Document downloaded from:

http://hdl.handle.net/10251/35905

This paper must be cited as:

Hernández Luz, C.; Roca Pérez, A.; Flich Cardo, J.; Silla Jiménez, F.; Duato Marín, JF. (2011). Fault-tolerant vertical link design for effective 3D stacking. IEEE Computer Architecture Letters. 10(2):41-44. doi:10.1109/L-CA.2011.17.

The final publication is available at

http://dx.doi.org/10.1109/L-CA.2011.17

Copyright Institute of Electrical and Electronics Engineers (IEEE)

Fault-Tolerant Vertical Link Design for Effective 3D Stacking

Carles Hernández, Antoni Roca, José Flich, Federico Silla, and José Duato
Grupo de Arquitecturas Paralelas
Departamento de Informática de Sistemas y Computadores
Universitat Politècnica de València
c/camí de vera s/n, 46022, Valencia
carherlu@gap.upv.es

***Abstract*—Recently, 3D stacking has been proposed to alleviate the memory bandwidth limitation arising in chip multiprocessors (CMPs). As the number of integrated cores in the chip increases the access to external memory becomes the bottleneck, thus demanding larger memory amounts inside the chip. The most accepted solution to implement vertical links between stacked dies is by using Through Silicon Vias (TSVs). However, TSVs are exposed to misalignment and random defects compromising the yield of the manufactured 3D chip. A common solution to this problem is by over-provisioning, thus impacting on area and cost. In this paper, we propose a fault-tolerant vertical link design. With its adoption, fault-tolerant vertical links can be implemented in a 3D chip design at low cost without the need of adding redundant TSVs (no over-provision). Preliminary results are very promising as the fault-tolerant vertical link design increases switch area only by 6.69% while the achieved interconnect yield tends to 100%.**

I. Introduction

The advance in manufacturing technologies enables the integration of a growing number of transistors on a single chip, devoted mainly to increase the number of cores and memories. As the number of cores increases, the communication infrastructure becomes critical. In this context, Networks on chips (NoCs) have been chosen as the interconnect choice for current and future Multiprocessor System-on-Chip (MPSoCs) and Chip Multiprocessor (CMPs) systems. The main benefits of NoC-based architectures are higher performance and predictability. Recent designs from major chip manufacturers include a NoC inside the chip to interconnect cores and memories ([8], [26]). Among the different topologies to perform core interconnection [4], 2D meshes are commonly accepted to be an appealing choice as they perfectly match the chip surface and are additionally aligned with the tile based design approach, where a tile containing a processing core with its associated L1 and L2 cache levels and the corresponding NoC switch is initially designed and later replicated all over the die.

As the number of cores increases, the access to memory (memory bandwidth) is predicted to become a major limiting factor for CMP and MPSoC designs. In this context, three dimensional Integrated Circuits (3DICs) arise as a promising technological solution to partially alleviate the problems caused by the pin-out limitations [21]. In 3DICs, dies are stacked on top of each other and vertical connections are established. One of the most promising technologies to enable vertical links between dies is the use of Through Silicon Vias (TSVs). However, one of the main drawbacks when manufacturing TSVs is their high defects rate, specially when compared with traditional 2D links [13]. TSV-based vertical links are exposed to misalignment and random defects. The first kind of failures are introduced in the alignment process of stacked wafers as a consequence of bonding pads shifting [20]. On the contrary, random defects are a consequence of several unpredictable phenomena where most of them are related with the thermal compression process used in the wafer stacking process [13].

For the reasons previously presented, efficient mechanisms and designs are needed to face the yield reduction experienced by TSVs in 3DICs. Most of the previous work on fault tolerant 3D architectures rely on the use of redundant TSVs [13], [17], adding redundancy to the vertical link. However, avoiding the use of extra TSVs (while keeping system performance) is specially interesting and needed. As the complexity of applications continues rising and the number of nodes (cores or memories) also increases, the impact of redundant TSVs on area footprint becomes impossible to neglect [19][10][17]. In particular, in [19], TSV footprint area of future 3DICs is predicted to be similar to the area of a computing core.

In this paper we propose a new vertical link design able to increase the yield of 3D TSV-based chips. The design provides high yield at a relatively low cost and without increasing the number of TSVs. Faulty TSVs are covered by adding an Omega network at both ends of vertical link (at the switch boundaries). This network allows a reprogrammed reconfigurability for efficient use of fault-free TSVs. Indeed, a subset of functional TSVs is obtained for any error pattern whenever the number of faulty TSVs is lower than N/2. The design does not impact performance as it takes advantage of the reduced delay of TSV interconnects and the fast reconfigurability of the Omega network.

This paper is organized as follows. In Section II the related work is analyzed. Section III presents the fault tolerant TSV-based vertical link design. Sections IV analyze the baseline 3D switch used throughput this paper, and shows the area and timing impact of the new vertical link design. Finally, Section V presents the conclusions and the future work.

II. Related work

Current and future fabrication processes enable the use of smaller feature sizes and new manufacturing technologies, as 3D stacking. Unfortunately, this poses an increase in reliability uncertainty. In this sense, several recent works in the literature have pointed out some of the problems affecting chip interconnects. Concretely, the implications of variability in NoC link interconnect are analyzed in [7] [6]. In [7] different measurements are provided of delay uncertainty for 45nm technologies down to 16nm. Similarly, in [6] delay variation of next technology nodes is analyzed. Both studies remark the importance of process variation in 2D communication links when technologies scale down. The presence of the increasing delay variability and the increase of defect densities in manufactured chips significantly increases the probability of having a faulty wire in a link. In this sense, there are several proposals that are able to tolerate faulty wires. Some of them tolerate infrequent run-time timing violations, where delay failures are tolerated at the cost of performance [3], [15].

Other proposals focus on increasing interconnect yield by using redundant hardware. The first work that proposed a compact hardware implementation to use spare wires in NoC links was [5]. Spare wires are added to tolerate a bounded number of faults without decreasing communication performance. To do so, a crossbar is used to chose a set of non-faulty wires to perform the transmissions. The same idea has been applied to vertical links in [13]. However, as stated in [19] and [17], the use of a high number of TSV wires leads to physical implementation challenges and overheads. Moreover, the proposal in [13] is unable to tolerate most of the failures caused by misalignments. To minimize the overhead, in [17] a combination of limited spare wires with the use of error correction codes is used to tolerate a high defect rate [17]. In the same way, serialization schemes have been proposed to reduce the TSVs footprint [19] [18].

In this paper we propose a new TSV-based fault tolerant vertical link design without increasing the number of TSVs (no hardware redundancy) and fault tolerance is provided regardless the failure pattern (thus, covering misalignment and random defects).

In addition, in order to evolve current 2D NoC designs to 3D stacked designs new switch architectures need to be developed. For example, in [11] the use of a modular switch architecture is proposed. The crossbar is divided into three independent decentralized switches that serve the different switch dimensions. Similarly, a multilayered chip is proposed in [16] where the switch design is spanned across the layers of the 3D chip. This is performed by classifying switch modules into separable and non-separable modules. In this paper, we connect in a transparent way the new vertical link design to the switch architecture proposed in [22]. This switch design was adopted as the best option due to its low latency and straightforward extension to the vertical dimension.

III. A Fault-Tolerant Vertical Link Design

A. Fault Tolerant 3D Link

Fault tolerance is provided by adding a hardware reconfiguration logic at both ends of the vertical link (Fig. 1). In particular, an omega network (Ω) is used at both ends to provide the reconfiguration of the vertical link at low cost. An Ω network is made of $log_2 N$ stages with $\frac{N}{2}$ 2×2 switches at each stage.

The Ω network has the concentration property as the network can be configured to concentrate any set of unrelated input signals (not necessarily consecutive) into a set of consecutive output signals and vice versa. In order to avoid faulty TSVs we will use such property. Indeed, during the test stage of the link faulty TSVs are identified and the proper configuration of the Ω network is computed and stored in a ROM memory associated with the network.

The link design relies on the assumption that the probability of having a number of faulty link wires larger than $\frac{N}{2}$ is negligible. With this assumption, transmission of one flit across the vertical link can be split in two phases using a subset of fault-free TSVs of size $\frac{N}{2}$. The Ω network at the input of the vertical link is used to de-concentrate half of the link into a set of $\frac{N}{2}$ fault-free TSVs, whereas the inverse Ω^{-1} network is used at the output of the vertical link to concentrate the transmitted data back to the corresponding link half. Figure 2 shows how the de-concentration and concentration operations are performed by the Ω and Ω^{-1} networks in a 4 bit width vertical link. Flits can be transmitted using the fault-free TSVs. The flit is transmitted then in two phases, both within the same clock cycle. Next, we deep into the timing constraints and analysis.

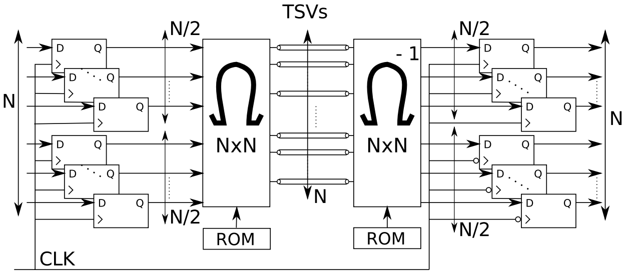

Fig. 1. Diagram of the proposed fault tolerant vertical link. Registers are within the switch

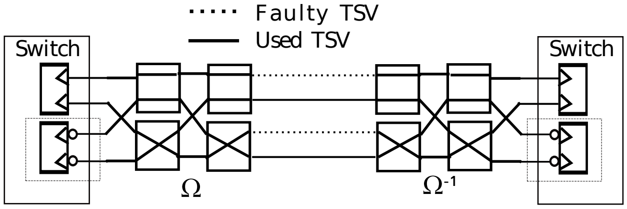

Fig. 2. Example of 4-bit link at the output of the link

B. Timing Analysis

In order to analyze the operating frequency of the network, we should differentiate between switches and links. Then, the slowest path (critical path) amongst all paths of a switch and links will fix the operating frequency of the network. Thus, the critical path of the network can be computed as

$$T = \max\{\text{switch stage delay}, \text{2D-link delay}, \text{TSV-link delay}\} \quad (1)$$

where the switch stage delay is the maximum delay of a functional stage of a switch. The 2D-link delay, is the delay that a packet suffers to cross a metalization wire in the 2D plane. That delay involve the wire delay and the setup and hold time of the registers that are connected to that wire. Finally, the TSV-link delay is the delay that involve the TSV via delay plus the setup and hold time of the registers connected to that via.

Normally, the critical path of a NoC is set by the critical path of a switch, as the 2D-link delay can be highly reduced by introducing an optimal number of high-sized repeaters along the wire [2]. Note that this comes at the expense of considerable increase in the power consumption of those wires. On the other hand, the TSV via presents a delay than is significant inferior with respect to the rest of the components of the network.

Then, the vertical link design relies on the fact that it is possible to transmit a flit within a clock cycle in two halves. Figure 3 shows the timing diagram for the link design. The first half ($\frac{N}{2}$ bits) is transmitted during the first half of the cycle meanwhile the second half is transmitted during the second half of the clock cycle. In order to keep clock frequency and hence the complexity of the circuit implemented low, the negative edge flank of the clock is sampled by half of the registers at both ends of the link (see Figure 1). Hence, all the link components work at the same baseline clock frequency.

Using both flank edges of the clock induce, however, tighter timing constraints to the link design. Figure 4 shows the delay of each component for a link with no fault tolerance support and for the vertical link design. T_s and T_h are the setup and hold time of the registers, respectively. T_Ω is the delay of the Ω network and T_TSV is the TSV delay (link delay). The latency of the link design is:

$$T_\text{s} + T_\text{h} + 2 * T_\Omega + T_\text{TSV} \leq \frac{T}{2} \quad (2)$$

TSV wires (vertical links) have an order of magnitude higher transmission rates than silicon-based links (horizontal links). In such scenario, the speed of messages traveling through horizontal links and then using vertical links will be bounded by the horizontal link

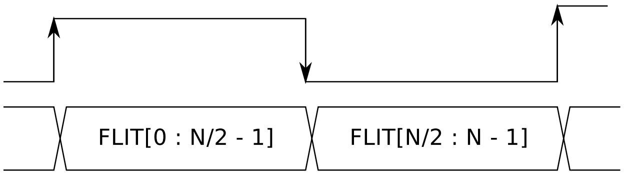

Fig. 3. Diagram of the proposed fault tolerant vertical link

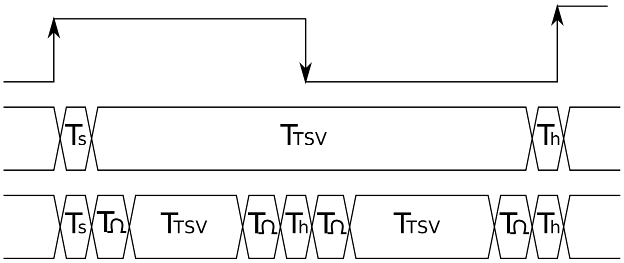

Fig. 4. Timing constraints of a standard TSV-based link design and the proposed fault tolerant vertical link design

speed, thus not achieving the potentials of the vertical TSV speed. This fact relaxes the timing constraints of our link design. We will see in the evaluation section how the link design is still faster than state-of-the-art horizontal silicon-based links.

C. Failure Pattern Tolerance

The reconfiguration capabilities of the Ω network enables the vertical link design to cover all the failure patterns (assuming at most $\frac{N}{2}$ failed TSV wires). This can be explained by analyzing the non-blocking permutations of the Ω^{-1} network. Note that properties that apply to the Ω^{-1} are also satisfied by the Ω network in the opposite direction. According to [12], a permutation π is passable by the Ω^{-1} network if and only if for all pairs

$$x \to \pi(x), y \to \pi(y) \tag{3}$$

in the permutation,

$$M(x, y) + L(\pi(x), \pi(y)) < n \tag{4}$$

where $n = log_2 N$. Note that functions $M(x, y)$ and $L(x, y)$ are defined as the number of most significant bits, and the least significant bits, respectively, which agree in the binary expansions of x, y. The non-blocking properties of the Ω^{-1} are thoroughly analyzed in [1]. In this study, it is demonstrated that Ω^{-1} is non-blocking for a set of monotonic inputs with concentrated output destinations. Note that this property matches the sorting requirements of TSVs in the proposed vertical fault tolerant link.

IV. Evaluation

In this Section results of area, timing, and yield are analyzed. Additionally, our approach is also compared with current state-of-the-art proposals. Before introducing the results obtained, a brief description of the the baseline 3D-switch is presented.

A. 3D-NoC Implementation

The baseline 3D-switch is a pipelined buffered wormhole switch with three stages, that is, an incoming packet needs three cycles to be routed, and forwarded to the next switch or node. Link delay is set to one cycle in any direction. The baseline 3D-switch performs the basic functions of a canonical NoC switch. That is, our 3D switch is able to buffer, route and forward packets to the next switch or node. The switch used throughout this paper presents two main properties. First, the buffer is spread through the stages, that is, any stage in the switch has its own outputs registered. Second, each output port is managed independently. That is, each output port has its own circuitry (output port controller) which is not connected to the circuitry of the rest of output ports. Each input port can reach any output port direction except the output port that goes in the same direction of the input port (u turns are not allowed). Also, the local port connecting the core and the switch can be reached from any input port. Thus, each output port works as a 6-to-1 switch. That is, each output port is able to buffer, route, and forward those packets that request that output port. The independence of the output port circuitry minimizes the impact of the increased number of input/output ports [4]. In fact, the baseline 3D-switch presented in this paper is a 3D-extension of a previous 2D-switch. A complete description and analysis of the 2D-switch can be seen in [22]. To allow the up/down directions of a 3D-mesh, two extra input/output ports must be added.

The switch has been implemented using the 45nm technology open source Nangate [24] with Synopsys DC. We have used M1-M3 metalization layers to perform the Place&Route with Cadence Encounter. Links are divided into data and flow control sublinks. Data sublink width is set to 8 bytes. Flow control sublink width is set to 2 bits. Then, the total link width is 66 bits. Flit size is set to data link width. Links are modeled using Virtuoso Analog Design Environment by Cadence. The area of the baseline 3D-switch is 39699.41 um^2. The critical path of the network (see III-B) is fixed by the switch, and it is equal to 0.53ns.

B. Area Overhead

In this section we analyze the costs of introducing the link design described in Section III. The area increment depends on the number of TSVs that should be interconnected. Set N to the number TSVs, which is equal to the data path size ($N = 64$). The area overhead introduced by the fault tolerant circuit is due to the transistors needed to implement the different Ω networks and the ROM memories used to store the link configuration. The Ω networks and the ROMs required have been designed using Analog Design Environment Virtuoso by Cadence. To achieve a high performance, each 2x2 switch of an Ω network has been implemented using CMOS transmission gates. Thus, the total number of transistors of a 2x2 switch of an Ω network is 8. Then, the total number of a whole Ω network is $8 * \frac{N}{2} * \log_2(N)$. Similarly, the number of transistor used by a ROM is $5 * \frac{N}{2} * \log_2(N)$. Table I shows the area occupied by a 3D baseline switch, and the extra area occupied by the link design inserted in the up and down input/output ports. Notice that, the extra area occupied by the link is only a 6.69% of the area of the whole switch.

	Area (um^2)
Baseline Switch	39699.41
Vertical Link Design	2655.74

TABLE I
Increment in area

C. Timing Analysis and Performance

Introducing the fault tolerant circuit increases the critical path of the TSV-link (see Section III-B). Table II shows the worst-case delays for the different components of the TSV based link design. We model the designed vertical link circuit using the Analog Design Environment Virtuoso by Cadence and the 45nm technology open source Nangate [24]. The TSV via has been modeled using an RC model, where the R and C parameters have been obtained from [23].

	Delay (ps)
Omega network	33
TSV via	31
Setup time	38
Hold Time	9

TABLE II
FAULT TOLERANT CIRCUIT COMPONENT DELAYS.

Using the results shown in Table II and Equation 2, the delay of the fault tolerant vertical link is 288 ps. Despite the increment in latency due to the fault tolerant mechanism, there is no penalty in the critical path of the whole network, as the critical path of the switch (0.53ns) is higher than the TSV-link delay with the fault tolerant circuit. As the critical path of the network is not affected, there is no loss in performance when introducing the fault tolerant vertical link.

D. Yield

As told before, our approach is based in the fact that the probability of having a vertical link with a number of faulty TSVs higher than $\frac{N}{2}$ is negligible. Assuming a probability P_{TSV} of having a faulty TSV, the probability of having a vertical link with a number of faulty TSVs lower than $\frac{N}{2}$ is given in Equation 5. In this equation C(N,i) represents the possible combinations of selecting i wires from a total of N wires.

$$Y \leq \sum_{i=0}^{N/2} C(N,i)(P_{TSV})^{i}(1-P_{TSV})^{N-i} \tag{5}$$

From [17] and [13] we have that the probability of having a faulty TSV is in the range of {0.0001, 0.001}. Assuming those TSV failure rates the yield of our approach tends to 100%. For example, our approach achieves a yield of 99.975 when P_{TSV} is 30%. Note that this failure probability is 10^3 higher than the expected TSV failure probabilities.

V. CONCLUSIONS

In this paper a new fault tolerant TSV link design is proposed to tolerate the presence of faulty TSVs. The design exploits the slack available in vertical links using TSVs to perform a splitted transmission of flits using the positive and negative edges of the clock. An Ω network is placed between the switch and the links allowing to transmit half of the flits by selecting a subset of $\frac{N}{2}$ fault-free TSVs.

One of the main advantages is the avoidance of extra TSVs. Additionally, results show that the ability to tolerate faulty TSVs is very superior to other proposals. Concretely, for the expected TSV failure rates, the yield of our approach tends to 100% regardless the origin of failures. Finally, results also confirm that the additional hardware cost of this proposal is affordable as the area of the 3D switch is increased only by 6.69%. As future work, we plan to reduce the number of TSVs below N while still guaranteeing reliable transmissions of N-bit flits. This is based on the fact that due to the large reconfiguration capabilities of the Ω network most of the TSVs of a link remain unused in the current proposal.

ACKNOWLEDGEMENT

This work was supported by the Spanish MEC and MICINN, as well as European Comission FEDER funds, under Grants CSD2006-00046 and TIN2009-14475-C04. It was also partly supported by the project *NaNoC* (project label 248972) which is funded by the European Commission within the Research Programme FP7.

REFERENCES

[1] V. Chandramouli and C.S Raghavendra, ”Nonblocking properties of interconnection switching networks”, IEEE Trans. on communications 1995.

[2] G. Chen and E.G. Friedman, ”Low-power repeaters driving RC and RLC interconnects with delay and bandwidth constraints”, IEEE Transactions on VSLI 2006.

[3] D.Ern, et Al., ”Razor: A Low-Power Pipeline Based on Circuit-Level Timing Speculation”, International Symposium on Microarchitecture 2003.

[4] J. Flich and D. Bertozzi, ”Designing Network On-Chip Architectures in the Nanoscale Era”, Chapman & Hall 2010.

[5] C. Grecu et al., ”NoC Interconnect Yield Improvement Using Crosspoint Redundancy”, Chapman and Hall 2010. in DFT 2006.

[6] F. Hassan et al., ”Impact of Device Variability in the Communication Structures for Future Synchronous SoC Designs”, International Symposium on SoC 2009.

[7] C. Hernandez, F. Silla, and J. Duato ”A Methodology for the Characterization of Process Variation in NoC Links”, *DATE 2010*.

[8] Y. Hoskote et al., ”A 5-GHz Mesh Interconnect for a Teraflops Processor,” in *IEEE Micro Magazine*, Sept-Oct. 2007, pp. 51-61.

[9] M.R.Kakoee, I.Loi, L.Benini, ”A New Physical Routing Approach for Robust Bundled Signaling on NoC Links”, GLSVLSI 2010.

[10] D. Kim and S. Lim. ”Through-silicon-via-aware delay and power prediction model for buffered interconnects in 3D ICs”. In SLIP 2010.

[11] J. Kim et al., ”A novel dimensionally-decomposed router for on-chip communication in 3D architectures”, ISCA 2007.

[12] N. Linial and M. Tarsi, ”Interpolation between bases and the shuffle exchange network”, European Journal of Combinatorics, January 1989.

[13] I. Loi, S. Mitra, T. H. Lee, S. Fujita, and L. Benini. ”A low-overhead fault tolerance scheme for TSV-based 3D network on chip links”. In Proceedings of the ICCAD 2008.

[14] M.Mondal et al., Provisioning On-Chip Networks under Buffered RC Interconnect Delay Variations, ISQED07

[15] S. Murali et al. ”Comparison of a Timing-Error Tolerant Scheme with a Traditional Re-transmission Mechanism for Networks on Chips” in NoCs 2006.

[16] D. Park et al., ”MIRA: A Multi-layered On-Chip Interconnect Router Architecture”, ISCA 2008.

[17] V. Pasca, L. Anghel, C. Rusu, R. Locatelli, and M. Coppola. ”Error resilience of intra-die and inter-die communication with 3D Spidergon STNoC”. In DATE 2010.

[18] V. Pasca, L. Anghel, and M. Benabdenbi. ”Fault tolerant communication in 3D integrated systems”. In the DSN Workshops 2010.

[19] S. Pasricha, ”Exploring serial vertical interconnects for 3D ICs”, In Proceedings of DAC 2009.

[20] R. Patti, ”Impact of Wafer-Level 3D Stacking on the Yield of ICs”, Future Fab Intl. Issue 23.

[21] V. F. Pavlidis and E. G. Friedman, ”Three-dimensional Integrated Circuit Design”, Morgan Kaufmann Publishers 2009.

[22] A. Roca, J. Flich, F. Silla, and J. Duato, ”A Low-Latency Modular Switch for CMP Systems”, available at http://www.disca.upv.es/jflich/tech_report_modular_switch.pdf

[23] I. Savidis et al., ”Electrical modeling and characterization of through-silicon vias (TSVs) for 3-D integrated circuits”, in Microelectronics Journal, January 2010.

[24] 45nm FreePDK. The Nangate Open Cell Library. (https://www.si2.org/openeda.si2.org/projects/nangatelib/)

[25] ”International Technology Roadmap for Semiconductors”, 2007 Edition, available online at http://www.itrs.net/Links/2007ITRS/Home2007.htm

[26] ”TILE-Gx Processors Family”, available at http://www.tilera.com/products/TILE-Gx.php