Contents lists available at ScienceDirect





Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

Impact of data serialization over TSVs on routing congestion in 3D-stacked multi-core processors

CrossMark

Giulia Beanato, Alessandro Cevrero, Giovanni De Micheli, Yusuf Leblebici

EPFL, Lausanne, Switzerland

ARTICLE INFO

Article history: Received 28 July 2014 Received in revised form 18 July 2015 Accepted 3 December 2015

Keywords: 3D integrated circuits Through Silicon Vias TSV CMP High speed serial link Routing congestion

1. Introduction

With the advent of deep-submicron CMOS technologies, on-chip interconnect wires have rapidly gained a lot of attention. With the scaling of the CMOS technologies, the parasitics effects due to the wiring do not exhibit the same scaling behavior as the CMOS logic. As the IC feature sizes shrink, device area shrinks roughly as the square of the scaling factor while the device propagation delay improves almost linearly with the decrease in feature size under constant field assumption. On the other hand, interconnect delay does not scale with feature size, and tend to gain importance as device dimensions are reduced and circuit speed is increased. Worsening the situation, as the silicon dies get larger, also the average length of the interconnects increases, hence their associated parasitic effects. As a consequence, the interconnects start dominating some of the most important metrics of digital ICs, such as speed, power consumption and reliability.

A promising solution to break through the interconnect wall emerged with the advent of 3D integration featuring TSVs. 3D ICs have the potential to reduce interconnects length and improve the system performance. Nevertheless, technological processes necessary for fabricating the TSVs connecting the superimposed layers cannot yet be regarded as mature. According to the ITRS roadmap [8], the TSV diameter will not shrink below 2–4 μ m for global interconnects. Using small vias is desirable to reduce the chip footprint, yet, as the diameter decreases, the TSV fabrication yield worsens. Since the silicon area occupied by a TSV is quite significant, it interferes with cells placement, spreading them out and limiting the achievable reduction of the average routing distance. In case of most via-last TSV technologies, the impact becomes even more severe since this TSVs

ABSTRACT

3D integration can alleviate routing congestion, reducing the wirelength and improving performances. Nevertheless, each TSV still occupies non-negligible silicon area: as the number of TSV increases, their effect on the chip routing is detrimental. The reduction in the number of 3D vias obtained with the adoption of serial vertical connections can relieve the routing congestion of the 3D system by reducing the average wirelength. In this paper we explore the impact of the serial approach on the chip routing of a 3D multi-processor platform to quantify the achievable wirelength reduction for a range of TSV technologies. The comparison between the serial and the parallel multi-processor configurations shows up to 12.4% wirelength improvement for the serial solution, with serious consequences on routing delay.

occupy all metal layers, becoming a routing obstacle. Hence, as the number of TSVs increases, the wirelength and form-factor benefits of 3-D ICs significantly reduce, as demonstrated by Kim et al. [9]. Serial vertical TSV interconnects have been proposed as an effective solution to keep under control of the TSV count. As demonstrated by Beanato et al. [2], the use of serial vertical interconnects can significantly reduce TSV area footprint with a reasonable power overhead and no performance loss. The aim of this paper is to explore the impact of serialization on the routing congestion of a 3D-CMP. A 3D Modular Multi-Core (3D-MMC) architecture has been designed and used as test case. 3D-MMC is based on the integration of identical multi-processor layers. Thanks to the homogeneous approach, the system performance can be augmented with minimal design cost compared to conventional planar IC designs.

2. Previous work

3D-ICs based on TSV technology provide high bandwidth interchip connections. The drawback is that most of the existing TSVs consume a large amount of silicon real estate, strongly affecting the performance of 3D ICs. The wirelength reduction varies depending on the number of TSVs. The impact of TSV size on the 3D wirelength distribution was first studied by Kim et al. [9] demonstrating that the wirelength increase due to TSV placement is not negligible. In high density 3D-ICs, routing congestion can cause routing failure or redesign from the beginning, to tackle this issue Ahn et al. [4] have proposed a precise routing congestion estimation method at the floorplan stage, which is beneficial to reduce the total design cost. A different approach focused on co-optimizing the TSV count and wirelength at the placement level was studied by Tsai et al. [14] and Cong et al. [5], while Lee et al. [11] have developed an algorithms for TSV resource sharing and optimization. While these previous works mostly focus on placement algorithms to reduce the TSV area impact on the design, this paper analyses the benefits of data serialization on the routing congestion of a 3D chip multi-processor system.

3. 3D modular multi-core architecture

In order to explore the potential of the serial data transmission through TSV, the 3D Modular Multi-Core (3D-MMC) architecture, proposed by Beanato et al. [1] and validated by Zhang et al. [16] has been used as test platform. Fig. 1 presents a basic block diagram of the stacked structure.

The 3D-MMC architecture has been specifically designed for stacking identical dies in order to form the 3D system. Without loss of generality, the proposed architecture can be expanded to include multiple identical layers that can communicate with each other. Each die can be considered as a planar multi-core architecture, composed by multiple Processing Elements (PEs), working in parallel. The cores exchange data through a shared memory implemented in the Peripheral Subsystem (PS) unit; the access of PE to the shared memory is arbitrated by a system of semaphores to avoid contention. The interaction between cores occurs through a specific source-routed NoC, composed of a 36-bit switch, in charge of the effective signals routing to and from 6 directions (North, South, East, West, Up, Down), and a Network-Interface (NI) for each logic block present on the layer. The network system has a 3D folded architecture in order to enable the management of the signals in both the horizontal and vertical directions. The intralayer communication is achieved through the introduction of a 3D connection macro, exploiting arrays of TSVs as vertical data bus.

3.1. 2D layer architecture

A single layer consists of four *Processing Elements* (*PE*) that exchange data through a shared memory, which is placed in the *Peripheral Subsystem* (*PS*) unit. A system of semaphores arbitrates

the access of PE's to the shared memory. The routing between each PE and the shared memory occurs through a specific source-routed NoC implemented to manage the signals routing to and from 6 directions (North, South, East, West, Up, Down). In the 3D stack, NoC on different layers are interconnected to enable the management of the signals in both the horizontal and vertical directions.

Fig. 2a and b illustrates the internal architecture of a PE and a PS, respectively. Each PE is built out of a 32-bit RISC processor, the open-source LEON3 unit from Aeroflex Gaisler, a general-purpose unit. The LEON3 unit is connected to slave modules through an AMBA bus. Each core utilizes privately addressable memory space, composed of a 32 kB ROM, containing the boot sequence, and a 32 kB RAM, as well as a common memory space composed by the system shared memories. The access of the cores to the shared space is regulated by a semaphore module present in the PS, able to avoid conflicts in case of simultaneous requests.

Multi-core interactions are managed by the *Network-Interface* (*NI*). The NI block is a master located within both PE and PS. It interfaces the AMBA bus to the NoC, and is responsible of forwarding/receiving data packets to/from the shared memory, which has an addressing space visible by each core. This NoC has been specifically adapted from [15] for the proposed CMP architecture. The 7×7 Switch is characterized by 5 horizontal interfaces (one for each PEs plus one for the PS) and 2 vertical ports (for the upper and lower dies), through which 36-bit FLIT (*FLow control unITs*) packets are transmitted. Similar to PEs, the PS contains NI and AHB JTAG acting as master modules whereas all the remaining units (semaphore, shared RAM) act as slaves.

Each PE in an N-layer system has access to N+1 different memory modules that can be accessed in parallel: a private-RAM contained in his own PE, a shared local-RAM located in the PS of its layer, and N-1 shared remote-RAMs situated in the PS of the other stacked layers. In the same way as proposed by Benini et al. [3], the proposed memory hierarchy with shared data memory for inter-processor communication simplifies the hardware complexity and avoids memory coherency overhead. The multi-core synchronization is handled at the software level.

TSVs in general have excellent high frequency properties, thanks to their low series resistance and limited parasitics, but their area footprint is bigger compared to the BEOL vias, therefore



Fig. 1. Block diagram of the 3D-MMC architecture. The generic 3D connection macro-block on each identical layer allows the inter-layer communication among multiple layers, with serial multiplexed TSV arrays.



Fig. 2. (a) Processing element (PE) internal architecture, with the LEON3 core and its private modules. Each unit is accessible through JTAG ports for debugging purposes. The network interface (NI) routes packets from PE to the shared memories in the (b) Peripheral Subsystems (PS).



Fig. 3. Schematic view of the 3D-MMC (a) parallel and (b) serial configurations. Note that the parallel configuration requires 174 signal TSVs while the serial configuration uses only 34 TSVs. Layout view of the 3D-MMC (c) parallel and (d) serial configurations design with 40 μm TSV channels. The red line depicts a sample path from CORE1 to the NoC. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

it is important to place the minimum number of TSVs without sacrificing vertical data-bandwidth. For this reason, a serializer– deserializer module can be introduced to optimize the trade-off between bandwidth and number of TSVs in the array. The 3D connection macro can be designed to send the inter-layer signals either as a parallel stream or serially.

Table 1

Physical design parameters and TSV count in parallel and serial case.

	Parallel configuration			Serial configur	Serial configuration		
Chip size (µm)	2050 imes 2650			2050×2650			
Signal TSVs	174			34			
Power TSVs	18			22			
KOZ (µm)	5			5			
TSV size (µm)	5	10	40	5	10	40	
TSV array dimension (µm) (TSVs+KOZ)	85 imes 245	125 imes 365	365 imes 1085	45 imes 145	65 imes 215	185×635	



Fig. 4. (a) 1:8 SER circuit and (b) 1:8 DES circuit.

3.2. Physical design

Both the serial and parallel version of 3D-MMC have been implemented in RTL. The designs have been synthesized with the UMC 90 nm CMOS technology library using Synopsys Design Compiler. The layouts have been placed and routed with Cadence Encounter. The functionality has been verified using Mentor Graphics ModelSim. The multiprocessor has been constrained to work at 200 MHz, with the serial vertical interconnection working at 1.6 GHz.

The routing analysis has been performed for 3D ICs based on three different TSV technologies assuming a via-last process. 5 μ m TSVs represent state-of-the-art for high density through silicon vias [15], 40 μ m TSVs are a more established technology which guarantees better reliability [8] while 10 μ m TSVs provide a fair compromise between the two extreme. All the TSVs are placed in a TSV array located in the center of the chip, as depicted in Fig. 3. A 5 μ m keep out zone and a minimum distance of 2.5 μ m from the metal interconnects has been used. Beyond the vertical signal connections, 22 TSVs has been added for the power and ground delivery in each design.

All designs have been constrained within a chip area of $2050 \,\mu\text{m} \times 2650 \,\mu\text{m}$. The serial configuration includes 10 serializers and 10 deserializers. The design parameters are summarized

in Table 1. The design placement of 3D-MMC in Fig. 3 clearly shows the TSV area reduction, and the consequent effect on the cell placement achieved by serializing the signal delivered by the 40 μ m TSV channels. The memory macros of each core, such as private RAM, register file, instruction cache, are placed all around the core area, while the TSVs are placed in a matrix in the center of the chip.

4. Serial vertical link

The architecture of the *Serializer–Deserializer (SERDES)* blocks used in the 3D connection macro depicted in Fig. 3d has been carefully chosen to fit the requirements of the system.

Different SERDES connections' topologies are already available in the literature. Asynchronous data links [7,6] are used to transfer data across different clock domains and employ handshake instead of clock signal for operation control. An acknowledgement signal should be added for each serial channel. System synchronous clocks or source-synchronous clock topologies require the clock to be transmitted, while another common timing mechanism for serial interconnects injects a clock into the data stream at the transmitting side and recovers the clock at the receiver.



Fig. 5. Waveforms describing the SERDES functionality.

Table 2SERDES characteristics.

	$F_{parallel}$ (MHz)	F_{serial} (GHz)	Area (μm^2)	Power (μW)
Ser 8:1	312	2.5	154	262
Deser 8:1	390	2.3	406	608



Fig. 6. Critical path of the serializer circuit.

The goal of this work is to reduce the number of TSVs in the system, therefore in the proposed high speed 3D serial vertical link, each serial vertical data connection transmits a single-ended signal using a single TSV channel. *N* slow speed channels at a frequency f_{par} are serialized into a high speed data stream at $f_{ser} = N \times f_{par}$, where *N* is the serialization level. The high speed stream is sent through a TSV and deserialized on the receiving layer. Asynchronous connections would require the addition of a second acknowledgement TSV for each connection, therefore a synchronous transmission has been chosen. Since the delay due to the TSV is negligible the clock phases are effectively matched [12], hence the clock can be distributed on each stacked die with identical distribution networks.

The transmitter circuit consists of a *serializer (SER)* followed by a buffering stage to drive the TSV. The architecture chosen for the serializer is based on the design proposed by Kurisu et al. [10]. Fig. 4a depicts the circuit diagram of the 8:1 SER: the circuit creates a pulse of width $1/f_{ser}$ and period $1/f_{par}$, which is then passed through a shift register synchronously to the fast clock CLK_{ser} in

order to produce N shifted pulses. These pulses are then used as enable signals for a combinational circuit that converts the N parallel signals, ParIN[7:0], into a serial stream, OUT_{ser}, that goes to the buffering stage driving the TSV.

The receiver architecture is depicted in Fig. 4b for a 1:8 deserialization scheme. The *deserializer* (*DES*) also exploits a shift register in order to produce *N* shifted pulses of width $1/f_{ser}$ and period $1/f_{par}$ that act as trigger for *N* latches. Each latch stores one bit of the incoming serial stream, IN_{ser}, until they are resynchronized with the system slow clock, CLK_{par}, through the output registers. The waveforms describing the SERDES functionality are depicted in Fig. 5.

In order to include the serial vertical connection into the semicustom digital design flow, the SERDES circuits have been implemented in RTL and synthesized with the UMC 90 nm CMOS technology library using the Synopsys Design Compiler. The functionality has been verified using Mentor Graphics ModelSim.

Table 2 summarizes the maximum working frequency, the gate area and the power consumed by the SERDES circuits. As expected, the maximum working frequency achievable by the semi-custom solution is limited by the clock-to-Q delay of the flip-flop available in the library. The serializer critical path is highlighted in Fig. 6.

As discussed in Section 3, the network system has a 3D folded architecture which extends its capability of managing the signal transmission also in the vertical direction through a 3D connection macro. Fig. 3a depicts the traditional parallel configuration exploiting one TSV for each inter-layer signal.

Instead, Fig. 3b depicts the serial configuration of 3D-MMC. The SERDES circuits have been integrated in the 3D connection macro at the vertical interface of the NoC. The 174 signal TSVs required for the parallel configuration are reduced to 34 after serialization: specific signals, like the clock, reset, layerID (2-bits) and JTAG debugging signals (TCK, TRST, TDI, TDO, TRST) are directly sent to the above and bottom layers, while the rest of the NoC data (144) and control signals (12) are grouped into bytes, serialized and then sent through the TSV channels to the upper and bottom layers. The number of TSVs marked in Fig. 3 considers both the TSV to the lower layer and the TSVs to the upper layer.



Fig. 7. Length trend, in μ m, of the longest 240 nets in the design for (a) 5 μ m (b) 10 μ m (c) 40 μ m TSVs for the parallel (blue) and serial (green) configuration. The serial solution results in a significant reduction of wire length, especially for the longer nets. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

5. Experimental results

The area-power trade-off for the serial configuration versus the parallel one has already been studied in [2]. Nevertheless, the impact of the TSV footprint on the chip area is not the only issue related to the TSV size. TSVs also contribute to routing congestion of each layer since they both interfere with cell placement and, in case of via-last TSVs, become a routing obstacle.

As the CMOS technology scales down, semi-global and global wires are becoming an increasingly important performance bottleneck since their typical wirelength does not scale [8]. Generally, the wire's parasitics define its performance, and both the resistance and the capacitance of a wire directly depend on its length *l*. Consequently the RC delay is proportional to l^2 , which becomes unacceptably large for long wires. Moreover, the switching of the interconnection capacitance of the wires causes dynamic power consumption following the relationship $P = A_f CV^2 f$, where *f* is the frequency of digital signal, A_f is wire activity factor, *V* stands for voltage swing between the two digital levels, and *C* is the total interconnect capacitance for a certain wirelength [8]. Since dynamic power is currently the main component of the power dissipation with approximately 50% of microprocessor power consumed by the interconnects [13], the designers should struggle to keep the routing congestion of a chip under control.

In this section we show how the proposed serial approach can reduce routing congestion improving the design performance using the 3D-MMC architecture as test vehicle and via-last TSVs for the inter-layer connections. As an example, we can consider the net depicted as a red line in Fig. 3 which connects CORE 1 to the NoC. A lower TSV count translates into a significant reduction of routing obstacles in the design, and allows the logic gates to be placed closer to each other. Hence in the serial configuration we can notice that the length of the net connecting CORE 1 to the NoC can be drastically reduced.

The following analysis has been performed on the routing of each placed and routed design. First we extract the length of each net in the designs focusing on the first 240 longest connections. Fig. 7 shows the trend of the considered nets for both the parallel (in blue) and the serial configuration (in green). The trends clearly show that the serialization causes a reduction of the wirelength, which is limited for the design utilizing the small 5 μ m TSVs, while becomes more marked for the designs featuring larger TSVs.

Focusing more on the length of the interconnects, we first define a lower bound to focus on the nets longer than that threshold. A typical 500 µm long on-chip copper connection in 90 nm CMOS technology at the intermediate level, metal2–metal6, is characterized by a resistance $R \sim 80$ m Ω and a capacitance that exceeds 70 fF. Consequently, the net delay can be approximated as $\frac{1}{2}RCl^2 = 717$ ps. The plots in Fig. 8 depict the histograms of the design's interconnections starting from the 500 µm threshold. In particular, it shows the number of nets in the design for each range of lengths. We can notice that for all the TSV technologies considered the number of long nets decreases in the case of serial configuration (green bars). As expected, there are few nets longer than 1300 µm in the design with 5 µm TSVs, while they are more numerous as the TSV size increases. The average length reduction for the considered designs are summarized in Table 3. A similar average length reduction has been



Fig. 8. Wirelength distribution for parallel (blue) and serial (green) solutions for different TSV diameters. A significant length reduction is observed in all cases as a result of serialization, with the most significant reduction in longer nets. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Tab	le 3	3
-----	------	---

Routing results and wirelength improvement.

TSV size (μm)	Average length reduction (μm)	Wirelength improvement (%)
5	60	5.3
10	152	12.4
40	150	11.2

observed for both 10 μ m and 40 μ m TSVs. This can be explained considering the fact that the routing algorithm can meet the given timing constraints hence there is no effort in further wirelength reduction. Stringent timing constraints should provide higher wirelength reduction for the 40 μ m TSVs. Alas, reducing cycle time results in negative slack in the memory access and as a consequence no further benefit has been observed. For the design using 5 μ m TSVs, the reduced number of TSVs needed after serializing leads to a 5.3% wirelength improvement. The benefits are more pronounced in the case of larger TSVs: for 40 μ m and 10 μ m TSVs, the overall wirelength improvement reaches respectively 11.2% and 12.4%.

6. Conclusion

This paper explores the impact of data serialization for interlayer communication on the chip's routing congestion. Adopting a serial vertical data communication approach allows for a reduction of the overall number of TSVs, therefore reducing the average onchip wirelength.

A modular 3D stacked multi-processor platform, 3D-MMC, consisting of identical dies has been introduced. The 3D-MMC architecture has been implemented using UMC 90 nm CMOS technology, and used as test vehicle. An 8-bit serialization of the 3D-MMC's inter-layer signals has been implemented and compared to the fully parallel solution for different TSV technologies. The wiring characteristics of each solution have been extracted from the placed and routed design.

Results show that the serial approach reaches up to 12.4% wirelength improvement compared to the fully parallel counterpart when using 10 μ m TSVs. Even for high end TSV technologies such as 5 μ m TSVs, the wirelength exhibits an average reduction of 5.3%.

References

- [1] G. Beanato, P. Giovannini, A. Cevrero, P. Athanasopoulos, M. Zervas, Y. Temiz, Y. Leblebici, Design and testing strategies for modular 3-d-multiprocessor systems using die-level through silicon via technology, IEEE J. Emerg. Select. Top. Circuits Syst. 2 (June (2)) (2012) 295–306.
- [2] Giulia Beanato, Alessandro Cevrero, Giovanni De Michel, Yusuf Leblebici, 3d serial tsv link for low-power chip-to-chip communication, In: 2014 IEEE International Conference on IC Design Technology (ICICDT), May 2014, pp. 1–4.
- [3] L. Benini, E. Flamand, D. Fuin, D. Melpignano, P2012: building an ecosystem for a scalable, modular and high-efficiency embedded computing accelerator, In:

2012 Design, Automation Test in Europe Conference Exhibition (DATE), March 2012, pp. 983–987.

- [4] Wenrui Li, Byung-Gyu Ahn, Jaehwan Kim, Jong-Wha Chong, Effective estimation method of routing congestion at floorplan stage for 3d ics, J. Semicond. Technol. Sci. 11 (2011) 344–349.
- [5] J. Cong, Guojie Luo, A multilevel analytical placement for 3d ics, In: Design Automation Conference, 2009, ASP-DAC 2009, Asia and South Pacific, January 2009, pp. 361–366.
- [6] R. Dobkin, M. Moyal, A. Kolodny, R. Ginosar, Asynchronous current mode serial communication, IEEE Trans. Very Large Scale Integr. (VLSI) Syst. 18 (July (7)) (2010) 1107–1117.
- [7] Bui Chinh Hien, Seok-Man Kim, Kyoungrok Cho, Design of a wave-pipelined serializer-deserializer with an asynchronous protocol for high speed interfaces, In: 2012 4th Asia Symposium on Quality Electronic Design (ASQED), July 2012, pp. 265–268.
- [8] ITRS, Interconnect, 2011, Accessed February 6, 2014.
- [9] Dae Hyun Kim, Saibal Mukhopadhyay, Sung Kyu Lim, Through-silicon-via aware interconnect prediction and optimization for 3d stacked ics, In: 2009 Proceedings of the 11th International Workshop on System Level Interconnect Prediction, SLIP'09, ACM, New York, NY, USA, pp. 85–92.
- [10] M. Kurisu, M. Kaneko, T. Suzaki, A. Tanabe, M. Togo, A. Furukawa, T. Tamura, K. Nakajima, K. Yoshida, 2.8 gb/s 176 mw byte-interleaved and 3.0 gb/s 118 mw bi-interleaved 8:1 multiplexers, In: 1996 42nd ISSCC on Solid-State Circuits Conference, Digest of Technical Papers, 1996 IEEE International, pp. 122–123.
- [11] Byunghyun Lee, Taewhan Kim, Algorithms for tsv resource sharing and optimization in designing 3d stacked ics, Integr. VLSI J. 47 (March (2)) (2014) 184-194.

- [12] Yong Liu, Wing Luk, D. Friedman, A compact low-power 3d i/o in 45nm cmos, In: 2012 IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012, pp. 142–144.
- [13] Nir Magen, Avinoam Kolodny, Uri Weiser, Nachum Shamir, Interconnectpower dissipation in a microprocessor, In: Proceedings of the 2004 International Workshop on System Level Interconnect Prediction, SLIP '04, ACM, New York, NY, USA, 2004, pp. 7–13.
- [14] Ming-Chao Tsai, Ting-Chi Wang, TingTing Hwang, Through-silicon via planning in 3-d floorplanning, IEEE Trans. Very Large Scale Integr. (VLSI) Syst. 19 (August (8)) (2011) 1448–1457.
- [15] G. Van der Plas, P. Limaye, I. Loi, A. Mercha, H. Oprins, C. Torregiani, S. Thijs, D. Linten, M. Stucchi, G. Katti, D. Velenis, V. Cherman, B. Vandevelde, V. Simons, I. De Wolf, R. Labie, D. Perry, S. Bronckers, N. Minas, M. Cupac, W. Ruythooren, J. Van Olmen, A. Phommahaxay, M. de Potter de ten Broeck, A. Opdebeeck, M. Rakowski, B. De Wachter, M. Dehan, M. Nelis, R. Agarwal, A. Pullini, F. Angiolini, L. Benini, W. Dehaene, Y. Travaly, E. Beyne, P. Marchal, Design issues and considerations for low-cost 3-d tsv ic technology, IEEE J. Solid-State Circuits 46 (January (1)) (2011) 293–307.
- [16] Tiansheng Zhang, Alessandro Cevrero, Giulia Beanato, Panagiotis Athanasopoulos, Ayse K. Coskun, Yusuf Leblebici, 3d-mmc: a modular 3d multi-core architecture with efficient resource pooling, In: 2013 Design, Automation Test in Europe Conference Exhibition (DATE), March 2013, pp. 1241–1246.