

A Different Approach to Fabricating Three-Dimensional Integrated Circuits

Rani S. Ghaida

*Department of Computer Engineering
Lebanese American University*

Byblos, Lebanon

rani@ece.unm.edu

Abstract

This work presents a new fabrication technology for three-dimensional integration. In 3D integrated circuits, transistors may be constructed on different device layers. Transistors are then wired to other transistors on the same layer using conventional wiring or to other transistors on another layer using vertical wiring. The proposed technology consists of fabricating n-type and p-type transistors each on a separate device layer using conventional 2D fabrication technology and then combine the two device layers with inter-wafer interconnect. The new proposed technology is capable of providing improvement for integrated circuits in total wire length and chip area, as well as easing chip design phase and eliminating chip latch-up problem. In this paper, the new 3D technology is presented, together with its advantages, feasibility, and challenges. In addition, solutions to the layout design as well as placement and routing software PRS-3D for the proposed 3D technology are offered. These tools were tested and, relative to conventional 2D circuits, 8.7% to 33% improvement in total wire length and 44% to 51.7% improvement in chip area were observed when 3D integration is used.

1. Introduction

3D integrated circuits refer to the fabrication of any single chip with multiple device layers stacked over each others. This is achieved by bonding multiple wafers using special inter-layer vias to achieve interconnections of the vertically staked chips. Different approaches for fabricating 3D integrated circuits are available. In these approaches, cells of the 3D IC exist in a single device layer and vertical wiring is used to connect cells belonging to different layers. The main objective of these approaches is to lower the overall interconnect length by using vertical wiring. In particular, existing 3D integration technologies make use of vertical wiring to reduce the maximum length of interconnects and the number of long interconnects.

Besides, 3D integration represents a way for yield enhancement. In fact, 3D integration reduces the size of dies since a single 3D chip is formed by multiple dies. As a result, the number of dies per wafer is increased while the number of defects per wafer is unaffected. Thus, the chip's manufacturing yield is improved.

2. Proposed 3D IC Fabrication Technology

In this section, a different approach to manufacturing 3D integrated circuits is proposed and the motivation for the work is presented.

2.1. Overview

There are many types of fabrication defects that are likely to occur while the chip is being manufactured. Design rules are formulated to minimize the probability of these defects to occur and make the production of faultless chips reach an acceptable level.

Design rules form constraints on minimizing the layout area. These rules differ from one fabrication technology to the other. However, the most limiting design rule for all technologies that allow the fabrication of n-type and p-type transistors on the same substrate is the isolation distance, which forces a minimum separation between n-diffusion and p-diffusion. For instance, in SCMOS technology, the isolation distance is of at least 10λ i.e. 5 times the minimum feature size allowed by the lithographic printing. In CMOS circuits, this separation represents an unexploited region, only a part of it is infrequently used for intra-cell wire routing. This fact was the main motivation for finding a solution to this problem.

Some studies showed that, as long as p-type and n-type transistors are built on the same wafer, there is no way to get rid of the unused region, separating the two types of transistors, or even reduce its area.

The idea of the proposed 3D fabrication technology is to have a single chip consisting of two device layers each holding either n-type or p-type transistors. In other words, each cell will have its n-type transistors on one

device layer and its p-type transistors on another. And then, connect the two types of transistors belonging to the same cell using vertical wiring. This wiring is done by building a series of vias to traverse metal layers and a special type of via “inter-wafer via” to cross wafers. This approach will completely isolate n-diffusion and p-diffusion and the design rule that forces a minimum separation between the two diffusions will be eliminated. Figure 1 shows a sample IC constructed using the proposed 3D technology.

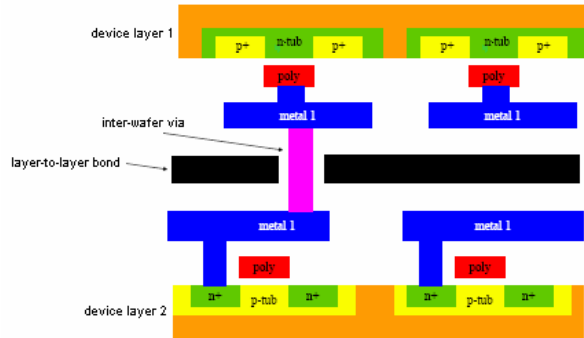


Figure 1: IC constructed using the proposed 3D technology.

2.2. Fabrication Process

The fabrication process for the proposed 3D technology consists of constructing the two device layers each by a separate process using the conventional 2D fabrication technologies. At that time, transistors are formed and all interconnects involving a single wafer are wired. In addition, every interconnect terminal that needs to be connected to a terminal in the other wafer, is wired to 3D contact points. Then, the wafers are bonded face-to-face, where the highest metal layers are adjacent, one on top of the other [1, 2], as shown in Figure 1, and 3D contact points are then aligned in order to form the inter-wafer vias.

3. Advantages of Proposed Technology

In addition to the advantage of yield enhancement common for all 3D integration because of the reduced die size, the proposed 3D integration technology has the following advantages.

3.1. Reducing Chip Area

As mentioned earlier, when the proposed technology is used, the area separating n-type and p-type diffusions is completely eliminated. So, cell heights are reduced and consequently chip area is diminished.

Also, building the different types of transistors on top of each other reduces significantly the planar size of the IC with negligible thickness increase. All 3D ICs fabrication technologies have this advantage over conventional 2D fabrication technologies.

3.2. Doubling the Maximum Number of Interconnect Layers

Development of wafer polishing techniques has allowed a continuing increase in the maximum number of interconnect layers constructed in ICs. However, currently used wafer polishing technique, called chemical mechanical planarization (CMP), is getting close to its limits and the maximum number of interconnect layers built on a single chip is not expected to increase significantly [3].

In the proposed 3D technology, each device layer is fabricated with a separate process using conventional 2D fabrication technologies and then the two device layers are bonded face-to-face. Thus, the produced IC contains double the maximum number of interconnect layers allowed for 2D ICs and consequently planar size of routing regions is reduced and chip area is diminished.

3.3. Avoiding Silicon-Controlled Rectifier

In ICs containing both types of transistor, p-tub region holding n-type transistors and n-tub region holding p-type transistors are not physically isolated. Therefore, current can flow between these regions resulting in a special circuit known as silicon-controlled rectifier (SCR) [4]. SCR turns on when triggered, conducting large amount of current. Once started, the current flow is usually impossible to interrupt without removing all power from the device. The amount of current drawn can be so high that it can either overload a power supply or destroy the device. This problem is known as the chip latch-up problem.

As long as n-type and p-type transistors are built on the same wafer, SCR cannot be eliminated but special methods are used to make it difficult to turn on.

In the proposed technology, n-tub and p-tub are on different wafers (different substrates) and the 2 types of tub are completely isolated. As a result, SCR can never occur and latch-up possibility is completely eliminated.

3.4. Easing Power/Ground Routing

Whenever the two types of transistors are constructed on the same wafer, power and ground routing is done on the same plane in order to supply the transistors present on the same level. This fact causes complexity in the power and ground routing.

In the proposed 3D technology, n-type and p-type transistors exist on two separate layers. Therefore, power and ground routing are done on separate planes and its complexity is significantly reduced.

3.5. Easing Pin Assignment Phase

When the proposed 3D technology is applied on CMOS circuits, I/O pins of every cell are available on all the 3D IC metal layers since these pins are wired vertically. This tremendously alleviates the pin assignment phase which is no more concerned in deciding on the metal layer availability for every pin.

4. Technical Feasibility

4.1. Wafer Stacking

Recent technological advances have permitted wafer stacking. Existing 3D technologies stack 2 to 5 wafers. These are bonded face-to-face (as in Figure 1), face-to-back where the highest metal layer of one wafer and the substrate of the other wafer are adjacent, or back-to-back where the substrate of both wafers are adjacent.

In the proposed 3D technology, only 2 wafers are stacked and the bonding of these wafers is a face-to-face bond which is the least difficult type of bond to fabricate. Thus, the fabrication of ICs using the proposed technology is the least complex among all other 3D technologies.

4.2. Wafer Alignment

3D contact points are performed in an Electronic Vision EV 450 Aligner [1, 5]. This system has an underlying $\pm 3 \mu\text{m}$ alignment tolerance. Therefore, 3D contact points are at least $3 \mu\text{m}$ in width which is large relative to conventional vias connecting metal layers. However, optical alignment systems are expected to improve which will decrease 3D contact point size down to 0.5 to $1 \mu\text{m}$.

4.3. Packaging

Existing 3D ICs packaging techniques, such as MCM-V [6, 7], can also be used for ICs employing the proposed fabrication technology. In these packaging techniques, I/O pads in each wafer or die are formed and then connected to the outside pins of the chip package using bonding wires.

Also, another method to pack 3D ICs using the proposed approach is available. This method constructs I/O pads only on the bottom die and then connect them to the outside pins of the package. Figure 2 better

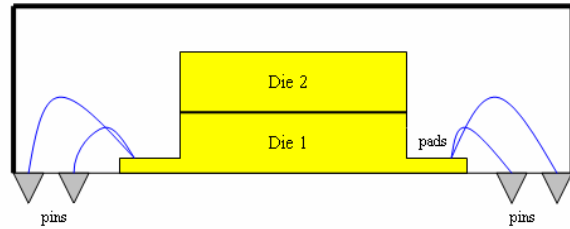


Figure 2: Possible packaging solution for the proposed 3D technology.

describes this packaging method. Note that it is enough to have I/O pads only on the bottom die since all cells (in CMOS circuits) are present on both dies and are internally connected.

4.4. Thermal Issues

A thermal study was conducted on the existing 3D ICs [8]. This study showed that the dissipation and removal of heat are significant issues for the design of integrated circuits, but with the development of cooling technologies the heat removal would be perfectly achieved. One cooling technology that may be used is the placement of fluid microchannels between the two stacked dies. The fluid in these microchannels absorbs the heat dissipated by the circuit and conducts it to a heat exchanger. This exchanger transfers the heat outside the IC and lowers the fluid temperature then the fluid is re-circulated through the microchannels.

5. Design Implementation

Integrated circuits that utilize the proposed 3D technology are designed in two phases: a different phase for each die. While doing the two phases, the designer has to make sure to connect all cells internal pins, which must be connected to the other die, to the 3D contact points. The major issue of this approach is the alignment of 3D contact point.

The alignment can be automatic using 3DMagic layout editor [9] which is based on Magic layout editor [10] and adapted to support 3D technologies.

The design of ICs applying the proposed technology can also be implemented using any conventional layout editor for 2D integration. However, this method is more difficult than the one using 3DMagic and requires an extra effort from the designer to align both device layers.

No matter what layout editor is used, the design of ICs applying the proposed 3D technology follows the steps described by the flowchart of Figure 3.

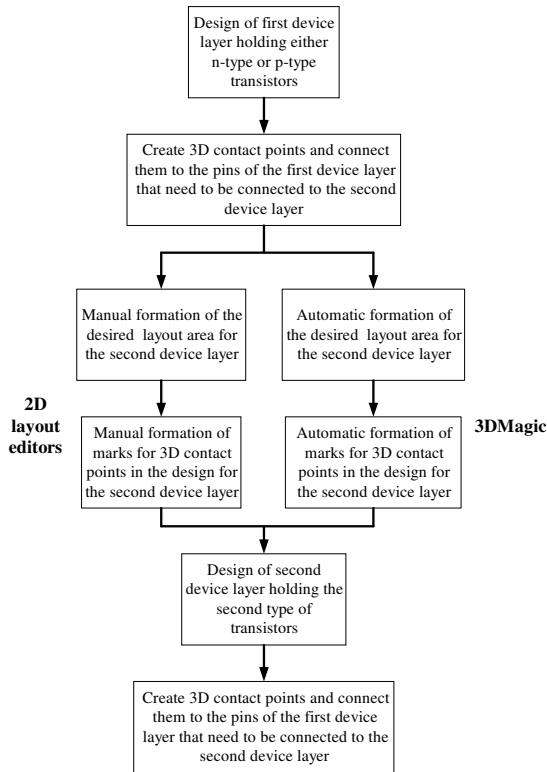


Figure 3: Design flow for cells applying the proposed 3D technology.

6. 3D Standard Cell Design Tools

In order to test the proposed 3D fabrication approach and compare it with the traditional 2D fabrication approaches, standard cell placement and routing software was developed and adapted for 2D and the proposed 3D technologies.

The software is implemented in about 5,000 lines of JAVA code. It is given the name Placement & Routing Software (PRS). Two versions of PRS are available: PRS-2D supporting 2D integration technologies, and PRS-3D supporting the proposed 3D technology.

PRS is presented: a) the cells' pitch, b) a set of cells with their corresponding widths and locations of possible over-the-cell routing, c) a set of cells' I/O pins and their locations, d) a list of two-terminal nets.

PRS outputs a display of an optimal placement of cells in rows minimizing IC's area and an optimal routing of nets minimizing overall wires' length.

Cells placement follows a topological model. It starts with a fixed initial placement of cells and then applies iterative cells' displacements (iterative placement) following the Simulated Annealing algorithm where the objective function is minimizing IC's area, as a first objective, and minimizing overall wire's length as a second objective. Cells' displacement is made by

choosing a cell at random, delete it from its current position, and then insert it at a new position selected arbitrary. The resulting placement is then compared to the Best placement (placement having the least cost of the objective function) ever achieved. If the new placement is better than Best placement and if it has no design rule violations, the new placement is set as Best placement. If the new placement is not better than Best placement achieved so far, the placement may still be chosen as the Current placement. Current placements are considered for the next iterations since they are likely to lead to improvements in the Best placement. Current placements are chosen according to a probability function that depends on the current temperature of the algorithm as well as the difference between the cost of the objective function for the new placement and the cost of the objective function for the actual Current placement.

In the proposed 3D technology, cells are divided into two parts: each belonging to a different device layer. A cell is assumed to have the height of its part that has the longest height, and the width of its part that has the largest width. As a result, 3D cells are characterized by only two dimensions. Thus, cell placement in PRS-3D is the same as in PRS-2D.

As for the routing of nets, it uses a gridless approach. It follows a reserved metal layer model (reserved to either horizontal or vertical direction). Since the proposed 3D technology consist of two 2D ICs abutted face-to-face, PRS-3D has double the number of metal layers as PRS-2D has. And, in PRS-3D, metal layers lying on the lower half layers are reserved the same way metal layers in PRS-2D are reserved, whereas metal layers lying on the upper half layers are reserved in the opposite way.

Two routing regions are involved: feedthrough routing region and channel routing region. Feedthroughs are located between the cells of the same row whereas channels are located between the cells of different rows. The routing process consists of two phases: global routing and detailed routing. During global routing, routes are sketched roughly. First, for every net, routing regions traversed by the net's route are determined. Then, the points at which routes cross the borders of routing regions are settled for all routes. The locations of these points are chosen with the purpose of minimizing the sum of Euclidean distances between the points associated with the same route and consequently minimizing the overall distance of the route. At this stage, detailed routing can take place. Feedthrough routing is done immediately, however channel routing requires the run of the Left Edge Algorithm to complete. First, nets are sorted in a descending order beginning by the net having the leftmost left edge. Then, each net is routed at a time by

allocating existing channel tracks for the routing of nets. Channel tracks can be used for the routing of more than one net while avoiding contacts between different nets. In the case where the usage of existent tracks results in short circuits of nets, a new track that can be used in later routing is created. Now, if the channel containing the newly created track experienced a change of height, everything below the track is pushed down with the intention of avoiding any design rule violation.

In addition, a design rule checker (DRC) is developed in order to look for design rules' violations and correct them. Possible violations that may occur and are corrected: minimum separations between vias and minimum separations between metal wires of the same layer.

The DRC consists of a preventive procedure and a post-routing procedure. The preventive procedure rejects cells' placement that is expected to cause the possible design rules' violations if routing takes place. Moreover, this procedure rejects points of intersection between routes and borders of routing regions if these points are expected to originate violations of the design rules. Furthermore, the preventive procedure is involved during channel routing and specifically when choosing the tracks so as to avoid the violations. On the other hand, the post-routing procedure is called when the routing is complete. This procedure searches for the violations described earlier and then separates the two vias or wires causing every violation by fixing one and moving the other down. Again, if the move induces new design rules' violations, they are directly corrected by following the same procedure.

7. Testing & Results

7.1. Comparison between 2D & 3D Circuits

The testing of PRS is done on different CMOS circuits. Circuits are either fictitious designs or real designs with actual cells drawn on Magic.

Three fictitious designs were tested on PRS. These designs contained a small number of cells (15 to 33 cells), a small number of cells' I/O pins (24 to 40 pins), and a small number of nets (12 to 20 nets). Here, cells were given imaginary dimensions. Each design was tested on PRS-2D and PRS-3D, with the only difference of cell pitch. For the latter, the cell part holding n-type transistors and the one holding p-type transistors are assumed to have the same dimensions. As a result, cell pitch was assumed to be equal to the pitch used in PRS-2D minus 10λ and then divided by 2. The observed results are 14% average reduction of total wire length and 51.7 % average reduction of total chip

Table 1: Results* of testing 3 fictitious designs each on PRS-2D and PRS-3D (results are shown in units of λ).

	Design 1		Design 2		Design 3	
	2D	3D	2D	3D	2D	3D
wire length	896	863	1,027	983	2,351	1826
Chip width	400	406	420	420	548	544
Chip height	354	171	358	167	321	159

area (area of one device layer for the 3D technology was considered).

Table 1 shows the results of testing the 3 designs each on PRS-2D and PRS-3D.

Reductions in the chip height and area are predictable because of three reasons: a) cell pitch for 3D is reduced by more than half of 2D cells, b) chip consists of two device layers placed one on top of the other, c) number of metal layers used for 2D is doubled for 3D technology which will reduce the heights of cells and consequently the area of the chip. Reduction in total wire length is due to the fact that the heights of channels and cells are reduced which will decrease distances between pins that need to be connected.

Figure 4 and Figure 5 show the output display of PRS-2D and PRS-3D when tested on a small fictitious design consisting of 15 cells.

Moreover, three real designs with actual cells were tested on PRS. All these cells are designed, for 2D and 3D technologies, in Magic layout editor. Then, cells dimensions and pins locations were measured and entered afterwards into the software for testing.

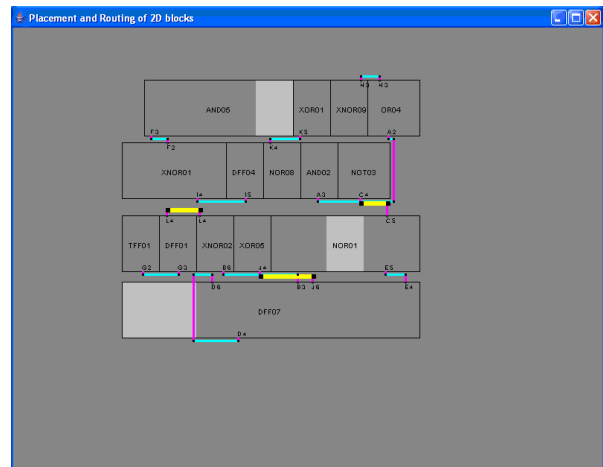


Figure 4: Output display of PRS-2D when tested on a fictitious example.

* Each result is an average result of more than 3 different tests

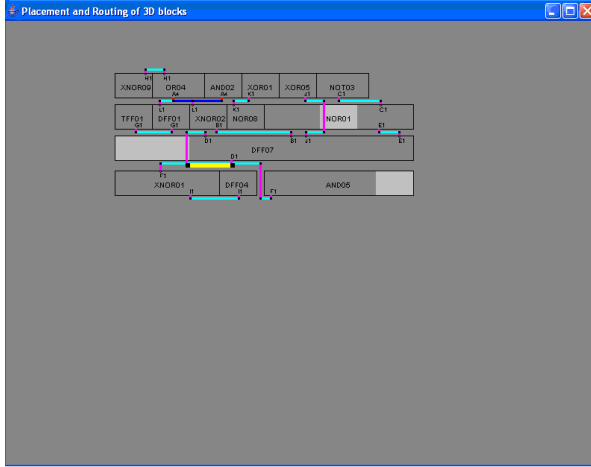


Figure 5: Output display of PRS-3D when tested on the same fictitious example as in Figure 4.

The first design is a circuit that computes the absolute value of the difference between two 2-bit positive numbers. This design is composed of 52 cells, 136 cells I/O pins, and 75 nets. Four types of standard cells are used in this circuit: XOR, NAND, NOR, and NOT gates. Average improvements of 10.5% in total wire length and 44% in chip area (area of one device layer for the 3D technology was considered) are observed.

The second real design is an 8 bits Left/Right barrel shifter. It consists of 86 cells, 254 cells I/O pins, and 114 nets. Three types of standard cells are used in this circuit: AND, OR, and NOT gates. Average improvements of 16.3% in total wire length and 44.5% in chip area (area of one device layer for the 3D technology was considered) are observed.

And finally, the third real design that is tested is a 16 bits comparator. It consists of 161 cells, 419 cells I/O pins, and 210 nets. Three types of standard cells are used here: NAND, NOR, and NOT gates. Average improvements of 8.66% in total wire length and 40.4% in chip area (area of one device layer for the 3D technology was considered) are noticed.

Results of testing the 3 real designs each on PRS-2D and PRS-3D are shown in Table 2.

Note that for real designs the reduction in chip area caused by 3D integration has dropped significantly compared to the reduction obtained in testing of fictitious designs. This is due to the fact that inter-wafer vias occupy a large space ($7 \times 7 \lambda$) in real cells affecting the overall dimension of cells. Therefore, the assumption taken for fictitious design stating that 3D cell has the same width as its corresponding 2D cell and a height equal to the height of its corresponding 2D cell – 10λ and then divided by 2 is no longer valid for real cells. Consequently, the expected reduction in the size of inter-wafer vias caused by further improvement in

Table 2: Results[†] of testing 3 real designs each on PRS-2D and PRS-3D (results are shown in units of λ).

	Design 1		Design 2		Design 3	
	2D	3D	2D	3D	2D	3D
Wire length	14.3k	12.8k	61.6k	51.6k	160k	146k
Chip width	370	358	902	801	1.8k	1.8k
Chip height	462	268	562	351	946	570

optical alignment systems will dramatically reduce the area of 3D chips.

It is noticed that percent reduction in total wire length and chip area caused by 3D integration is still significant even for large designs.

The routing tool is again tested separately from the testing of the placement tool. This is done by testing the routing tool for 2D and 3D technologies on the same placement. So, PRS-2D and PRS-3D are tested on a single placement of each of the 3 fictitious designs and the first real design that were previously used for testing the placement tool. For the fictitious designs, improvement of 9.3% in total wire length is noticed, whereas for the real design, improvement of 33% in total wire length is observed. The improvement in total wire length for the real design is impressive. It is much larger than the improvement observed for the fictitious example. This is justified by the fact that the real design has a lot more nets (75 nets for real example compared to a maximum of 20 nets for fictitious examples) and as a result, the effect of doubling the maximum number of interconnect layer is more impinging.

7.2. Comparison between 2D & 3D Cells in real designs

In order to compare between 2D designs and 3D designs, 2D and 3D cells were both designed using the same layout editor (Magic layout editor) and the same technology (SCMOS 1.0 $\mu\text{m}/\lambda$), and both cells were optimally reduced in size.

Table 3 shows a comparison between the area of 2D cells and the area of 3D cells as well as a comparison between the ratio of empty spaces over the total space in 2D cells and the ratio of empty spaces over the total space in 3D cells.

The comparison indicates that in real designs the average size of 3D cells is 55.5% of the average size of 2D cells in contrast with fictitious designs where the

[†] Each result is an average result of more than 3 different tests

Table 3: Comparison of area and ratio of empty spaces over total space between 2D and 3D cells (results are shown in units of λ).

Design	Cell	Area (squares)		Ratio of empty/total (squares)	
		2D	3D	2D	3D
1	NOT	918	486	0.52	0.50
1	NAND	1350	650	0.41	0.41
1	NOR	1404	650	0.47	0.41
1	XOR	2916	1595	0.37	0.42
2	NOT	608	460	0.40	0.42
2	AND	1444	740	0.33	0.28
2	OR	1330	740	0.31	0.28
3	NOT	608	418	0.41	0.40
3	NAND	912	570	0.34	0.38
3	NOR	912	570	0.34	0.38

size of a 3D cell is assumed to be less than half the size of its corresponding 2D cells. Besides, the comparison shows that on average the ratio of empty squares to the total squares in 2D cells is approximately the same as in 3D cells. In other words, the effort made to optimize the size of 2D and 3D cells is roughly the same.

7.3. Assumptions Affecting Testing Results

Some assumptions were made during testing that restricted the improvements in total wire length and chip area when 3D integration is used.

The first assumption is the size of inter-wafer vias which is huge ($7 \times 7 \lambda$) compared to the minimum transistor's gate ($2 \times 3 \lambda$). This affected the dimension of 3D cells and consequently the overall chip area.

The second assumption is in determining the dimensions of 3D cells. The assumption states that the width of 3D cells is determined by the part of the cell having the largest width and the height of 3D cells is determined by the part of the cell having the largest height. This assumption generates empty areas on the part of the cell with the smallest dimension that are neither used for transistors nor for routing. These vacant areas could have been used for routing which reduces the chip area.

The third assumption is deactivating the DRC during all testing in order to reduce the run time of tests. In some cases, the DRC affects the total wire length and overall area of the chip. But these cases are more likely to occur in 2D design rather than in 3D design since 3D integration has double the maximum number of interconnect layers in 2D integration and consequently less interconnect layer overlaps will occur when more layers are available for routing of nets.

8. Conclusion

Three-dimensional integrated circuits are expected to offer significant benefits over conventional two-dimensional integrated circuits. A number of 3D integration fabrication technologies are already available. This paper presents a different approach to fabricating 3D integrated circuits and illustrates some of the potentials, challenges, and feasibility of this method.

In addition, the paper offers solutions to the layout design of the new technology, as well as placement and routing tools for these circuits. Using these tools, performance results of ICs using the proposed 3D technology are analyzed. Improvement of 8.7% to 33% in overall wire length and 44% to 51.7% in chip area are observed.

Relatively to 2D integration technologies, the 3D fabrication technology proposed in this work provide less improvement in overall wire length compared to other 3D fabrication technologies. However, the proposed technology offers additional improvement in chip area, eliminates the chip latch-up problem, and reduces the complexity of power/ground routing and pin assignment phase.

9. References

- [1] A. Fan, A. Rahman, and R. Reif. Copper wafer bonding. *Electrochemical and Solid-State Letters*, 2:534–536, 1999.
- [2] Y. Kwon, A. Jindal, J. J. McMahon, J.-Q. Lu, R. J. Gutmann, and T. S. Cale. Dielectric glue wafer bonding for 3D ICs. In *Proc. MRS*, Spring 2003.
- [3] A. S. Brown. Flat, cheap, and under control: Applied materials' new polishing technology could be the key to the coming generation of microchips. *IEEE Spectrum*, 34-39, January 2005.
- [4] W. Wolf. *Modern VLSI design: System-on-chip design*, Third Edition, Upper Saddle River, NJ: Prentice Hall PTR, 2002.
- [5] K. N. Chen, A. Fan, and R. Reif. Microstructure examination of copper wafer bonding. *J. Electron Mat.*, 30(4):331-335, 2001.
- [6] C. W. Eichelberger. Three-dimensional multichip module system. United States Patent 5,111,278, May 1992.
- [7] C. Val. 3-D packaging – applications of vertical multichip modules (MCM-V) for microsystems. In *Proc. IEEE/CPMT IEMT Symp.*, 1994.
- [8] S. Das. Design automation and analysis of three-dimensional integrated circuits, *PhD Dissertation*, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2004.
- [9] S. M. Alam, D. E. Troxel, and C. V. Thompson. A comprehensive layout methodology and layout-specific circuit analyses for three-dimensional integrated circuits. In *Proc. ISQED*, March 2002.

- [10] John K. Ousterhout, Gordon T. Hamachi, Rober N. Mayo, Walter S. Scott, and George S. Taylor. The Magic VLSI Layout System. *IEEE Design and Test*, pages 19-30, February 1985.