Cost-Effective Integration of Three-Dimensional (3D) ICs Emphasizing Testing Cost Analysis

Yibo Chen, Dimin Niu and Yuan Xie
Department of Computer Science and Engineering
The Pennsylvania State University, University Park, PA 16803
Email: {yxc236, dun118, yuanxie}@cse.psu.edu

Krishnendu Chakrabarty
Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708
Email: krish@ee.duke.edu

***Abstract*—Three-dimensional (3D) ICs promise to overcome barriers in interconnect scaling by leveraging fast, dense inter-die vias, thereby offering benefits of improved performance, higher memory bandwidth, smaller form factors, and heterogeneous integration. However, when deciding to adopt this emerging technology as a mainstream design approach, designers must consider the cost of 3D integration. IC testing is a key factor that affects the final product cost, and it could be a major portion of the total IC cost. In 3D IC design, various testing strategies and different integration methods could affect the final product cost dramatically, and the interaction with other cost factors could result in various trade-offs. This paper develops a comprehensive and parameterized testing cost model for 3D IC integration, and analyzes the trade-offs associated with testing strategies and testing circuit overheads. With the proposed testing cost model, designers can explore the most cost-effective integration and testing strategies for 3D IC chips.**

I. Introduction

With the continuous technology scaling, interconnect has emerged as the dominant source of circuit delay and power consumption. Three-dimensional (3D) ICs have been proposed as a promising means to mitigate the interconnect-related problems [1], [2]. In a 3D IC, multiple device layers are stacked together with direct vertical interconnects through substrates. 3D ICs offer a number of advantages over traditional two-dimensional (2D) design, such as (1) Higher packing density and smaller footprint; (2) Shorter global interconnect due to the short length of through-silicon vias (TSVs) and the flexibility of vertical routing; (3) Higher performance because of the interconnect wire length reduction and bandwidth improvement; (4) Lower interconnect power consumption due to reduced interconnect capacitance; (5) Support of heterogenous integration: each single die can have different technologies. Consequently, 3D IC designs have drawn a lot of attention from both academia and industry in recent years.

The majority of the existing research efforts on 3D ICs is focused on how to take advantage of the performance, power, smaller form-factor, and heterogeneous integration benefits offered by 3D integration, with emphasis on manufacturing, design tools, and novel architectures. However, all the advantages of 3D ICs ultimately have to be translated into cost savings when a design strategy has to be decided [3]. Consequently, studies on 3D IC cost analysis have been carried out recently to help designers make early design decisions from cost perspective [3]–[5]. For example, Mercier *et al.* [4] demonstrated a yield model of 3D chip stacks using wire bonding. Ferri *et al.* [5] discussed the parametric yield management for 3D ICs. Dong *et al.* presented a system-level 3D-IC cost analysis framework, which provided an early-stage design estimation and cost prediction for key quantities in 3D ICs. Weerasekera *et al.* [6] discussed realistic metrics for performance and cost tradeoff analysis in both 2D and 3D heterogeneous systems. However, the design-for-testability (DFT) issues related to 3D IC design have not been addressed by most of the existing research work [3], [7]–[14]

This work is supported in part by NSF 0903432, 1017277, 1017391, and SRC grants.

DFT plays a key role in reducing the test efforts and improving test quality. Since it also incurs certain overhead for the design, the economics of incorporating DFT has been explored [15]. In 3D IC design, various testing strategies and different integration methods could also affect the testing cost dramatically, and the interaction with other cost factors could results in different trade-offs. For example, wafer-to-wafer stacking could save the Known-Good-Die (KGD) testing cost and improve the throughput, compared to die-to-wafer stacking; however, the overall yield could be much lower, and therefore the total IC cost could be higher. It is therefore very important to estimate the test economics with a testing cost model, and integrate it into a cost driven 3D IC design flow to strike a balance between cost and other benefits (such as performance/power/area) [16].

To understand the impact of testing strategies on 3D IC design and explore the tradeoffs between testing strategies and integration options, in this paper, we propose a testing cost analysis methodology with a break-down of testing cost for different integration strategies. We address several unique issues raised by 3D integration, including the overall yield calculation, the impact of new defect types introduced by 3D integration, and the interactions with fabrication cost. With the testing cost model we perform a set of trade-off analysis and show that the design choices could be different after testing cost is taken into account.

II. Preliminaries and Motivations

This section provides some preliminaries on 3D IC stacking, and presents a simple case discussion which motivates the work in this paper.

A. Preliminaries on 3D IC Stacking

3D ICs can provide advanced system integration by stacking different dies into a single chip. The layers could be connected with wire bonding, TSV, microbump, or even inductive/capacitive contact [1]. TSV-based 3D technology provides the possibility for high density interconnection between the layers by the mean creating vertical connections through the silicon substrate, and consequently is the focus of the majority of current research on 3D integration technologies. *Die-to-wafer (D2W)* and *wafer-to-wafer (W2W)* are two different ways to bond multiple dies in TSV-based 3D integration. W2W bonding stacks all layers of wafer before a single 3D chip is sliced and packaged, while D2W bonding mount different layer of dies onto the base wafer sequentially.

B. Motivations

Partitioning a large 2D chip to be two smaller dies and stacking them together may incur extra manufacturing cost due to the additional steps for 3D integration with extra mask cost, and extra KGD testing cost (for D2W or D2D stacking). For example, in Figure 1, there are two different approaches to integrate the Boundary Scan Latches (BSL) to improve the BIST: 1) For each die there is a separate BSL and I/O circuitry so that we can have a full diagnostic capability

978-1-4244-8192-7/10/$26.00 ©2010 IEEE

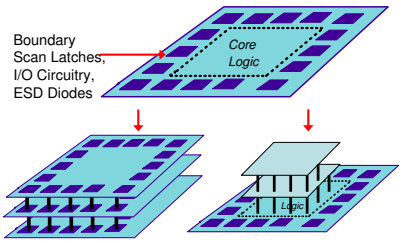

Fig. 1. Two scenarios for testability and cost trade-off analysis.

for completed assembly with independent unit BIST; 2) Only the bottom layer has BSL and I/O circuitry, and therefore results in die area saving with lower cost. However, the diagnostic capability is potentially compromised because the diagnosis signals would have to go through TSVs to reach the top two layers. This calls for a comprehensive testing cost model that captures the impacts from all aspects. Of course, such cost models are timing sensitive, when a technology becomes mature, it is more accessible and cheaper, and the testing becomes easier. Therefore, we plan to develop parametric cost models, so that users can adjust various process parameters to reflect more accurate estimation.

III. Testing Cost Modeling For 3D ICs

In this section, we first present a basic cost model that captures the economics of testing in 3D IC design. We then discuss the impact of new issues raised by 3D stacking, including the use of different stacking strategies, new defect types during stacking, as well as the interaction between testing costs and fabrication costs of 3D ICs.

A. Basic Cost Model of 3D Testing

Our testing cost model is adapted from the 2D DFT model in [15]. For the sake of brevity, we mainly address the uniqueness due to 3D integration. The testing cost of 3D ICs consists of four main components-preparation cost, execution cost, test-related silicon cost, and imperfect-test-quality cost. Hence, we compute testing cost C_{test} (per good chip) as [15]:

$$C_{test} = C_{prep} + C_{exec} + C_{silicon} + C_{quality} \quad (1)$$

C_{prep} captures fixed costs of test generation, tester program creation, and any design effort for incorporating test-related features (all nonrecurring costs, including software systems). Note that in 3D ICs, this may include the preparation cost for pre-bond testing of TSVs.

C_{exec} consists of costs of test-related hardware such as probe cards and cost incurred by tester use. Tester-use cost depends on factors including tester setup time, test execution time (as a function of die area and yield), and capital cost of the tester (as a function of die area and number of I/O pins) and capital equipment depreciation rate. This should also be compensated if pre-bond TSV testing is applied.

$C_{silicon}$ is the cost required to incorporate DFT features. We model this cost as a function of wafer size, die size, yield, and the extra area required by DFT.

$C_{quality}$ is the penalty due to the faulty chips that escape the test. This component exhibits increasing importance for 3D integration with the relatively low yield and high cost of 3D ICs.

Test preparation cost for 3D ICs. For the cost of test preparation for 3D ICs, we start from the model presented in [15] for 2D cases and augment it for 3D IC integration. Generally, the preparation cost contains the test-pattern generation cost (C_{test_gen}), tester-program preparation cost (C_{test_prog}), and the additional design cost for (C_{DFT_design}).

In 3D IC designs, different bonding technologies might be used to assemble die or wafer layers together into a single chip. For example, if W2W bonding is used, no pre-bonding test will be performed. In this case, all of the test-related preparations, including the DFT design, test-pattern generation and tester-program preparation, should target at the whole chip, instead of a single die. Therefore, the testing cost only applies to the final test after wafer bonding and sorting. The preparation cost for a W2W-bonded 3D chip can be calculated as:

$$C_{prep}^{overall} = \frac{1}{Y_{W2W}}(C_{DFT\_design}^{overall} + C_{test\_gen}^{overall} + C_{test\_prog}^{overall}), \quad (2)$$

where Y_{W2W} denotes the final yield of the chip after bonding and will be discussed in the next sub-section.

If D2W bonding is used, each layer should be tested before bonding, so the preparation cost will be changed to:

$$C_{prep}^{overall} = \frac{1}{Y_{D2W}}(Y_N C_{prep}(N) + \sum_{i=1}^{N-1} C_{prep}(i)). \quad (3)$$

where

$$C_{prep}(i) = \frac{1}{Y_i}(C_{DFT\_design}(i) + C_{test\_gen}(i) + C_{test\_prog}(i)). \quad (4)$$

where $1 \ldots N$ denote the N layers in the chip, Y_i is the yield for each die before bonding, and $C_{DFT\_design}(i)$, $C_{test\_gen}(i)$, $C_{test\_prog}(i)$ are calculated with regard to the area of each die using the 2D model in [15]. Note that layer N is the bottom layer in the stack that is closest to IO pads, and it is usually tested only after bonding.

Test execution cost for 3D ICs. The test execution cost is the total cost per chip of hardware consumption and the testing devices expense. In 2D IC testing, all individual ICs on a wafer should be tested before they are packaged. The wafer testing is performed by using a hardware called wafer prober or probe card. Therefore, the hardware cost for testing is mainly determined by the cost of probe cards [15]. We use N_{probe} to represent the number of wafers a probe card can test and use C_{probe} to denote the price of a probe card. Thus the hardware cost is calculated as:

$$C_{hw} = C_{probe}\lceil V/N_{probe}\rceil/V. \quad (5)$$

In plain terms, a chip with n layer should be tested n times for D2W stacking while only 1 time for W2W stacking. Thus, the hardware cost for 3D ICs can be summarized as

$$\begin{cases} C_{W2W,hw} = C_{probe}\lceil V_{chip}/N_{probe}\rceil/V_{chip}, \\ C_{D2W,hw} = C_{probe}\lceil NV_{chip}/N_{probe}\rceil/V_{chip}. \end{cases} \quad (6)$$

Besides the cost of probe cards, the cost of testing devices should also be considered in the execution cost. The cost of the testing device can be calculated as:

$$C_{device} = R_{test}T_{test} + R_{idle}T_{idle}. \quad (7)$$

where R_{test} is the cost rate of the testing devices when they are operating, containing the cost of operator, electricity, device depreciation, etc. R_{idle} denotes the cost rate of the testing device when they are inactive, which is much smaller than R_{test}. T_{test} and T_{idel} are the testing time and the idle time of the testing device.

We denote the relationship between the testing rate and idle rate as: $R_{test} = \eta R_{idle}$, where η is the proportion between the testing cost and idle cost and satisfies: $\eta \gg 1$. Also, the testing time and idle time satisfy : $T_{test} + T_{idle} = T_{year}$, where T_{year} is the total second number of one year. Also, the testing rate and idle rate are in

proportion to the testing device's price. A testing device's unit price is :

$$Q = K_{capital} \cdot K_{pin} \cdot \sqrt{A_{die}}. \quad (8)$$

where $K_{capital}$ is the average device price per pin and K_{pin} is the average density of pins in a die.

Therefore, the cost rate of the testing devices for a 3D chip can be summarized as

$$R_{D2W,test,i} = \beta_{depr} K_{capital} \cdot K_{pin} \cdot \sqrt{A_i + A_{DFT,i}}\ , \quad (9)$$

$$R_{W2W,test,i} = \begin{cases} \beta_{depr} K_{capital} K_{pin} \sqrt{A_i} & if\ i < N \\ \beta_{depr} K_{capital} K_{pin} \sqrt{A_N + A_{DFT}} & if\ i = N \end{cases} \quad (10)$$

where β_{depr} is the annual depreciation rate of the testing device.

The testing time can be simply modeled as:

$$T_{test} = T_{setup} + K_{ave\_t} A_{die}^2. \quad (11)$$

where K_{ave_time} is a constant multiplier that relates testing time to the die area. Thus, the testing time for a 3D chip is:

$$T_{D2W,test,i} = T_{setup} + K_{ave\_t}(A_i + A_{DFT,i})^2. \quad (12)$$

$$T_{W2W,test,i} = \begin{cases} T_{setup} + K_{ave\_t} {A_i}^2. & if\ i < N \\ T_{setup} + K_{ave\_t}(A_N + A_{DFT})^2. & if\ i = N \end{cases} \quad (13)$$

Testing circuit overhead. Testing circuits usually occupy a certain fraction of the total silicon area. In 3D integration with KGD testing (D2W or W2W stacking), since test structures have to be implemented in each layer, the total area overhead of the testing circuits will be larger than that of the 2D case, and the relative area overhead of the testing circuits increases as the number of total layers goes up. For a given die with DFT structures, we denote the area ratio between DFT circuits and the whole die by α_{DFT}. For W2W stacking, we have $A_{DFT,total} = \alpha_{DFT} \cdot A_{2D}$, while for D2W stacking:

$$A_{DFT,i} = \alpha_{DFT} \cdot A_i = \alpha_{DFT} \cdot \frac{A_{2D}}{N} \quad (14)$$

$$\begin{aligned} A_{DFT,total} &= \alpha_{DFT} \cdot A_{2D} + \sum_{i=1}^{N-1} A_{DFT,i} \\ &= \alpha_{DFT} \cdot A_{2D} \cdot \frac{2N-1}{N} \end{aligned} \quad (15)$$

Therefore, the DFT circuit overhead $A_{DFT,total}$ for D2W stacking is an increasing function of layer count N. The silicon cost for the DFT circuit overhead is given by:

$$C_{silicon,D2W} = \frac{Q_{wafer}}{\pi R_{wafer}^2 \beta_{waf\_die}} \left[\frac{A_{DFT,total}}{Y_{overall}} \right] \quad (16)$$

Imperfect test quality. For simplicity in modeling, we only model escape cost due to imperfect test quality. The test escape rate is a function of fault coverage and the fault occurrence rate (or simply the yield) at a particular stage of test. Williams and Brown [17] give the escape rate, the ratio of the defective ICs that pass testing to all the ICs that pass testing, as

$$E_r = 1 - Y^{1-fc} \quad (17)$$

where f is fault coverage. The same relation stands for W2W stacking of 3D ICs. Similarly, if D2W stacking is used, the error rate will be changed to:

$$E_{r,D2W} = 1 - \prod_{i=1}^{N} Y_i^{1-fc_i} \quad (18)$$

where Y_i and fc_i are the yield and fault coverage for die in layer i, respectively.

The cost of imperfect test quality is then calculated by:

$$C_{quality} = C_{penalty} \cdot E_r \quad (19)$$

where $C_{penalty}$ is the economic penalty for allowing a defective IC to escape testing. Depending on the importance attached to escaping ICs, the penalty might be as much as 10 times of the manufacturing cost of a good IC.

B. The Impact of 3D Stacking Strategies on Chip Yield

3D stacking strategies have significant impacts on the overall chip yield, which in turns affects the average cost per chip. The yield model for a single die has been well investigated [18], [19]. Assuming that the defects are randomly distributed on a wafer, the probability of the number of defect dies in a wafer can be described as a binomial random variable and can be approximated by a Poisson random variable. Therefore, the yield of a die can be simply modeled as:

$$Y_P = e^{-D_0 A_{die}}. \quad (20)$$

where D_0 is the average density of the defect and A_{die} is area of the certain chip.

However, it has been shown that the defects are usually not randomly distributed across the chip, but are always clustered. In this case, the die yield should be higher than the result of Equation (20). Therefore, a compound Poisson distribution is proposed by Murphy [18], which applies a weighting function to modulate the original Poisson distribution. A widely used weighting function is the Gamma function [4]. The Gamma function based yield model is:

$$Y_G = \left(1 + \frac{D_0 A_{die}}{\alpha}\right)^{-\alpha}. \quad (21)$$

The parameter α is defined as $\alpha = (\mu_D/\sigma_D)^2$ and depends upon the complexity of the manufacturing process. D_0 is the silicon defect density parameter. Obviously, the yield decreases exponentially with the increase of die area. Thus, the smaller dies in 3D IC design may result in higher yield than that of a larger 2D die and therefore reduce the cost.

To calculate the overall yield of a 3D chip, the following aspects should be taken into consideration:

- For 3D ICs, both the die yield and the stacking yield should be consider at the same time. The defects exist in each die will certainly affect the overall chip yield of after stacking. At the mean time, an unsuccessful stacking operation can also cause a chip to fail.
- In 3D ICs, dies at different layer are not necessary to have the same size. As shown in Equation (21), different die sizes will result in different yields.
- Stacking yield should be defined carefully. There are two sources of the stacking failure: (1) Failure results from imperfect TSVs; (2) Failure results from imperfect stacking operation. However, since the TSV-related failures have already been captured when calculating the stacking yield, they should not be counted as defects of a single die.

- Different stacking methods, such D2W stacking and W2W stacking, require different yield models.

Yield model for W2W stacking. During W2W stacking, each die cannot be tested until the stacking is finished. The DFT circuitry is located at the bottom layer of the chip, which is closer to the IOs. Based on these considerations, the yield for W2W stacking can be modeled as:

$$Y_{overall,W2W} = [1 + \frac{D_0}{\alpha}(A_N + A_{DFT})]^{-\alpha} \cdot \prod_{i=1}^{N-1} Y_{S,W2W,i}(1 + \frac{D_0 A_i}{\alpha})^{-\alpha} \quad (22)$$

where $Y_{S,W2W,i}$ denotes the stacking yield between layer i and layer $i-1$, and A_{DFT} is the area of DFT circuitry on the bottom layer.

Yield model for D2W stacking. For D2W stacking design, a higher yield can be achieved by introducing KGD test. The dies in layer 1 to layer $N-1$ are tested separately before the stacking operation. Thus, during the stacking step, all these dies can be considered as "good dies". Thus the chip yield will be:

$$Y_{overall,D2W} = [1 + \frac{D_0}{\alpha}(A_N + A_{DFT,N})]^{-\alpha} \cdot \prod_{i=1}^{N-1} Y_{S,D2W,i} \quad (23)$$

where $Y_{S,D2W,i}$ is the D2W stacking yield factor, which will be discussed in the next subsection.

C. New Defect Types Due to 3D Stacking

Because 3D integration incurs additional processing steps, such as TSV forming and bonding, new defect mechanisms (unique to 3D integration) must be addressed as part of a test strategy. First, in TSV-based 3D ICs, TSVs under manufacturing suffer from conductor open defects and dielectric short defects, thus the TSV failure rate is unwillingly high [10], [13]. Moreover, the thinning, alignment, and stacking of the wafers add extra steps to the manufacturing process. During bonding, any foreign particle caught between the wafers can lead to peeling, as well as delamination, which dramatically reduces bonding quality and yield. In addition, to maintain good conductivity and minimize resistance, the interconnect TSVs and micropads between wafers must be precisely aligned.

Due to the new defect types brought in by 3D stacking, the stacking yield factor in Equations (22)-(23) can be modeled as:

$$\begin{aligned} Y_S &= Y_{bonding} \cdot Y_{TSV} \\ &= Y_{bonding} \cdot (1 - f_{TSV})^{N_{TSV}} \end{aligned} \quad (24)$$

where $Y_{bonding}$ captures the yield loss of the chip due to faults in the bonding processes, f_{TSV} is the TSV failure rate, N_{TSV} is the total number of TSVs, and Y_{TSV} describes the rate of loss due to failed TSVs.

In current TSV process technology, f_{TSV} varies from 50 *ppm* to 5%. A simple calculation shows that, given a design with 200 TSVs and a typical TSV failure rate of 0.1%, the yield loss due to failed TSV will be as much as 20%, which is barely acceptable. For this reason, designers have proposed several techniques, including pre-bond TSV testing [13] and redundant TSV insertion [20] to mitigate the impact of high TSV failure rate.

Pre-bond TSV testing. The current interconnection test proposed for 3D IC is done with two or more dies in a stack, which is good only for TSVs after bonding [13]. Since the yield of TSVs has a dramatic influence on overall chip yield, some pre-bond TSV testing schemes are needed in order to reduce the risk of bonding dies that have irreparable TSV failures. As via-first TSVs have one end that is not only floating but also buried in the wafer substrate before thinning and bonding, double-end probing for TSV resistance measurement becomes incapable. Alternatively, a single-end probing based TSV test method was proposed [13] to diagnose pre-bond TSVs with the parasitic capacitance measurements. The TSV is modeled as a conductor with its parasitic capacitance in a given range. If there is any defect in the TSV or in the insulator surrounding TSV, the measured parasitic capacitance will deviate from the nominal value. In this way, the faulty TSVs can be detected and the failure rate f_{TSV} is greatly reduced. Correspondingly, the area overhead of the testing circuit will be translated to testing cost, as it increases A_{DFT} in Equations (14)-(16). The trade-off between f_{TSV} and the increase of A_{DFT} will then be explored in our 3D testing cost model.

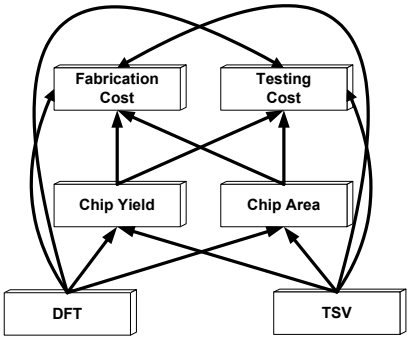

Fig. 2. Dependency graph of the key quantities in 3D IC cost model.

Redundant TSV insertion. Another effective way to reduce the impact of faulty TSVs is to create TSV redundancy. A straightforward doubling redundancy of TSVs can tolerant any single TSV failure. However, it fails if both of the TSVs fail. The degree of redundancy m is defined as how many TSVs are used for a single signal. The TSV yield and the TSV area with redundancy is then given by:

$$Y_{TSV} = \left(1 - (f_{TSV})^m\right)^{N_{TSV}}, \quad (25)$$

$$A_{TSV,redundant} = m \cdot A_{TSV} \quad (26)$$

D. Interaction with Fabrication Cost

The fabrication cost for 3D ICs has been analyzed in [3]. However, DFT-related cost concerns have not been touched yet. According to the discussions in previous sub-sections, incorporating DFT in 3D IC design will have significant impact on chip fabrication and total cost of chips. Figure 2 shows the dependencies between the key quantities in 3D IC cost model considering testing cost. We can see that testing cost is closely related to fabrication cost through the inclusion of TSVs and DFT structures. Both of the fabrication and testing costs have strong correlations with chip area and yield, and the trade-offs between all the quantities in the graph are to be explored by our 3D cost model.

To facilitate the total cost analysis, we first update the total chip area to be:

$$A_{overall} = A_{2D} + A_{DFT} + A_{TSV} * N_{TSV} \quad (27)$$

where A_{TSV} is the chip area occupied by a single TSV. N_{TSV} is the TSV count, which can be estimated from A_{2D} using a methodology similar to the one presented in [3]. With this we calculate A_is in Equations (9)-(16).

The total cost of 3D ICs is then calculated by:

$$C_{total} = C_{fabrication} + C_{test} \quad (28)$$

TABLE I
CONSTANT VALUES USED IN OUR COST MODEL.

Eqn No.	Const Name	Value	Remarks
(21)	D_0	$0.004/mm^2$	The density of point-defects per unit area.
(21)	α	2.0	The yield model parameter
(17)	fc	0.99	The default fault coverage
(24)	$Y_{bonding}$	0.95	The yield loss due to wafer thinning, etc.
(25)	f_{TSV}	10 ppm	The default TSV defect rate.
(27)	A_{TSV}	10 μm^2	The TSV cross-sectional area.

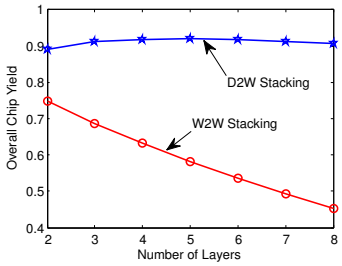

Fig. 3. Overall chip yield with different number of layers and stacking strategies.

We reuse the fabrication cost analysis in [3] for estimation of $C_{fabrication}$ and calculate C_{test} using the model presented in this paper.

IV. COST ANALYSIS FOR 3D ICS

In this section we apply our testing cost model to 3D IC designs and analyze the interactions between IC cost and the key quantities in 3D ICs, including the layer count, the DFT alpha factor, and the TSV faulty rate reduction strategies.

A. Model Parameters

We assume various input values to perform experiments with our testing cost model. Essentially, the model uses three categories of parameters: independent variables, 3D integration factors, and constants.

- Independent Variables: The key independent variables used in our studies are production volume V and die area A_{2D}. We explore a spectrum of values for V from small, ASIC-like volumes (100,000) to large, microprocessor-like volumes (100 million). For A_{2D}, we explore die areas ranging from 0.5 cm^2 to 4 cm^2.
- 3D alpha factors: The impact of 3D integration on testing is difficult to estimate without specific attributes of the IC design. Since very little data on the modeling of specific attributes has been published, we explore a range of values that are reasonable to the best of our knowledge. We argue that the proposed 3D IC test model is parametric-based and the analysis can be applied on any other values of parameters without loss of credibility.
- Constants: The constants used in this paper are from both our industry partners[1] and data in previous literatures. Table I lists the constants used in this paper.

All the cost analysis is performed with IBM $65nm$ Common Platform technology. The TSV parameters are scaled down from the data in MIT Lincoln Lab $130nm$ 3D technology.

B. Trade-off Analysis

The impact of layer count and stacking strategies. Given a design with specified equivalent 2D chip area, we first address the questions in a DFT perspective, that how many layers the design should be partitioned to, and what stacking strategy should be used. As discussed earlier, the layer count has significant impact on the overall chip yield as well as the DFT circuit overhead. To verify this, we perform simulations on a benchmark design with silicon area of 200 mm^2 and production volume of $100,000$. Figure 3 shows the overall chip yield results with different design partitioning and stacking strategies. In W2W stacking, the overall chip yield decreases rapidly as the number of layers increases. Although using more layers lead to smaller die area for each layer, the corresponding yield improvement cannot compensate the yield loss due to stacking untested imperfect dies. On the other hand, the overall chip yield in D2W stacking slightly improves as number of layers increases from 2 to 5, thanks to the KGD testing before stacking. When the layer count further increases, the yield loss on the stacking operations starts to dominate the yield change.

[1]Arbitrary unit (A.U.) is used to present the cost value due to the non-disclosure agreement with our industry partners.

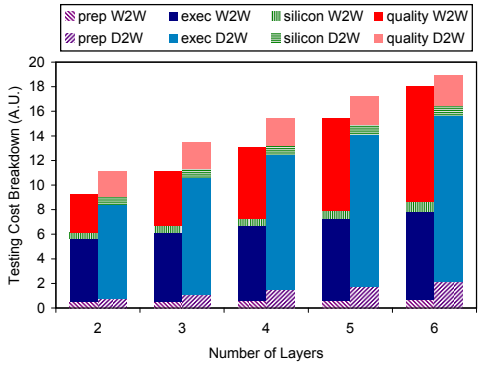

Fig. 4. Testing cost with different number of layers and stacking strategies on IBM 65nm technology.

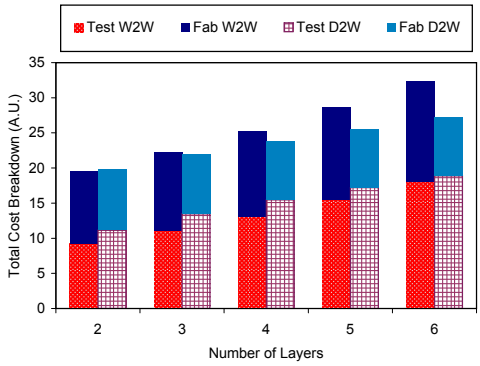

Fig. 5. Chip total cost with different number of layers and stacking strategies on IBM 65nm technology.

In Figure 4, a breakdown of the testing cost for the benchmark design is depicted. We can see that for W2W stacking, the penalty for test escapes ($C_{Quality}$) rises as the overall yield goes down. In D2W stacking, the test execution cost occupies the largest portion as testing is carried out on each die.

Figure 5 summaries the total cost of the benchmark design with different layer counts and stacking strategies, including both the testing cost and the silicon fabrication cost. When only 2 layers are used for the design, W2W stacking has almost the same cost as that of D2W stacking. Both the fabrication and the testing cost increase rapidly with layer count, due to the drop of the overall yield. It suggests that W2W stacking might be more favorable in design with low layer counts, because it can achieve higher production throughput

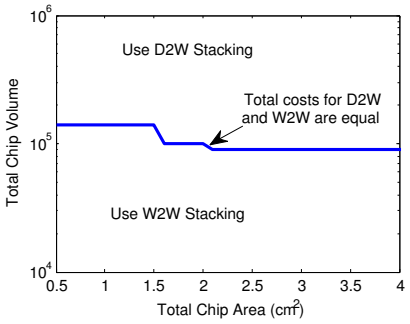

Fig. 6. D2W and W2D domains in the chip-area and chip-volume plane.

without performing KGD testing, and therefore reduce the time-to-market of the design.

V. Design Space Exploration for Cost-Effective 3D ICs

Based on the 3D test cost model described in previous sections, we investigate the impact of different design options on the 3D IC cost, with several given specifications. Since the design space exploration is conducted before the real design, the results help designers make the final decision.

For a design with equivalent 2D die area A and production volume V, the evaluation process is then formulated as:

Minimize:

$$Total\_Cost = f_{(A,V)}(l, m) \tag{29}$$

Subject to:

$$\begin{array}{rcl} p_{(A,V)}(l,m) & \geq & Performance\_Constraint \\ q_{(A,V)}(l,m) & \leq & Quality\_Constraint \\ \ldots & & \ldots \end{array} \tag{30}$$

where l is the number of layers used in the 3D design, and m is the variable indicating the 3D integration methodologies. Several constraints such as the performance constraint on maximum clock frequency, or the quality constraint on the overall error rate, can be imposed to guide the design space exploration.

A case study is then performed for a 3-layer 3D design to find out a better integration strategy between D2W and W2W stacking. The chip area spans from $0.5\ cm^2$ to $4\ cm^2$, while the production volume spans from 10^4 to 10^6. Figure 6 depicts the D2W and W2W domains in the chip-area and chip-volume plane. The curve shows the input parameters with which the total costs of D2W and W2W stacking are almost equal. In the domain above the curve, D2W stacking is more favorable with lower average cost, while under the curve W2W stacking is a better choice. The reason behind is that, in design with small production volumes, the average testing cost occupies a large portion in the total cost for D2W stacking, since D2W stacking comes with large overhead on DFT. As the production volume goes up, the testing cost for D2W stacking is amortized, and at the same time it benefits from the higher overall yield.

From the case study we get a different conclusion on stacking strategy selection from that in [3], where it was claimed that D2W stacking usually has cost advantage over W2W stacking. This means when testing cost is taken into account in the cost analysis, design choice might be different from the one obtained in conventional fabrication-only cost analysis. This again demonstrates the importance of the work in test cost modeling for 3D ICs.

VI. Conclusion

In 3D IC designs, various testing strategies and different integration methods could affect the final product cost dramatically, and their interactions with other cost factors could result in various trade-offs. This paper develops a testing cost model for 3D ICs and analyzes the trade-offs associated with stacking options and test strategies for complex 3D IC chips. The model can help designers make early stages decision on selection of stacking strategies as well as testing methodologies. For future work, if the analysis can be applied during design time with accurate data, a cost-driven 3D IC flow might become possible. The integration of the 3D IC cost model into design flows guides designers to optimize their 3D IC designs and eventually to manufacture low-cost products.

References

[1] W. R. Davis, J. Wilson, and et al., "Demystifying 3D ICs: the pros and cons of going vertical," *IEEE Design & Test of Computers*, vol. 22, no. 6, pp. 498–510, 2005.

[2] Y. Xie, G. H. Loh, and et al., "Design space exploration for 3D architectures," *J. Emerg. Technol. Comput. Syst.*, vol. 2, no. 2, pp. 65–103, 2006.

[3] X. Dong and Y. Xie, "System-level cost analysis and design exploration for three-dimensional integrated circuits (3D ICs)," in *Proc. Asia and South Pacific Design Automation Conf. (ASP-DAC)*, 2009, pp. 234–241.

[4] P. Mercier, S. R. Singh, K. Iniewski, B. Moore, and P. O'Shea, "Yield and cost modeling for 3D chip stack technologies," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, 2006, pp. 357–360.

[5] C. Ferri, S. Reda, and R. I. Bahar, "Strategies for improving the parametric yield and profits of 3D ICs," in *ICCAD*, 2007, pp. 220–226.

[6] R. Weerasekera, D. Pamunuwa, L.-R. Zheng, and H. Tenhunen, "Two-dimensional and three-dimensional integration of heterogeneous electronic systems under cost, performance, and technological constraints," *IEEE TCAD*, vol. 28, no. 8, pp. 1237–1250, 2009.

[7] D. L. Lewis and H.-H. S. Lee, "Testing circuit-partitioned 3D IC designs," in *Proc. IEEE Computer Society Annual Symp. on VLSI (ISVLSI)*, 2009, pp. 139–144.

[8] M.-S. Chan, C.-Y. Wang, and Y.-C. Chen, "An efficient approach to SIP design integration," in *Proc. of International Symposium on Quality of Electronic Design (ISQED)*, 2009, pp. 241–247.

[9] V. F. Pavlidis and E. G. Friedman, "Interconnect-based design methodologies for three-dimensional integrated circuits," *Proceedings of the IEEE*, vol. 97, no. 1, pp. 123–140, 2009.

[10] E. J. Marinissen and Y. Zorian, "Testing 3D chips containing through-silicon vias," in *Proc. Int. Test Conf. (ITC)*, 2009, pp. 1–11.

[11] X. Zhao, D. L. Lewis, H.-H. S. Lee, and S. K. Lim, "Pre-bond testable low-power clock tree design for 3D stacked ics," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, 2009, pp. 184–190.

[12] A. J. Walker, "A manufacturing cost model for 3-D monolithic memory integrated circuits," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, no. 2, pp. 268–275, 2009.

[13] P.-Y. Chen, C.-W. Wu, and D.-M. Kwai, "On-chip TSV testing for 3D IC before bonding using sense amplification," in *Proc. of Asian Test Symp (ATS)*, 2009, pp. 450–455.

[14] M. Tsai, A. Klooz, A. Leonard, J. Appel, and P. Franzon, "Through silicon via (TSV) defect/pinhole self test circuit for 3D-IC," in *Proc. IEEE Int. Conf. 3D System Integration Conference (3D-SIC)*, 2009, pp. 1–8.

[15] P. K. Nag, A. Gattiker, S. Wei, R. D. Blanton, and W. Maly, "Modeling the economics of testing: a DFT perspective," *IEEE Design & Test of Computers*, vol. 19, no. 1, pp. 29–41, 2002.

[16] H.-H. S. Lee and K. Chakrabarty, "Test challenges for 3D integrated circuits," *IEEE Design & Test of Computers*, vol. 26, no. 5, pp. 26–35, 2009.

[17] T. W. Williams and N. C. Brown, "Defect level as a function of fault coverage," *IEEE Transactions on Computers*, no. 12, pp. 987–988, 1981.

[18] B. T. Murphy, "Cost-size optima of monolithic integrated circuits," *Proceedings of the IEEE*, vol. 52, no. 12, pp. 1537–1545, 1964.

[19] T. L. Michalka, R. C. Varshney, and J. D. Meindl, "A discussion of yield modeling with defect clustering, circuit repair, and circuit redundancy," *IEEE Transactions on Semiconductor Manufacturing*, vol. 3, no. 3, pp. 116–127, 1990.

[20] D.-M. Kwai and C.-W. Wu, "3D integration opportunities, issues, and solutions: a designers perspective," in *Proc. SPIE Lithography Asia*, 2009.