3D-IC Technologies and 3D FPGA

Xin Wu, XILINX INC
San Jose, California, 95124 USA

***Abstract*—An overview of various 3D-IC technologies is provided. The suitableness of several of these technologies for FPGA will be discussed, from XILINX' own experiences**

Keywords—3D-IC technology, Programmable logic, FPGA

I. INTRODUCTION

Silicon technologies have been grown rapidly in past 50 years following Moore's Law. This brought the prosperity of semiconductor industry. However in recent several technology nodes it has been facing more technical challenges, design complexity and cost increases.

In mean time, many 3D-IC technologies have been becoming available gradually in past several years. These 3D-IC technologies provide alternative of future growth opportunities. They provide more functionality, better power-performance and reduction of design cycles, by integrating different technologies (for example memory vs. logic vs. analog vs. sensors, etc.) and different functional chips, as well as allowing die partitions.

XILINX has participated since 2006 in 3D-IC technology development. Today there are more than 7 3D-IC products from 2 generations of FPGA (Field Programmable Gate Arrays) family nodes in shipping to customers. In following author will provide a brief overview of various 3D-IC technologies and select a few of them to discuss their suitableness on application to FPGA.

II. 3D-IC TECHNOLOGIES OVERVIEW

Today in silicon technology, there is in general only one main baseline in whole industry. For example every fab is doing high-k / metal gate, multi-patterning optical lithography, now FinFET, etc. This is mostly due to extremely high cost of development (several thousands of engineers-years), of equipment (for example lithography tool nowadays easily cost 50 to hundreds of million dollars) and of materials, as well as the high complexity of technical challenges (for example gate length is less than 20nm while optical litho tool light source wavelength still 193nm).

Unlike silicon technology, 3D-IC has many different technology flavors available and more in development [1] [2]. The reasons are similar, the development cost, tools, materials and challenges are relatively moderate. The availability of varieties of 3D-IC technologies provides opportunities for choosing more suitable technology to specific products.

Following is an overview of different 3D-IC technologies.

A. Passive 3D-IC technologies

The "passive" defines as the connecting components, such as package substrates, interposer or fan-outs, etc. do not contain active transistors.

Fig 1 shows a very brief landscape of passive 3D-IC.

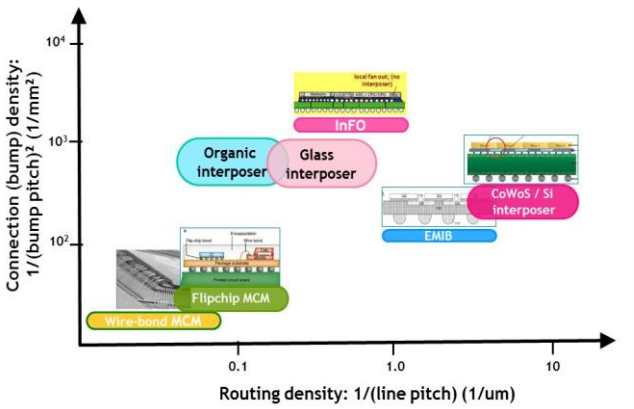

Figure 1 Various passive 3D-IC technology comparison. The vertical axis is connection density per mm^2; the horizontal axis is routing (connecting line) density per um.

There are many different ways to compare different passive 3D-IC technologies. In Fig.1 we use "connection density", i.e. the "bump" density per mm^2, vs. "routing density" (connecting line density) which is inversion of minimum line pitch, as measure. Please note this comparison does not counting the total possible layer count. Some passive 3D-IC technologies have limited layer count due to technology constrain, for example some fan-out situations.

From the lower-left corner of Fig.1 there is a group of relatively mature technologies, refer as MCMs, including wirebond MCM, flip chip MCM and hybrids of them, In which multiple chips are connected on a package substrate. MCMs mostly utilize existing packaging technologies with connecting density less than 100 per mm2 (bump pitch larger or equal than 100 um) and a connection line pitch (width + space) larger than 10um. They are in general relatively lower-cost

In the middle group of Fig.1 are several of fan-out technologies and glass interposers (we left the organic interposer in lower-left group since achieving better than 10um routing line pitch is more challenging there). In this group connection (bump) pitch can be less than 100um, most like using uBumps. The routing line width and space in this group are in general equal or larger than 1um and 1um, which is perhaps the limit of non-damascene CMP Cu process. There are several fan-out technologies belong to this group, we use InFO as representative

here. Although in theory its connection density can be very high, its routing line pitch falls into this group range. Plus we feel 2 layers of fan-out routing in InFO is doable but more than 3 layers would be much more challenging, due to planarity concern. Therefore total routing capability (routing density per layer multiplies number of layers) is moderate. The cost in this group will typically be higher than those in the MCM group.

The right side group of Fig.1, mostly utilize Si fab technology. It can achieve 0.5/0.5um line width / spacing easily, with much room to advance to even nano-meter range as today's capability. On the other hand its connecting density, is still similar to the middle group, in a pitch range ~35-60um range, because today's uBump capability limit. TSV (through-silicon-via) is readily available in Si interposers. There are multiple technologies using Si process, TSV and uBumps, we pick CoWoS as representative here. Although in general the manufacture cost of this group is higher than that of middle group, if counting in yield (the biggest cost factor!), maturity (also yield), capability and volume readiness, this group might not always has cost disadvantage from actual product point of view. EMIB [3] has been placed also in this group but slightly less dense in both connection density and routing capability (or one might consider it as in-between middle and right groups). The reasons will be discussed more in details in next section.

The right group allows die partitioning (i.e., connection without going through I/Os) more easily. Therefore, they not only offer higher content integration, but also can provide large power-performance benefits because which reduces the number of connections required for I/Os. Xilinx' 3D FPGA (SSIT – stacked silicon interconnect technology) are using CoWoS.

B. Active 3D-IC technologies

The "active" 3D-IC defines as the connecting components contain active transistors.

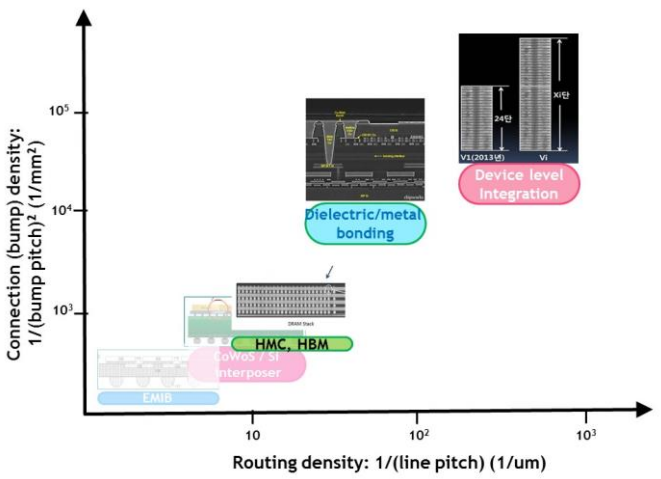

Figure 2 Active 3D-ICs, using similar vertical and horizontal axis's as Fig.1

Fig.2 is an illustration of active 3D-ICs categories. Similar to the Fig.1 list of passive 3D-ICs, in Fig.2 it still uses the connection density vs. routing density as measure. Once again there are many different ways to describe 3D-ICs, each has its own merit.

Compared with passive ones, active 3D-ICs are so far all using silicon technology as base, thus both their connection and routing densities are equal or much higher than that of passive 3D-ICs. Plus with active devices (transistors), active 3D-ICs in general are more capable and denser. Nevertheless this does not mean when active 3D-IC becomes vastly available in future, passive 3D-IC would disappear. On contrary, author believes both passive and active 3D-ICs will co-exist.

Note in Fig.2, the scale is different than that of Fig.1 passive 3D-IC situation. At the lower-left corner, it is the "stacked DRAM memory" group, i.e. HMC and HBM, etc. They use the similar uBump and TSV as those passive 3D-ICs such as CoWoS, therefore having similar connection and routing density. Fig.2 showed the passive 3D-IC right-most group there as reference. One can see these two groups, the highest density in passive 3D-IC and the lowest group in active 3D-IC, are in the similar position there. Today these stacked memories are in early sampling / production.

The middle group of Fig.2, moved away from uBump technology which is believed very difficult to achieve much better than today's connection density, i.e. ~1000 per mm^2. The connection technology in the middle group is dielectric or metal bonding based, which can easily reach one or two more orders of magnitudes higher than that of uBumps, in $10^4 \sim 10^5$ or more ranges. Most of these bonding technologies today are in wafer level, i.e. it requires stacking one wafer to another wafer. The key to success of these bonding technologies is the surface preparation and flatness. It is much easy to handle surface preparation in wafer level process, therefore makes these bonding technologies more achievable. It also brings challenges to product design - it requires foot-print compatibility between stacked wafers, which demands much closer co-design of these wafers. This also makes integration 3rd party IP more challenging. A typical example of product in this middle group is integrated logic and image sensors. Recent publication showed their success [4].

The right-most group, is stacking at device level during silicon process, for example multi-stacking Flash which is already in early production. Because it is in full silicon wafer process, the connection capability should be the same as today's silicon process, i.e. it can be the same as contacts, vias and MEOL layers, as small as ~50nm in pitch. It is the same to its routing capability, also in ~50nm pitch range today. Although device-level 3D-IC integration is the most capable technology, in terms of its connection and routing density, there is so far only one type of product, Flash memory, be able to utilize it. The reason is quite straightforward; it is because only Flash can stack memory cells without the need of metal connections in between. One can not produce FEOL (front-end-of-line) devices after BEOL (back-end-of-line) metals, due to temperature and contamination constrains.

III. ADVANTAGES AND DISADVANTAGES OF 3D-IC TECHNOLOGIES IN 3D FPGA APPLICATION

In this section, comparison of several passive 3D-IC technologies on their application to 3D FPGA will be conducted. In Fig.1 passive 3D-IC technologies, author picked MCM, EMIB and COWOS for this study. InFO was originally as a part of consideration, but because it is more suitable to a size equal or less than 15 x 15 mm^2, and more practical to have 2 or less routing layers. While 3D FPGAs are in general much larger and require more layers of routing, thus InFO is not included here for the comparison.

A. 3D-IC FPGA partitions

Fig.3 shows a typical FPGA nowadays in block view.

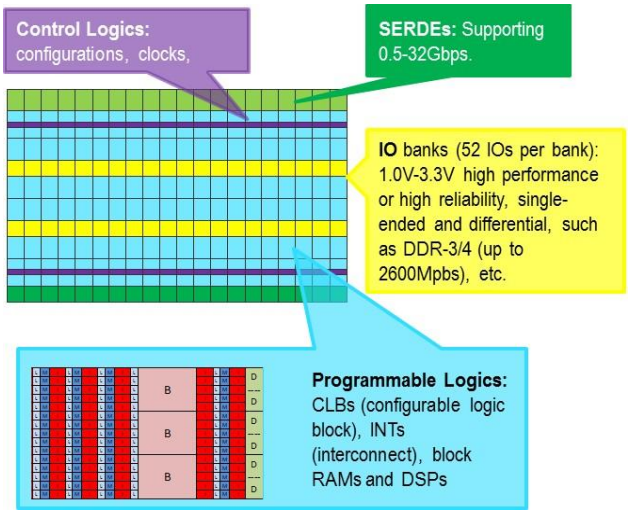

Figure 3 FPGA block diagram, which includes Programmable Logics, IOs, SERDEs and Control Logics, etc.

A FPGA consists many blocks regularly arranged. In Fig.3 4 major blocks are listed for later discussion convenience (there are actually many more blocks exist in FPGA architecture). These 4 blocks are Programmable Logic blocks including CLBs (configurable logic blocks), INTs (interconnects), RAMs and DSPs; there are also IOs, SERDEs and Control Logics.

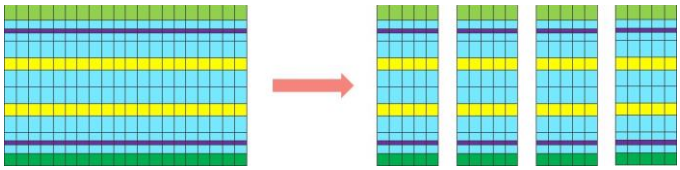

Figure 4a A FPGA is partitioned into several homogeneous dies, each contains all 4 basic blocks

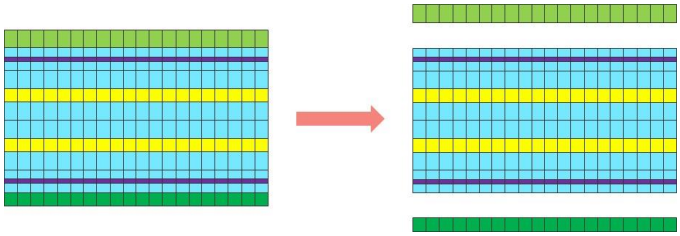

Figure 4b A FPGA is partitioned into SERDE and rest FPGA dies, heterogeneously

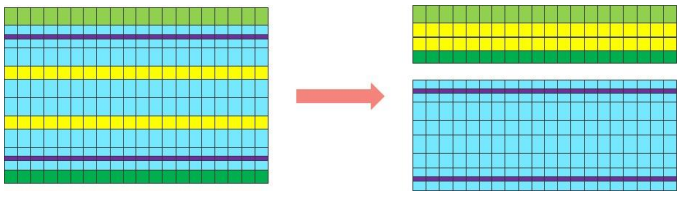

Figure 4c A FPGA is further partitioned into IO and SERDE in one die and rest FPGA dies, heterogeneously

Depends on different purpose and technologies, a 3D FPGA can be seen as different FPGA partitions. Fig.4a, Fig.4b and Fig 4c are 3 easily thinkable ways.

In Fig.4a a FPGA is partitioned into several identical sub dies homogeneously. In this case it requires approximately 10k to 15k of connections between each pair of sub dies. At the same time this partition allows best yield advantage (of small die vs. large die) [5], as well as allowing larger than standard photo lithography scanner field size 3D-IC FPGAs [6].

In Fig.4b's situation, partitioning FPGA into SERDE dies and rest. It requires less dense connections between them. For example in Xilinx' 28nm heterogeneous 3D-IC 7vh580T, the number of connections between logic die to a 2-quat (8 channels) 28Gbps SERDEs were in 2k ~ 4k range [7]. This way it allows design core logic and SERDEs in different technologies. However it does not allow yield benefit and larger-than photo lithography scanner field size FPGAs.

The 3rd partition approach in Fig.4c requires much higher connection density than that of Fig.4a situation, thus all today's passive 3D-IC technologies listed in Fig.1 would not be able to support. This situation will thus not be further discussed in this article.

B. Technologies suitableness

As mentioned above, because of InFO's limited size and routing layers count as of now, its application to 3D FPGA will not be discussed in following section. Only MCM, EMIB and CoWoS will be studied.

1. MCM

Capability: Fig.4a type of 3D FPGA where 10~15k connections between dies needed, would most likely not doable in MCM; However small quantity of SERDEs die partition in Fig.4b type of 3D FPGA, for example 1 quat / 4 channels, which needs 1k to 2k of connections, would be possible in MCM. Nevertheless even such a relatively "small" number of additional connections (besides originally needed C4s), it would make that MCM package substrate design more complicated thus more expensive. Also due to larger size of connection (C4) and routing dimension, it would be somewhat more challenge for high speed signals.

Matureness of the technology: MCM has been quite mature, after many years of volume production.

Cost: manufacture cost should be relatively low compared with other two candidates (EMIB and CoWoS). However for overall 3D FPGA cost, yield is the number one factor. MCM 3D FPGA would need equal amount of effort of other technologies on KGD (known good die), DFT (design for testing), redundancy and repairing capability, etc., which needs good amount of effort and resources (as part of cost).

Major challenge: by the end, the economic benefit of MCM 3D FPGA needs to be proven. Since it can only do small quantity of SERDEs die partition in Fig.4b situation, whether ultimately MCM 3D FPGA can be cheaper than just integrate these small number of SERDEs into a monolithic chip which requires simpler package substrate and possibly better yield (than 2 chips and MCM integration), is really a question.

2. EMIB

Capability: because using silicon connection (Si bridge) and uBumps, EMIB should be able to do both Fig.4a and Fig.4b types of 3D FPGAs. However following several factors might somewhat reduce EMIB's capability. Such as (a) because of CTE (coefficient of thermal expansion) difference between silicon vs. organic substrate, which could make alignment more challenging; (b) no TSV in silicon bridge could make power (Vcc and GND) connection not easy in bridge area, which translate to signal integrity challenge; (c) different size of bumps (uBumps on silicon bridge while C4 on organic substrate) further challenge alignment; etc. Thus overall EMIB's capability on 3D FPGA might still be OK but not as good as following CoWoS.

Matureness of the technology: not yet having production experience in industry. It might take time and volume to learn yield, reliability, mufactureability and to ramp.

Cost: the manufacture cost of EMIB 3D process would be lower than that of CoWoS but higher than that of MCM, obviously. Nevertheless of the total cost of a 3D FPGA product, as previously mentioned, yield is the number one factor and total BOM cost must be considered. Because CTE difference and its at relatively early stage, the yield of EMIB will take time to ramp; The embedded silicon bridges, especially when at larger number, will make package substrate manufacture more challenge thus higher cost. These two factors should be considered as a part of EMIB 3D FPGA cost.

Major challenge: author believes the CTE difference and different size of bumps in one die attach process, could be most challenging part of EMIB. If the yield is not good enough, then it needs large effort to achieve the benefit of EMIB.

3. CoWoS (Si interposer, TSV and uBumps)

Capability: CoWoS which represents silicon interposer with TSV and uBump technology can do both Fig.4a and Fig.4b types of 3D FPGA without problem. Actually both have been done and are now in volume production. Other products, such as HBM and SOC integration, etc. have been announced and are coming up soon. Please also note that today's stacking DRAMs, such as HMC and HBM, although as active 3D-IC, they are essentially the same TSV and uBump based technology.

Matureness of the technology: relatively mature, in volume production in past several years. If counting stacking DRAMs, the volume is relatively high today.

Cost: the manufacture cost of CoWoS is higher than both EMIB and MCM, because of TSV and CoWoS 3D integration steps which were not present in other two situations. However if including yield and package substrate cost together, the gap might not be as significant. Plus all CoWoS types of technology providers are working on cost reduction too.

Major challenges: CoWoS technologies are relatively mature, i.e. major road block problems have been solved in past several years. Perhaps how to further reduce cost in order to attract more products, are the major challenge [8].

IV. WHAT ARE NEXT

Above presentation discussed various 3D-IC technologies and their applications on 3D FPGAs. Author expects more 3D FPGAs (in number and percentage of products) in future because more mature 3D-IC technologies and continual increase of top die wafer cost each node, thus 3D FPGA making more sense.

To IC industry, 3D-IC is in general still in early stage. The key to its further growth is not only more variety of new 3D-IC technologies, more capable and less costly; but also finding more suitable products with architectures which will benefit from 3D-IC technologies. The latter is more important than that of former in author's view.

REFERENCES

1. M Bamal, S List, M Stucchi, A Verhulst1, M V Hove, R Cartuyvels, et al., "Performance comparison of interconnect technology and architecture options for deep sub-micron technology nodes", Proc. 2006 IEEE International Interconnect Tech. Conf. (IITC), Burlingame, 2006, pp. 202–204.

2. A W Topol, D C La Tulipe Jr., L Shi, D J Frank, K Bernstein, S E Steen, et al., "Three-dimensional integrated circuits," IBM J. Res. And Dev., Vol. 50, No. 4/5 July/Sept. 2006,

3. H Braunisch, A Aleksov, S Lotz, and J Swan, "High-Speed Performance of Silicon Bridge Die-to-Die Interconnects", Proc Electrical Performance of Electronic Packaging and Systems (EPEPS), 2011 IEEE 20th Conference, pp 95-98

4. Sony Product Technology Release ISX014, "Diagonal 4.6 mm (Type 1/4) 8.08M-effective pixel stacked CMOS image sensor (SOC) with built-in high-performance image-processing engine and camera system functions," http://www.sony.net/Products/SC-HP/cx_news/vol70/pdf/isx014.pdf

5. B Banijamali, S Ramalingam, K Nagarajan, R Chaware, "Advanced reliability study of TSV interposers and interconnects for the 28nm technology FPGA," Proc. 2011 IEEE 61st Electronic Components and Tech. Conf. (ECTC), Lake Buena Vista, 2011, pp. 285-290.

6. B Banijamali, H Liu, S Ramalingam, I Barber, T Lee, J Chang, et al., "Reliability evaluation of an extreme TSV interposer and interconnects for the 20nm technology CoWoS IC-package," Proc. 2015 IEEE 65st Electronic Components and Tech. Conf. (ECTC), San Diego, 2015, pp. 276-280

7. L Madden, E Wu, N Kim, B Banijamali, K Abugharbieh, S Ramalingam, et al., "Advancing high performance heterogeneous integration through die stacking," in Proc. IEEE 38th European Solid-State Circuits Conf. (ESSCIRC), Bordeaux, 2012, pp. 18–24.

8. W-S Kwon, S Ramalingam, X Wu, L Madden, "Cost effective and high-performance 28nm FPGA with new disruptive Stack Silicon Interconnect (SSI) technology," Proc. 2014 International Symposium on Microelectronics (IMAPS), San Diego, 2014, pp. 599 – 605