Fortune: A New Fault-Tolerance TSV Configuration in Router-Based Redundancy Structure

Qi Xu, *Member, IEEE*, Hao Geng, *Graduate Student Member, IEEE*, Tianming Ni, *Member, IEEE*, Song Chen, *Member, IEEE*, Bei Yu, *Member, IEEE*, Yi Kang, *Member, IEEE*, and Xiaoqing Wen, *Fellow, IEEE*

***Abstract*—In three-dimensional integrated circuits (3D-ICs), through silicon via (TSV) is a critical technique in providing vertical connections. However, yield is one of the key obstacles to adopt the TSV-based 3D-ICs technology in the industry. Various fault-tolerance structures using redundant TSVs to repair faulty functional TSVs have been proposed in the literature for yield and reliability enhancement. However, the TSV repair paths under delay constraint cannot always be generated due to the lack of appropriate repair algorithms. In this article, we propose an effective TSV repair strategy for the router-based TSV redundancy architecture, taking into account the delay overhead. First, we prove that the router-based fault-tolerance structure configuration (RFSC) with the delay constraint is equivalent to the length-bounded multicommodity flow (LBMCF) problem. Then, an integer linear programming (ILP) formulation with acceptable scalability is presented to solve the LBMCF problem. The experimental results demonstrate that, compared with state-of-the-art fault-tolerance designs, the proposed ILP model can provide higher yield and lower delay overhead.**

***Index Terms*—Fault tolerance, three-dimensional integrated circuit (3D-IC), through silicon via (TSV) repair, yield enhancement.**

I. Introduction

As DEVICE feature sizes continue to rapidly decrease, the interconnecting delay is becoming a bottleneck limiting IC performance. The three-dimensional integrated circuits (3D-ICs) technology involving vertically stacking multiple dies connected by through silicon vias (TSVs) provides a promising way to alleviate the interconnecting problem and achieves a significant reduction in chip area, wire length, and interconnect power [1]. Study indicates that the average wire length of a 3D-IC varies according to the square root of the number of layers. Moreover, 3D-ICs have the potential for heterogeneous integration, which is essential for the More than Moore (MtM) technology. 3D integration has already seen several commercial applications in the form of 3D memory but there are still existing significant open problems in both academia and industry [2]. In this work, we will focus on the TSV reliability problem.

In general, there are two types of yield losses in 3D-ICs: 1) defects in stacked dies and 2) defects that occurred during the assembling process [3]. For the former case, prebond testing is critical to avoid the stacking of defective dies. For the latter case, adding redundant TSVs (referred to **r-TSVs**) to repair faulty functional TSVs (termed as **f-TSVs**) is an effective method for increasing yield.

One fundamental problem in the TSV fault-tolerance design is the fault-tolerance structure configuration, that is, how to generate the TSV replacing paths in the fault-tolerance structure. Hsieh and Hwang [4] proposed a regular TSV replacing chain structure. Jiang *et al.* [5] proposed a router-based TSV redundancy structure to repair clustered TSV faults. Furthermore, Xu *et al.* [6] presented a switch-based TSV fault-tolerance structure during floorplanning. Besides, to effectively repair clustered TSV faults, Lee *et al.* [7] developed a group-based TSV architecture. Recently, with considering the 1-hop delay constraint, a cellular- and a honeycomb-based fault-tolerance structure were proposed by Wang *et al.* [8] and Ni *et al.* [9], respectively.

From previous works, we notice that although some TSV fault-tolerance structures can effectively handle clustered TSV faults, there is no appropriate fault-tolerance structure configuration algorithm to generate the TSV replacing paths. To tackle the above issue, in this article, we propose an effective TSV repair strategy for the router-based TSV redundancy architecture, taking into account the delay overhead. Our main technical contributions are listed as follows.

1) We prove that the router-based fault-tolerance structure configuration (RFSC) with the delay constraint is equivalent to the length-bounded multicommodity flow (LBMCF) problem.
2) We present an integer linear programming (ILP) formulation with acceptable scalability to solve the LBMCF problem.

Manuscript received 23 April 2021; revised 26 July 2021 and 5 October 2021; accepted 18 November 2021. Date of publication 7 December 2021; date of current version 20 September 2022. This work was supported in part by the National Key R&D Program of China under Grant 2019YFB2204800; in part by the National Natural Science Foundation of China (NSFC) under Grant 61904047, Grant 62174001, Grant 61904001, Grant 61874102, Grant 61732020, Grant 61931008; in part by the Guangdong Provincial Key Laboratory under Grant 2020B121201001; and in part by the Key Research and Development projects in Anhui Province under Grant 202104b11020032. This article was recommended by Associate Editor P. Gupta. *(Qi Xu and Hao Geng contributed equally to this work.) (Corresponding authors: Song Chen; Bei Yu.)*

Qi Xu, Song Chen, and Yi Kang are with the School of Microelectronics, University of Science and Technology of China, Hefei 230026, China (e-mail: xuqi@ustc.edu.cn; songch@ustc.edu.cn).

Hao Geng and Bei Yu are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, SAR (e-mail: hgeng@cse.cuhk.edu.hk).

Tianming Ni is with the College of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China (e-mail: timmyni126@126.com).

Xiaoqing Wen is with the Department of Computer Science and Networks, Kyushu Institute of Technology, Fukuoka 804-8550, Japan.

Digital Object Identifier 10.1109/TCAD.2021.3133484

3) The experimental results demonstrate that, compared with state-of-the-art fault-tolerance designs, the proposed ILP model can provide higher yield and lower delay overhead.

The remainder of this article is organized as follows. Section II presents the preliminary and gives the problem formulation. The proposed ILP model is described in Section III. Section IV lists experimental results, followed by the conclusion in Section V.

II. Preliminaries

A. TSV Defect Model

TSV defect-distribution models are divided into two types in the previous literature, namely, uniform defect distribution and clustered defect distribution. For the uniform defect-distribution model, each TSV fails independently. This model is valid for certain random defects, such as void formation and lamination due to thermal-induced stress [10]. However, many types of TSV defects appear during the imperfect bonding process. Besides, the surface roughness, the cleanliness of dies, and the height variation of the TSVs also influence the bonding process. Consequently, the presence of a TSV fault increases the probability of more defective TSVs in close vicinity. This is called the clustered defect-distribution model [11].

In our work, we take the clustered defect distribution into account. A compound Poisson distribution is adopted to model the clustering effect. In this model, the number of existing defects (regarded as cluster centers) follows a Poisson distribution, and the distribution of defect density is described by a Gamma function. If we assume N_c cluster centers, the defect probability of TSV_i, P_i, is expressed as [11]

$$P_i = p \cdot \left(1 + \sum_{j=1}^{N_c} \left(\frac{1}{d_{ij}}\right)^{\alpha}\right) \tag{1}$$

where p is the single TSV failure rate, while d_{ij} is the distance between TSV_i and the jth cluster center. Meanwhile, α denotes the clustering coefficient, which indicates the clustering extent (i.e., larger α implies higher clustering).

B. Router-Based TSV Fault-Tolerance Structure

By adding the multiplexers and demultiplexers (i.e., Muxes and Dmuxes) and carefully designing the reconfigurable TSV replacing paths, a TSV fault-tolerance structure can be generated, where the r-TSVs can be used to transfer signals in the presence of faulty f-TSVs. To repair the clustered faults, Jiang *et al.* [5] developed a router-based TSV redundancy architecture. As shown in Fig. 1, the f-TSVs are regularly distributed in a uniform 4×4 grid structure, and the r-TSVs are placed on the right and bottom boundaries of the structure. Thus, the signals are transferred from two directions (from left to right or from top to bottom). Besides, each f-TSV is connected to a router, which contains six ports and three 3-to-1 multiplexers. The signal and its corresponding f-TSV occupy two ports in the router, while the remaining four ports are linked to other routers in four different directions. Therefore,

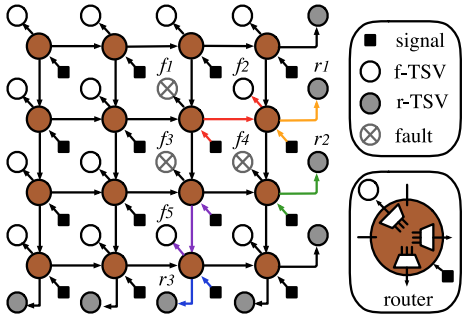

Fig. 1. Router-based TSV redundancy architecture, where the repair paths for the faulty f-TSVs are shown in colored lines. $f_1 : \{f_1 \rightarrow f_2\}$ (red lines), $f_2 : \{f_2 \rightarrow r_1\}$ (orange lines), $f_3 : \{f_3 \rightarrow f_5\}$ (purple lines), $f_5 : \{f_5 \rightarrow r_3\}$ (blue lines), and $f_4 : \{f_4 \rightarrow r_2\}$ (green lines), while other fault-free f-TSVs transfer their corresponding signals.

the signal port and two linking ports (left and top) can connect to the TSV port and the remaining linking ports (right and bottom) through the multiplexers. To minimize the delay overhead, a heuristic search algorithm is developed to generate replacing paths for each faulty TSV. As a result, the signals related to faulty TSVs can be reallocated to fault-free TSVs that are distant rather than to the neighboring TSVs, and thus the router-based redundancy structure achieves a high yield.

Actually, the signal path allocation in the TSV fault-tolerance structure will not introduce routing congestion. The reason is that in the router-based fault-tolerance architecture, TSVs and control circuits (i.e., Muxes and Dmuxes) are preplaced at uniform grid structures. By utilizing a congestion estimator at the placement stage, the infeasible r-TSV locations can be screened [12], and thus the routing problems can be tackled before the fault-tolerance structure configuration. Moreover, although each router in the router-based structure requires six control signals to configure the TSV replacing paths, the control signals can be stored in a 6-bit register in each router instead of being generated from a global controller. We handle all registers in the structure by the bus sharing mechanism. As a result, the control signals of routers will not cause severe routing congestions. Based on the SMIC 40 nm library, the area of a 6-bit register is 22.2264 μm^2. In addition, if other fault-tolerant structures [6]–[9] also use this way to eliminate routing congestion, corresponding types of registers will be introduced. So we need to take the incurred area overhead of registers into account.

C. Problem Formulation

In this work, we adopt the router-based TSV fault-tolerance structure. We denote the size of architecture as $R \times C$, where R and C indicate the rows and columns of the uniform f-TSV grid. Since the r-TSVs are placed on the right and bottom boundaries of the grid, the number of r-TSVs is $R + C$. The redundancy ratio of the router-based TSV redundancy architecture equals to $R + C : R \times C$.

The existing TSV testing technique can be directly adopted to achieve the faulty TSV distributions [13]. Since the fault-free TSV is now connected to the previous reallocated signal, its corresponding signal needs to be reallocated as well. This

TABLE I
NOTATIONS USED IN ILP

k, n_f, n_r	number of signals, fault-free f-TSVs, and fault-free r-TSVs in structure
V_s	set of signals, $V_s = \{s_i \mid i = 1, \cdots, k\}$
V_p	set of router ports, $V_p = \{p_i \mid i = 1, \cdots, 3k\}$
V_f	set of fault-free f-TSVs, $V_f = \{f_i \mid i = 1, \cdots, n_f\}$
V_r	set of fault-free r-TSVs, $V_r = \{r_i \mid i = 1, \cdots, n_r\}$
V_{rp}	set of r-TSV ports, $V_{rp} = \{rp_i \mid i = 1, \cdots, n_r\}$
V_t	set of all fault-free TSVs, $V_t = V_f \cup V_r$
e	end vertex
E_{sp}	set of edges from signal $s_i \in V_s$ to router port $p_j \in V_p$
E_{pp}	set of edges from router port $p_i \in V_p$ to other port $p_j \in V_p$
E_{pf}	set of edges from router port $p_i \in V_p$ to f-TSV $f_j \in V_f$
E_{pr}	set of edges from router port $p_i \in V_p$ to r-TSV port $rp_j \in V_{rp}$
E_{rr}	set of edges from r-TSV port $rp_i \in V_{rp}$ to r-TSV $r_j \in V_r$
E_{te}	set of edges from TSV $t_j \in V_t$ to end vertex e
l_{s_i}	maximum length constraint value for each signal $s_i \in V_s$
l_b	balanced length constraint value
$x_{uv}^{(s_i,e)}$	binary variable; if a unit flow from $s_i \in V_s$ to e goes through edge (u, v), then $x_{uv}^{(s_i,e)} = 1$, otherwise $x_{uv}^{(s_i,e)} = 0$

process continues until a redundant TSV on the boundaries is used. As a result, a TSV replacing path for each f-TSV is generated. A fault-tolerance configuration solution is feasible only if the replacing paths for each f-TSV are vertex disjoint.

Definition 1 (Vertex-Disjoint Path): A set of paths are vertex disjoint if no two of them have vertices in common. Obviously, they also have no intersecting edges.

To guarantee the timing correctness of the circuit after repairing, the additional delay overhead incurred by signal reallocation should be taken into account.

Definition 2 (Delay): The delay overhead is modeled as the length or hop between the original f-TSV and its reallocated fault-free TSV.

Based on the above terminologies, we define the problem of the RFSC as follows.

Problem 1 (RFSC): Given a router-based TSV redundancy architecture with faulty TSV distributions, we configure the TSV fault-tolerance structure, where each f-TSV finds a vertex-disjoint TSV replacing path to a fault-free TSV under the delay constraint.

Theorem 1: The RFSC problem is equivalent to the LBMCF problem.

The detailed proof is provided in the Appendix. As proved in [14], the LBMCF problem is $\mathcal{NP}$-complete, which can be handled by ILP.

III. INTEGER LINEAR PROGRAMMING FORMULATION

In this section, we discuss how the LBMCF problem can be formulated as an ILP problem. For convenience, some notations used in this section are listed in Table I.

Given a router-based TSV fault-tolerance structure with faulty TSV distribution, we construct a directed graph $G(V, E)$ consisting of TSV replaceable relations. The vertex set V contains five portions, $V = V_s \cup V_p \cup V_t \cup V_{rp} \cup e$. Note that each router can be simplified into three port vertices in the directed graph G. One of which is connected to the corresponding f-TSV, and the other two are linked to the port

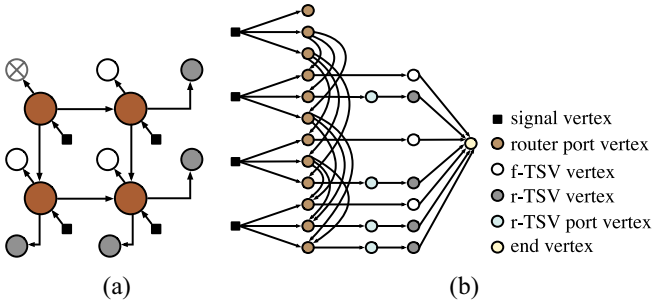

Fig. 2. (a) Example of a router-based structure with one faulty f-TSV. (b) Corresponding directed graph G.

vertices of the right or bottom neighboring router or r-TSV ports. The edge set E is composed of six sets of edges, $E = E_{sp} \cup E_{pp} \cup E_{pf} \cup E_{pr} \cup E_{rr} \cup E_{te}$. For convenience, we use E_1 to represent the union of the five edge sets, E_{sp}, E_{pp}, E_{pf}, E_{pr}, and E_{rr}. For instance, given the router-based structure with three fault-free f-TSVs and four r-TSVs in Fig. 2(a), the corresponding directed graph is shown in Fig. 2(b).

Based on the above notations, the LBMCF problem can be formulated as the following integer linear programming (2):

$$\min \sum_{s_i \in V_s} \sum_{(u,v) \in E_1} x_{uv}^{(s_i,e)} \tag{2a}$$

$$\text{s.t.} \sum_{(u,v) \in E} x_{uv}^{(s_i,e)} - \sum_{(v,u) \in E} x_{vu}^{(s_i,e)} = \begin{cases} 1, & \text{if } u = s_i \\ 0, & \text{if } u \in V - \{s_i, e\}\ \forall s_i \in V_s \\ -1, & \text{if } u = e \end{cases} \tag{2b}$$

$$\sum_{s_i \in V_s} \sum_{(u,v) \in E_1} x_{uv}^{(s_i,e)} \le 1 \quad \forall v \in V_p \cup V_{rp} \cup V_t \tag{2c}$$

$$\sum_{(u,v) \in E_1} x_{uv}^{(s_i,e)} \le 2l_{s_i} + 1 \quad \forall s_i \in V_s \tag{2d}$$

$$\sum_{(u,v) \in E_1} x_{uv}^{(s_i,e)} - \sum_{(u,v) \in E_1} x_{uv}^{(s_j,e)} \le l_b \quad \forall s_i, s_j \in V_s, i > j \tag{2e}$$

$$\sum_{(u,v) \in E_1} x_{uv}^{(s_i,e)} - \sum_{(u,v) \in E_1} x_{uv}^{(s_j,e)} \ge -l_b \quad \forall s_i, s_j \in V_s, i > j \tag{2f}$$

$$x_{uv}^{(s_i,e)} \in \{0, 1\} \quad \forall (u, v) \in E, s_i \in V_s. \tag{2g}$$

The objective function (2a) is to minimize the total delay overhead incurred by signal reallocation. The number of binary variables $x_{uv}^{(s_i,e)}$ is $k \times |E|$, where k is the number of signals, while $|E|$ is the number of edges in directed graph G. The constraint (2b) defines a unit flow from $s_i \in V_s$ to end vertex e, which corresponds to a replacing path from s_i to e. The number of this set of constraints is k. The constraint (2c) ensures that the TSV replacing path for each f-TSV is vertex disjoint. For example, considering the structure in 2(a), we have to search for four vertex-disjoint paths which start from each f-TSV and end at a fault-free TSV. The number of this set of constraints is $(3k + n_f + 2n_r)$. Constraint (2d) restricts the maximum length of the TSV replacing path for

each f-TSV. In addition, the constraints (2e) and (2f) consider the balanced delay overhead among any pair of signals. As a result, the balanced delay of the TSV replacing path for each f-TSV can meet the design requirement.

For instance, in Fig. 1, there are three clustered f-TSV faults in the structure (i.e., f_1, f_3, and f_4). With considering the delay constraint, a fault-tolerance structure is generated as shown in colored lines. That is, the vertex-disjoint paths for the faulty f-TSVs are $f_1 : \{f_1 \rightarrow f_2\}$, $f_2 : \{f_2 \rightarrow r_1\}$, $f_3 : \{f_3 \rightarrow f_5\}$, $f_5 : \{f_5 \rightarrow r_3\}$, and $f_4 : \{f_4 \rightarrow r_2\}$, while other fault-free f-TSVs transfer their corresponding signals.

IV. Experimental Results

A. Simulation Setup

The proposed algorithms have been implemented in C++ language and tested on a 12-core 2.0 GHz Linux server with 64-GB RAM. To verify the effectiveness of our algorithms, we employ the same industrial benchmark set as applied in [5], which consists of three sample chips and the number of TSVs is 1024 (`Chip1`), 16384 (`Chip2`), and 131072 (`Chip3`), respectively. The TSV cell size including the keep-out zone is 10 μm $\times$ 10 μm [5]. The distance between TSVs is set to 50 μm [7]. According to the RC delay model described in [15], the wire delay is assumed to be $\sim$ 5 Ns per 10 mm. In 3-D ICs, the wire delay overhead should be considered in both upper and lower dies. The area and delay of used multiplexers and demultiplexers are evaluated by Synopsys Design Compiler based on the SMIC 40 nm technology. GUROBI [16] is used as the ILP solver with a time limit of 1800 s. For the sake of simplicity, in this work, we assume that all signals have a unified maximum length constraint value. The balanced length constraint value is chosen as 1. A Monte-Carlo simulation is exploited to evaluate the performance of the proposed TSV repair algorithms on different benchmarks. When all f-TSVs find a length-bounded vertex-disjoint replacing path, we denote it as a repairable case. On the other hand, an irreparable case is one in which the fault-tolerance solution cannot be achieved within the time limit. Therefore, the TSV yield can be calculated by

$$\text{Yield } (\%) = \frac{\#\text{repairable case}}{\#\text{repairable case} + \#\text{irreparable case}}. \quad (3)$$

B. Impact of Parameters on Performance

In the first experiment, we analyze the impact of the TSV failure rate p on the performance. The experiment is performed on all benchmarks. Three options for the TSV failure rate p are considered: 0.01, 0.005, and 0.001. The size of the TSV redundancy architecture is set to 32 $\times$ 32, while the maximum length constraint value is set to 1 in the experiment. Fig. 3 illustrates the statistic results averaged over 1000 independent experiments. As shown in Fig. 3(a), with the increase of the TSV failure rate, the chip yield drops. The reason is that the number of faulty TSVs is proportional to the TSV failure rate [11]. Since a TSV redundancy architecture can be repaired only if each f-TSV finds a length-bounded vertex-disjoint TSV replacing path to a fault-free TSV, the more faulty TSVs, the less likely the TSV replacing paths can be found. We also observe

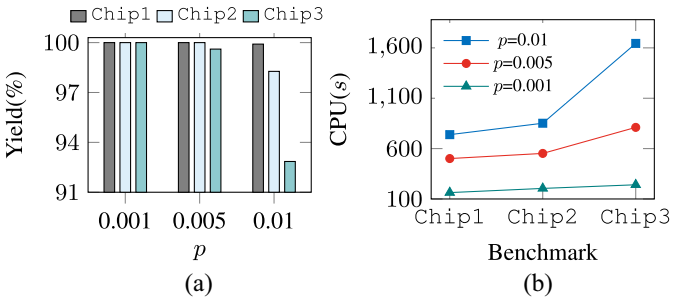

Fig. 3. (a) Effect of the TSV failure rate p on yield. (b) Scalability of the proposed ILP model. (TSV numbers are `Chip1`: 1024, `Chip2`: 16384, and `Chip3`: 131072).

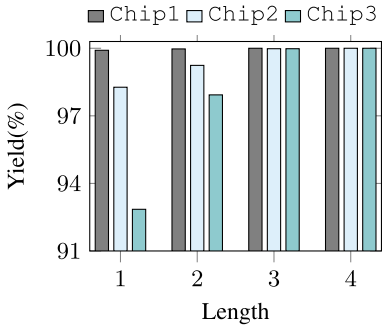

Fig. 4. Impact of the maximum length constraint on the yield.

from the figure that the more f-TSVs in chip, the smaller yield. For example, under a higher failure rate ($p = 0.01$), the yield of the ILP model on "`Chip1`" can still reach 99.91%, while the yields on "`Chip2`" and "`Chip3`" drop nearly 1.64% and 7.06%. That is, because given the same TSV failure rate, the more f-TSVs, the more faulty TSVs occurred in chip. Thus with f-TSV numbers increasing, the chip yield drops. Besides, as depicted in Fig. 3(b), the runtime of the ILP model increases with the size of the benchmark. But, our ILP formulation has good scalability, i.e., with TSV numbers increase by 128 times from "`Chip1`" to "`Chip3`," the runtime only rises by 6.8 times at the TSV failure rate of 0.01. Even for the largest chip, the ILP model can still obtain fault-tolerance solutions in an acceptable time.

In the second experiment, we investigate the impact of the maximum length constraint on the performance of the proposed ILP model. The experiment is performed on all benchmarks, and we choose four different maximum length constraints, 1, 2, 3, and 4. The TSV redundancy architecture with the size of 32 $\times$ 32 is employed in the experiment. For each benchmark, we execute the proposed ILP method 1000 times independently. Fig. 4 shows results of the yield with respect to the changing of the length constraint value. With an increment of the length constraint value, yield ascends. In addition, from the experimental data, we notice that the balanced length constraint will not affect the result under 1-hop length constraint. However, when the maximum length constraint value is set to greater than 1, the yield of the ILP model with the balanced length constraint is slightly lower than that without the balanced length constraint. The main reason is that the balanced length constraint imposes a strong restriction

TABLE II
COMPARISON WITH STATE-OF-THE-ARTS (AREA UNIT: $\times 10^4$ μm^2, DELAY UNIT: *ps*, AND YIELD UNIT: %)

Benchmark	α	Router_LBHS [5]			Group_MF [7]			Cellular_MCMF [8]			Honeycomb_MCMF [9]			Ours		
		Area	Delay	Yield	Area	Delay	Yield	Area	Delay	Yield	Area	Delay	Yield	Area	Delay	Yield
Chip1	1	4.000	190	83.71	4.650	400	89.43	4.282	190	94.21	5.545	270	97.21	4.000	190	99.91
	2			83.53			89.35			93.83			96.97			99.82
	3			83.27			89.26			93.47			96.59			99.68
Chip2	1	63.997	190	76.65	74.275	400	84.39	68.510	190	92.34	88.714	270	95.38	63.997	190	98.27
	2			76.37			84.27			91.86			95.12			98.12
	3			75.92			84.16			91.39			94.83			98.01
Chip3	1	511.972	190	64.25	594.134	400	78.28	548.076	190	89.52	709.035	270	90.17	511.972	190	92.85
	2			63.87			78.12			89.03			89.93			92.74
	3			63.43			77.98			88.58			89.84			92.59
Average	–	193.323	190	74.56	224.353	400	83.92	206.956	190	91.58	267.765	270	94.00	**193.323**	**190**	**96.89**

on the TSV replacing path. Therefore, in other experiments, the maximum length constraint value is chosen as 1 with the balanced length constraint value 1 as well.

C. Comparison With Previous Works

Here, we compare five state-of-the-art works [5], [7]–[9] to demonstrate the superiority of the proposed algorithm. The compound Poisson distribution is adopted to model the clustered defect distribution. The TSV failure rate p is set to 0.01. In order to see the impact of the clustering coefficient α on performance, three different α values (1, 2, and 3) are chosen.

We first compare the proposed ILP model with "Router_LBHS" [5], where a length-bounded heuristic search algorithm is iteratively performed to repair the faulty TSVs in the router-based TSV fault-tolerance architecture. The size of the router-based architecture is set to 32 $\times$ 32, and the maximum length constraint value is chosen as 1. We execute the two methods independently for each benchmark, and list the average statistic results in Table II. Column "Area" is the extra area overhead incurred by the fault-tolerance structure, which contains the used r-TSVs and extra Muxes and registers. Column "Delay" lists the delay overhead of the fault-tolerance solution, whist column "Yield" denotes the TSV yield. As shown in Table II, compared with Router_LBHS, the proposed ILP can improve the yield by 22.33% on average. That is, because in the heuristic search algorithm [5], the paths generated for preceding signals are no longer available for the latter signals, and thus the solution space is reduced. Consequently, the yield of the heuristic search algorithm is dropped. The delay overhead of the TSV redundancy architecture can be calculated with the longest TSV repair path. For the router-based TSV fault-tolerance structure, the length of the longest TSV repair path is 1. Therefore, the total wire delay is 50 ps. Moreover, two 3-to-1 Muxes are needed to transfer signals on the repair path. Based on the SMIC 40 nm library, the delay of a 3-to-1 Mux is 70 ps. As a result, the total delay overhead is 190 ps. Since Router_LBHS and the proposed ILP method are based on the same router structure, the area and delay overheads are the same.

We further compare the proposed ILP model with "Group_MF" [7], "Cellular_MCMF" [8], and "Honeycomb_MCMF" [9]. In Group_MF, a group-based redundancy structure is presented and a max-flow-based algorithm is performed to repair the faulty TSVs. We set 12 f-TSVs and 4 r-TSVs in the group-based architecture, and the 12 f-TSVs are divided into four groups. In Cellular_MCMF, a cellular structure is proposed, and a min-cost max-flow-based algorithm is performed to repair the faulty TSVs. We set the size of the cellular architecture to 8$\times$8 with 10 r-TSVs. Besides, in Honeycomb_MCMF, a min-cost max-flow-based heuristic method is developed to generate the TSV repair paths in the honeycomb-based redundancy structure. A two-layer honeycomb-based structure is considered in the experiment. Thus, 25 f-TSVs and six r-TSVs are included in each honeycomb-based structure.

The area and yield results of each structure are shown in Table II. Compared with "Group_MF," "Cellular_MCMF," and "Honeycomb_MCMF," the ILP model can improve chip yield by 12.97%, 5.31%, and 2.89%, respectively. That is, because the ILP model can fully explore the solution space of the fault-tolerance structure. As a result, the vertex-disjoint replacing path for each f-TSV are constructed optimally. It can also be seen that the area cost of the proposed architecture is lower than that of the cellular structure [8], the group-based [7], and honeycomb-based structures [9]. The reason is that the redundancy ratios of the cellular and the group-based architectures are larger than that of the router-based architecture (1:16), the required r-TSV numbers are significantly increased. Besides, in the two-layer honeycomb-based design, two 2-to-1 Muxes, 19 3-to-1 Muxes, four 4-to-1 Muxes, 15 1-to-4 Dmuxes, seven 1-to-5 Dmuxes, six 6-to-1 Muxes, and three 1-to-6 Dmuxes are needed. As a result, the area costs of the cellular, the group-based and the honeycomb-based structures are high. Since the signals can only be reallocated to TSVs within one hop in the cellular [8] and honeycomb-based structures [9], the wire delay overhead is 50 ps. The longest repair path in the cellular-based structure contains one 1-to-4 Dmux (60 ps) and one 4-to-1 Mux (80 ps), whilst one 1-to-6 Dmux (100 ps) and one 6-to-1 Mux (120 ps) exist on the longest path in the honeycomb-based structure. Thus, the total delay overheads of the cellular and honeycomb-based architectures are 190 and 270 ps, respectively. For the group-based structure, the additional delay overhead is 400 ps, which is the sum of 200 ps wire delay and 120 ps 6-to-1 Mux delay of the upper die and 80 ps 4-to-1 Mux delay of lower die.

D. Discussion

TSVs are usually bundled together in a 3D-IC design, and the router-based TSV fault-tolerance structure can be directly adopted to such designs if we treat each TSV bundle as a grid [5]. Therefore, for very large TSV clusters, we first group TSVs into TSV bundles (e.g., 32×32 grid), and then perform the proposed ILP model for each grid to generate the TSV replacing paths. Besides, in the future, we can further develop a Lagrangian relaxation-based heuristic method to speed up the fault-tolerance structure configuration process.

V. Conclusion

In this article, we have focused on the TSV repair strategy for the router-based TSV redundancy architecture under delay overhead. An ILP-based model has been proposed to repair clustered TSV faults, with minimizing both the delay overhead and the hardware cost. The experimental results demonstrate that, compared with state-of-the-art fault-tolerance designs, the proposed method can provide a higher yield and lower delay overhead. As continuing growth of technology node, 3D-IC turns out to be a promising solution to further scaling, we believe this article will stimulate more research on yield-aware 3D-IC design.

Appendix
Proof of Theorem 1

Proof of $\Rightarrow$ Part of Theorem: In the RFSC problem, we have to find a unit flow between each signal vertex s_i and sink vertex e, with minimization of the total delay overhead incurred by signal reallocation. A feasible solution should satisfy the following constraints.

1) Independent flow conservation constraints for each pair of s_i and e.
2) The sum of flows passing through each TSV and router port vertex cannot exceed 1 (vertex disjoint).
3) The length of all TSV replacing flows (paths) is bounded by length constraint value l_{s_i}.

Therefore, the vertex-disjoint length-bounded s_i-e flows problem can be formulated as the LBMCF problem.

Proof of $\Leftarrow$ Part of Theorem: In the case of multiple commodities, we are given k source-sink vertex pairs $(s_1, e), \ldots, (s_k, e)$ called commodities, where s_i and e are the signal vertex and sink vertex in directed graph $G(V, E)$ as defined in Section III. A multicommodity flow m_f is a set of s_i-e-flows m_{f_i}, for $i = 1, \ldots, k$. In addition to the flow conservation constraint, the multicommodity flow is feasible only if each vertex $v \in V - e$ also holds the capacity constraint, that is, $\sum_{i=1}^{k} m_{f_i}(v) \leq c_v$. Thus, the LBMCF is a multicommodity flow such that each commodity flow f_i is a length bounded s_i-e-flow. If we set the vertex capacity constraint c_v to 1, the feasible multicommodity flow solution can be mapped into the feasible TSV fault-tolerance solution on $G(V, E)$.

Acknowledgment

The authors would like to thank the Information Science Laboratory Center of USTC for hardware and software services.

References

[1] S. J. Souri, K. Banerjee, A. Mehrotra, and K. C. Saraswat, "Multiple Si layer ICs: Motivation, performance analysis, and design implications," in *Proc. DAC*, Los Angeles, CA, USA, 2000, pp. 213–220.
[2] T. Lu, C. Serafy, Z. Yang, S. K. Samal, S K. Lim, and A. Srivastava, "TSV-based 3-D ICs: Design methods and tools," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 10, pp. 1593–1619, Oct. 2017.
[3] Q. Xu, L. Jiang, H. Li, and B. Eklow, "Yield enhancement for 3D-stacked ICs: Recent advances and challenges," in *Proc. ASPDAC*, Sydney, NSW, Australia, Feb. 2012, pp. 731–737.
[4] A.-C. Hsieh and T. T. Hwang, "TSV redundancy: Architecture and design issues in 3-D IC," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 4, pp. 711–722, Apr. 2012.
[5] L. Jiang, Q. Xu, and B. Eklow, "On effective through-silicon via repair for 3-D-stacked ICs," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 4, pp. 559–571, Apr. 2013.
[6] Q. Xu, S. Chen, X. Xu, and B. Yu, "Clustered fault tolerance TSV planning for 3-D integrated circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 8, pp. 1287–1300, Aug. 2017.
[7] I. Lee, M. Cheong, and S. Kang, "Highly reliable redundant TSV architecture for clustered faults," *IEEE Trans. Rel.*, vol. 68, no. 1, pp. 237–247, Mar. 2019.
[8] Q. Wang, Z. Liu, J. Jiang, N. Jing, and W. Sheng, "A new cellular-based redundant TSV structure for clustered faults," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 27, no. 2, pp. 458–467, Feb. 2019.
[9] T. Ni *et al.*, "LCHR-TSV: Novel low cost and highly repairable honeycomb-based TSV redundancy architecture for clustered faults," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 10, pp. 2938–2951, Oct. 2020.
[10] A. P. Karmarkar, X. Xu, and V. Moroz, "Performanace and reliability analysis of 3D-integration structures employing through silicon via (TSV)," in *Proc. IRPS*, Montreal, QC, Canada, Apr. 2009, pp. 682–687.
[11] S. Wang, K. Chakrabarty, and M. B. Tahoori, "Defect clustering-aware spare-TSV allocation in 3-D ICs for yield enhancement," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 38, no. 10, pp. 1928–1941, Oct. 2019.
[12] C. Li, M. Xie, C.-K. Koh, J. Cong, and P. H. Madden, "Routability-driven placement and white space allocation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 26, no. 5, pp. 858–871, May 2007.
[13] B. Noia and K. Chakrabarty, *Design-for-Test and Test Optimization Techniques for TSV-based 3D Stacked ICs*. Cham, Switzerland: Springer, 2014.
[14] G. Baier *et al.*, "Length-bounded cuts and flows," *ACM Trans. Algorithms*, vol. 7, no. 1, p. 4, 2010.
[15] H. Kitada *et al.*, "The influence of the size effect of copper interconnects on RC delay variability beyond 45nm technology," in *Proc. IITC*, Burlingame, CA, USA, 2007, pp. 10–12.
[16] Gurobi Optimization Inc. *Gurobi Optimizer Reference Manual*. (2014). [Online]. Available: http://www.gurobi.com