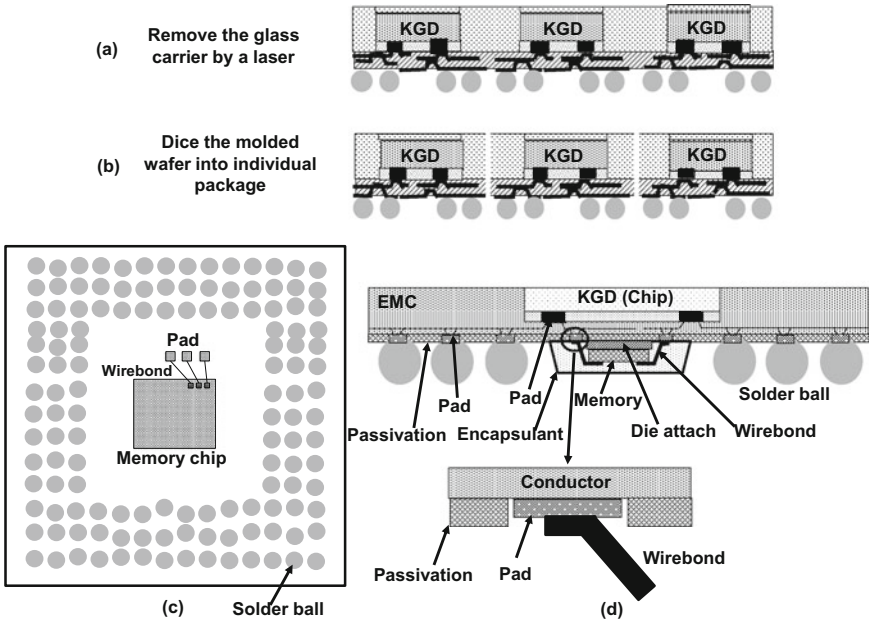



**Fig. 11.30** Wire bonding memory chip at the bottom of the individual application processor package

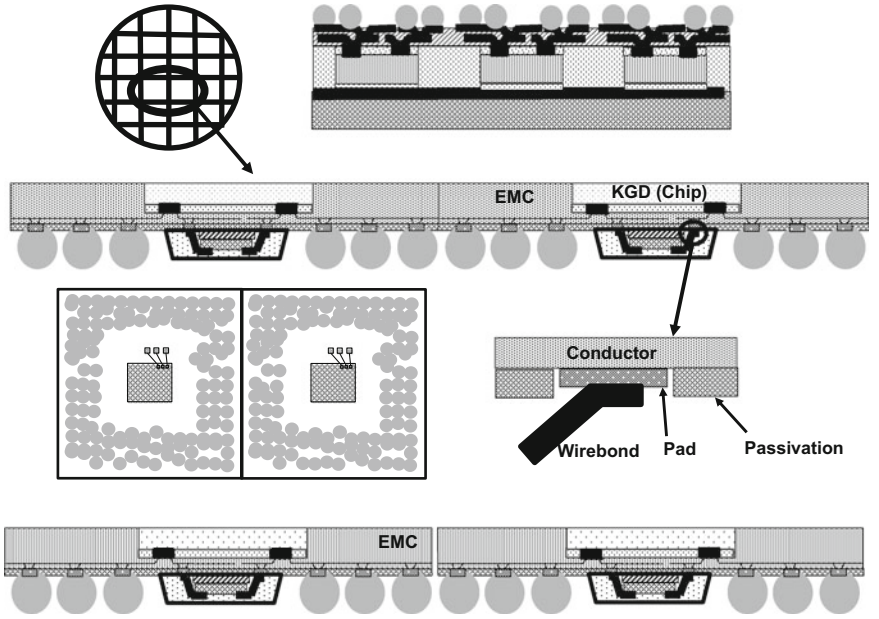



**Fig. 11.31** Wire bonding memory chip at the bottom of the application processor package on a wafer

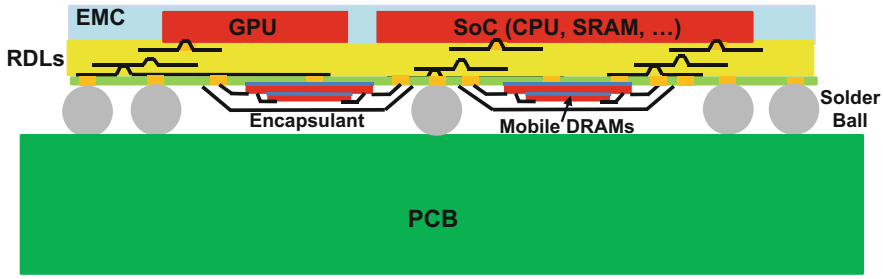



**Fig. 11.32** 3D IC heterogeneous integration to package the application processor chipset

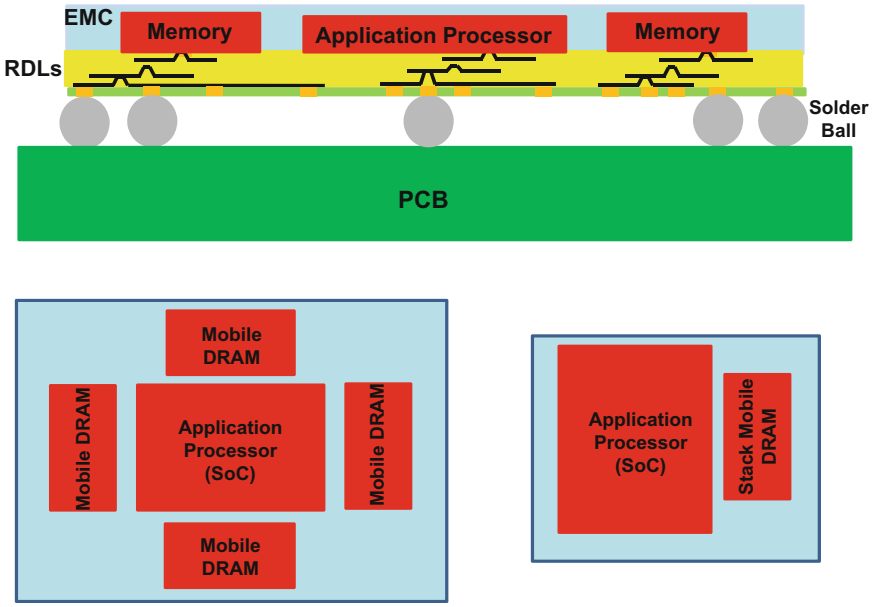



**Fig. 11.33** 2D/3D IC heterogeneous integration to package the application processor chipset

In [21], we asked the question: "What if there is no PoP for the application processor chipset?" We proposed to place the application processor and the mobile DRAMs side-by-side on a build-up package substrate. The memory chips can be either cross-stacked or individually placed by wire bonding. Also, the memory chips can be placed individually by solder-bumped flip chips. The memory chips can even be stacked and have TSVs. In this study, because we used the FOWLP method to construe the RDLs for the interconnections between the SoC and mobile DRAMs as shown in Fig. 11.33, the build-up package substrate was eliminated.

## 11.10 3D IC High-Performance Heterogeneous Integration by FOWLP

A high-performance 3D IC heterogeneous integration of CPU, GPU, FPGA, ASIC, HBM, etc., is presented in this section. The emphasis is placed on the manufacturing process.

### *11.10.1 High-Performance 3D IC Heterogeneous Integration System*

Figure 11.34 schematically shows a 3D IC high-performance heterogeneous integration by FOWLP technology. It can be seen that it consists of a GPU, a FPGA (field-programmable grid array), CPU, or a high-performance application-specific integrated circuit (ASIC), and is surrounded by high-bandwidth memory (HBM) cubes. Each HBM cube consists of four DRAMs and a logic base with through-silicon vias (TSVs) [2, 3] straight through them. Each DRAM chip has >500 TSVs. The interconnections between the GPU/FPGA/CPU/ASIC and HBMs are through the RDLs. The major heat path of this structure is from the backside of the GPU/FPGA/CPU/ASIC to the heat spreader. A heat sink can be added on top of the heat spreader if it is necessary.

### *11.10.2 Manufacturing Process*

In this case, the emphasis is placed on the manufacturing method (process) of this structure. This method comprises these steps: (1) testing for KGD of device wafers; (2) sputtering UBM; (3) electroplating the Cu contact pad; (4) spin coating a polymer on top of the device wafers; and (5) painting a thermal interface material (TIM) on the bottom (backside) of the device wafers (Fig. 11.35). The last step is different from the conventional method (the first case) which is laminating a DAF on the bottom of the device wafers.

After the steps outlined above are completed, the following are done: (1) the individual KGDs are picked and placed face-up on a metal such as copper, aluminum, steel, and an alloy 42 (with thermal expansion coefficient = 8 to $10 \times 10^{-6}$/°C) carrier about 1 mm thick; (2) molding the EMC on the reconstituted wafer is accomplished by using the compression method and then post mold curing (PMC) of the EMC; (3) backgrinding the EMC and polymer to expose the Cu contact pad; (4) building up the RDLs; and (5) mounting the solder balls. Then, the reconstituted wafer is diced into individual packages (Fig. 11.35). (Note: this process is different from the conventional method, which used a glass carrier and was coated with an LTHC release layer).

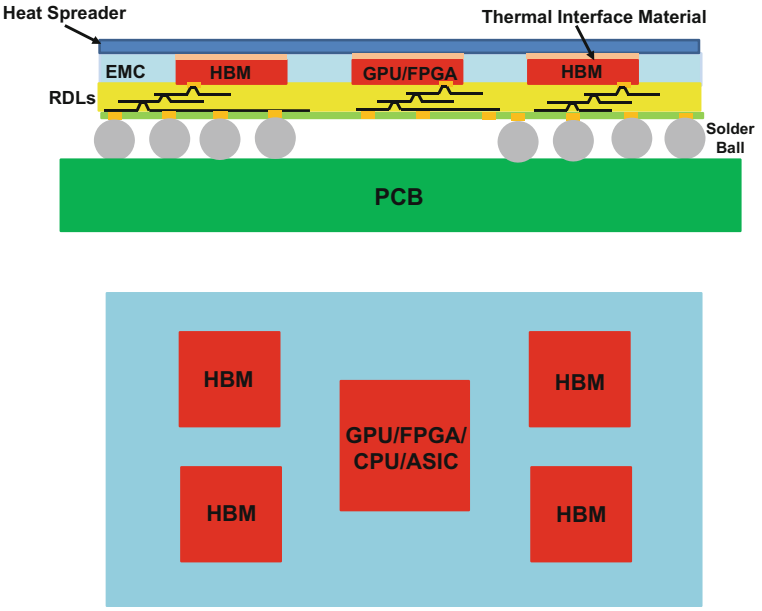



**Fig. 11.34** 3D IC high-performance heterogeneous integration by FOWLP

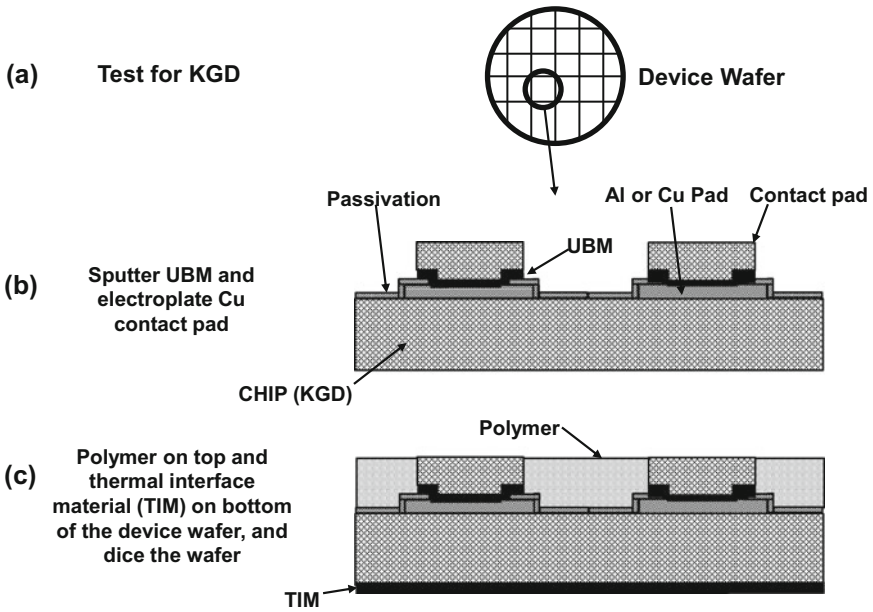



**Fig. 11.35** Manufacturing method for 3D IC high-performance heterogeneous integration

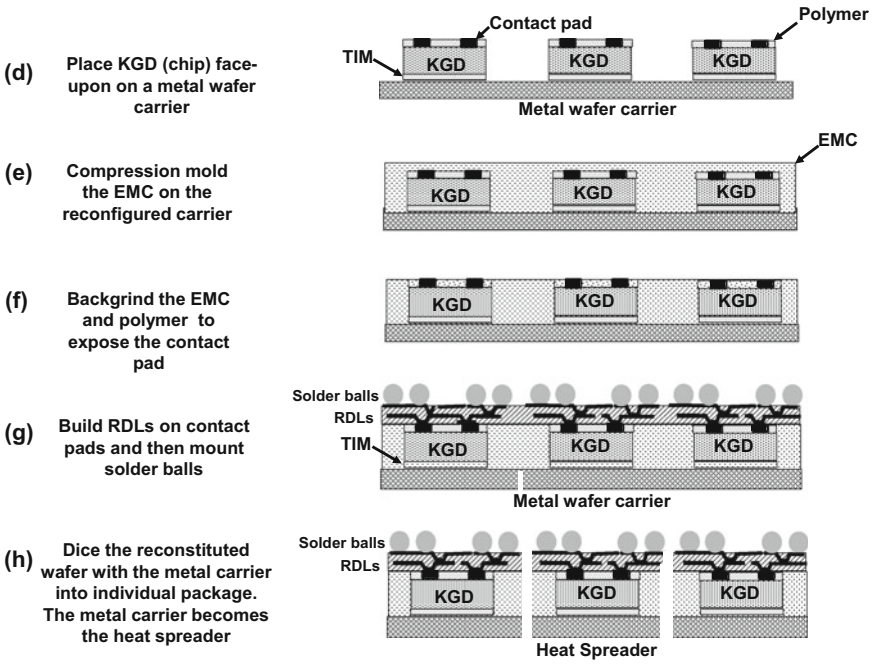



**Fig. 11.35** (continued)

### *11.10.3 Advantages of the New Manufacturing Process*

It should be emphasized that unlike the conventional method, there is no debonding of the carrier. The metal carrier becomes the heat spreader of the individual high-performance heterogeneous integration package. This new method of manufacturing high-performance chips and memory cubes in a heterogeneous integration scheme with the FOWLP technology results in fewer assembly steps, lower cost, faster time-to-market, and higher assembly yield. Also, because of the metal carrier, the warpage is reduced during all the process steps. Furthermore, because of the metal carrier, the individual package size can be larger.

## 11.11 Summary and Recommendations

Two 3D IC heterogeneous integrations by FOWLP technology have been presented. The first 3D IC heterogeneous integration is emphasized on the design and the other 3D IC high-performance heterogeneous integration is on the manufacturing method. Some important results and recommendations are as follows:

- A 3D IC heterogeneous integration of the application processor chipset has been proposed. The interconnections between the application processor and mobile DRAMs are through the RDLs, which are fabricated using the FOWLP method. The manufacturing processes for making the 3D IC heterogeneous integration have also been presented.
- When it is difficult and costly to reduce the feature size to make the SoC, one way is not to integrate some of the functions (for example, the GPU) into the SoC and instead place the GPU chip side-by-side with the SoC.
- The simplest heterogeneous integration of the application processor chipset is to place the application processor and the mobile DRAMs side-by-side on RDLs. One consideration is that the package size could be too large to be reliable. One of the alternatives is to stack up the mobile DRAMs by wire bonding (for lower cost) or TSV (for wider bandwidth.)
- A 3D IC high-performance heterogeneous integration of GPU/FPGA/CPU/ASIC and HBM/HBM2 by FOWLP technology has been proposed. Emphasis is placed on a simple and effective manufacturing method to fabricate the structure. Unlike the conventional method, there is no debonding of the temporary metal carrier. The metal carrier becomes the heat spreader of the individual high-performance heterogeneous integration package.
- The advantages of heterogeneous integration are time-to-market, performance, form factor, power consumption, signal integrity, and cost.
- In order to lower the package profile and enhance the electrical and thermal performance of the application processor chipset for mobile applications such as smartphones and tablets, the current PoP format should be eliminated.
- The recent advances of heterogeneous integrations on organic substrates, silicon wafers, and RDLs have been briefly mentioned.

## References

1. Lau, J.H. 1994. A Brief Interlocution to Wire Bonding, Tape Automated Bonding, and Flip Chip on Board for Multichip Module Applications. In *Chip on Board Technologies for Multichip Modules*, ed. J.H. Lau, 1–100. New York: Van Nostrand Reinhold.
2. Lau, J.H. 2013. *Through-Silicon Via (TSV) for 3D Integration*. New York: McGraw-Hill.
3. Lau, J.H. 2016. *3D IC Integration and Packaging*. New York: McGraw-Hill.
4. Lau, J.H. 2015. Semiconductor and Packaging for Internet of Things. *Chip Scale Review* 19: 25–30.
5. Moore, G. 1965. Cramming More Components Onto Integrated Circuits. *Electronics* 38 (8).
6. Jokerst, N.M. 1997. Hybrid Integrated Optoelectronics: Thin Film Devices Bonded to Host Substrates. *International Journal of High Speed Electronics and Systems* 8 (2): 325–356.
7. Vrazel, M., J. Chang, I. Song, K. Chung, M. Brooke, N. Jokerst, A. Brown, and D. Wills. 2001. Highly Alignment Tolerant InGaAs Inverted MSM Photodetector Heterogeneously Integrated on a Differential Si CMOS Receiver Operating at 1 Gbps. In *Proceedings of IEEE/ECTC, May 2001*, 1–6.
8. Jokerst, N.M., M.A. Brooke, S. Cho, S. Wilkinson, M. Vrazel, S. Fike, J. Tabler, Y. Joo, S. Seo, D. Wills, and A. Brown. 2003. The Heterogeneous Integration of Optical

Interconnections Into Integrated Microsystems. *IEEE Journal of Selected Topics in Quantum Electronics* 9 (2): 350–360.
9. Lau, J.H., P. Tzeng, C. Lee, C. Zhan, M. Li, J. Cline, et al. 2014. Redistribution Layers (RDLs) for 2.5D/3D IC Integration. In *IMAPS Proceedings*, *October 2013*, 434–441. Also, *IMAPS Transactions, Journal of Microelectronic Packaging* 11 (1): 16–24.
10. Li, L., P. Chia, P. Ton, M. Nagar, S. Patil, J. Xie, et al., 3D SiP with Organic Interposer of ASIC and Memory Integration. In *IEEE/ECTC Proceedings*, *2016*, 1445–1450.
11. Souriau, J., O. Lignier, M. Charrier, and G. Poupon. 2005. Wafer Level Processing of 3D System in Package for RF and Data Applications. In *IEEE/ECTC Proceedings*, 356–361.
12. Henry, D., D. Belhachemi, J.-C. Souriau, C. Brunet-Manquat, C. Puget, G. Ponthenier, J. Vallejo, C. Lecouvey, and N. Sillon. 2006. Low Electrical Resistance Silicon Through Vias: Technology and Characterization. In *IEEE/ECTC Proceedings*, *2006*, 1360–1366.
13. Selvanayagam, C., J.H. Lau, X. Zhang, S. Seah, K. Vaidyanathan, and T.C. Chai. 2009. Nonlinear Thermal Stress/Strain Analyses of Copper Filled TSV (Through Silicon Via) and Their Flip-Chip Microbumps. In *Proceedings of IEEE/ECTC*, *May 2008*, 1073–1081. Also, *IEEE Transactions on Advanced Packaging* 32 (4), 720–728 (November 2009).
14. Chai, T.C., X. Zhang, J.H. Lau, C.S. Selvanayagam, D. Pinjala, Y. Hoe, Y. Ong, V. Rao, E. Wai, H. Li, E. Liao, N. Ranganathan, V. Kripesh, S. Liu, J. Sun, M. Ravi, C. Vath III, and Y. Tsutsumi. 2011. Development of Large Die Fine-Pitch Cu/low-*k* FCBGA Package with Through Silicon Via (TSV) Interposer. *IEEE Transactions on CPMT* 1 (5): 660–672.
15. Banijamali, B., S. Ramalingam, H. Liu, and M. Kim. 2012. Outstanding and Innovative Reliability Study of 3D TSV Interposer and Fine Pitch Solder Micro-bumps. In *Proceedings of IEEE/ECTC*, *May 2012*, 309–314.
16. Chaware, R., K. Nagarajan, and S. Ramalingam. 2012. Assembly and Reliability Challenges in 3D Integration of 28 nm FPGA Die on a Large High Density 65 nm Passive Interposer. In *IEEE/ECTC Proceedings*, *2012*, 279–283.
17. Banijamali, B., C. Chiu, C. Hsieh, T. Lin, C. Hu, S. Hou, S. Ramalingam, S. Jeng, L. Madden, and D. Yu. 2013. Reliability Evaluation of a CoWoS-Enabled 3D IC Package. In *Proceedings of IEEE/ECTC*, *May 2013*, 35–40.
18. Yu, A., J.H. Lau, S. Ho, A. Kumar, W. Hnin, W. Lee, M. Jong, V. Sekhar, V. Kripesh, D. Pinjala, S. Chen, C. Chan, C. Chao, C. Chiu, C. Huang, and C. Chen. 2011. Fabrication of High Aspect Ratio TSV and Assembly With Fine-Pitch Low-Cost Solder Microbump for Si Interposer Technology With High-Density Interconnects. *IEEE Transactions on CPMT* 1 (9): 1336–1344.
19. Lau, J.H. 2011. Overview and Outlook of TSV and 3D Integrations. *Microelectronics International* 28 (2): 8–22.
20. Lau, J.H., and G.Y. Tang. 2012. Effects of TSVs (Through-Silicon Vias) on Thermal Performances of 3D IC Integration System-In-Package (SiP). *Microelectronics Reliability* 52 (11): 2660–2669.
21. Lau, J.H. 2014. The Future of Interposer for Semiconductor IC Packaging. *Chip Scale Review* 18 (1): 32–36.
22. Lau, J.H., S. Lee, M. Yuen, J. Wu, J. Lo, H. Fan, and H. Chen. 2013. Apparatus Having Thermal-Enhanced and Cost-Effective 3D IC Integration Structure With Through Silicon Via Interposer. *U.S. Patent No. 8,604,603*, Filed February 29, 2010 and Issued December 10, 2013.
23. Bajwa, A., S. Jangam, S. Pal, N. Marathe, T. Bai, T. Fukushima, M. Goorsky, and S.S. Iyer. 2017. Heterogeneous Integration at Fine Pitch ($\leq 10$ μm) Using Thermal Compression Bonding. In *IEEE/ECTC Proceedings*, *2017*, 1276–1284.
24. Lau, J.H. 2016. TSV-Less Interposers. *Chip Scale Review* 20: 28–35.
25. Kwon, W., S. Ramalingam, X. Wu, L. Madden, C. Huang, H. Chang, et al. 2014. Cost-Effective and High-Performance 28 nm FPGA with New Disruptive Silicon-Less Interconnect Technology (SLIT). In *Proceedings of International Symposium on Microelectronics*, 599–605.

26. Liang, F., H. Chang, W. Tseng, J. Lai, S. Cheng, M. Ma, et al. 2016. Development of Non-TSV Interposer (NTI) for High Electrical Performance Package. In *IEEE/ECTC Proceedings*, *2016*, 31–36.
27. Hiner, D., M. Kelly, R. Huemoeller, and R. Reed. 2015. Silicon Interposer-Less Integrated Module—SLIM. In *International Conference and Exhibition on Device Packaging*, *2015*.
28. Chiu, C., Z. Qian, M. Manusharow. 2014. Bridge Interconnect with Air Gap in Package Assembly. *US Patent No. 8,872,349*.
29. Mahajan, R., R. Sankman, N. Patel, D. Kim, K. Aygun, Z. Qian, et al. 2016. Embedded Multi-die Interconnect Bridge (EMIB)—A High-Density, High-Bandwidth Packaging Interconnect. In *IEEE Proceedings of Electronic Components and Technology Conference*, *2016*, 557–565.
30. Pendse, R.D. 2011. Semiconductor Device and Method of Forming Extended Semiconductor Device with Fan-Out Interconnect Structure to Reduce Complexity of Substruate. *Patent Publication No. US2013/0161833A1*, Filed on December 23, 2011.
31. Yoon, S.W., P. Tang, R. Emigh, Y. Lin, P.C. Marimuthu, and R. Pendse. 2013. Fanout Flipchip eWLB (embedded Wafer Level Ball Grid Array) Technology as 2.5D Packaging Solutions. In *Proceedings of IEEE/ECTC*, *2013*, 1855–1860.
32. Lin, Y., W. Lai, C. Kao, J. Lou, P. Yang, C. Wang, et al. 2016. Wafer Warpage Experiments and Simulation for Fan-Out Chip-On-Substrate. In *IEEE Proceedings of Electronic Components and Technology Conference*, *2016*, 13–18.
33. Chen, N., T. Hsieh, J. Jinn, P. Chang, F. Huang, J. Xiao, et al. 2016. A Novel System-In-Package with Fan-Out WLP for High-Speed SERDES Application. In *IEEE Proceedings of Electronic Components and Technology Conference*, *2016*, 1495–1501.
34. Lau, J.H. 2015. Patent Issues of Fan-Out Wafer/Panel-Level Packaging. *Chip Scale Review* 19: 42–46.
35. Lau, J.H., N. Fan, M. Li. 2016. Design, Material, Process, and Equipment of Embedded Fan-Out Wafer/Panel-Level Packaging. *Chip Scale Review* 20: 38–44.
36. Tseng, C., C. Liu, C. Wu, D. Yu. 2016. InFO (Wafer-Level Integrated Fan-Out) Technology. In *IEEE/ECTC Proceedings*, *June 2016*, 1–6.
37. Hsieh, C., C. Wu, D. Yu. 2016. Analysis and Comparison of Thermal Performance of Advanced Packaging Technologies for State-of-the-Art Mobile Applications. In *IEEE/ECTC Proceedings*, *June 2016*, 1430–1438.
38. Lau, J.H., M. Li, D. Tian, N. Fan, E. Kuah, K. Wu, et al. 2017. Warpage and Thermal Characterization of Fan-Out Wafer-Level Packaging. In *IEEE/ECTC Proceedings*, *2017*, 595–602. Also, *IEEE Transactions on CPMT* 7 (10), 1729–1938 (October 2017).
39. Lau, J.H., M. Li, N. Fan, E. Kuah, Z. Li, K. Tan, T. Chen, et al. 2017. Fan-Out Wafer-Level Packaging (FOWLP) of Large Chip with Multiple Redistribution-Layers (RDLs). In *IMAPS Proceedings*, *October 2017*, 576–583. Also, *IMAPS Transactions, Journal of Microelectronics and Electronic Packaging* 14 (4), 123–131 (October 2017).
40. Li, M., Q. Li, J.H. Lau, N. Fan, E. Kuah, K. Wu, et al. 2017. Characterizations of Fan-Out Wafer-Level Packaging. In *IMAPS Proceedings*, *October 2017*, 557–562.
41. Lau, J.H. 2018. 3D IC Heterogeneous Integration by FOWLP. *Chip Scale Review* 22: 16–21.
42. Lau, J.H. 2018. FOWLP for 3D IC Heterogeneous Integration. In *IEEE/CSTIC Proceedings*, VII. 1–6.
43. Kuah, E., W. Chan, J. Hao, N. Fan, M. Li, J.H. Lau, et al. 2017. Dispensing Challenges of Large Format Packaging and Some of Its Possible Solutions. In *IEEE/EPTC Proceedings*, *December 2017*, S27_1-6.
44. Hua, X., H. Xu, Z. Li Zhang, D. Chen, K. Tan, J.H. Lau, et al. 2017. Development of Chip-First and Die-Up Fan-Out Wafer-Level Packaging. In *IEEE/EPTC Proceedings*, *December 2017*, S23_1-6.