3D Embedded Multi-core: Some Perspectives

Fabien Clermidy, Florian Darve, Denis Dutoit, Walid Lafi, Pascal Vivet
CEA-LETI, Minatec Campus, 38054 Grenoble, FRANCE
{firstname.lastname@cea.fr}

***Abstract*—3D technologies using Through Silicon Vias (TSV) have not yet proved their viability for being deployed in large-range products. In this paper, we investigate three promising perspectives for short to medium terms adoption of such technology in high-end System-on-Chip built around multi-core architectures: the wide bus concept will help solving high bandwidth requirements with external memory. 3D Network-on-Chip is a promising solution for increased modularity and scalability. We show that an efficient implementation provides an available bandwidth outperforming classical interfaces. Finally, we put in perspective the active interposer concept which aims at simplifying and improving power, test and debug management.**

Keywords: 3D, TSV, Through Silicon Via, Network-on-Chip, NoC, power management, test, debug

I. INTRODUCTION

Due to increasing flexibility, performance and power consumption constraints, multi-core architectures have to face many issues. On the other hand, 3D stacking is seen as one of the most interesting technologies for System-on-Chip (SoC) developments. However, the matching between 3D technologies possibilities and multi-core constraints is not straightforward. This is due to some limitations of 3D stacking technologies as well as a difficult economical equation which can be summarized as "who will be responsible in case of failure of a two chips assembly?".

The main limitation of 3D stacking is the TSV diameter which cannot probably go further beyond the µm scale, far from via density: the thickness step required for the wafer is directly related to the TSV diameter for via filling reason. Few µm for this thickness is hard to perform for the assembly step due to physical constraints. The consequence is that fine-grain partitioning leads to solutions with limited power and performance gains [1] and considerable area loss. However, higher grain partitioning is still a technically viable solution. In terms of economical view, it is requested to create a new commercial sharing between assemblers, founders and IP providers. In terms of architectural view, it requires thinking about new partitioning between different on-chip or on-board components. In this paper, we investigate three objectives for partitioning we think are pertinent for future 3D embedded SoC based on multi-core architectures.

Communication of the SoC with a fast high-storage memory (DRAM or SDRAM type) has always been a challenge for multi-core architectures. Such memory is off-chip due to density versus price reasons. High resolution TV, image processing and augmented reality are some typical examples of applications which require drastically increasing this bandwidth with the working memory. Fast serial links have emerged for solving this issue as well as their low-power counterparts, such as DDR and LP-DDR standards. Hence, these standards have extrinsic limitations when the bandwidth must be further increased mainly due to high frequency in a noisy (variable) environment. Reducing this frequency is possible if the bus width can be enlarged. Current SoC cannot afford such a solution due to pad number limitation. In this context, the number of TSV available per area unit can help solving this issue. This aspect will be further discussed in section III.

Scalability and modularity of the performance is another advantage brought by 3D TSV technology. The idea is to use stacking for reducing the die size, but obtaining similar performance than a large die thanks to stacking of simple dies. It will then reduce the fixed costs and open the way to a range of low-end to high-end markets for a single die. One example is given in Multiple Input Multiple Output (MIMO) telecommunication applications: a single chip can be used for a single antenna, and stacking of multiple chips will upgrade the performance and the number of antennas that can be processed in parallel. In that case, efficient communication between computing cores in the stacked system is essential. 3D Network-on-Chip is the fundamental element for dealing with this issue, and will be discussed in section IV.

Advanced technology nodes, 32 nm and above, are very expensive technologies and are prone to variability. Including passive, analogue IPs, test or debug is, in a way, a waste of this costly area. With high-density TSV, we can envisage deporting a part of such functions as well as components of the board, in a less aggressive, less costly technology. This is the concept of active interposer which will be developed in section V.

II. 3D TSV TECHNOLOGY: SOME FIGURES

In 3D stacking technologies, TSVs are the basic building elements providing connections between the different stacked dies. The idea is very simple, and consists in making a hole in the silicon which will be filled in another process step. However, multiple technological options can be investigated. In this section, we present a short overview of various 3D technologies [2].

On a top level fabrication process, three main techniques can be used to stack integrated circuits: Wafer-to-Wafer (W2W), Die-to-Wafer (D2W) and Die-to-Die (D2D). In W2W, two complete wafers are stacked together and dies are extracted after assembling. This solution provides a high throughput process but a huge constraint: dies of the different wafers must have the same dimension. Respectively, in D2W technique a die from one wafer is picked and placed on the top of another

978-3-9810801-7-9/DATE11/©2011 EDAA

die integrated in a wafer. This solution allows testing dies individually before 3D assembly. This is commonly called the Known-Good-Die assembly process (KGD) which provides a better yield compared to the W2W technique. Moreover, it deletes the constraint of identical dimensions for the two dies and heterogeneous technologies (distinct CMOS nodes, analogue, memories, MEMs …) can be used. D2D is the last solution, more costly in terms of fabrication, but mandatory if the assembly step is done by the manufacturer of the lower size die or by an external assembler.

Three principal manufacturing steps are usually used for the 3D process: via drilling and filling, wafer thinning and backside metallisation. These three steps may be processed together or separately. Depending on these steps positioning in the global manufacturing process, three technology options are possible: TSV first, TSV middle or TSV last. TSV are said "firsts" if they are processed before the front-end CMOS steps (transistors masks) [5]. The so-called "Via Middle" TSVs [3] are processed after the front-end transistor process but before metal layer fabrication. Wafer thinning and backside metallisation can be process during the assembling steps. Finally, "Via Last" [4] (or post backend TSVs) are realized after the metallization stage. In such a case, TSVs can be processed by packaging fabrication lines.

Different TSV integration densities are today available. Low density TSVs have a pitch larger than 100 µm. Medium density TSVs [6] have a pitch of about 50µm which represents an integration density of about 500 TSV/mm². High Density TSV [7] consists of really high aspect ratio TSVs with a diameter lower than 10µm which offers an interesting integration density of 10 000 TSV/mm² .

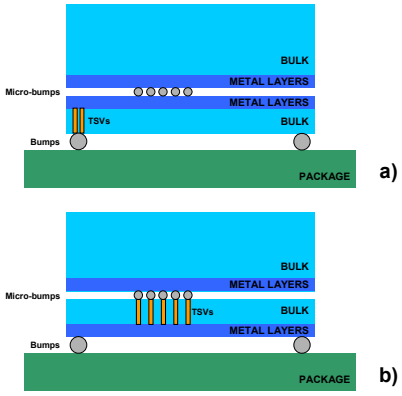

Figure 1. Stacking Options: a)Face-to-Face, b)Face-to-Back

Once the process is defined, different options can be used for the assembly step. The most natural one comes from flip-chip techniques. It consists of solder micro-balls which perform both electrical conductivity and mechanical links between dies. In order to reduce existing flip-chip bumps dimensions to be compatible to TSV sizes, copper pillars also known as micro-bumps have been developed [9]. Like BGA and flip-chip packaging techniques, copper pillars are fabricated using a brazing (soldering) process operation. Copper-pillar diameters have been demonstrated down to 25 µm [6]. A more advanced packaging solution consists in direct copper bonding between two dies without any brazing element by using molecular copper bonding. This type of process is still active R&D and is not yet available due to surface roughness constraints, but will allow higher integration density probably less than 10µm [10].

Finally, depending on the side of the two dies (back side is the passive silicon, front side is the active layer), different assembly can be realized (Figure 1). For the Face-to-Face (F2F) connection, two dies ended by a high level metal interconnection (called "Alu Cap") can be assembled without thinning step. Nevertheless, TSV process steps are required to interconnect the stack with the package and the thinning process is performed on the assembled dies. For the Face-to-Back (F2B) solution, one of the two dies must be thinned before assembly to reach the TSVs contact. Note that this last solution is the only one which allows stacking more than two dies.

III. Improving Memory Communication Bandwidth

With the increasing demand for processing power in multi-core SoC, memory-interconnect bandwidth is becoming the performance bottleneck of the overall system. Tera-scale memory bandwidth will likely be required in the coming years. Today's off-package memory interconnects have reached their limit in terms of cost, bandwidth capability, power consumption and implementation complexity. Three-dimensional (3D) stacking, which enables low-latency, high bandwidth and very dense die-to-die interconnects, is a unique opportunity to propose new connection schemes between the multi-core SoC and the memory.

A. The Memory-Link Bottleneck

Typical computer architecture [14] as described in Figure 2 consists of a microprocessor (CPU), a chipset and the main memories connected through the system bus (interconnect). The system and processing memories are split to provide the require bandwidth.

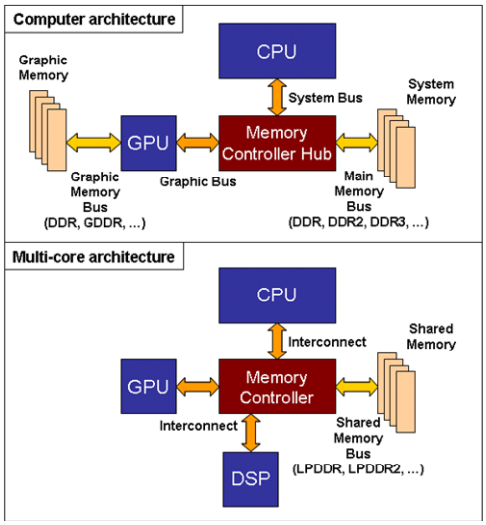

Figure 2. System architectures and memory interconnects

On the contrary, architectures targeting mobile applications must fulfil hard area and power consumption constraints. As a consequence, modern SoC are integrating the chipset (memory

controller, graphic component) as IP blocks with the CPU and the modem processor in a single multi-core SoC. Such organization puts more pressure on the memory bus as all the data and control traffics are merged on the same physical bus.

B. *Bandwidth and power consumption requirements*

Memory bandwidth is defined as the product of the number of data bits in the memory bus and the speed of a single bit [14]

BW = nb of bits x single bit rate

A typical value reached by the current generation of multi-core SoC used for handheld devices is the following one: 32 bits wide bus; LPDDR2 JEDEC standard providing 533MHz frequency. Consequently, the memory bandwidth expressed in bytes per second is:

4 Bytes x 533MHz x 2 (Double Data Rate) = 4.264 GB/s

The power consumption of such an interface can be expressed as the product of the memory bus capacitance, the square of the signalling voltage range and the signal toggle rate. In other words, the power consumption is proportional to the interconnect technology capacitance, the interconnect activity (linked to the frequency of the memory bus) and the square of the voltage excursion.

In a typical multi-core SoC, many operators are requesting data from the same shared memory: the CPU, the multimedia processors and the modem processor. Complex use cases could occur in such a computing scheme and lead to a heavy demand on memory bandwidth. It is foreseen that the terabyte-per-second [15] memory bandwidth will soon be reached. The following section will see the impact on actual memory interfaces and their evolutions.

C. *High-Bandwidth Memory Interfaces Overview*

In computer architecture, a widely used memory interconnect is the DDR3 protocol. It is a parallel 32 bits dual clock edge (DDR) synchronous protocol. The maximum frequency is 1066 MHz and the IO-voltage is 1.5 V. This interface achieves approximately 10 GB/s memory bandwidth with a power consumption evaluated at 30 mW per Gbit/s. The devices (processor and memories) are placed in different packages on a PCB (Figure 3). The package and PCB technologies used in computers permit a good transmission line quality to ensure the required signal integrity for 1.5 V-range signalling level at 1 GHz.

In mobile computing multi-core SoC, the cost-driven PCB does not allow the required signal integrity for such signalling level and frequency. A dedicated protocol has been standardized in JEDEC to address mobile computing specific environment targeting lower cost and lower consumption than computer architecture. The most widely used memory interconnect is the LPDDR2 interface. It is derived from the DDR3 protocol and keeps the main features, such as the parallel 32 bit dual clock edge (DDR) synchronous protocol. The main differences are coming from the signalling voltage (1.2 V) and the lower frequency (533 MHz). These specificities are aiming at less power consumption with the drawback of a more limited memory bandwidth.

D. *Memory-Interface Evolution*

As previously noted, the next generations of memory interfaces have to cope with increasing memory-bandwidth demands while keeping the power consumption at a reasonable limit. The memory bandwidth formula expressed above shows two solutions to improve the bandwidth: increase the number of data bits or increase the bit rate thanks to the frequency.

The power consumption as indicated above can be in three ways: by limiting the interconnect capacitance, by limiting the signal voltage range (quadratic impact) or by limiting the frequency. Finally, each memory interface evolution has to keep the latency stable for performance reasons. With these simple equations in mind, an evolutionary trend is emerging for high bandwidth: wide data memory interface.

E. *Wide Data Interface*

For this interface proposal, the bandwidth improvement is achieved through an unprecedented extension of the bus width between the multi-core SoC and the memory up to 512-bits. This evolution has been made possible thanks to the 3D stacking [16]. This 3D technology provides a much higher connection density that makes the overall silicon cost (area) of the wide data memory interface in the same range as the typical LPDDR2 interface located in the SoC pad ring.

Lower power consumption is mainly achieved thanks to the lower capacitance of the TSV interconnect. In conjunction with wider data bit, a Single-Data Rate (SDR) is used and the frequency reduced from 533MHz for the LPDDR2 interface down to 200MHz. The lower bit rate significantly simplifies the design of the pad interface. Signal-integrity issues are also singularly minimized with SDR signals at 200MHz through TSVs between dies. In a whole, the bandwidth improvement provided by the wider data bus interface (x16 vs. LPDDR2) compensates and exceeds the performance loss coming from the lower bit rate (/5.33 vs. LPDDR2). The resulting wide data interface throughput is three times the LPDDR2 one. With regards to package integration (Figure 3), the wide data memory is stacked on top of the processor SoC. The connections are located in the centre area of the memory die and consist of an array of micro-bumps whose pitch is in the range of 40µm to 50µm.

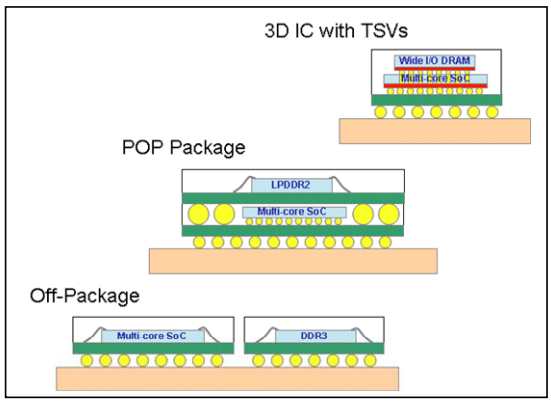

Figure 3. Proposed roadmap of package architecture to address the memory bandwidth challenge in mobile world

F. Conclusion

Table I gives a summary of the three different memory interfaces described in this section for the 1TBps target. It clearly proves the need to go to large bit-width interfaces for power consumption reason. In this context, 3D stacking has a clear advantage compared to other interfaces solutions. DRAM memory on processor vertical integration with wide data interconnects through TSV is seen as the driving 3D technology for improving the multi-core computing performances.

TABLE I. MAIN FEATURES OF DIFFERENT HIGH-BANDWIDTH MEMORY INTERFACES

	Memory link, peak bandwidth and power consumption efficiency			Cost for 1TBps memory bandwidth	
				Number of data IO pins	Interface power consumption
Computing memory IF standard	Multi-core SoC → DDR3 → DRAM 1066 MHz I/O bus clock, 32 bits, 1.5 V, Double Data Rate	8.532 GBps 30 mW/Gbps		3800	240 W
Mobile memory IF standards	Multi-core SoC → LPDDR2 → DRAM 533 MHz I/O bus clock, 32 bits, 1.2 V, Double Data Rate	4.264 GBps 20 mW/Gbps		7700	160 W
	Multo-core SoC → Wide I/O → DRAM 200 MHz I/O bus clock, 512 bits, 1.2 V, Single Data Rate	12.8 GBps 4 mW/Gbps		41000	32 W

IV. MULTI-CORE SCALABILITY USING ASYNCHRONOUS 3D NOC

A. Introduction

Even if 3D technology introduces a whole set of application possibilities and can solve some architectural issues as proved in the previous section, it also aggravates some current VLSI design problems. For example, delivering clock to each die and dealing with clock synchronization is a critical problem in the context of 3D integrated circuits. Remembering that even in two dimensional chips constructing a highly symmetric fractal structured clock distribution network (e.g. an H-tree), which is essentially needed to route the clock signal to all parts of a chip with equal propagation delays, is very hard, we can imagine that constructing a three-dimensional clock distribution network (tree) is very difficult, as a huge number of vertical interconnects will be needed [11]. Furthermore, run-time temperature variations and stresses associated to thermal issues significantly increased in 3D integration will introduce additional time-varying clock skew.

As a consequence, Globally Asynchronous Locally Synchronous (GALS) approaches seem to be mandatory to be used in 3D integrated systems. In GALS system, several physically independent clusters are clocked by a different clock signal. Moreover, GALS paradigm enables the implementation of various forms of Dynamic Voltage and Frequency Scaling DVFS methods [17] which will be essential for future 3D-SoCs to limit the power consumption and the healing of the stacked structure.

In order to provide the necessary computational power and communication throughput, Networks-on-Chip (NoC) offers a structured solution to construct fully scalable Multi-Core architecture, while benefiting from a GALS paradigm. Since a NoC spans the entire chip, the network can be used as the globally asynchronous part of the system while the subsystem modules - individual cores or cluster of cores - are locally synchronous parts [12].

For 3D technologies, a structured interconnect such as a 3D NoC using a GALS scheme will benefit from the third dimension and will bring a scalable and modular communication infrastructure which can be easily exploited.

B. 3D NoC infrastructure

Figure 4 shows an example of a 3D NoC architecture. The 3D NoC can be easily built as a regular 3D-mesh, using NoC routers and NoC links. In order to reduce the high complexity of a 7x7 3D NoC router and preserve its performance, a 3D router can be implanted as a hierarchical router [13]: the 7x7 3D router is then decomposed in one 5x5 "*2D*" router dedicated to intra-tier communication and one 4x4 "*3D*" router for vertical communication. Regarding performances, the throughput is preserved due to smaller routers, and latency is only increased for NoC paths when changing from the 2D plan to the 3D link or to the local processing unit. For 2D paths only or 3D paths only, the latency is not increased. Such a regular 3D NoC topology offers a scalable interconnect for Multi-Core.

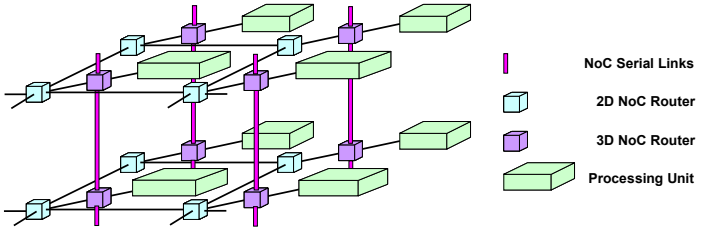

Figure 4. 3D NoC topology using hierarchical routers and serial links

C. Asynchronous logic implementation

As previously shown, the clocking strategy of such a communication infrastructure must be carefully analyzed. An interesting solution is to implement the 3D NoC architecture using Quasi Delay Insensitive (QDI) asynchronous logic, as presented in [12]. As a result, no clock will go through the two dies, simplifying the clock scheme to two classical 2D problems. The network low level handshake signals are encoded using a QDI 4-phase 4-rail asynchronous protocol, also called 1-of-4, as presented in figure 5. The four data rails encode a two bit value, with one acknowledgment signal.

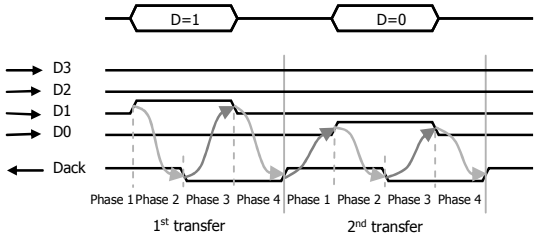

Figure 5. QDI 4-Phases Asynchronous Protocol

In such QDI logic, all along any individual asynchronous handshakes, any signal transition must be acknowledged by a new signal transition. This very strict logic level property ensures that the functional correctness of the logic is insensitive to any delay variation caused by voltage, temperature, or fabrication process variations. Moreover, this asynchronous logic property allows mitigating any timing variations due to 3D stacking: thermal impact, TSV delays, Micro-bumps delays, variable or unknown delays due to heterogeneous dies stacking, chips from distinct corners, etc.

D. 3D Serial Link

Die-to-die interconnect pitch imposes a large area overhead compared to the corresponding horizontal wires. There is a trade-off between the great benefit of exploiting high-speed short-length TSVs and a demand to limit their number. On a technological view, TSV resistance and capacitance are very low, allowing a transition frequency much higher than in 2D plan. Consequently, serializing the vertical links [19] in the asynchronous 3D-NoC can help minimizing the number of TSVs without bandwidth loss: an asynchronous 4-phase handshake may sustain about 2Gbit/s, which is 4 times faster than a similar asynchronous 2D NoC average throughput (about 500MHz).

When designing a 3D serial link, the main design tradeoff concerns the serialization ratio, which defines the number of 3D connection per 3D link. For the proposed 4-phase 4-rails asynchronous encoding, each NoC bidirectional 36-bit link represents 180 signals (including a few control signals). For instance, if serialization ratio is two the number of bits is reduced to 18-bits, which reduce the number of TSV connections down to 90 signals, but also divide by 2 the NoC vertical link throughput.

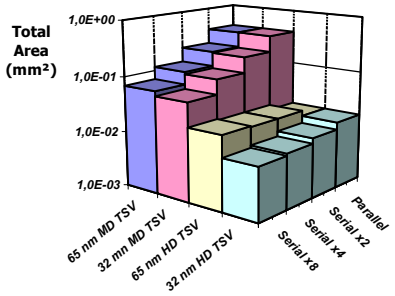

Figure 6. Serialization area reduction tradeoffs

Figure 6 shows the serialization area reduction tradeoffs according to the serialization ratio, the TSV density and the CMOS technology. TSV technology is independent from the CMOS technology and thus TSV sizes and CMOS density can be mixed. In this example, two TSV technologies are evaluated: middle density [6] (MD TSVs) and high density [7] (HD TSVs). Their integration density are respectively 500 TSV/mm² and 10 000 TSV/mm². Similarly, two CMOS technologies are used for the serialization logic, 65nm and 32nm. For conservative MD TSV, we can observe that serialization can save significant area, without degrading throughput. This is less true for aggressive HD TSV where higher density CMOS technology is necessary to compensate the cost of serialization logic.

E. Comparison between WideIO and 3D NoC

3D NoC gives the opportunity to go further in terms of performance than a memory interface thanks to more limited constraints. In order to give an idea of the potential throughput which can be obtained thanks to 3D technologies, we made a comparison between the wideIO standard and its 3D NoC counterpart.

TABLE II. COMPARISON BETWEEN WIDEIO AND 3D NOC

4 x Bi-dir channels 128 bits	#TSV	#bits & frequency	Through-put	Power Consumption
WideIO	1104 TSVs	512 bits @ 200 MHZ	12.8 GB/s	32 mW/GBps
ANoC *(x4 serial link)*	712 TSVs	2 * 512 bits @ 500MHz	64.0 GB/s	26.6 mW/GBps

Table II summarizes the results. Serial link allows to reduce the overall number of TSVs, while still offering aggressice throughput. The potential of 3D asynchronous NoC technology is clearly demonstrated with 5 times the throughput of the wideIO for similar range of power consumption per GBps.

V. DEPORTING SOME FUNCTIONS

The notion of active interposer is seen as an interesting integration technology for 3D stacking of heterogeneous systems. The idea is quite simple: a SoC is partitioned into several dies, each of one being processed with the most relevant technology node in terms of performance and cost. Multi-core SoC can take advantage of such strategy by splitting the processing parts, the memories and the "service" units (power, test, debug, legacy I/Os). In such an integration scheme, the interposer can be passive and becomes the backplane of the platform by implementing the signal as well as the power distribution interconnects. Going to an active interposer is a natural evolution that will permit to embed some legacy peripherals and power management functions (Fig. 7).

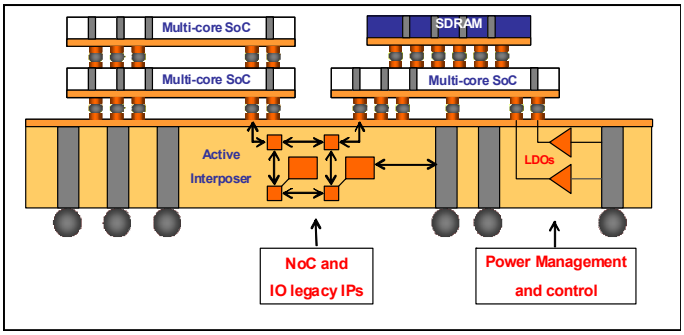

Figure 7. Multi-core integration in an active interposer based 3D stack

A. Legacy IOs

In a conventional SoC, input/output pads are placed at the periphery of the die. As a result, when the number of IOs becomes high the multi-core SoC area is fixed by the pad ring area and is said pad limited. Some complex technical solutions such as double row pads and embedded pads can be sparsely used for limiting this effect. However, with new advanced technology nodes, pad cells don't shrink as much as the logic

and the memory cells. The equivalent cost of pad area is then increasing with the technology node evolution.

With 3D IC, we explore new possibilities of system level partitioning where the IO pads can be placed in an active interposer using a mature, cost effective, technology. The new interface between the Multi-core SoC and the interposer can then be limited to an efficient and unique interconnect like the 3D NoC. In this context, the interposer embeds the peripherals which are connected to a large number of IOs. Deporting the IOs from the multi-core SoC to the interposer will save between 5 to 10% area of the high performance digital SoC and save time-to-market by avoiding the redesign of these IPs.

B. Power management

In a conventional 2D planar digital system, the energy management SoC is placed on the Printed Circuit Board (PCB) close to the Multi-core SoC. Power delivery in such a system is deeply impacted by parasitic on the supply path: PCB, package, Multi-core SoC IOs. This long path degrades signal and power integrity (SI/PI) and does not allow any accurate power management feedback loop. One of the solutions proposed to improve signal integrity is to increase the number of supply IOs of the Multi-core SoC. The power integrity is improved by applying decoupling capacitors on the supply delivery path. However, whatever the solution, 2D approach is drastically limiting the precision of feedback loop for power delivery management. This accurate control is now needed to optimize energy efficiency in presence of in-die variability [18].

Interposer based power delivery is seen as an opportunity to explore new power distribution architecture needed for fine-grain control to cope with in-die variability. Passive interposer with embedded passive components is a first step for improving SI/PI of Multi-core SoC. Becoming active, the interposer will act as a power hub of the Multi-Core SoC thanks to its central place in a 3D stacking based design. Voltage regulators can also be embedded in the interposer in order to shorten the power delivery path and decrease parasitic. Vertical layout arrangement through simultaneous floorplanning between dies is also seen as an opportunity to make the power control devices in close relationship with the power domains of the Muti-core SoC.

VI. CONCLUSION

3D stacking of Multi-core heterogeneous system opens a new era of architecture exploration with new partitioning of the overall system on chip. In this paper, three perspectives have been presented using 3D stacking: The WideIO memory scheme is probably the most advanced for providing high throughput memory connection; this can be further extended to the 3D NoC concept, using asynchronous logic, for providing a 1TBps bandwidth at reasonable power and area budgets. Going further, the concept of interposer, passive or active, is another parallel path which can solve some issues of new technology nodes around the active interposer : IOs, legacy IPs, power and variability management, test and debug facilities could be deported in an active interposer for improving performance and cost of 3D embedded Multi-core.

REFERENCES

[1] J. Ouyang, G. Sun, Y. Chen, L. Duan, T. Zhang, Y. Xie, and M. J. Irwin, "Arithmetic Unit Design Using 180nm TSV-based 3D Stacking Technology", *Proc. of the Inlt Conference on 3D System Integration*, 3DIC'09, Sept. 2009

[2] G. Poupon, M. Scannell, L. di Cioccio, D. Henry, P. Leduc, L. Clavelier and N. Sillon, "3D Integration : a technological toolbox" *3rd International Conference on Microsystems, Packaging, Assembly & Circuits Technology,* IMPACT'08, 2008.

[3] G. Pares et al.; "Mid-process through silicon vias technology using tungsten metallization: Process optimazation and electrical results", *Electronics Packaging Technology Conference*, EPTC '09, 2009

[4] J. Charbonnier et al., "Wafer level packaging technology development for CMOS image sensors using Through Silicon Vias", *2nd Electronics System-Integration Technology Conference*, ESTC'08, 2008

[5] D. Henry et al.. "Via First Technology Development Based on High Aspect Ratio Trenches Filled with Doped Polysilicon", *Proc. of 57th Electronic Components and Technology Conference*, ECTC '07, 2007

[6] P. Chausse, M. Bouchoucha, D. Henry, N. Sillon, L.L. Chapelon, "Polymer filling of medium density through silicon via for 3D-packaging" *Electronics Packaging Technology Conference,* EPTC '09. pp 790 – 794, 2009

[7] L. Cadix et al., "Integration and frequency dependent electrical modeling of Through Silicon Vias (TSV) for high density 3DICs", *Proc. Of Interconnect Technology Conference*, IITC'10, 2010

[8] J. Charbonnier et al., "Development and Characterisation of a 3D Technology Including TSV and Cu Pillars for High Frequency Applications", *Proc. of 60th Electronic Components and Technology Conference*, ECTC'10, 2010

[9] D. Henry et al., "3D integration technology for set-top box application", *Proc. of conf. on 3D System Integration*, 3DIC'09, 2009

[10] L. Di Cioccioa et al., "Enabling 3D Interconnects with Metal Direct Bonding" *Proc. of conf. on Integrated Circuit Design and Technology*, ICICDT '07, 2007

[11] V. F. Pavlidis, I. Savidis, and E. G. Friedman "Clock Distribution Networks for 3-D Integrated Circuits", *Proc. of IEEE Custom Intergrated Circuits Conference*, CICC'08, 2008

[12] Y. Thonnart, P. Vivet and F. Clermidy, "A Fully-Asynchronous Low-Power Framework for GALS NoC Integration", *Proc. of Design And Test in Europe*, DATE'10, March 2010

[13] W. Lafi, A. Jerraya, D. Lattard, "An efficient hierarchical router for 3D NoC architecture", *proc. of IEEE International Symposium on Rapid System Prototyping* , RSP'10, 2010

[14] L. A. Polka, H. Kalyanam, G. Hu, S. Krishnamoorthy, "Package Technology to Address the Memory Bandwidth Challenge for Tera-scale Computing", *Intel Technology Journal*, vol. 11, no. 3, Aug. 22, pp. 197–205, 2007

[15] P. Franzon, "Creating 3D-Specific Systems-Architecture, Design, CAD", *Proc. of Design, Automation and Test in Europe,* DATE'10.

[16] G. H. Loh, "3D-Stacked Memory Architectures for Multi-Core Processors", *Proceedings of the 35th Intl. Symp. on Computer Architecture*, ISCA'08, 2008

[17] E. Beigné, F. Clermidy, J. Durupt, H. Lhermet, S. Miermont, Y. Thonnart, T. Tran Xuan, A.Valentian, D. Varreau, P. Vivet, C. Chancel, X. Popon, H. Lebreton, "An Asynchronous Power Aware and Adaptive NoC based Circuit", IEEE Journal Of Solid State Circuits, April 2009, vol.44, pp.1167-1177.

[18] E. Beigne, P. Vivet, S. Miermont, "An Innovative Local Adaptive Voltage Scaling Architecture for On-chip Variability Compensation", IEEE International Symposium on Circuits and Systems (ISCAS) 2011, Rio, Brazil, May 2011.

[19] F. Darve, P. Vivet, "An Asynchronous Serial Link for a 3D Network-on-Chip", Friday Workshop on 3D Integration - Applications, Technology, Architecture, Design, Automation, and Test -, DATE'2010, Dresden, Germany, 13 March 2010