Panel: “Will 3D-IC Remain a Technology of the Future… Even in the Future?”

Organizer: Marco Casale-Rossi, Synopsys, USA
Invited Speaker: Patrick Leduc, LETI, France
Panel Moderator: Prof. Giovanni De Micheli, EPFL, Switzerland
Panelists: Patrick Blouet, STMicroelectronics, France; Brendan Farley, Xilinx, Ireland;
Anna Fontanelli, Monozukuri, Italy; Dragomir Milojevic, IMEC, Belgium; Steve Smith, Synopsys, USA

Abstract

If asked “who needs faster planes?” the vast majority of the 2.75 billion airline passengers (source: IATA 2011) would say that they do need faster planes, and that they need them right now. Still, the commercial aircrafts cruising speed has remained the same – 800 km/hour – over the last 50+ years, and after the sad end of the Concorde project, neither Airbus nor Boeing are seriously working on the topic. Along the same lines, when asked “who needs 3D-IC?”, most IC designers say that they desperately need 3D-IC to keep advancing electronic products performance, whilst addressing the needs of higher bandwidth, lower cost, heterogeneous integration, and power constraints. Still, 3D-IC continues to be the technology of the future. What are the road blocks towards 3D-IC adoption? Is it process technology, foundry or OSAT commercial offering, or EDA, or the business economics that is holding 3D-IC on the ground? In the introductory presentation of this panel session, LETI Patrick Leduc will illustrate the state-of-the-art of commercial, mainstream 3D-IC. EPFL Professor Giovanni de Micheli will then moderate an industry and research panel, to understand what are the key factors preventing 3D-IC from becoming the technology of today.

Invited Talk: Is 3D a Success Story?

Patrick Leduc, LETI, France

TSV and 3D chip stacking were imagined more than 20 years ago, but TSV technology is very recent. The first use of TSV in production for MEMS and image sensors dates only from less than 5 years. Prodigious advanced has been done during the last 5 years.

5 years ago, the main debate between technologists was on the possibility to modify the CMOS process to integrate TSV middle, a huge and deep Cu contact, without impacting the yield of the front end process. The choice between TSV last and TSV middle options was unclear. At that time, tungsten was often preferred to copper because of the unknown impact of thermal expansion mismatch between copper and silicon. The quality of copper filling in high aspect ratios was not good. Last but not least, the handling of thin wafer was not mastered and temporary bonding techniques were only at advanced development phase.

Today, all these technological challenges were solved and there is no serious technology impediment left on the road to 3D. Furthermore, several interposer (2,5D) and 3-D prototypes have demonstrated the feasibility of the technology. Xilinx was the first to launch a real product using a silicon interposer, the Virtex-7, well-known in the 3D community. Analysts expect a take-off of interposer business in the next years. Concerning 3-D, a relative deception was the Wide I/O memory for mobile application. Considering the very good results obtained with the technology, a production was expected this year but was postponed because classical DDR memories have done significant progress. However, the need for more bandwidth continues to growth and parallelism will remain a good solution. Heterogeneous partitioning has been demonstrated and is ready for production in several companies. First 3D standards have been proposed by SEMI. Questions remain on test, costs, and business models. These questions concern the industrialization phase, which shows that 3D ideas have gaining ground. Anyway, everybody is waiting the first revenue with 3D.

5 year of developments seems long when high level of investments are done and when huge hope is put on 3-D. Considering the modification it induces on semiconductor ecosystem, 5 years are probably not enough to really mature the technology. Moreover, improvements will continue appearing to reduce the TSV size and cost. Ideally, TSV pitch should approach global interconnect pitch in a logic stack. High density 3D stacking solutions are in development. New EDA tools for 3D partitioning are being developed. More value will put in interposer with embedded passives and actives devices. 3-D technology maturity is increasing and the number of 3-D vendors will continue to increase as well. 3-D adoption

978-3-9815370-0-0/DATE13/©2013 EDAA

will become easier. In parallel, in the near future, the cost of advanced CMOS technology node, as well as design cost, will be so high that heterogeneous partitioning will become a necessity. The question is: do we have alternatives to 3-D?

Panel Moderator's Introduction

Giovanni De Micheli, EPFL, Switzerland

3-Dimensional (3D) integration is just one embodiment of packaging electronic circuits, which has been and will be different in various product lines. Indeed, there is not yet a single, preferred solution to 3D integration, and possibly a plurality of solutions may coexist in the future. Large-scale memory systems span multiple dies, and their efficient packaging in 3D is key to their commercial viability. Processor-memory systems, and specifically manycores with hierarchical caches, will deliver increasingly superior performance only if connectivity can exploit features like wide-I/O for bandwidth and vertical connections for latency reasons. Thus the high-performance computing market demands 3D solutions. Finally, the increasingly more challenging design of required analog components within digital systems can be a showstopper to use downscaled technologies. The need or advantage to combine analog/RF circuits, antennae, and sensors in close proximity with digital processing is yet another driver towards smart packaging and 3D integration.

Panelists' Position Statements

Patrick Blouet, STMicroelectronics, France

3D integration is a technology on the road for a while, nevertheless it is far to be in use in mass production applications. There are several reasons to this situation and even if a lot of progress were made toward massive usage there are still important barriers which need to be overcome.

A certain number of successes are now visible in the FPGA or memory domain but there is still a long way to go to achieve an eco-system of die able to be stacked to produce very advanced system form different silicon suppliers.

The most important question is to understand why a system should go 3D. There are a lot of reasons to go in a 3D structure:

- Technology partitioning
- Performance partitioning
- Function partitioning
- End of efficiency of Moore's law

There are a lot of barriers to adoption of 3D, the main one being economical but it is not the only one. Among others are:

- Supply chain
- Lack of standard for 3D stacking
- Test complexity

3D with homogeneous suppliers is becoming a solution but 2D with heterogeneous suppliers is still out of reach. In parallel 2.5D solutions are getting momentum as it is for now good trade-off between technical requirements and business constraints.

Brendan Farley, Xilinx, Ireland

Today, the driving forces in all critical applications are power and bandwidth. This requires optimized and separate processes for digital, memory and analog while maintaining low power, low latency, wideband interconnect performance normally associated with monolithic implementation. Xilinx has recognized that this can only be achieved using 3D-IC technology and in 2012 Xilinx announced shipments of the world's first 3D heterogeneous FPGA product using stacked silicon interconnect.

Digital, Memory and Analog processes have not converged. For example, analog does not scale on advanced digital processes and cannot co-exist with large amounts of noisy digital circuitry. In addition, the cost of leading processes is increasing due to advanced lithography, complex devices and a limited number of suppliers. Therefore, companies must innovate with existing technologies. The only economic option is some form of heterogeneous integration.

Xilinx heterogeneous 3D devices deliver 2.8 Tb/s of bandwidth to meet the requirements of the wired communications market. This is enabled by power efficient, wide digital bus architecture between the FPGA and analog die within the package. For such applications, one cannot consider traditional packaging as a solution as it is not scaling nor is not expected to scale. This bandwidth imperative is not confined to wired applications: requirements for bandwidth in air is increasing by 3X every 18months, far exceeding even traditional Moore's Law improvements.

3D technology is economic where it delivers value over existing MCM technology in areas such as extreme bandwidth at low power and / or low latency interconnect. In 2011, Xilinx used SSIT to deliver more than Moore with the launch of the 6.8 billion transistor 2000T FPGA device - the world's largest. The many thousands of low latency, power efficient interconnect lines allow the 4 separate FPGA die within the package to interact as if a single, monolithic unit with no disadvantage in power or performance. Extreme bandwidth will become the norm

in numerous applications and this will create economies of scale to sustain and accelerate 3D penetration

The key impediment to accelerated deployment is one of business rather than technology: the business model to market 3D devices comprising die from multiple vendors is immature. Also, a lack of industry knowledge and established strategies to deliver Known Good Die (KGD), in particular for high performance devices, presents a related technical impediment. In addition, the limited number of technology vendors in the 3D ecosystem will delay widespread access and the benefits of economies of scale in the short term. These limitations may drive both horizontal and vertical integration in the industry, to allow companies to accelerate their own deployments and to gain competitive advantage in the market.

The next step in the evolution of the technology will be to add more value to the interposer through the integration of passive and active components. Various challenges remain before this can happen including: thermal simulation and management; and mitigation of Through Silicon Via (TSV) stress effects. The benefits will be higher levels of integration and / or reduced costs which will result in more widespread adoption.

3D-IC technology is here today: Xilinx is in production with 3D devices that would be otherwise impossible to manufacture. As the economics of the traditional semiconductor progression become more challenged while increasing numbers of applications drive extreme bandwidth at low power, 3D-IC technology will mature and gain widespread industry adoption.

Anna Fontanelli, Monozukuri, Italy

As technology nodes advance, limitations in bandwidth begin to curtail the number of cores that we can effectively integrate on a single die, as well as to limit the multimedia performance that we can reasonably achieve. The quest for higher bandwidth is possibly the strongest force driving us towards more than 2D-IC integration to improve performance while at the same time reducing power consumption.

This brings the I/O issue to the forefront: while the theoretical integration capacity keeps increasing from one technology node to the next, the number of I/O pins per million of transistors keeps decreasing, thus harming our ability to exploit a given technology node integration capacity to its full extent. Over five technology nodes, i.e. from 180 to 32/28 nanometers, the number of transistors per square millimeters has increased by 32 times, while the number of I/O pins per million of transistors has decreased by… 45 times; I/O pins do not comply with Moore's Law at all, and have become by now a precious resource.

There is one missing ingredient that would ease making 3D-IC the technology of today: I/O management, planning, and optimization. Only by handling the interface among package, silicon interposer(s), and die at the same time, in a homogeneous, integrated environment it is possible to make effective decisions about whether 2.5D-IC, 3D-IC, or a hybrid configuration is the most appropriate solution. As I write, there is almost an order of magnitude difference in terms of size and pitch, among a C4 bump, a micro-bump, and a TSV; silicon interposer features are one to three orders of magnitude smaller than package and TSV features, making "horizontal" connections much more effective than "vertical" ones; decisions about "horizontal" and "vertical" resources utilization have a profound influence on the overall performance and feasibility itself of the target system, and are often not trivial.

I am convinced that mixing and matching 2.5D-IC and 3D-IC in a hybrid configuration that some call 5.5D-IC, is the most effective solution moving forward, as it provides the advantages of both worlds, especially if silicon interposers become active as far as I/O is concerned. However, deciding the relative placement of the different die on a silicon interposer or in a stack remains a challenge; with computational complexity in the order of magnitude of $n!$ where n is the number of die, slicing/partitioning of more than 2D-IC systems, and "unraveling" thousands of "vertical" and "horizontal" interconnects demand for dedicated software tools.

Unfortunately, there is a historical barrier between the package design tools and the IC ones, which makes I/O management incredibly difficult. Due to the very low level of standardization, the strong dependence on each and every customer different requirements, and the very limited TAM, general purpose solutions from leading EDA vendors can be ruled out.

Personally, I believe that a "haute couture" solution from the "shop around the corner" is much more effective than "prêt-à-porter" from "department stores". Of course, this requires some sort of "engagement", but there are a lot of opportunities for the braves.

Dragomir Milojevic, IMEC, Belgium

Over the past few years both research and industry communities were heavily promoting 2.5 and 3D integration technologies as one of the possible paths to enable further IC scaling. But still today these technologies haven't been used to manufacture mainstream mobile or high-performance products. Often, the cost has been cited as one of the main blockers: expensive extra processing steps (TSV/µbumps, back-side metallization, die stacking, etc.), questionable yield and not fully developed supply chain. Further, the actual performance benefits are not as good as expected,

especially when taking the cost into account. And finally, not all of the practical aspects linked with the physical design of such systems are fully mastered. Although considerable efforts have been made to solve various design issues (thermal, mechanical, etc.), there is a general opinion that we lack design methodologies and EDA support. Or these are mandatory to enable cost-effective 3D products.

Having in mind all the above, the question whether the 3D technology will remain a valid path for further technology development remains open. Despite all the negative feelings, we believe that 3D integration does have a future. Even more, we believe that the 3D will be the only way to enable further scaling and practical usage of very Advanced Process nodes (those bellow 10nm). The future of 3D might be in the heterogeneous integration.

Today, manufacturing of integrated circuits is based on homogeneous integration solely, where one specific process is used to manufacture the complete circuit. To scale the circuit, all devices within the circuit need to scale equally. Recently we began to observe that it is becoming more and more difficult to keep-up with homogeneous scaling of all components in the technology library in a homogeneous manner. It is therefore possible to imagine that in one point in time we will hit, what we could call, the limit of cost-effective homogeneous scaling. After this limit, not all components required to build a usable IC will scale cost-effectively in the next generation process nodes. And as a worst case we could even imagine that bellow certain feature size it will be impossible to integrate certain functionalities, typically analogue, IO drivers, ESD, buffers for long interconnect etc.

If this hypothesis is shown to be true, the only way to enable further cost-effective scaling of ICs will be to introduce heterogeneous integration in which Advanced and Standard Process nodes are combined in the same circuit. Such combination will appear as non-separable one, since the Advanced Process part of the circuit will require Standard Process to enable communication with the external world, ESD protection, connectivity with distant blocks etc.

Heterogeneous circuits could be built using traditional 2D integration, but 3D looks like a perfect, and maybe the only cost-wise viable, candidate since it can naturally combine multiple process within the same circuit package. If 3D integration is adopted, we could have a choice between Die-to-Wafer (D2W) and Wafer-to-Wafer (W2W) stacking. D2W integration has been traditionally promoted because of the improved compound yield. After processing, wafers are subject to pre-bond tests that are used to determine good dies, and only known good dies are then individually stacked on the wafers. But if we assume heterogeneous integration with Advanced Process in which the IOs are not manufacturable any more, the die testing becomes difficult. The yield argument for D2W stacking doesn't hold any more and we could then use W2W, with face-to-face integration. Heterogeneous Advanced and Standard Process stack will have the usual FEOL and BEOL with the inter-die connections made using Cu-Cu pads. These pads can be small for Advanced Process node, typically 1µm2. For Standard Process node, the Cu pads need to be bigger to ensure the correct alignment precision typically 4X4µm2 and result in 5µm inter-die interconnect pitch (and these could potentially become more aggressive). In Standard Process wafers we could use less expensive via-middle TSVs to enable the connection between the stack and the package. Because of such high-density inter-die wiring capabilities, the BEOL appears now as a common, shared beck-end with full, or reduced metal layer stack on one, or both dies. This will depend on application and more precisely on routing requirements for a given partition and the floorplan.

Weather such integration approach will become a mainstream technology and a possible path for wide spread deployment of 3D stacking is still to be seen. Meanwhile we will continue the exploration of both technology aspects of W2W heterogeneous integration and the impact on the system level design. We see there a very interesting opportunity for the next generation system design.

Steve Smith, Synopsys, USA

Over the past few years, we have carefully looked at the technology and market landscape, in close collaboration with our partners. We have concluded that: an evolutionary transition from 2D- to 3D-IC integration is possible, and is steadily progressing as we write; silicon interposers (a.k.a. 2.5D-IC) play – and will continue to play – a fundamental role, due to their finer interconnect features as compared to both package substrates and TSV; and finally that 3D- and 2.5D-IC can indeed be the technologies of today.

Contrary to some claims, I am convinced that the sky is not falling: on the contrary, the eco-system is in place, and equipment manufacturers, silicon foundries, OSAT, and EDA companies have demonstrated the full technical feasibility of more than 2D-IC integration.

Our R&D investments and recent announcements both stem from these conclusions: leveraging our existing implementation and verification technologies, we address the roadmap and the requirements of our partners as they emerge, incrementally: physical implementation, RC extraction, DRC/LVS, static timing analysis (STA), design-for-test (DFT) and SPICE simulation have been enhanced to support 3D-IC and 2.5D-IC design.

As I write, we have not found any evidence of technical roadblocks. The enhanced EDA tools are available now, and we are collaborating with engineering teams that are ready to take advantage of the opportunities offered by 3D- and 2.5D-IC technologies as they emerge.

If there is one ingredient that is proving critical in moving 3D-IC from being the technology of the future to being the technology of today, this is collaboration, i.e. the willingness of all involved parties to work together, sometimes sharing the burden of debugging and improving the existing features, testing them again and again, accepting temporary workarounds while waiting for a new feature to become available, always acknowledging that technology is only seldom perfect since its inception, and knowing that this joint effort will pay off.

Moving forward, there is much more to 3D-IC than just implementation or verification, and I believe there is a great deal of opportunities to extend the traditional EDA solutions to support this emerging technology. For example, virtual and fast prototyping are ideal for highly heterogeneous systems such as 3D-IC; thermo-electro-mechanical simulation is also critical, and TCAD – while originally meant for process technology development – is an extremely powerful technology for modeling the thermo-electro-mechanical effects introduced by 3D structures.

3D-IC represents a technically and economically viable alternative to scaling, as well as a complement to scaling in boosting integration and performance to unprecedented levels. All the ingredients are within our reach.