

Characterization and Implementation of Fault-Tolerant Vertical Links for 3-D Networks-on-Chip

Igor Loi, Federico Angiolini, Shinobu Fujita, *Member, IEEE*, Subhasish Mitra, *Member, IEEE*,
and Luca Benini, *Senior Member, IEEE*

Abstract—Through silicon vias (TSVs) provide an efficient way to support vertical communication among different layers of a vertically stacked chip, enabling scalable 3-D networks-on-chip (NoC) architectures. Unfortunately, low TSV yields significantly impact the feasibility of high-bandwidth vertical connectivity. In this paper, we present a semi-automated design flow for 3-D NoCs including a defect-tolerance scheme to increase the global yield of 3-D stacked chips. Starting from an accurate physical and geometrical model of TSVs: 1) we extract a circuit-level model for vertical interconnections; 2) we use it to evaluate the design implications of extending switch architectures with ports in the vertical direction; moreover, 3) we present a defect-tolerance technique for TSV-based multi-bit links through an effective use of redundancy; and finally, 4) we present a design flow allowing for post-layout simulation of NoCs with links in all three physical dimensions. Experimental results show that a 3-D NoC implementation yields around 10% frequency improvement over a 2-D one, thanks to the propagation delay advantage of TSVs and the shorter links. In addition, the adopted fault tolerance scheme demonstrates a significant yield improvement, ranging from 66% to 98%, with a low area cost (20.9% on a vertical link in a NoC switch, which leads a modest 2.1% increase in the total switch area) in 130 nm technology, with minimal impact on very large-scale integrated design and test flows.

Index Terms—3-D integration, fault tolerance, network-on-chip (NoC).

I. INTRODUCTION

ONE OF THE LIMITING factors to performance scaling of silicon chips under the 130 nm node is attributable

Manuscript received November 20, 2009; revised June 3, 2010; accepted July 11, 2010. Date of current version December 17, 2010. This work was supported by the European Project JTI ENIAC, under Grants END 120214 and PRO-3D FP7-ICT-3.6-248776 for DEIS, and by the PRO3D Project financed by the European Community 7th Framework Programme (ref. FP7-ICT-248776). This paper was recommended by Associate Editor V. Narayanan.

I. Loi is with the Department of Electronic Engineering, University of Bologna, Bologna, Italy (e-mail: igor.loi@unibo.it).

F. Angiolini is with EPFL, Lausanne, Switzerland (e-mail: federico.angiolini@epfl.ch/unibo.it).

S. Fujita is with Toshiba, San Jose, CA 95131 USA, and Kawasaki, Kanagawa, Japan (e-mail: shinobu.fujita@toshiba.co.jp).

S. Mitra is with the Department of Electrical Engineering and the Department of Computer Science, Stanford University, Stanford, CA 94305 USA (e-mail: subh@stanford.edu).

L. Benini is with the Department of Electronics and Computer Science, University of Bologna, Bologna, Italy (e-mail: luca.benini@unibo.it).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCAD.2010.2065990

to interconnect scaling. Key metrics of interconnect delay and energy dissipation now tend to dominate with respect to switching devices [1].

To tackle interconnect and architectural scalability challenges, two major trends are emerging. On one hand, 3-D integrated circuits (ICs) alleviate the interconnect I/O bandwidth and latency bottlenecks, by leveraging the third axis to minimize communication distances and to provide more connectivity among blocks. 3-D ICs may also enable heterogeneous integration and new classes of applications through significantly improved performance and energy efficiency of complex system architectures (e.g., technologies from Tezzaron Semiconductor Corporation [2], IMEC, Leuven, Belgium, MIT Lincoln Labs, Lexington, MA, and IBM [3]). One of the most promising technologies for 3-D IC integration is based on through silicon vias (TSVs), pillars manufactured across thinned silicon substrates to establish inter-die connectivity after die bonding. Salient TSVs features include fine pitches, high densities, and high compatibility with the standard complementary metal-oxide-semiconductor (CMOS) process.

On the other hand, architectures based on the network-on-chip (NoC) design paradigm are receiving increasing technical consensus. In particular, 3-D NoCs combine the benefits of short vertical interconnects of 3-D ICs and the scalability of NoCs. A vertical link can be physically implemented as a cluster of TSVs. Unfortunately, currently available processes for TSV fabrication have low yields relative to standard 2-D processes. Fig. 1 shows the yields of chips containing TSVs manufactured in three different process technologies: HRI [4], IMEC [5], and IBM [6]. Thus, fault-tolerance schemes are needed. A fine-grained, single-TSV-oriented redundancy approach would intuitively lead to doubling of the TSV count and severe logic overhead. The use of a redundancy scheme at the link level on the other hand would alleviate reliability concerns while taking advantage of the NoC architecture to reduce overhead.

Our work moves from a circuit-level model for vertical TSV-based interconnects, including accurate 3-D parasitic extraction. Comparative analysis demonstrates that not only vertical interconnects are usable but also that they are highly competitive with horizontal wires in delay and power, with a reasonable area overhead. As a second main contribution, we extend a 2-D NoC switch architecture to deal with vertical

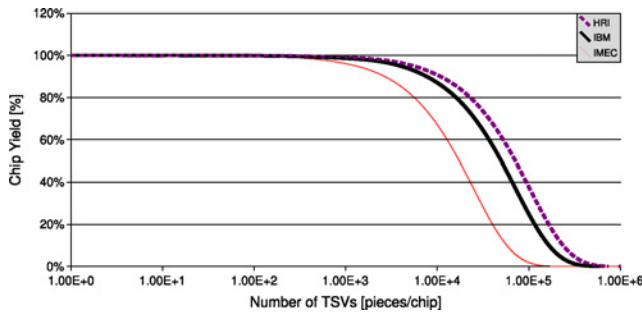


Fig. 1. Yield for TSV-based chips in three different processes: IBM, HRI, and IMEC. Only random (complete or partial) open defects are considered in this figure, since misalignments are well controlled during the bonding phase. Yield is evaluated using the Poisson distribution.

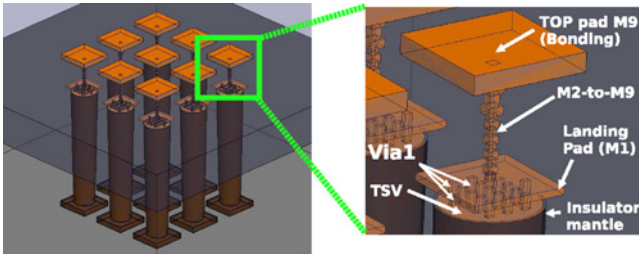


Fig. 2. TSV bundle in Si-bulk technology and detail of routing. A TSV traversing the substrate is available for routing as a large plate in Metal 1 called landing pad. Connectivity is provided by standard routing, just taking care of cell placement and metal obstructions.

links. Our third contribution is the development of a prototype design flow for automatic instantiation of 3-D NoCs. As a fourth contribution, we then describe a defect-tolerant multi-bit vertical link which enables significant yield improvement with respect to random defects at an extremely low cost. Like traditional defect-tolerance techniques (such as those used for memories), our technique also relies on redundancy. Our fifth contribution is an efficient physical design of such defect-tolerant TSV-based links, featuring low cost and minimal disruption of the overall design, production, and test flows. While this link design is generally applicable to both 3-D NoCs and other 3-D interconnects, it is especially useful for the former as it can take advantage of the NoC switch architecture to minimize the system-level area cost.

We show a case study where a planar NoC topology is folded and implemented across two chip layers in two variants—with redundant links and without redundant links. We also present a detailed analysis to evaluate benefits, feasibility, and hardware cost, estimated at the layout level. In our experiments, we achieve significant yield improvements (from 66% to 98% for different configurations) for random open defects, a major challenge for TSVs. Our layout results demonstrate the feasibility of this approach and its low cost (20.9% area overhead for a single vertical link, i.e., a negligible increase with respect to the whole layout).

II. RELATED WORK

Interconnect scaling has become one of the most crucial challenges in chip design, and is only expected to get worse in the future. 3-D integration and NoC design methodologies are expected to overcome many of these challenges. NoCs

have been suggested as a scalable communication fabric [7], [8]. 3-D integration has been proposed in different ways (e.g., Tezzaron Semiconductor Corporation [2], IMEC, MIT Lincoln Labs, and IBM [3]) providing promising solutions to enable connectivity along the vertical direction.

Recently, some research has been undertaken on 3-D NoCs. In [9], the authors proposed a dimension decomposition scheme to optimize the cost of 3-D NoC switches, and presented some area and frequency figures derived from a physical implementation. The fundamental assumption of their work is that a regular, homogeneous NoC is the best solution for a 3-D design, and, therefore, the next logical step is to reduce the cost of each required building block. However, we believe that, for such complex designs as stacked 3-D chips, which are likely to mix logic layers with memory layers and even more uncommon functionality, heterogeneity will likely be significant, especially along the vertical axis. For this reason, we propose a more general approach, where the designer is allowed to choose among planar and vertical communication on a switch-by-switch basis, without any topological constraint. Post-silicon nano-scale 3-D interconnections have also been recently investigated [10], but large-scale availability of these technologies in the near future is uncertain.

Only few works partially address the characterization the vertical interconnects for use in 3-D NoCs with respect to physical implementation and timing requirements. In [11], the authors present various possible 3-D topologies for 3-D NoC, considering power an latency cost. In [12], a comparison of three 3-D clock distribution network topologies is presented. In [13] and [14], the author presents a design tool to synthesize application-specific 3-D NoCs. This tool is able to find the best NoC topology, and to assign the network components on to the 3-D layers and performs a placement of them in each layer.

To the best of our knowledge, no previous work fully characterizes the vertical interconnections for use in NoCs, especially with respect to physical implementation and timing requirements.

As technology scales, fault tolerance is becoming a key concern in on-chip communication. Optical proximity correction and redundant via placement [15] have solved a huge number of cases of faults related, mainly, to interconnects. Roving STARS [16] have been proposed for field-programmable gate array testing, diagnosis, and fault-tolerance, where spare resources are always present in the neighborhood of the located fault, simplifying fault bypassing. Several fault-tolerant algorithms for on-chip interconnects have been presented by [17], but as the authors emphasize, this approach is not well suited to the NoC context due to the large area cost. Experiments by HRI on 3-D ICs report very high yields of over 60%; the redundancy scheme used realizes each vertical interconnect expensively as a pair of vias (twins) [4].

Several significant achievements have been announced in the last few months, confirming the rapidly increasing industrial research and development effort in this area. In [18], an 8 GB 3-D DDR3 using TSVs to stack four dynamic random access memory dies is presented. This memory uses a set of redundant TSVs with a check-and-repair scheme to increase

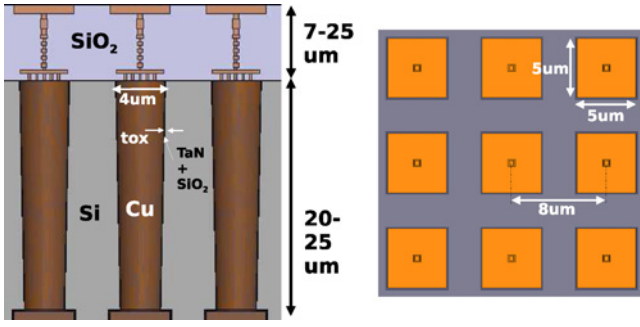


Fig. 3. Schematic representation of a bundle of 3-D vias.

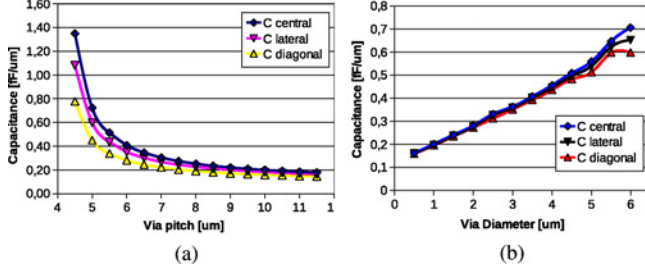


Fig. 4. Capacitance trend when (a) sweeping the pitch of vias having a constant diameter and (b) sweeping the diameter of vias having a constant pitch. C_{Central} refers to the central TSV capacitance of the bundle, C_{Diagonal} stands for the capacitance of one of the TSV placed in one of the four corners, while C_{Lateral} means the capacitance for one of the four lateral TSVs.

the chip yield. Vertical vias are over-provisioned by 2:1 and 4:2, achieving a target yield of 95% and 99.8%, respectively.

No previous work fully addresses the TSV yield loss and the yield improvements for vertical links of 3-D NoCs. In this paper, we propose both a detailed characterization of the vertical links and switches, and a novel scheme to overcome the yield limitation. The starting point of this paper is [19], [20], where a thorough physical and timing analysis of the vertical links has been conducted on a real 3-D NoC. Further, it is worth stressing that the proposed scheme can also be applied successfully to alternative interconnection schemes, such as buses.

III. 3-D NOC DESIGN

In this section, we model the performance of vertical interconnects to assess 3-D NoC implementation tradeoffs. We first quantify the delay of a group of signals implemented as a bundle of TSVs, then we analyze the performance of the whole structure of a 3-D communication channel comprising two switches on different planes connected by TSVs.

A. TSV Link Modeling

Even in 2-D point-to-point connections, due to variability and timing constraints, each wire of a NoC link should be kept in the same routing group or “bus.” This requirement is of utmost importance when the link is routed through one or more layers of a 3-D stack. Therefore, vertical links should not be used in isolation but it is highly recommended to create buses of such wires.

The geometry of a TSV bus connecting adjacent stacked wafers is shown schematically in Fig. 2 for bulk-silicon technologies. Given the physical proximity of the TSVs, concerns

related to capacitive coupling and signal integrity within such buses may arise. Further, all parasitics must be quantified in order to assess the propagation delay of such TSV buses.

We use an electromagnetic field solver tool [21] to extract the resistance inductance capacitance parasitics of the 3-D structure. This makes the study of signal integrity and delay possible. The starting point of our analysis is a simple configuration composed of nine TSVs placed in a 3×3 grid structure. The baseline configuration refers to the standard Si-bulk technology, and can be summarized as (Fig. 3) follows:

- 1) copper vias;
- 2) $4 \mu\text{m}$ via diameter;
- 3) $5 \mu\text{m} \times 5 \mu\text{m}$ pads at via extremities;
- 4) $8 \mu\text{m}$ via pitch;
- 5) $0.5 \mu\text{m}$ oxide thickness (t_{OX});
- 6) $25\text{--}50 \mu\text{m}$ layer thickness (substrate + metallizations).

It is important to note that the TSV process does not scale with the CMOS technology. TSV diameters and pitches are two to three order bigger than transistor gate lengths. This implies that, even moving to newer technologies, the intrinsic cost for vertical interconnect do not change. For this geometry and sizing, the TSV inductance and inductive coupling becomes negligible.¹ For this reason, in the following, we assume that the intrinsic TSV delay is a function of resistance and capacitance only.

Resistance can be described with a single parameter as a function of via length and cross-section. For example, copper vias with $4 \times 4 \mu\text{m}$ diameter show a resistance per micrometer around $1.18 \text{ m}\Omega/\mu\text{m}$. Skin effect is negligible at few gigahertz with these dimensions. A comparison between TSVs and top metal wires (Metal 8, 130 nm technology node), which have $0.4 \times 0.8 \mu\text{m}$ cross-section shows that TSV resistance per μm is 50 times smaller than Metal 8; this is significant, especially since TSVs are typically much shorter than global 2-D wires.

Capacitance, on the contrary, due to coupling effects, poses several more modeling issues. In Fig. 4, we report extraction results for the parasitic capacitance of TSVs while sweeping TSV pitch and diameter. The capacitance toward the ground plane is a dominant element in the overall capacitance (common cost in all the three cases), while the cross-capacitance between TSV is mainly due to the bonding pads on top of the stack. In Fig. 4(a), it is shown the capacitance trend when sweeping the distance between the TSVs. As the distance increases, the coupling capacitance between TSVs decreases (term proportional to $1/p^2$ where p is the pitch). This capacitance is related only to the top part of the TSV (from M1 to top metal).

In Fig. 4(b), we swept the TSV diameter. As can be seen, increasing the diameter means to increase the TSV surface exposed to the bulk region (with fixed oxide thickness). Therefore the capacitance increases linearly with the diameter. Due to the reference structure symmetry only three TSV cases are relevant: central, lateral, and diagonal. Both figures report the capacitance trend for these three contributions.

¹Extracted self-inductances are in the order of 6 pH, while mutual inductances are in the order of 1 nH. SPICE simulations do not show any significant delay variation.

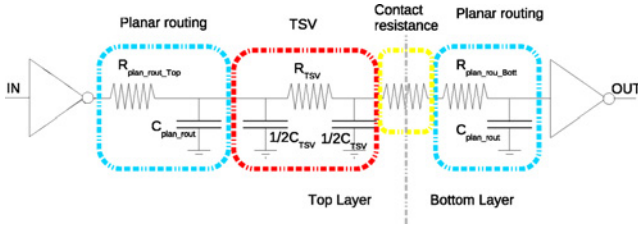


Fig. 5. Electrical model of a single bit vertical link, between two tiers.

To put these results in perspective, the maximum un-repeated planar line length in Metal 2 and Metal 3 is 1.5 mm in 130 nm. If we take 1.5 mm as a reasonable planar inter-switch link length, we observe that vertical links exhibit roughly one order of magnitude lower capacitive load. Roughly, the same ratio can be found for resistance.

To evaluate dynamic properties and signal integrity, we create a Π -network electrical model of the TSV bundle (Fig. 5). Crosstalk is modeled by the cross-capacitance between TSV. Since we suppose that TSVs provides electrical inter-die connectivity through metal bonding, we introduce in our model a contact resistance. To characterize the link delay, we built a communication channel composed by drivers, planar routing, and TSVs. We perform two iterations (up to post-layout level), the first to characterize the planar routing, and the second to resize the drivers with the back-annotated capacitance and resistance of the 3-D link (multi-corner analysis). Delay estimates through a SPICE simulation result in 18.5 ps for TSVs of $4\mu\text{m}$ diameter and $8\mu\text{m}$ pitch. This latency is dominated by planar routing since resistances of wires are up to two order of magnitude greater than the TSVs, while capacitances are comparable. As a consequence, even after taking coupling effects of tightly packed TSV bundles into account, vertical links turn out to be substantially faster and more energy efficient than moderate size planar links.

B. 3-D NoC Architectural and Physical Design

NoC components and NoC design tools require modifications to support vertical links made of TSVs. As discussed in Section II, 3-D designs are likely to expose a large degree of heterogeneity, especially along the vertical axis. Therefore, we choose to base our integration effort on the \times pipes [22] NoC library, which supports arbitrary connectivity, and on its instantiation toolchain [23]. Thus, we can leverage a semiautomatic design flow, from register transfer level (RTL) description to layout-level verification.

\times pipes switches come in two variants, conceived to best match two flow control protocols. The first is ACK/NACK, a retransmission-based protocol featuring increased error resilience. The second is STALL/GO, a simple variant of credit-based flow control allowing for pipelined links to be transparently deployed. In the ACK/NACK case [Fig. 6(b)], output buffers need to be inserted within switches, since any transmitted packet should be stored for potential retransmission. This implies a hardware cost, but it also means that NoC links are enclosed between two clocked buffers at the sending and receiving ends (dual-stage pipeline router). Hence, a whole clock period (or more in case of pipelining) is available for signal

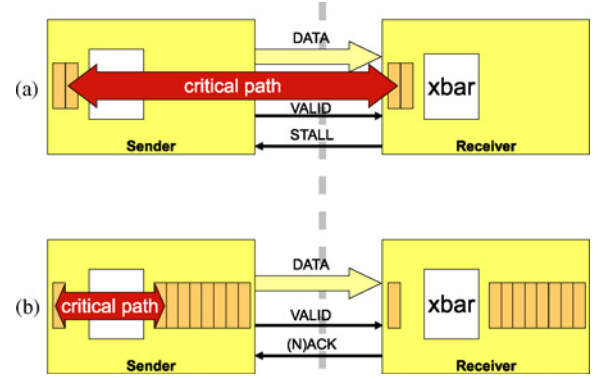


Fig. 6. (a) STALL/GO and (b) ACK/NACK switches and link. In (a), only the switch input is buffered, and the critical timing path is across the link. In (b), the switch is buffered both at inputs and outputs, therefore the critical path is split.

propagation along the wires of the inter-switch links. The link and the switch logic are decoupled by the output buffer.

In contrast, in STALL/GO, low switch latency and reduced buffer cost are the main goals. STALL/GO switches, therefore, adopt a lean architecture [Fig. 6(a)], where only switch inputs are buffered (single stage pipeline router). The switch logic and the link propagation time (up to the following switch or to the first link pipeline stage) contribute to a same timing path, which becomes the bottleneck for the system. While ACK/NACK routers with small buffer depth (four in Fig. 8) transparently allows for links of arbitrary propagation time, with STALL/GO the link propagation time directly impacts the maximum operating frequency of the switches and so of the whole NoC.

We leverage the information gathered in the beginning of this section to build library exchange format (LEF) and liberty (LIB) descriptions of vertical vias and top pads. LEF macros are standard hardware descriptions at the layout level, including information about process technology, routing blockage, keep-out areas and pin/pad locations; the LIB files carry the timing and power information to ensure timing convergence. Based on these models, TSVs can be taken into account within the design during the placement and routing stage; the corresponding obstructions are positioned at the input or output pins of a switch port, just as a horizontal wire would. At the RTL level, the design remains completely unchanged with respect to a 2-D implementation, and the load corresponding to the TSV electrical model are added for accurate timing analysis. This brings several advantages: 1) the presence of vertical wires is totally transparent to the architectural and functional views of the architecture; 2) a chip may feature any degree of connectivity heterogeneity since vertical links can be added or exchanged for horizontal ones; 3) vertical bandwidth can be added only where needed in the chip, saving switch ports everywhere else; and 4) building upon the savings brought by the previous item, the set of switches with vertical ports, i.e., the ones located where vertical bandwidth is really needed, can have ideal performance because they can be implemented as full crossbars.

Thanks to this approach, a complete design flow is achieved; this includes the ability to extract and simulate a 3-D layout,

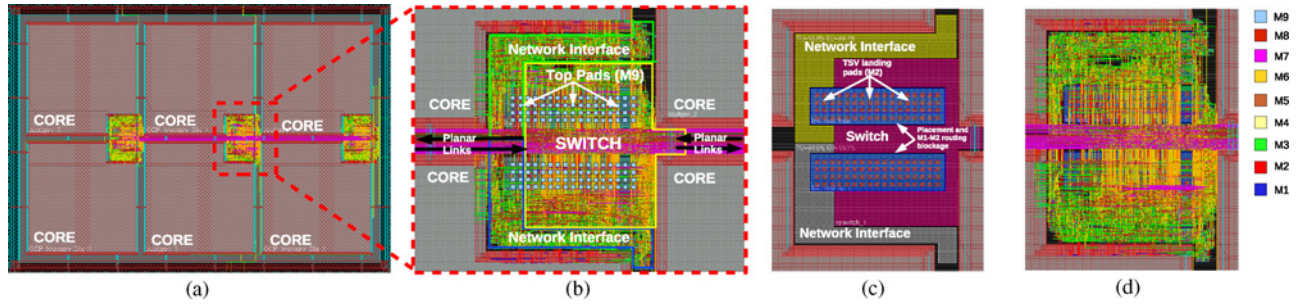


Fig. 7. Layout details (65 nm) of a NoC topology and switches with 3-D ports. (a) Topology where switches feature the UP port. (b) Detail of a switch with an UP port (I/O pads on M9): Metal 8 and Metal 9 are reserved for the vertical link routing and bonding. (c) Floorplan of a switch with a DOWN port. The TSV hard macros are placed close to the switch. (d) Post-place and route detail of a switch with a DOWN port.

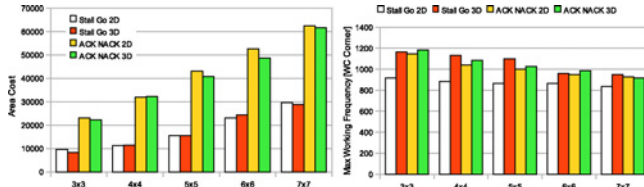


Fig. 8. Silicon cost and maximum frequency achievable by STALL/GO versus ACK/NACK switches in 2-D and 3-D flows, for varying switch cardinalities, in 65 nm.

where all switch ports are exposed to proper timing constraints and load information is available for both horizontal and vertical connections. A depiction of a sample layout featuring a 5×5 switch with vertical ports (UP direction with 64-bit data width on a two-tier layout) is presented in Fig. 7. The TSV macros (or top pads) are placed close to the pin-out of the switch block, shortening the wires leading to the base of the via, thus reducing parasitic and improving timing.

The choice of a NoC topology must be performed by taking into account available performance information. Therefore, it is important to build a timing model of the switches. In Fig. 8, we explore the frequency that STALL/GO and ACK/NACK switches of different cardinalities can achieve when driving horizontal (1.5 mm) or vertical ($50 \mu\text{m}$) links in 65 nm technology. STALL/GO is, in general, slightly slower than ACK/NACK due to the contribution of link delay on critical paths; however, when used in combination with TSVs, it regains 100–250 MHz (an average of 21%), while maintaining its low-overhead properties (and single-cycle latency) as shown in Fig. 8. In other words, the NoC can be clocked faster when the slower horizontal links are replaced by fast vertical links. As predicted at the beginning of this section, ACK/NACK switches do not gain any performance benefits when moving to 3-D, since vertical interconnects do not influence the critical path, which is enclosed within the router.

IV. YIELD ENHANCEMENT

As seen in Section I, the main limiting factor to reach high yield levels in 3-D ICs is directly related to the TSVs. In this section, we discuss the nature of the defects that induce yield loss, and then we formulate a novel scheme to identify and replace faulty TSVs. Finally, we will discuss and quantify the cost of the hardware resources for the testing of the whole 3-D structure.

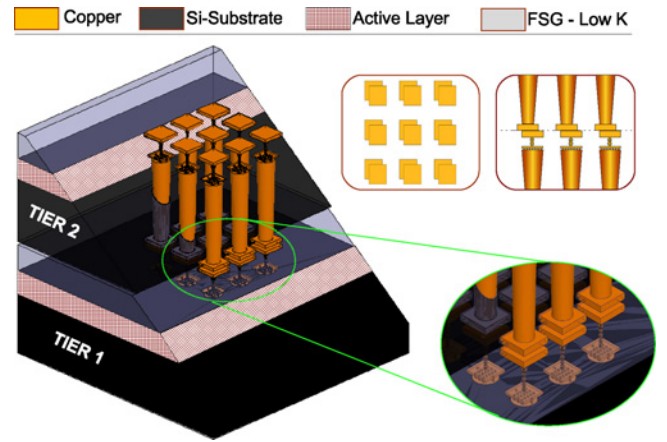


Fig. 9. Cross-section of a vertical link across two tiers. The figure also shows the worst-case misalignment scenario.

A. Reliability Analysis of 3-D NoC Links

The primary failure mechanisms for TSVs are misalignments and random (complete or partial) open defects [24]. Misalignments are due to imprecise wafer alignment prior to and during wafer bonding (Fig. 9), which results in shifts of the bonding pads from their nominal positions. Random defects comprise a variety of physical phenomena during, e.g., the thermal compression process used in wafer stacking, eventually leading to opens along TSVs.

Starting from these considerations and based on [19], we have conducted a detailed study to quantify the impact of TSV failures on overall chip yield. To this end, we use our electrical model of tiers interconnected by TSVs (Fig. 5). The vias are driven by an inverter followed by a stretch of planar interconnect (global routing). The contact resistance depends on the quality and area of bonding.

In case of misalignments (e.g., top wafer shifts along the X or Y axes or small rotations), the bonded area decreases. This phenomenon has been modeled as a variable resistance (central resistor in Fig. 5) after the Π network, and the outcome is summarized in Table I. As can be seen, misalignments of even noticeable entity do not normally compromise functionality and have a minimum impact on delay, which is usually dominated by the overall planar routing parasitics [19]. Extreme misalignments, like in the last row of Table I, are highly unlikely in state-of-the-art wafer bonding processes [2], [3],

TABLE I

PAD CONTACT RESISTANCE AND DELAY INCREASE FOR CU-CU WAFER METAL BONDING UNDER DIFFERENT MISALIGNMENT CASES [25], [26]

Misalignment (μm) in X-Y	Contact Area (μm^2)	Contact Resistance (Ω)	Δ Delay (%)
0	4×4	10 m	0
1	3×3	19 m	< 1%
2	2×2	40 m	< 1%
3	1×1	160 m	< 1%
3.98	0.02×0.02	1 K	22%

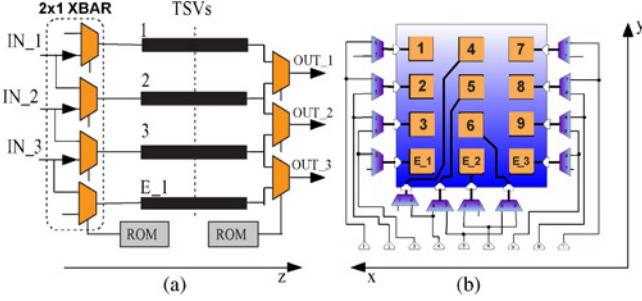


Fig. 10. Redundant TSV mapping scheme. (a) Simplified crossbar functional scheme for dynamic routing. (b) TSV obstructions (the orange squares are the TSV pads) and routing. Extra pads (E_n) are spread around, permitting the bypassing of faults by $2 \times$ multiplexers.

[27]. This motivates special emphasis on workarounds for the other main source of yield losses, random defects.

Random (complete or partial) open defects affect single vias or a small area of the interface because of failure mechanisms such as dislocations, O_2 trapped on the surface, void formation, or even mechanical failures in TSVs [3], [25], [28], [29]. To model the effects of these defects, we assumed an uniform TSV defect distribution and performed several Monte-Carlo simulations. Based on our results (Section V), we concluded that random (complete or partial) open defects are far more relevant compared to misalignment problems. For this reason, we focus on these defects in the following sections.

B. Yield Enhancement Approach

Among the numerous techniques to increase wafer yield of very large-scale integrated designs, we focus on hardware redundancy, deployed at design time, with some amount of post-manufacturing configuration. We use active redundancy in the form of spare pads and reconfigurable routing hardware (Fig. 10). We then implement a link re-routing solution, designed to leverage post-manufacturing configurability of the TSV interconnect map. This allows us to achieve high yield while minimizing the overhead in terms of the number of pads and extra logic. Combining testing resources (e.g., scan chains²) with such reconfigurability plays a key role in achieving high yield. This solution allows us to test each vertical interconnect and diagnose defects, to isolate any failed TSV, and finally to restore functionality through reconfiguration by routing the affected signals over to the spare pads.

As we see in Fig. 10(a), in our proposed dynamic routing scheme, all pads are driven by a 2×1 crossbar, and each signal

can be routed to two different TSVs. Our approach is capable to tolerate an arbitrary number of faults per link, simply increasing the number of extra TSVs. We define as “cluster” a group of signals (with the same direction), that share one extra TSV, therefore more extra TSVs means more clusters. For each cluster, a single defect is tolerated. Since the link is bidirectional, at least two extra TSVs are needed. This means that two faults are tolerated (the first in the forward cluster, the second in the backward). The opposite corner case is realized duplicating every TSVs (every TSV of the link has a backup). This second case is capable to handle N faults, where N is the width of the link. When more than one defect per cluster is detected, the entire link is disabled. To do that, we clamp the flow control signals to a safe value and then we reprogram the routing tables, in order to bypass the link [30]. To increase the resilience due spot defects, the cluster size must be reduced increasing the extra TSVs (2 to 38 as depicted in Fig. 14), and the TSVs within the cluster must be spread in order to maximize the mutual distance. In case of stacked TSVs (providing connectivity between routers from tier X to tier $X+i$ with $i > 1$), the repair approach is re-conducted to the baseline case where only one fault per “vertical cluster” is tolerated. In this case, the repair capability decreases as compared with the baseline case, because the fault probability of a structure made of stacked TSVs is greater than the single TSV.

The routing crossbar is extremely small, as a strategic choice to keep the area overhead as low as possible—for each additional re-routing degree of freedom, the crossbar size increases linearly, while the dynamic performance of the link degrades. With this lean architecture, faults are recovered by shifting affected signals to the neighboring pads, and further shifting the displaced connections over to other adjacent pads until all connections are across safe electrical structures. To clarify the recovery scheme, we shall consider Fig. 10(b). Supposing that pad 2 is affected by some defects (resulting, e.g., in an open circuit), we route signal 1 normally through its associated pad 2, while signal 2 gets rerouted through pad 3, and therefore signal 3 gets re-mapped to pad E_1 . Signals outside this column are not shifted since the defect is contained inside the first cluster; the recovery process is performed locally. The proper routing information is elaborated off-chip (to minimize hardware complexity and overhead) during chip testing, and is then stored on-chip into a small one-time programmable (OTP) memory (e.g., a fuse read only memory).

The importance of the testing stage is evident, as it determines all the necessary inputs to correctly set the crossbar up. To test the physical interconnect, we reuse the scan chains which are normally inserted anyway in the design, thus incurring no overhead for this. Fig. 11 illustrates the hardware facilities used to test the TSVs. The TSVs are tested by injecting test vectors (TVs) in one tier (e.g., the bottom one). The TV is propagated to the destination tier (e.g., the top one), where it is captured and transmitted off-chip. In summary, the approach is split into five steps as follows:

- 1) inject TVs (e.g., bottom tier);
- 2) propagate TVs across TSVs and capture them (e.g., top tier);

²The use of scan chains does not normally imply any extra cost, as they are typically integrated in every design.

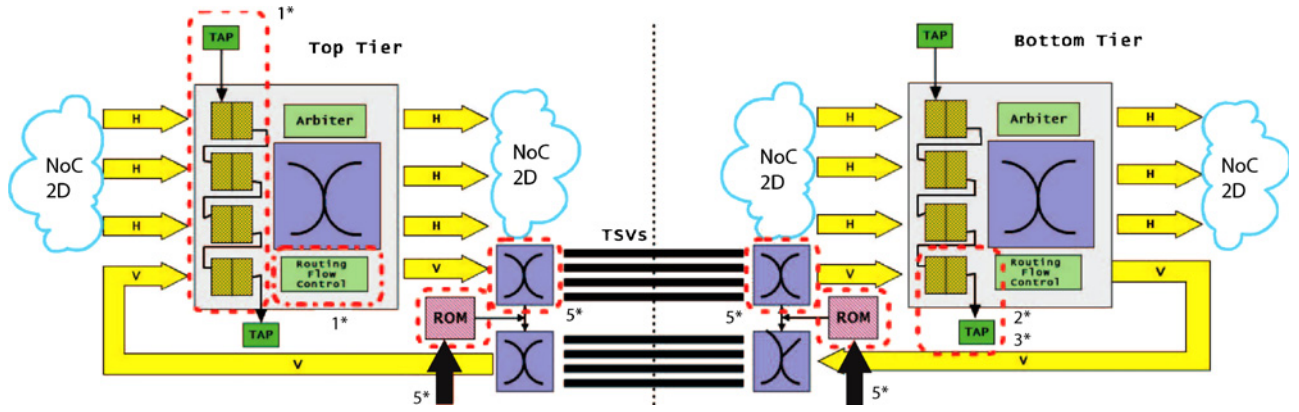


Fig. 11. TSV NoC test environment. In test mode, TVs are injected from the test access point (TAP) (1*) into the switch input buffer (scan), then the path through the crossbar is enabled (1*) and flow control is disabled. After some cycles, the stimuli reach the next tier where they are captured (2*) from the input buffer, and then shifted out through the TAP (3*). This stream is analyzed off-chip then, based upon the failure map the OTP memories are programmed (5*), reconfiguring the crossbar to isolate failed structures. Labels are inserted accordingly with the testing operations described in Section IV.B (label 4 is omitted since it represents the off-chip elaboration).

- 3) scan out the captured data (e.g., top tier);
- 4) elaborate off-chip the interconnect map;
- 5) reconfigure the crossbar (both bottom and top tier).

The process can be performed at any speed allowed by the external I/O pins. Since the interconnect map is devised off-chip, minimal logic is required on-chip for the mapping procedure—mostly, the OTP memory to store the crossbar configurations.

V. EXPERIMENTAL RESULTS

In this section, we discuss the experimental results for the 3-D NoC, the TSV fault-tolerant scheme, and the achieved yield improvement. We first quantify the cost of adding a 3-D port to switches, when moving from 2-D to 3-D, considering both switch cardinality and flow control. Then, we quantify the cost of the fault-tolerant interface and the yield improvement when the number of spare TSV increases. To get these results, we synthesized the NoC with the TSMC 65 nm technology library (general-purpose process). The front-end flow (multi- V_{th}) has been performed with Synopsys Design Compiler in topographical mode, while the back end with Cadence SoC Encounter. The sign off has been made with both VoltageStorm and PrimeTime, while functional verification is performed with Mentor Graphis ModelSim.

A. Implementation of TSV-Based 3-D NoCs

In this section, we present a NoC implementation based on a 2-D 3×2 quasi-mesh (called simply mesh in the following) and migrate it to a 3-D arrangement (Fig. 12). The 3-D mapping is achieved by splitting in two halves the mesh and overlapping them in separate chip layers, with communication achieved through TSVs. The stacked topology has exactly the same functionality of the planar implementation.

As a first step, we leverage SunFloor [23] to instantiate the 2-D mesh (NoC with deterministic source routing and STALL-GO flow control policy). There is no need to modify the RTL output of SunFloor in any way. Next, we identify the best partitioning for mapping onto the layer stack. This

task for this simple topology is done manually in our case (however, SunFloor 3-D [13] can be used to create directly a 3-D topology, partitioning, and 3-D floorplan). The partitioning criteria include manufacturing limitations, chip pin-out, area considerations, bandwidth demands, and thermal requirements. For example, our test 3×2 mesh connects three processors and three memories. To better balance thermal gradients across the die, we assume that processors cannot be stacked on top of each other; to avoid the formation of hot spots, we interleave processors and memories. The links connect either two different switches or a switch and a network interface; our choice is to cut bi-dimensional topologies across switch-to-switch links, replacing the latter with an upstream and a downstream port.

Then we perform synthesis, placement, and routing of the RTL in two separate runs, one per design partition. During placement, we insert TSV macros at the proper switch boundaries. We initially do not include any fault tolerance. We choose the minimum TSV diameter ($4 \mu\text{m}$) and pitch achievable in current technologies. The area overhead of each vertical via is $64 \mu\text{m}^2$ ($8 \mu\text{m} \times 8 \mu\text{m}$). For each of the UP and DOWN switch ports, $2 \times (2 + Flit_Width) + 4$ TSVs are needed, where 2 is the number of control signals for flow control, etc., $Flit_Width$ is the flit width, 4 is the number of signals used synchronization, and the multiplicative factor 2 is introduced since we consider a bidirectional connection. When $Flit_Width$ is set to 64, the area overhead of a 5×5 switch with a vertical port with respect to a fully planar 5×5 switch is about 15% with ACK/NACK and 60% with STALL/GO. In exchange for this area cost, switches run around 21%.³

The total power consumption for the 65 nm NoC (Fig. 13) running at 200 MHz⁴ (excluding the hard macros) is 53 mW for the planar implementation, and 38 mW for the entire 3-D stack, which leads to a 28% power reduction. Sequential power

³We assume that no additional critical paths are present, and that switch bottleneck is located on 2-D long link or 3-D link.

⁴Synthesis and place and route have been targeted for 200 MHz in order to have a fair comparison between 2-D and 3-D NoC implementation. 2-D NoC is limited due to the long planar links while 3-D NoC critical path is in the mid-link between the two switches.

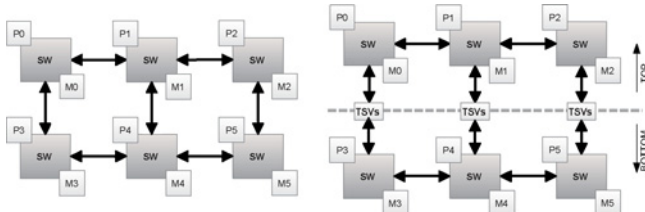


Fig. 12. 2-D 3×2 mesh NoC topology and one possible 3-D re-implementation.

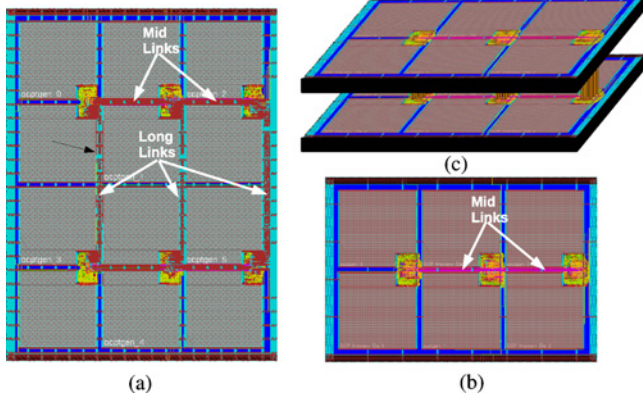


Fig. 13. Layouts for (a) 2-D 3×2 mesh, (b) one of the halves of its 3-D re-implementation, and (c) 3-D view of the two stacked halves (vertical axis not to scale).

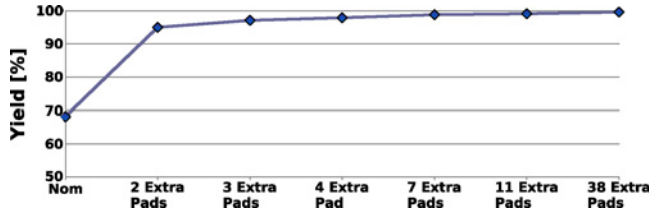


Fig. 14. Yield over seven different hardware configurations: no redundancy, 2, 3, 4, 7, 11, and 38 extra pads, which correspond to 38, 40, 41, 42, 45, 49, and 76 TSVs per 32-bit 3-D link. A fixed defect frequency of 9.75 defects per million opportunities (5.77 SIGMA) is assumed, and a design with 4.2 million TSVs is analyzed.

consumption is roughly the same in both designs, as expected, but combinational power is 34 mW in the 2-D NoC (due to buffering on long horizontal links) and only 19 mW in the 3-D (42% better). The clock tree achieves a smaller skew, 33 ps in the 3-D implementation compared to 84 ps in 2-D.

B. Yield and Hardware Cost of the Redundant Solutions

The proposed fault-tolerant solutions, and a non-redundant baseline case, have been implemented up to the layout level. Placement, routing, and post layout verification have been performed. As depicted in Fig. 15, the planar topology has been partitioned in two parts (dotted line), between the central routers. The topology under test (Fig. 15) includes six processors and six memories, placed on two layers. Vertical communication is achieved through the two central switches which act as a gateway for 3-D NoC traffic. The reconfigurable crossbars have been inserted between the TSV pads and the switch. For a 32-bit link, the NoC protocol uses 38 bits, where the remaining 6 bits belong to flow control signaling and mesochronous handling (i.e., the clock and reset signals which are forwarded along with the data).

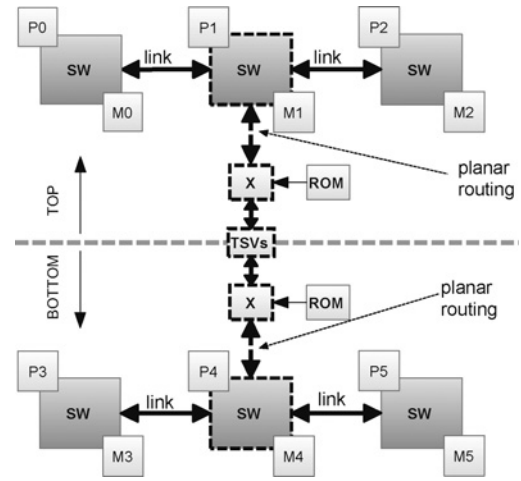


Fig. 15. 3-D NoC topology. Dashed boxes indicate the resources involved in the TSV test process.

The nature of the reference NoC switches, namely, the flow control, has influenced the adopted testing solution. During testing, a portion of the hardware works in scan mode (inject) and the other in capture mode; the flow control has to be explicitly managed to avoid the formation of communication stalls. Four scan chain groups have been inserted, driven by a simple finite state machine (FSM), accomplishing high efficiency and reliability. The overhead of this approach is mainly due to the crossbar logic around the via bundles, to the OTP memory and to the small FSM. The scan chain cost is not taken into account since, as mentioned before, the design must be testable anyway, and this contribution is present as well on planar ICs.

Several experiments have been conducted, especially with the dynamic routing technique, in order to evaluate how many extra pads and area may be needed for implementation, and in order to explore the tradeoffs between yield and cost. We implemented six different configurations, respectively, with 2, 3, 4, 7, 11, and 38 extra pads. It is worth noting that, in each unidirectional link of 38 signals, spare pads are separately needed for incoming (mostly, flow control) and outgoing (mostly, data) wires, hence, the need for at least 2 spares. The outgoing group typically features many more wires than the incoming one (35 versus 3 in our example), so the correction performance is maximized with an asymmetric assignment of spares to the two groups. With only 2 extra pads, one spare must be assigned to the 35 outgoing signals, while the 3 incoming wires share the second spare. With 4 spares, the optimal arrangement is to assign 3 to the outgoing bundle, and the fourth to the incoming bundle. In the extreme case of 38 spares, each TSV has a backup.

Fig. 14 illustrates the yield improvement in case of 2, 3, 4, 7, 11, and 38 extra pads and based on experimental data, assuming a fixed defect frequency of 9.75 defects per million opportunities (HRI TSV process [4]). We emulated 100 K TSV links with and without redundancy. Without post-manufacturing processing, the system is unable to recover from damaged vias, and tolerates only small misalignments, thus exhibiting a yield of only 68%. When dynamic routing

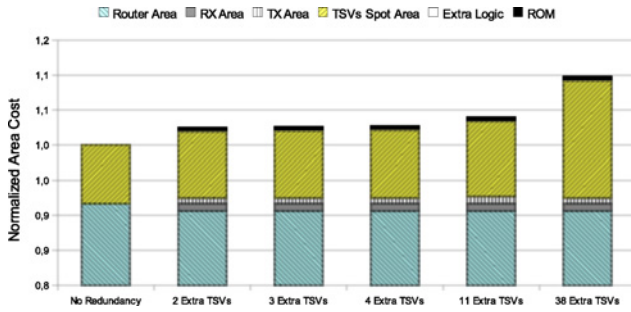


Fig. 16. Normalized area cost in case of no redundancy versus dynamic routing with 2, 3, 4, 7, 11, and 38 extra pads. The proposed contribution shows only 1.6% area overhead for 2 extra pads (second bar), 2.1% for 4 extra pads, and 10.5% for full redundancy (38 extra pads).

TABLE II

AREA OVERHEAD (μm^2) OF DYNAMIC ROUTING WITH 4 EXTRA PADS IN 130 NM AND 65 NM

		130 nm					
		Switch Area	TSV Area	Routing Hardware	Total Area	Link Area Increase	Total Area Increase
No redundancy		54 000	4864	—	58 864	—	—
With redundancy		53 000	5376	1713	60 090	20.9%	2.1%
		65 nm					
		Switch Area	TSV Area	Routing Hardware	Total Area	Link Area Increase	Total Area Increase
No redundancy		13 500	4864	—	18 364	—	—
With redundancy		13 250	5376	430	19 056	14.2%	3.8%

redundancy is adopted, the recovery algorithm shows excellent results, especially with 27 extra pads. Increasing the number of extra pads further brings minimal yield benefits, and the increase in cost of TSV obstructions, TSV crossbar and the OTP memory may be unjustified. With only 4 extra pads per 3-D link, yield increases from 68% to 98%.

Concerning the silicon cost, Fig. 16 shows the normalized area cost in case of different degrees of redundancy applied to a single 3-D link. As the number of extra TSVs increases, the TSV and routing logic areas grow linearly. The area overhead, with reference to the baseline non-redundant 3-D switch, is 1.6% in case of 2 extra pads, 2.1% for 4 extra pads, and 10.5% for 38 extra pads. As a stand-alone component, disregarding the rest of the switch, the redundant link with 4 extra pads, as shown in Table II, has a modestly 20.9% larger area than a non-redundant link. Timing performance along the fault-tolerant link, as outlined at the beginning of Section IV-B, are degraded because two multiplexer are inserted. This involves few additional gates in the path, leading an overall latency increase up to 90 ps in 65 nm. This degradation, despite not negligible, is in the order of 5–7% (depending on the target frequency), thus, can be tolerated.

To evaluate the impact of the dynamic routing solution using advanced technology nodes, we performed an experiment using a 65 nm technology library. As Table II shows, by scaling the technology, the dynamic routing area scales as well. However, we conservatively assume that TSVs may not shrink, as the TSV process may be independent from the technology node. In this pessimistic assumption, the overhead of our solution with 4 extra pads is still just 3.8%. In Table II,

switches with redundant links show smaller area with respect to the baseline scenario (non-redundant) because the TSV drivers are moved in the re-routing stage (multiplexers).

VI. CONCLUSION AND FUTURE WORK

In this paper, we have studied the performance and system-level impact of TSV as one of the possible ways to implement high-density vertical NoC links. We have showed that, even when accounting for the coupling effects in dense vertical link bundles, the parasitics associated with TSVs are one order of magnitude smaller than traditional horizontal wires, making 3-D NoCs a very promising approach. We have showed how to design NoC switches with vertical ports. We have showed that our flow is capable of generating layouts of 3-D NoCs which are fully compatible with accurate post-layout timing, area, and power analysis. We have proposed a novel fault-tolerant dynamic routing approach, based on post-manufacturing study and reconfiguration of the electrical resources, leveraging a small amount of on-chip spares. The scheme proved capable of yields up to 98% with a minimum silicon cost of just 20.9% per TSV link in 130 nm. This cost was further projected to decrease to just 14.2% in the newest 65 nm technologies.

Research on 3-D NoCs is just now beginning, and much work remains to be done. Among the areas requiring more attention, we plan on focusing on serialization–deserialization (SER–DES) of the data at the 3-D interface (reducing the number of TSVs, therefore, increasing the overall yield and decreasing the via area), and how to over-clock efficiently the data transfers of the serialized bus in order to overcome the throughput penalty of the SER–DES. Future work also may revolve around timing faults, which are an often underestimated source of failures.

ACKNOWLEDGMENT

This paper is the result of a collaboration between the University of Bologna, Bologna, Italy, Toshiba Corporation, Tokyo, Japan, and the Center for Integrated Systems, Stanford Univ., Stanford, CA.

REFERENCES

- [1] *International Technology Roadmap for Semiconductors*. (2009) [Online]. Available: <http://public.irst.net>
- [2] R. S. Patti, "Three-dimensional integrated circuits and the future of system-on-chip designs," *Proc. IEEE*, vol. 94, no. 6, pp. 1214–1224, Jun. 2006.
- [3] A. Topol, D. L. Tulipe, L. Shi, D. Frank, K. Bernstein, S. Steen, A. Kumar, G. Singco, A. Young, K. Guarini, and M. Jeong, "Three-dimensional integrated circuits," *IBM J. Res. Develop.*, vol. 50, nos. 4–5, pp. 491–506, Jul.–Sep. 2006.
- [4] N. Miyakawa, E. Hashimoto, T. Maebashi, N. Nakamura, Y. Sacho, S. Nakayama, and S. Toyoda, "Multilayer stacking technology using wafer-to-wafer stacked method," *ACM J. Emerging Technol. Comput. Syst.*, vol. 4, no. 4, Oct. 2008.
- [5] B. Swinnen, W. Ruythooren, P. D. M. L. Bogaerts, L. Carbonell, K. D. Munck, B. Eyckens, S. Stoukatch, D. S. Tezcan, D. Sabuncuoglu, Z. Tokei, J. Vaes, J. V. Aelst, and E. Beyne, "3-D integration by Cu–Cu thermo-compression bonding of extremely thinned bulk-Si die containing 10 μm pitch through-Si vias," in *Proc. IEDM*, Jan. 2006, pp. 1–4.

- [6] A. Topol, D. L. Tulipe, L. Shi, S. Alam, D. Frank, S. Steen, J. Vichiconti, D. Posillico, M. Cobb, S. Medd, J. Patel, S. Goma, D. DiMilia, M. Robson, E. Duch, M. Farinelli, C. Wang, R. Conti, D. Canaperi, L. Deligianni, A. Kumar, K. Kwietniak, C. D'Emic, J. Ott, A. Young, K. Guarini, and M. Jeong, "Enabling SOI based assembly technology for three dimensional integrated circuits," in *Proc. IEEE IEDM Tech. Dig.*, 2005, pp. 352–355.
- [7] W. J. Dally and B. Towles, "Route packets, not wires: On-chip interconnection networks," in *Proc. 38th Des. Automat. Conf.*, Jun. 2001, pp. 684–689.
- [8] L. Benini and G. De Micheli, "Networks on chip: A new SoC paradigm," *IEEE Comput.*, vol. 35, no. 1, pp. 70–78, Jan. 2002.
- [9] J. Kim, C. Nicopoulos, D. Park, R. Das, Y. Xie, N. Vijaykrishnan, M. S. Yousif, and C. R. Das, "A novel dimensionally-decomposed router for on-chip communication in 3-D architectures," in *Proc. 34th ISCA*, 2007, pp. 138–149.
- [10] S. Fujita, K. Nomura, K. Abe, and T. Lee, "3-D on-chip networking technology based on post-silicon devices for future networks-on-chip," in *Proc. Nano-Netw. Workshops*, Sep. 2006, pp. 1–5.
- [11] V. Pavdilis and E. Friedman, "3-D topologies for networks-on-chip," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 15, no. 10, pp. 1081–1090, Oct. 2007.
- [12] V. Pavdilis, I. Savidis, and E. Friedman, "Clock distribution networks for 3-D integrated circuits," in *Proc. CICC*, 2008, pp. 651–654.
- [13] C. Seiculescu, S. Murali, L. Benini, and G. D. Micheli, "Sunfloor 3-D: A tool for networks on chip topology synthesis for 3-D system on chips," in *Proc. DATE Conf.*, Apr. 2009, pp. 9–14.
- [14] S. Murali, C. Seiculescu, L. Benini, and G. D. Micheli, "Synthesis of networks on chips for 3-D systems on chips," in *Proc. ASP-DAC*, 2009, pp. 242–247.
- [15] M. Rencher and F. Schellenberg, "Why interconnect and lithography modeling impacts yield," in *What's Yield Got to Do with IC*, vol. 1. 2002 [Online]. Available: http://i.cmpnet.com/eedesign/2003/inside_eedesign7.pdf
- [16] M. Abramovici, C. Stroud, C. Hamilton, S. Wijesuriya, and V. Verma, "Using roving stars for on-line testing and diagnosis of FPGAs in fault-tolerant applications," in *Proc. IEEE Int. Test Conf.*, Sep. 1999, pp. 973–982.
- [17] M. Pirretti, G. M. Link, R. R. Brooks, N. Vijaykrishnan, M. Kandemir, and M. J. Irwin, "Fault tolerant algorithms for network-on-chip interconnect," in *Proc. ISVLSI*, vol. 26. 2004, pp. 46–51.
- [18] U. Kang, H.-J. Chung, S. Heo, S.-H. Ahn, H. Lee, S.-H. Cha, J. Ahn, D. Kwon, J. H. Kim, J.-W. Lee, H.-S. Joo, W.-S. Kim, H.-K. Kim, E.-M. Lee, S.-R. Kim, K.-H. Ma, D.-H. Jang, N.-S. Kim, M.-S. Choi, S.-J. Oh, J.-B. Lee, T.-K. Jung, J.-H. Yoo, and C. Kim, "8 Gb 3-D DDR3 DRAM using through-silicon-via technology," in *Proc. IEEE Int. Solid-State Circuit Conf.*, Feb. 2009, pp. 130–132.
- [19] I. Loi, F. Angiolini, and L. Benini, "Supporting vertical links for 3-D networks-on-chip: Toward an automated design and analysis flow," in *Proc. Nano-Netw. Conf.*, 2007, pp. 23–27.
- [20] I. Loi, F. Angiolini, and L. Benini, "Developing mesochronous synchronizer to enable 3-D NoCs," in *Proc. DATE Conf.*, 2008, pp. 1414–1419.
- [21] Ansoft. (2007). *Q3D Extractor* [Online]. Available: <http://www.ansoft.com/>
- [22] F. Angiolini, P. Meloni, S. Carta, L. Benini, and L. Raffo, "Contrasting a NoC and a traditional interconnect fabric with layout awareness," in *Proc. Des., Automat. Test Eur. Conf. Exhibit.*, 2006, pp. 124–129.
- [23] S. Murali, P. Meloni, F. Angiolini, D. Atienza, S. Carta, L. Benini, G. D. Micheli, and L. Raffo, "Designing message-dependent deadlock free networks on chips for application-specific systems on chips," in *Proc. VLSI-SoC*, 2006, pp. 158–163.
- [24] R. Patti. (2007, Sep.). Impact of wafer-level 3-D stacking on the yield of ICs. *Future Fab Int.* [Online]. Available: http://www.futurefab.com/documents.asp?d_id=4415
- [25] K.-N. Chen, A. Fan, and R. Reif, "Microstructure examination of copper wafer bonding," *J. Electron. Mater.*, vol. 30, no. 4, pp. 331–335, Apr. 2001.
- [26] K. N. Chen, A. Fan, and R. Reif, "Interfacial morphologies and possible mechanisms of copper wafer bonding," *J. Mater. Sci.*, vol. 37, no. 16, pp. 3441–3446, Aug. 2002.
- [27] N. Miura, D. Mizoguchi, M. Inoue, T. Sakurai, and T. Kuroda, "A 195-GB/s 1.2-W inductive inter-chip wireless superconnect with transmit power control scheme for 3-D-stacked system in a package," *IEEE J. Solid State Circuits*, vol. 41, no. 1, pp. 23–34, Jan. 2006.
- [28] K.-N. Chen, C. Tan, A. Fan, and R. Reif, "Morphology and bond strength of copper wafer bonding," *Electrochem. Solid-State Lett.*, vol. 7, no. 1, pp. 14–16, 2004.
- [29] A. Papanikolaou, M. Miranda, H. Wang, F. Catthoor, M. Satyakiran, P. Marchal, B. Kaczer, C. Bruynseraede, and Z. Tokei, "Reliability issues in deep deep sub-micron technologies: Time-dependent variability and its impact on embedded system design," in *Proc. Int. Conf. Very Large Scale Integr. IFIP*, Oct. 2006, pp. 342–347.
- [30] I. Loi, F. Angiolini, and L. Benini, "Synthesis of low-overhead configurable source routing tables for network interfaces," in *Proc. Des., Automat. Test Eur. Conf. Exhibit.*, 2009, pp. 262–267.



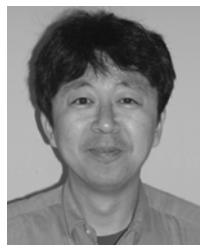
Igor Loi received the B.S. degree in electrical engineering from the University of Cagliari, Cagliari, Italy, in 2005, and the Ph.D. degree from the Department of Electronics and Computer Science, University of Bologna, Bologna, Italy, in 2010.

He is currently in a post-doctoral position with the Department of Electronic Engineering, University of Bologna. His current research interests include 3-D integrated circuit technologies and networks-on-chip.



Federico Angiolini received the M.S. degree (summa cum laude) in electrical engineering from the University of Bologna, Bologna, Italy, in 2003, and the Ph.D. degree from the Department of Electronics and Computer Science, University of Bologna, in 2008.

He is currently the Vice President of Engineering with the INOCs, Lausanne VD, Switzerland. His current research interests include memory hierarchies, multiprocessor-embedded systems, and networks-on-chip.



Shinobu Fujita (M'03) received the B.S., M.S., and Ph.D. degrees in applied physics from the University of Tokyo, Tokyo, Japan, in 1984, 1986, and 1989, respectively.

He was with Toshiba Corporation, Tokyo, Japan, in 1989, where he initially focused on the development of high-speed compound semiconductor devices. He has been engaged in nanotechnology and systems based on silicon nanoscale devices since 1994. Currently, he is with Toshiba America Research, Inc., San Jose, CA, as a Visiting Researcher collaborating with the Center for Integrated Systems, Stanford University, Stanford, CA.



Subhasish Mitra (SM'06) is currently with the Robust Systems Group, Department of Electrical Engineering, Stanford University, Stanford, CA, and with the Department of Computer Science, Stanford University. He was a Principal Engineer with Intel Corporation, West Babylon, NY. His current research interests include robust system design, very large-scale integrated (VLSI) design, computer-aided design (CAD), validation and testing, and emerging nanotechnologies. His X-Compact technique for test compression has been used in more than 50 Intel products, and has influenced major CAD tools. His instruction footprint recording and analysis technology for post-silicon validation, created jointly with his students, was characterized as a breakthrough in the communications of the Association for Computing Machinery (ACM). His work on the first demonstration of imperfection-immune carbon nanotube VLSI circuits, jointly with his students and collaborators, was selected by the National Science Foundation, Arlington, VA, as a Research Highlight to the U.S. Congress, and was highlighted as a significant breakthrough by the Semiconductor Research Corporation, Durham, NC, and the *MIT Technology Review*, Cambridge, MA.

His major honors include the Presidential Early Career Award for Scientists and Engineers, the highest U.S. honor for Early-Career Outstanding Scientists and Engineers, the ACM SIGDA Outstanding New Faculty Award, the IEEE CAS/CEDA Pederson Award for the IEEE Transactions on CAD Best Paper, the IEEE/ACM Design Automation Conference Best Paper Award, the Terman Fellowship, and the Intel Achievement Award, Intel's highest corporate honor. He also serves as an invited member on DARPA's Information Science and Technology Board.



Luca Benini (S'94–M'97–SM'04) received the B.S. degree (summa cum laude) in electrical engineering from the University of Bologna, Bologna, Italy, in 1991, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1994 and 1997, respectively.

He is currently an Associate Professor with the Department of Electronics and Computer Science, University of Bologna, and is a Visiting Professor with Stanford University. His current research interests include aspects of the computer-aided design

of digital circuits, with special emphasis on low-power applications, and the design of portable systems. On these topics, he has published more than 200 papers in international conferences and journals. He is the co-author of three books.

Dr. Benini is a member of the technical program committees for several technical conferences, including the Design Automation Conference, the International Symposium on Low Power Design, and the International Symposium on Hardware–Software Codesign. He has been the Program Chair of the Design, Automation, and Test in Europe Conference since 2005.