

Reconfigured test architecture optimization for TSV-based three-dimensional SoCs

Kele Shen^{1a)}, Dong Xiang², and Zhou Jiang²

¹ Department of Computer Science, Tsinghua University,
Beijing 100084, People's Republic of China

² School of Software, Tsinghua University,
Beijing 100084, People's Republic of China

a) shenkele@aliyun.com

Abstract: The technology of three-dimensional (3D) SoCs is emerging as a promising approach for extending Moore's Law. Managing test architecture design and optimization of 3D integration are crucial challenges. In this paper, we propose a reconfigured test architecture optimization for 3D SoCs, including a novel scheme to minimize the pre-bond test time and Known-Good Stack (KGS) test to guarantee the yield of 3D SoCs. Experimental results on ITC'02 SoC benchmark circuits show that our scheme reduces the total test time by around 23% on average and nearly 30% in maximum compared with one baseline solution.

Keywords: 3D SoC, optimization, DfT, test access mechanism (TAM), test time

Classification: Integrated circuits

References

- [1] X. Yuan, L. Gabriel, B. Bryan and B. Kerry: ACM J. Emerging Technol. Comput. Syst. **2** (2006) 65. DOI:10.1145/1148015.1148016
- [2] D. Sergej, K. Brion, C. Vivek, M. Subhasish, S. Navdeep, G. S. Kumar, Ch. J. Jan, M. Ashok, L. Frank and M. E. Jan: IEEE ITC Dig. (2012) 1.
- [3] Z. Yervant, M. E. Jan and D. Sujit: IEEE ITC Dig. (1998) 130. DOI:10.1109/TEST.1998.743146
- [4] D. L. Lewis, S. Panth, X. Zhao, S. K. Lim and H.-H. S. Lee: IEEE ICCD Dig. (2011) 90. DOI:10.1109/ICCD.2011.6081381
- [5] Teradyne Technologies: Advanced Digital with Silicon Germanium Technology (2006) <http://www.teradyne.com/>.
- [6] S. Anuja and C. Krishnendu: IEEE Trans. Comput. **56** (2007) 120. DOI:10.1109/TC.2007.250628
- [7] I. Vikram, C. Krishnendu and M. E. Jan: J. Electron. Test. **18** (2002) 213. DOI:10.1023/A:1014916913577
- [8] G. Subhash, H. Mark, H. Sangki and P. Robert: Proc. of the 21st International VLSI Multilevel Interconnection Conference Dig. (2004).
- [9] J. Li, H. Lin and X. Qiang: IEEE DATE Dig. (2009) 200. DOI:10.1109/DATE.2009.5090661

1 Introduction

As semiconductor technology develops rapidly, semiconductor feature size becomes tinier and the interconnects in the System-on-Chip (SOC) turn into a performance bottleneck. Three-dimensional (3D) integration technology which provides a new architecture to solve this problem, is therefore a promising option to continue the Moores Law.

The advantages of 3D integration are as follows. First of all, interconnect length can be reduced typically by using through-silicon-vias (TSVs) to link the layers. Secondly, power consumption decreases with the reduction of wire length. Thirdly, 3D integration supports mixed-technology between die [1].

Although 3D integration has many advantages, there are still critical challenges. Firstly, there are not enough CAD tools in 3D area, even some existing 3D IC tools are from 2D versions with little modification [2]. Secondly, test time of 3D integration needs to be managed efficiently. Thirdly, heat dissipation becomes an urgent issue and may cause the chip to fail.

In this paper, we propose a reconfigured test architecture optimization for 3D SoCs to meet these demands, the main contributions of the paper are:

- We propose the DfT mechanism of core partitioning to further optimize the pre-bond test time compared with one baseline solution.
- We next introduce KGS test in test architecture design to guarantee the yield of 3D SoCs.

The rest of our paper is organized as follows. Section 2 reviews related work and describes our motivation in this paper. In Section 3, test architecture optimization problem is shown. We detail our reconfigured test architecture for the above problem in Section 4. Experimental results are then listed in Section 5. Finally, we conclude the paper in Section 6.

2 Preliminaries and motivation

For 3D SoC devices, modular testing should be associated with test access mechanisms (TAMs) and test wrappers [3]. TAMs deliver test patterns to cores, test wrappers are used interface between the internal core-under-test (CUT) and several bits of the TAM by turning them into wrapper chains. Each TAM has its own test width. Designing a TAM architecture should first consider minimizing the total test time. As in [4], test time is calculated by $T = (l + 1) \times (p + 1)$, in the above equation, p is the number of test patterns which can't be changed in wrapper design. l is the length of the longest wrapper chain. As a result, to reduce test time, minimizing the length of the longest wrapper chain must be in the most important position.

Fig. 1 shows an example for pre-bond testing of three-layer SoC. Assume that the pre-bond TAM width for each layer is a single bit. There are three cores on layer 1, one core on layer 2 and two cores on layer 3. The small rhombuses in the cores represent scannable cells, wrapper chains are depicted as the thick black lines.

In Fig. 1(a), the pre-bond wrapper chains on each layer are assigned by using the algorithm in [4], the longest wrapper chain on each layer is 16, 4 and

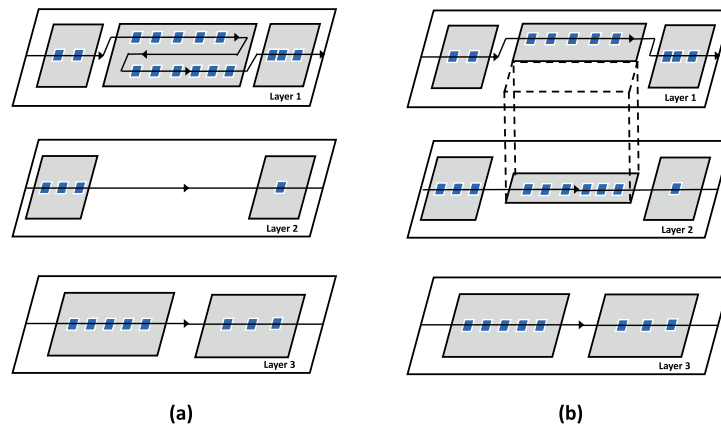


Fig. 1. An example of 3D SoC for pre-bond test (a) without core partitioning (b) after core partitioning.

8. As the length of the longest wrapper chain is positive correlated to the test time of each layer and ATE can test each layer simultaneously during pre-bond test of 3D SoCs [5], minimizing the longest one of these wrapper chains is able to reduce the pre-bond test time [6]. In Fig. 1(a), the longest wrapper chain for pre-bond test is 16.

In prior work, cores in SoC are often placed on only one layer, however, using core partitioning DfT mechanism will make 3D SoCs into fine granularity, so as to further minimize the test time of 3D SoCs. Not just using TSVs to connect each core externally, we can place cores onto two different layers and connect them via TSVs at the same time. For instance, in Fig. 1(a), we choose the core which has the largest number of scan chains, then we partition a part of it and map them onto layer 2, as depicted in Fig. 1(b).

Fig. 1(b) illustrates that the largest core on layer 1 is mapped onto two layers, to clearly introduce our motivation, we use imaginary lines to describe that a part of the core has been placed onto layer 2. Now the longest wrapper chain of three layers for pre-bond test is 10 instead of 16. Since the longest wrapper chain is proportional to test time, our proposal gets a better optimization. If we partition the core into several layers, the core can be tested in pre-bond test as we can test internal scan chain of the core in different layers. In particular, the test in our paper is structural testing, not functional testing. Functional testing can only operate when all layers related to the core are bonded, while structural testing concerns the open/short faults in the core. Therefore, different parts of a core can be tested in pre-bond test. Note that in Fig. 1(a) and 1(b), TSVs are omitted for the sake of explanation, however, in the real circuits, two parts of the core should be connected via TSVs. The wire length of wrapper chains is also an important component for the optimization, but test time is the primary goal. Moreover, our method will not make wire length increase too much.

3 Problem formulation

The test architecture design and optimization problem for 2D SoCs has been proved to be NP-hard [7], thus the problem for 3D SoCs is also NP-hard, since

the optimization problem for 2D SoCs is the sub-problem of 3D SoCs.

We formulate the problem in this paper as follows:

Problem: test-architecture optimization for 3D SoCs

- given a set of cores C , each core $c \in C$ has the number of inputs (in_c), the number of outputs (out_c), the number of bidirectionals (bid_c), the number of scan chains ($scnum_c$) and the scan chain length (l_c);
- given 3D partition of cores, that is, the number of layers ($layernum$), each core belongs to which layer;
- given the set of pre-bond TAM widths ($prewidth_{layer}$), where the subscript represents the serial number of layer, the set of mid-bond TAM widths ($midwidth_{layer}$), for example, if layer one and layer two are under mid-bond testing, the TAM length is depicted as $midwidth_{1|2}$, post-bond TAM width is written as $postwidth$;

Determine the assignment of I/Os, bidirectionals and scan chains in pre-bond, mid-bond and post-bond tests in order to minimize total test time, note that the assignment should not introduce too large wire length.

4 Proposed scheme

As mentioned above, the problem of 3D SoC optimization is NP-hard. We choose *Best Fit Decreasing* (BFD) heuristic to produce near-optimal TAM assignment.

As explained in section 2, the longest wrapper chain is in direct proportion to the total test time, therefore, we focus on minimizing the length of the longest wrapper chain during each stage of test.

Firstly, we define a metric called longest wrapper chain (LWC_{layer}) as follows:

$$LWC_{layer} = \max_{i=0}^n(wc_i) \quad (1)$$

In equation (1), wc_i represents the length of the i th wrapper chain, the number of wrapper chain is n , LWC_{layer} is described as the length of the longest wrapper chain on the $layer$. Note that the $layer$ can be one layer or more.

Secondly, we introduce another metric called whole test length (WTL), because minimizing WTL can directly minimize the total test time:

$$WTL = \max_{i=0}^n(LWC_i) + \sum_{i=2}^{ln} LWC_{1|i} \quad (2)$$

In equation (2), minimizing $\max_{i=0}^n(LWC_i)$ is our objective in pre-bond test, that is, it is the maximum length of each layer's LWC for pre-bond test. Since ATE can test each layer simultaneously during pre-bond test of 3D SoCs, $\max_{i=0}^n(LWC_i)$ is proportional to the test time for pre-bond test. $\sum_{i=2}^{ln} LWC_{1|i}$ indicates the LWC for mid-bond and post-bond tests in the same way.

4.1 Select the *partlayer*

To decide which layer should operate core partitioning mechanism, we need to select the *partlayer* which has longest wrapper chain for pre-bond test by the algorithm of BFD.

Fig. 1 shows an example of three-layer SoC. For the sake of simplicity, suppose each layer is distributed one bit TAM width for pre-bond test in the figure. After using the algorithm of BFD to compute the LWC of each layer, we can find that LWC_1 is 16, LWC_2 is 4 and LWC_3 is 8. Thus layer 1 is the layer to be selected, that is, *partlayer* is 1.

4.2 Core partitioning-BFD scheme

Core partitioning-BFD scheme is aim to further minimize the $\max_{i=0}^n(LWC_i)$ for pre-bond test. Our scheme to tackle the pre-bond testing optimization problem for 3D SoCs is shown in Fig. 2. The input of Fig. 2 is *partlayer*, the set of cores C and the test parameters for each core, TAM width of each layer is also made as an input. We also choose the smaller adjacent layer of *partlayer* in order to better minimize the longest wrapper chain.

Core Partitioning-BFD

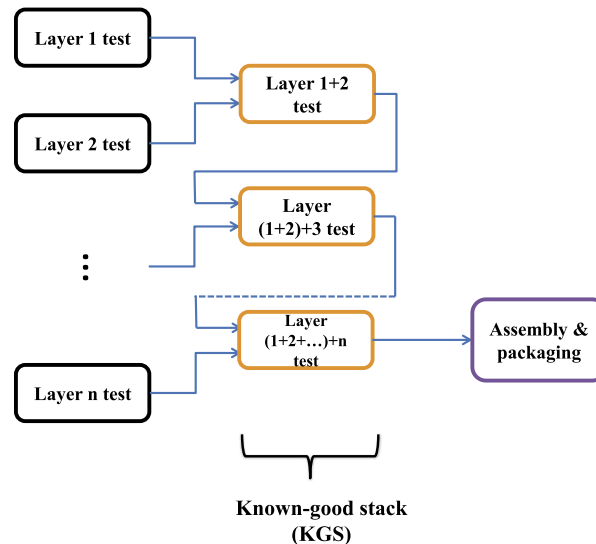
- ```
{
 1) Choose the parlayer and select the adjacent layer adlayer
 which LWC is smaller
 2) Using the same algorithm in pre-bond test to assign the
 wrapper chains of two layers as if two layers are merged
 together, TAM width is the sum of two layers' TAM widths
 which is similar to cores. Calculate the new LWC_{me} for two
 layers
 3) For each wrapper chain i obtained in step 2
 4) If i includes two layers' internal scan chains
 5) Count the number m of internal scan chains on parlayer
 6) Count the number n of internal scan chains on adlayer
 7) If ($m \geq n$)
 8) Partition cores which contain the internal scan chains
 in n onto two layers
 Transfer the internal scan chains in n from adlayer
 to parlayer
 9) else
 10) Partition cores which contain the internal scan chains
 in m onto two layers
 Transfer the internal scan chains in m from parlayer
 to adlayer
 11) Finish core partitioning process
 12) Replace $LWC_{parlayer}$ and $LWC_{adlayer}$ with LWC_{me} and
 recalculate $\max_{i=0}^n(LWC_i)$
}
```

**Fig. 2.** Core partitioning-BFD scheme for pre-bond test.

#### 4.3 Known Good Stack-BFD scheme

Once the wrapper chain assignment of pre-bond test is finished, the final step is using the BFD algorithm in mid-bond and post-bond tests. As shown in Fig. 3, the first column represents the pre-bond testing, the second column lists KGS tests. From the first line to the last but one line in column 2, we operate mid-bond test from layer  $1 + 2$  to layer  $1 + 2 + \dots + n - 1$ . For each mid-bond test, we choose BFD algorithm to assign the wrapper chains. The

last line in column 2 is the post-bond testing using the same BFD algorithm. Therefore, we can obtain the minimal *WTL* by finishing all the process as mentioned above so as to minimize the total test time of 3D SoCs. Note that the number of TSVs is omitted because the size of TSVs can be fabricated into micron level [8]. Therefore, TSVs shouldn't be a constraint for the 3D SoC test any more [9].



**Fig. 3.** Known-Good Stack flow for mid-bond and post-bond tests.

## 5 Experiments

### 5.1 Experimental setup

To demonstrate our optimization scheme, we use four ITC'02 benchmark SoCs to realize our experiments. Firstly we place these four SoCs onto three silicon layers randomly while balancing the number of cores on each layer. As mentioned above, we use the architecture of Test Bus in our experiments. Commercial tools are used to obtain 3D placement of the scan chains. All the programs are finished within only a few seconds.

As *WTL* is directly proportional to the total test time, we compare our proposed scheme's test time optimization with a baseline solution called All BFD in [4] using the metric of *WTL*. For the two above schemes, the BFD algorithm is used to design all the pre-bond, mid-bond and post-bond wrapper chains. In order to analyse the performance of core partitioning scheme instead of discussing a more superior algorithm, we need to guarantee that two schemes are both using BFD algorithm. The difference is that our scheme adds core partitioning mechanism to optimise the pre-bond wrapper chains assignment. Both of these two schemes involve KGS process.

To test our optimization scheme under different circumstances, we vary the total number of TAM width. Assume that the post-bond TAM width is equal to the sum of pre-bond TAM widths assigned to each layer for the sake of ease. In pre-bond test, the sum of the pre-bond TAM width from each layer

**Table I.** Experimental results comparison between proposed and the original scheme

| Width | p93791 |       |           | p34392 |      |           | t512505 |       |           | p22810 |      |           |
|-------|--------|-------|-----------|--------|------|-----------|---------|-------|-----------|--------|------|-----------|
|       | CP     | NCP   | Ratio (%) | CP     | NCP  | Ratio (%) | CP      | NCP   | Ratio (%) | CP     | NCP  | Ratio (%) |
| 12    | 24505  | 25816 | 5.08      | 6439   | 8551 | 24.70     | 22620   | 27999 | 19.21     | 7314   | 8585 | 14.80     |
| 18    | 21483  | 29539 | 27.27     | 5566   | 7371 | 24.49     | 18698   | 25266 | 26.00     | 6619   | 8966 | 26.18     |
| 24    | 16351  | 22302 | 26.68     | 4433   | 5700 | 22.23     | 14382   | 19154 | 24.91     | 4995   | 6831 | 26.88     |
| 30    | 12991  | 17824 | 27.12     | 3685   | 4596 | 19.82     | 12305   | 15702 | 21.63     | 4250   | 5576 | 23.78     |
| 36    | 10856  | 14827 | 26.78     | 3208   | 4020 | 20.20     | 11045   | 13971 | 20.94     | 3675   | 4750 | 22.63     |
| 42    | 9452   | 12829 | 26.32     | 2770   | 3612 | 23.31     | 9891    | 12681 | 22.00     | 3150   | 4129 | 23.71     |
| 48    | 8295   | 11309 | 26.65     | 2418   | 3228 | 25.09     | 8155    | 10255 | 20.48     | 2647   | 3505 | 24.48     |
| 54    | 7208   | 9893  | 27.14     | 2418   | 3152 | 23.29     | 7600    | 9401  | 19.16     | 2447   | 3241 | 24.50     |
| 60    | 6570   | 8962  | 26.69     | 2418   | 3007 | 19.59     | 7501    | 9107  | 17.63     | 2293   | 3064 | 25.16     |
| 66    | 6017   | 8227  | 26.86     | 2418   | 2949 | 18.01     | 6380    | 8241  | 22.58     | 2186   | 2846 | 23.19     |
| 72    | 5539   | 7567  | 26.80     | 2418   | 2813 | 14.04     | 5923    | 7950  | 25.50     | 2106   | 2738 | 23.08     |
| 78    | 5151   | 7032  | 26.75     | 2418   | 2693 | 10.21     | 5639    | 7667  | 26.45     | 1884   | 2496 | 24.52     |
| 84    | 4804   | 6580  | 26.99     | 2418   | 2594 | 6.78      | 5422    | 7369  | 26.42     | 1932   | 2425 | 20.33     |

is the total pre-bond TAM width, and we divide the total TAM width to each layer for pre-bond test as evenly as possible.

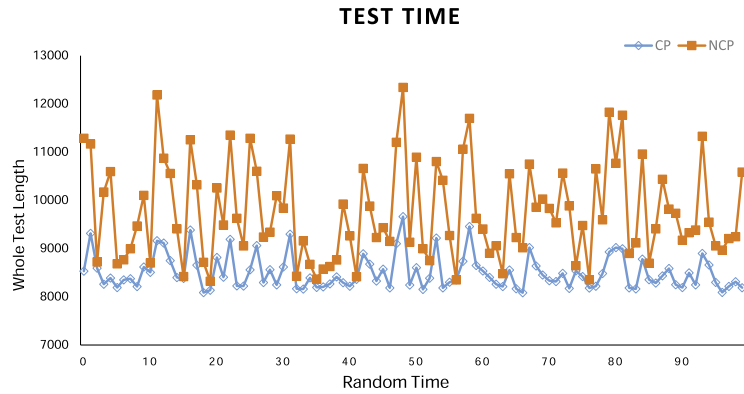
## 5.2 Experimental results

Table I presents our experimental results compared with the original scheme in [4] on four benchmarks of ITC'02. In the table, CP represents our proposed scheme using core partitioning. NCP means the original scheme without core partitioning. Ratio is the reduction percentage of *WTL*, which can be obtained by  $(NCP - CP)/NCP$ . Table I illustrates that our proposed scheme on all the benchmarks reduces total test time by nearly 30% in maximum and around 23% on average.

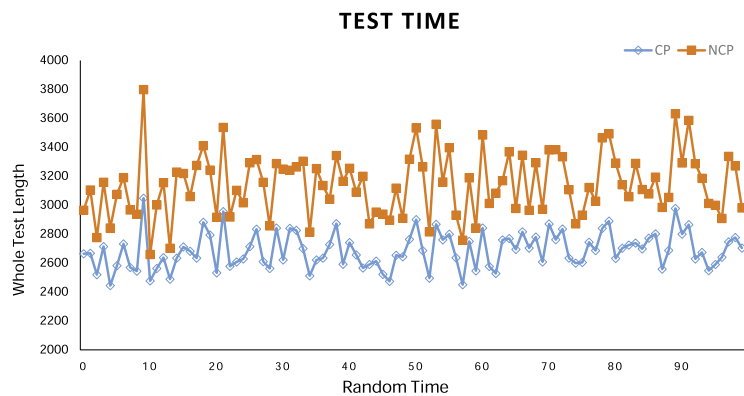
We can find that when the total TAM width gets larger, the *WTL* in both of two schemes decreases for p93791, t512505 and p22810. However, for p34392, as the total TAM width reaches 48, the *WTL* stops reducing and maintains at 2418. This is due to a bottleneck core in p34392, and we can deduce that as the total TAM width increases, other three SoCs's *WTL* will also go into a bottleneck.

Fig. 4 and 5 show the *WTL* results for 3-layer p93791 and p22810 respectively. Specifically, we randomly place each SoC onto three silicon layers for one hundred times, so as to demonstrate the stability of the reduction of test time in our proposed scheme. We provide only two *WTL* graph because the results for all the other benchmarks and configurations prove almost the same trends. In Fig. 4 and 5, the total TAM width of two benchmark are both 48. Our core partitioning scheme (CP) is always superior to the original scheme (NCP) in reduction of test time. In particular, CP in p93791 obtains more than 10 percentage of *WTL* reduction during 63% of





**Fig. 4.** *WTL* reduction of p93791 in one hundred random times.



**Fig. 5.** *WTL* reduction of p22810 in one hundred random times.

total random times. In the best cases, we reduce 24.4% *WTL*. Similarly in p22810, CP gains more than 10 percentage of *WTL* reduction during 87% of total random times, and the best case is 20.1%. Therefore, we come to a conclusion that our proposed scheme has a steady reduction of total test time of 3D SoCs.

## 6 Conclusion

In this paper, we have presented a reconfigured methodology to optimize test wrapper design for TSV-Based 3D SoCs. We introduce a new scheme called core partitioning-BFD to further minimize the test time, KGS test is used to guarantee the yield of 3D SoCs with more than two layers. Experimental results show that the proposed scheme can steadily reduce test time by nearly 30% in maximum and around 23% on average when compared with one basic solution.