Evaluation for Stacked-Layer Data Bus Based on Isolated Unit-size Repeater Insertion

Chia-Chun Tsai*

Department of Computer Science and Information Engineering, Nanhua University, Chiayi, Taiwan, ROC

Abstract

The data bus of a stacked-layer chip always supports that data of a program are frequently running on the bus at different timing periods. The data access on the bus determines the average access time of the program that dominates its performance. In this paper, we proposed an algorithm based on the insertion of isolated unit-size repeaters to evaluate the average access time of a stacked-layer data bus with a complete timing period. For a timing period, the algorithm is trying to insert a number of unit-size repeaters into bus wires along the path of a source-sink pair for isolating extra capacitive loadings to reduce their access time as possible. Then, the average access time of a data bus with a complete timing period can thus be reduced in advance. If the saving has at least 10% in average access time, the stacked-layer data bus can be reconstructed with inserted unit-size repeaters for reducing their average access times to most of programs ran on the bus. Experimental results for six tested cases show that our algorithm has the saving in average access time up to 50.81% with 70 unit-size repeaters in a number of milliseconds on average. The proposed approach is a rapid evaluation for the reconstruction of a stacked-layer data bus.

Keywords: Stacked-layer chip, 3D data bus, Unit-size repeater, Average access time.

1. Introduction

For a stacked-layer chip [1], each layer has own local data bus and a number of TSVs is used to vertically connect these local data buses located on different layers to integrate them to be a global data bus. The 3D global data bus consists of a number of 2D local data buses. Data are frequently running on the 2D local data bus or 3D global data bus for executing multiple programs. A data access time is defined and calculated from a source to at least one sink at a timing period. For a program with a number of hundreds or thousands timing periods, its average access time is defined as the total data access times divided by the total timing periods and it will dominate the program performance.

In nanotechnology, a longer interconnection wire always dominates the propagation delay even over than a gate delay because of their increased wire resistance and capacitance. Fig. 1(a) shows a 2D local data bus and there is a bidirectional data access between terminals p1 and p6 at two different timing periods. From the figure, obviously, these extra loading capacitances, C_2, C_3, C_4, and C_5 will increase the data access time of the source-sink pair terminals p1 and p6. Each extra loading capacitance comes from their branch wire capacitance and terminal capacitive loading along the path of the source-sink pair p1-p6 or p6-p1. As shown in Fig. 1(b), most of these extra capacitive loadings can be isolated just by inserting a unit-size bidirectional repeater into each branch wire. That is, these extra loading capacitances are dramatically reduced to be C'_2, C'_3, C'_4, and C'_5, and $C'_2 < C_2$, $C'_3 < C_3$, $C'_4 < C_4$, and $C'_5 < C_5$. This results the data access time between the source-sink pair of terminals p1 and p6 can be reduced clearly.

The above concepts can expand to other source-sink pairs on a data bus for isolating unnecessary capacitive loadings by

* Corresponding author. E-mail address: chun@nhu.edu.tw
Tel.: +886-5-2721001#50309

inserting unit-size repeaters into their branch wires to reduce their access times. Thus, we can recalculate the average access time for a program with whole timing periods that data run on the bus. If the saving in average access time with inserted unit-size repeaters is at least 10% (here, we call it the basic time-space ratio depending on the user's definition) than that of without inserted any unit-size repeaters. Then, the data bus can be reconstructed with inserting a number of unit-size repeaters for reducing the average access time to most of multiple programs ran on the bus. Here, we emphasis the inserted repeater with just a unit size due to the limited space of chip area and the minor reconstruction of the data bus.

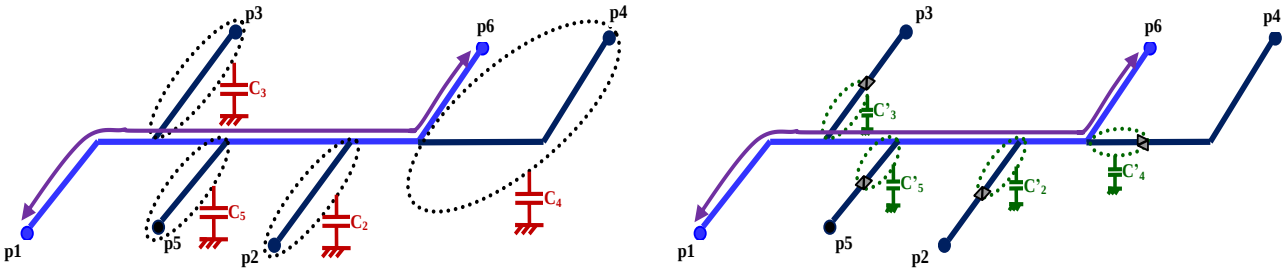


Fig. 1 (a) Extra loading capacitances occurred on the data access of a source-sink pair of terminals p1 and p6 on a local data bus topology and (b) these extra capacitive loadings are isolated and reduced by inserting unit-sized repeaters.

How to evaluate the data bus of a stacked-layer chip to run well for reducing the average access time for multiple programs? A few papers were conducted to this topic. It is the valuable problem for investigation in advance.

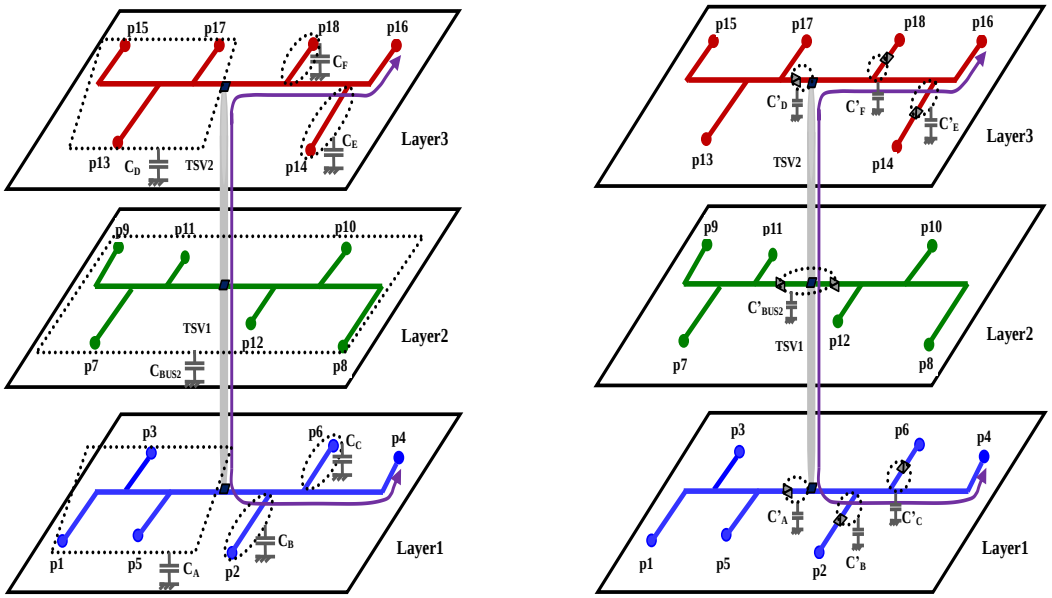
Daneshtalab et al. [2] proposed an appropriate bus isolation strategy for a 3D stacked-layer chip and had a high-performance inter-layer bus structure (HIBS). The HIBS can reduce the complexity of bus arbitrators and make the saving in the propagation delay of data communication. Thakkar et al. [3] introduced a new architecture called 3D-Wiz that is used for reducing the interaction overloading between data bus of DRAMs. The architecture can reduce their access times among any DRAMs. Cho et al. [4] presented the analysis of system bus considering the interconnection of TSVs on a stacked-layer SoC. They found the maximum throughput of the system bus of a 3D stacked-layer chip depending on the data bandwidth. Tsai [5] first conducted repeater insertion to minimize the propagation delay for a 3D data bus based on RC delay model, but they are not to consider the capacitive loading effect of unnecessary local data bus. Tsai [6-7] created some approaches for the effectiveness of critical access time with embedded isolated switches [8-9] and inserted repeaters for a 3D data bus, but no any considerations about the pre-evaluation of data bus.

This work we propose an effective evaluation approach to the 3D data bus of a stacked-layer chip by estimating multiple programs their average access time ran on the data bus whether can be reduced in advance or not. The approach is trying to insert a number of unit-size repeaters to isolate most of extra capacitive loadings to reduce their access times of all the source-sink pairs at different timing periods. The emphasis in unit-size repeater is due to the limitation of chip area and the light construction for the data bus. The demonstrated results show that the most of 3D stacked-layer data buses at a complete timing period their average access times with inserted unit-size repeaters into bus wires can be dramatically reduced for the performance promotion.

2. Problem Formulation

A 3D stacked-layer data bus as shown in Fig. 2 extended from Fig. 1, there is a number of n terminals and exists a maximal number of $n\times(n\text{-}1)$ timing periods among n terminals as well as $n\times(n\text{-}1)$ access times. The number of $n\times(n\text{-}1)$ timing periods is called a complete timing period for a data bus with n terminals. Generally, an executed program has a number of hundreds or thousands timing periods that data are frequently running on the data bus and these timing periods may cover whole complete timing period. If most of data access times at different timing periods for the program can be reduced a little, then its whole average access time will be decreased, that is, the program performance in execution time can thus be promoted.

Fig. 2(a) shows the bidirectional data access between two terminals p4 located on layer1 and p16 located on layer3 on the different timing periods of a 3D stacked-layer data bus. Obviously, their data access times, $T_{p4\text{-}p16}$ and $T_{p16\text{-}p4}$, between p4 and p16 cover those extra loading capacitances, C_A, C_B, C_C, C_D, C_E, C_F and C_{BUS2}. Especially, the total capacitance of local bus on layer2, C_{BUS2}, will be a bigger capacitive loading for their access time. As shown in Fig. 2(b), if we insert a number of unit-size repeaters to some bus wires, then most of each extra loading capacitance can be reduced to be C'_A, C'_B, C'_C, C'_D, C'_E, C'_F and C'_{BUS2} by isolating their branch wire capacitive loadings. Thus the data access times, $T_{p4\text{-}p16}$ and $T_{p16\text{-}p4}$, can be largely reduced.



(a) (b)

Fig. 2 (a) Data access with extra capacitive loadings from the source-sink pair of terminals p4 and p16 and (b) those extra loading capacitances can be isolated and reduced by inserting a number of unit-size repeaters for reducing the access time.

Based on Elmore Π-RC delay model [10], the access time T_{ij} (T_{ji}) from source i (j) to sink j (i) along the path of p4-p16 pair in Fig. 2(a) is represented as below.

$$T_{ij} = \sum_{(f,g)\in \mathrm{path}(i,j)} (R_{d_i} + r_{fg})\left(\frac{c_{fg}}{2} + c(T_g)\right) \quad (1)$$

where R_{di} is the output driving resistance of source i, r_{fg} and c_{fg} are resistance and capacitance of a bus wire (f,g), respectively, and $c(T_g)$ is the lumped capacitance of branch rooted at node g. It is noted that $c(T_g)$ contains those extra capacitive loadings, C_A, C_B, C_C, C_D, C_E, C_F and C_{BUS2}.

Fig. 3 shows the equivalent Π-RC circuit based on Elmore delay model of Fig. 2(b) between terminals p4 located on layer1 and p16 located on layer3 with two TSVs and a number of inserted unit-size bidirectional repeaters for isolating extra loading capacitances. From the figure, extra loading capacitances C_{11}, C_{12}, and C_{13} on layer1 are isolated from the corresponding inserted unit-size repeaters RP_{11}, RP_{12}, and RP_{13}; extra capacitances C_{21} and C_{22} on layer2 are isolated from the corresponding inserted unit-size repeaters RP_{21} and RP_{22}; and extra capacitances C_{31}, C_{32}, and C_{33} on layer3 are isolated from the corresponding inserted unit-size repeaters RP_{31}, RP_{32}, and RP_{33}. A unit-size bidirectional repeater has two sets of input capacitance c_B, intrinsic delay t_B, and output resistance r_B that are inversely connected in parallel. The access time is the scaled-50% propagation delay based on Elmore RC delay model. Likely a bus wire, a TSV has also the equivalent RC mode [11] with the resistance $rTSV$ and two half capacitances of $cTSV$/2. The access time T'_{ij} (T'_{ji}) from source i (j) to sink j (i) along the path of p4-p16 pair with isolated unit-size repeaters is represented as below.

$$T'_{ij} = \sum_{(f,g)\in \mathrm{path}(i,j),\mathrm{RP}_k \notin \mathrm{path}(i,j)} (R_{d_i} + r_{fg})\left(\frac{c_{fg}}{2} + c(T_g)\right) \quad (2)$$

where $c(Tg)$ is the lumped capacitance of branch rooted at node g, but just including the capacitances within those isolated unit-size repeaters RP_k.

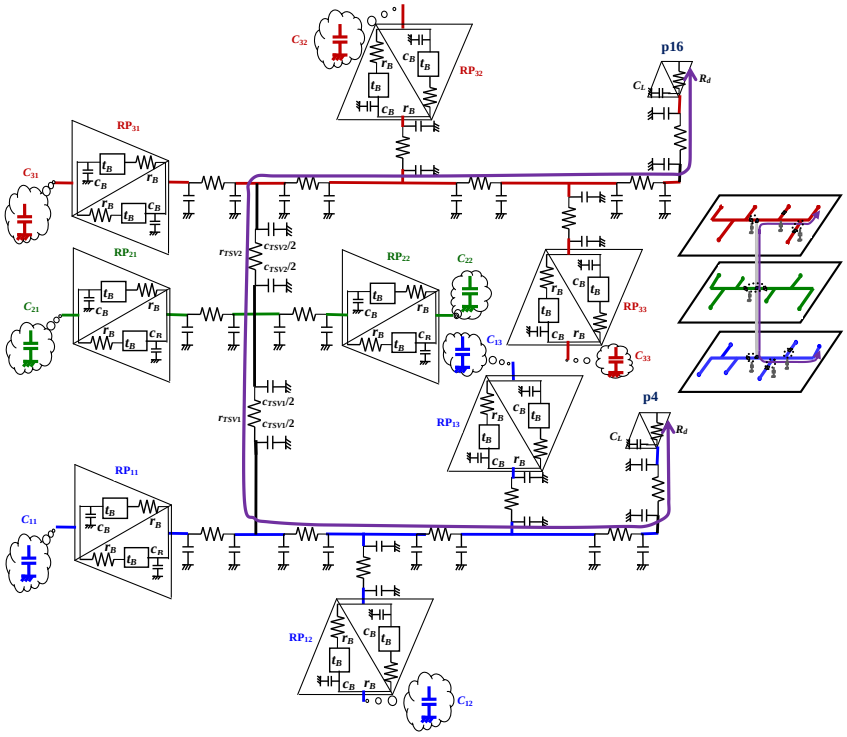


Fig. 3 The equivalent Π-RC circuit of Fig. 2(b) along the path of a source-sink pair of terminals p4 and p16 with two TSVs and a number of isolated unit-size bidirectional repeaters.

For a data bus with a complete timing period, data run on the data bus for whole timing period. Like the above introduction, other data access times among terminals have also the same consideration by inserting a number of unit-size repeaters into the bus wires for reducing their access times. The average access time for a complete timing period is represented its performance. Since a unit-size repeater has also including the input capacitance, output resistance, and intrinsic delay, the access times for all the source-sink pair with inserted repeaters will affect with each other. Thus, we need to evaluate a data bus with inserting unit-size repeaters whether can decrease the average access time of a complete timing or not. That is, we can estimate the saving in the average access time for data running whole complete timing period of a data bus without/with inserting unit-size repeaters into bus wires. If yes, the data bus can be reconstructed with inserting a number of unit-size repeaters for reducing the average access time to most of multiple programs.

Therefore, the problem for evaluating the 3D data bus of a stacked-layer chip in average access time at a complete timing period can be defined as below.

Given the topology of a stacked-layer data bus that has a number of n terminals, a number of q bus wires, and a complete timing period, the objective is to evaluate the possible reconstruction of the data bus by inserting unit-size repeaters into bus wires such that the saving in average access time with inserted unit-sized repeaters is at least the basic time-space ratio than that of without any inserted repeaters, where the basic time-space ratio depending on the user's definition, such as 10 %.

3. Evaluation of a Stacked-Layer Data Bus

3.1. Estimation of a Unit-size Repeater Insertion

To understand the effects in data access time, it is required to make the estimation before/after inserting a unit-size bidirectional repeater into a bus wire. As shown in Fig. 4(a), the access time T_{ij} from source i to sink j along the bus wire l_1 based on the Elmore delay model can be got. If the sink connects a subtree wire segments, then it exists the extra loading capacitance C_S and the access time T_{ij} will be increased and represented as

$$T_{ij} = r_1(c_1/2+C_{Lj}+C_S) + R_{di}(C_{Li}+c_1+C_{Lj}+C_S) \quad (3)$$

where $r_1$and c_1 are the resistance and capacitance of wire l_1, respectively, C_{Li} and C_{Lj} are the input loading capacitances of source i and sink j, respectively and R_{di} is the output driving resistance of source i.

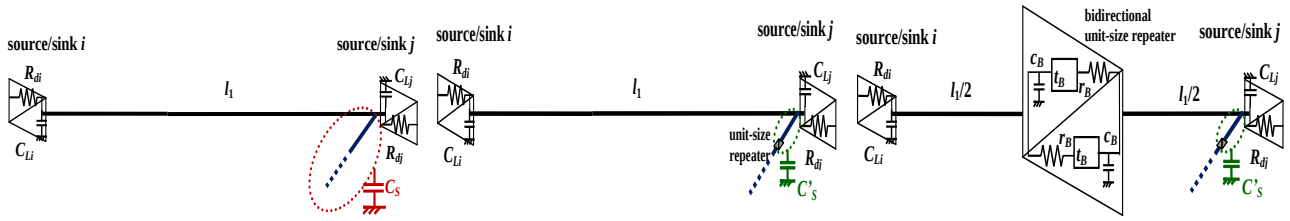


Fig. 4 (a) A subtree wires connected the sink to form an extra capacitance Cs, (b) An inserted unit-size repeater to reduce extra capacitance to be C's, and (c) An inserted bidirectional repeater is inserted into the bus wire for reducing the access time.

To reduce the access time T_{ij}, we can insert a unit-size repeater to isolate the subtree wires that can largely decrease the extra loading capacitance C_S to be C'_S, $C'_S < C_S$, as shown in Fig. 4(b), that is, the equation (3) is updated to be T'_{ij}

$$T'_{ij} = r_1(c_1/2+C_{Lj}+C'_S) + R_{di}(C_{Li}+c_1+C_{Lj}+C'_S) \quad (4)$$

As shown in Fig. 4(c), the access time T'_{ij} from source i to sink j can be reduced in advance by inserting a unit-size bidirectional repeater into the middle of the bus wire l_1 if it was enough longer, that is, the equation (4) is updated as

$$T'_{ij} = r_1/2\ (c_1/4+C_{Lj}+C'_S) + r_B(c_B+c_1/2+C_{Lj}+C'_S) + t_B + r_1/2\ (c_1/4+c_B) + R_{di}(C_{Li}+c_1/2+c_B) \quad (5)$$

where r_B, c_B, and t_B are the output resistance, input capacitance, and intrinsic delay of a unit-size bidirectional repeater, respectively.

For simplification, we assume that C_S is the multiple times of the wire capacitance c_1, that is, $C_S = mc_1$, $m \geq 0$. And C'_S is sum of the half of capacitance c_1 and the input capacitance c_B, i.e., $C'_S = c_1/2+c_B$ if $m > 0$ and $C'_S = 0$ if $m = 0$. If the source and sink are also a bidirectional unit-size repeater, then $R_{di} = r_B$ and $C_{Li} = C_{Lj} = c_B$. The access times T_{ij} and T'_{ij} from source i to j without/with inserted repeaters are respectively derived as follow.

$$\begin{aligned} T_{ij} &= r_1(c_1/2+c_B+C_S) + r_B(2c_B+c_1+C_S) = r_1(c_1/2+c_B+mc_1) + r_B(2c_B+c_1+mc_1) \\ &= (m+1/2)r_1c_1 + r_1c_B + 2r_Bc_B + (m+1)r_Bc_1\,, \quad m \geq 0 \end{aligned} \quad (6)$$

$$\begin{aligned} T'_{ij} &= r_1/2\ (c_1/2+2c_B+C'_S) + r_B(4c_B+c_1+C'_S) + t_B = r_1c_1/4 + r_1c_B + 4r_Bc_B + r_Bc_1 + t_B\,, \ m = 0 \\ &= r_1c_1/2 + 3r_1c_B/2 + 5r_Bc_B + 3r_Bc_1/2 + t_B\,, \quad m > 0 \end{aligned} \quad (7)$$

If m is progressively large, then the access time T_{ij} will be increased, but the access time T'_{ij} always keeps a fixed value that is independent of m. If T'_{ij} with inserted a unit-size repeater into the wire l_1 is always less than T_{ij}, then the reduced access time of $(T_{ij}-T'_{ij})$ is obviously meaningful. Here, we want to know how the wire length l_1 can be inserted a unit-size repeater for effectively reducing the access time.

Case 1: m = 0,

$$T_{ij}-T'_{ij} = r_1c_1/4 - 2r_Bc_B - t_B = (r_wc_w)l_1^2/4 - (2r_Bc_B + t_B) > 0$$

where the units of r_w and r_B are **Ω**, the units of c_w and c_B are pF, and the unit of t_B is ps, and the unit of l_1 is μm. We can derive the wire length l_1 (μm) is

$$l_1 > 2\sqrt{\frac{2r_Bc_B+t_B}{r_wc_w}} \quad (8)$$

Case 2: m > 0,

$$T_{ij} - T'_{ij} = mr_1c_1 - r_1c_B/2 - 3r_Bc_B + (m-1/2)r_Bc_1 - t_B = mr_wc_wl_1^2 - (r_wc_B/2+(1/2-m)r_Bc_w)l_1 - (3r_Bc_B+t_B) > 0$$

The wire length l_1 (μm) can be formulated as

$$l_1 > \frac{0.5r_wc_B+(0.5-m)r_Bc_w+\sqrt{(0.5r_wc_B+(0.5-m)r_Bc_w)^2+4mr_wc_w(3r_Bc_B+t_B)}}{2mr_wc_w} \quad (9)$$

3.2. Access Time Effects with Inserted Unit-size Repeaters

Due to the strategy of isolating capacitive loadings by inserting unit-size repeaters for a source-sink pair of two terminals, the access time of a source-sink pair with the shorter path has larger reduction in extra capacitances than that of the source-sink pair with the longer path. For a data bus with a complete timing period, all the bus wires are almost inserted full unit-size repeaters. The data access time of a source-sink pair with the longer path may increase. Fig. 5 shows its extended data access of a source-sink pair of two terminals p4 and p16 in Fig. 2(b) that has up to inserted six unit-size repeaters, RP_{14}, RP_{15}, RP_{16}, RP_{34}, RP_{35},and RP_{36}, along their longer path. Repeaters RP_{14} and RP_{16} are inserted for isolating extra capacitive loading due to the p2-p6 path, RP_{15} is inserted for the isolation due to the p4-p6 path, RP_{34} and RP_{36} are inserted for the isolation due to the p14-p18 path, and RP_{35} is inserted for the isolation due to the p14-p16 path. The access time T'_{ij} from source i to sink j along the path of p4-p16 pair with inserted unit-size repeaters is formulated as below.

$$T'_{ij} = \sum_{(f,g),RP_x \in path(i,j),RP_k \notin path(i,j)} (R_{d_i} + r_{fg} + r_{B_x})(\frac{c_{fg}}{2} + c_{B_x} + c(T_g)) + t_{B_x} \quad (10)$$

where RP_k is the number of isolated unit-size repeaters that are not located on the path of p4-p16 pair and RP_x is the number of inserted repeaters that are located on the path of p4-p16 pair.

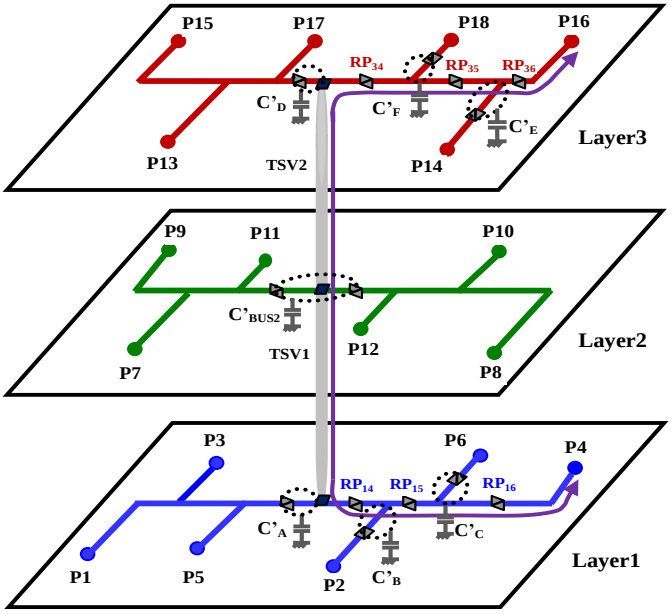


Fig. 5 The data access of the source-sink pair of p4-p16 extended from Fig. 2(b) that has six inserted unit-size repeaters, RP_{14}, RP_{15}, RP_{16}, RP_{24}, RP_{25}, and RP_{26}.

3.3. Algorithm for Evaluating Stacked-Layer Data Bus with Isolated Unit-size Repeater Insertion

The algorithm Evaluate_Stacked-layer_DataBus() is introduced in Fig. 6 for solving the above defined problem. The initial step is to read a 3D data bus topology to construct their data structure. Then, we calculate the average access time T_{av} of an original 3D data bus without any inserted repeaters for a complete timing period using the function Find_AverageAccessTime(). In step3, the *for* loop for each timing of a complete timing period to insert a number of unit-size bidirectional repeaters into the middle of all the branch bus wires along the path of each source-sink pair estimated by Eqs. (8) and (9) for isolating the branch capacitive loadings, but at most a repeater is inserted into the middle of a bus wire. The new average access time *U-*T_{av} of a 3D data bus with inserted unit-size repeaters for a complete timing period using the same function Find_AverageAccessTime() in step4. Finally, if the saving *U-saving* in average access time defined as (*Tav* – *U-Tav*) / *Tav**100% is larger than the basic time-space ratio 10%, then, the 3D data bus can be reconstructed by inserting a number of unit-size bidirectional repeaters in the space depending on the limited chip area. Otherwise, give up the reconstruction of a 3D data bus topology.

```
Evaluate_Stacked-layer_DataBus()
{   /* A 3D data bus topology with the number of n terminals,
       q bus wires, and a complete timing period. */
   step1: Scan a 3D data bus topology and construct its data
          structure.
   step2: Compute each source-sink access time of n(n-1) timing
          periods and get the whole average access time Tav
          by the function of Find_AverageAccessTime().
   step3: for (each timing of a complete timing period)
     {   insert a number of unit-size repeaters to isolate those all
         the extra capacitive loadings form the source-sink pair;
         but at most a repeater is inserted the middle of a bus wire.
     }
   step4: Estimate each source-sink access time of n(n-1) timing
          periods with inserted unit-size repeaters and calculate
          the whole average access time U-Tav.
   step5: if (U-saving = (Tav – U-Tav) / Tav * 100%
                         > basic time-space ratio 10%)
       then, the 3D data bus can be reconstructed by inserting
             a number of unit-size bidirectional repeaters into
             the space depending on a limited chip area.
       else, give up the reconstruction of a 3D data bus topology.
}
```

Fig. 6 An algorithm is to evaluate a stacked-layer data bus with isolated unit-size repeater insertion at a complete timing period for the average access time reduction.

The time complexity of the proposed algorithm is $O(n^2)$ because the $n(n\text{-}1)$ timing periods are executed, where n is the number of terminals.

4. Experimental Results

We have implemented the proposed algorithm in C language on an i7 CPU@2.7GHz, dual cores with 8GB RAM, running MS-Windows 10. Table 1 shows the parameters of 45nm technology [12] based on Elmore RC delay model. Terms r_w and c_w represent the resistance and capacitance of a unit-length wire, respectively. *rTSV* and *cTSV* are the resistance and capacitance of a TSV, respectively. r_B, c_B, and t_B denote the output resistance, input capacitance, and intrinsic delay of a unit-size repeater, respectively.

Table 1 Parameters based on 45nm technology.

a unit-length wire		a TSV		a unit-size repeater		
r_w	c_w	*rTSV*	*cTSV*	r_B	c_B	t_B
0.1Ω	0.2fF	0.035Ω	15.48fF	122Ω	24fF	17ps

We refer six 3D data bus topologies with 3 stacked layers from [5-7] and reduce them in total length by five times for testing our proposed algorithm. For a data bus, the driving resistance R_d of all the sources and the loading capacitance C_L of all the sinks are assumed to be those of a unit-size repeater. The inserted repeaters to the wires are also fixed by a unit size due to the limited space of chip area and the minor reconstruction of the data bus.

Table 2 shows the evaluations in average access time for six 3D data bus topologies that their total lengths are reduced by 5 (marked with r5) with a complete timing period (marked with -*nxn*) and 2000 timing periods (marked with -2k), respectively. In the table, *#Term* , *#Loc*, *Tlength*, and *#Peri* are the number of terminals, number of bus wires, total wire length, and number of timing periods, respectively, of a 3D data bus. *Tav* and *U-Tav* are the average access times without/with inserting a number of *U-size* unit-sized repeaters, respectively. *U-Saving* is the saving ratio defined as (*Tav- U-Tav*)/*Tav**100%. We always insert a bidirectional unit-size repeater into each bus wire for conducting the timing periods, thus their *U-size* is double to the bus wires *#Loc*. For all the cases with complete timing periods (marked with $r5-*nxn*) and 2000 timing periods (marked with $r5-2k), their

corresponding *U-Tav* and *U-saving* are almost equivalent with each other, for example, the *U-saving*s of Test0r5-*nxn* and Test0r5-2k are 37.09% versus 37.27%. These average access times, *U-Tav*s, have better savings, *U-saving*s, in the range of 37.09% to 60.88% and they are always larger than the basic time-space ratio 10%. The results show that all the cases are suitable to reconstruct their data bus by inserting a number of unit-size repeaters for reducing the average access time to any programs with a number of hundred or thousand timings running on the data bus.

Table 2 Evaluation in average access times *Tav* and *U-Tav* of six 3D data buses without/with inserted unit-sized repeaters for their total length reduced by 5 with a complete timing period and 2000 timing periods.

Example	*#Term*	*#Loc*	*Tlength*	Complete timing period ($r5-*nxn*)					2k timing periods ($r5-2k)				
				#Peri	*Tav*(ns)	*U-Tav*(ns)	*U-size*	*U-saving*	*#Peri*	*Tav*(ns)	*U-Tav*(ns)	*U-size*	*U-saving*
Test0r5-*	18	38	7510μm	306	0.3625	0.2281	76	37.09%	2000	0.3627	0.2275	76	37.27%
CaseFr5-*	15	29	12069μm	210	0.6078	0.2378	58	60.88%	2000	0.6099	0.2386	58	60.88%
CaseGr5-*	10	21	8797μm	90	0.4419	0.2270	42	48.62%	2000	0.4405	0.2240	42	49.16%
CaseHr5-*	9	20	8538μm	72	0.4393	0.2380	40	45.73%	2000	0.4392	0.2398	40	45.40%
CaseJr5-*	21	44	11166μm	420	0.6117	0.2917	88	52.31%	2000	0.6104	0.2903	88	52.43%
CaseKr5-*	30	58	13776μm	870	0.7686	0.3059	116	60.20%	2000	0.7556	0.3075	116	59.30%

*: *nxn* or 2k

We extend the evaluation for all the cases that their total lengths are reduced by 10 (marked with r10). Table 3 shows their corresponding *U-saving*s of the cases with a complete timing period (marked with $r10-*nxn*) and 2000 timing periods (marked with $r10-2k). Like the evaluations in Table 2, they are almost equivalent with each other. It is noted that three cases Test0r10-*nxn* (Test0r10-2k), CaseGr10-*nxn* (CaseGr10-2k), and CaseHr10-*nxn* (CaseHr10-2k) with the complete timing periods (2000 timing periods), their corresponding *U-saving*s have -8.58% (-9.79%), 7.24% (7.27%), and 0.33% (-0.1%) under the basic time-space ratio 10%. Obviously, these three cases of data bus are not suitable for inserting a number of unit-size repeaters for reducing their access times.

Table 3 Evaluation in average access times *Tav* and *U-Tav* of six 3D data buses without/with inserted unit-sized repeaters for their total length reduced by 10 with a complete timing period and 2000 timing periods.

Example	*#Term*	*#Loc*	*Tlength*	Complete timing period ($r10-*nxn*)					2k timing periods ($r10-2k)				
				#Peri	*Tav*(ns)	*U-Tav*(ns)	*U-size*	*U-saving*	*#Peri*	*Tav*(ns)	*U-Tav*(ns)	*U-size*	*U-saving*
Test0r10-*	18	38	3736μm	306	0.1856	0.2015	76	**-8.58%**	2000	0.1857	0.2038	76	**-9.79%**
CaseFr10-*	15	29	6020μm	210	0.2703	0.1889	58	30.11%	2000	0.2703	0.1884	58	30.28%
CaseGr10-*	10	21	4388μm	90	0.1952	0.1810	42	**7.27%**	2000	0.1942	0.1801	42	**7.24%**
CaseHr10-*	9	20	4259μm	72	0.1908	0.1901	40	**0.33%**	2000	0.1907	0.1909	40	**-0.10%**
CaseJr10-*	21	44	5561μm	420	0.2826	0.2479	88	12.27%	2000	0.2830	0.2480	88	12.35%
CaseKr10-*	30	58	6859μm	870	0.3622	0.2601	116	28.18%	2000	0.3521	0.2610	116	25.89%

*: *nxn* or 2k

Table 4 presents the evaluation in average access time for the data bus topology CaseK that the total length is reduced by 5 or 10 (marked with r5 or r10) with a complete timing period (marked with *-nxn*) or different number of timing periods (marked with -0.1k to -5k). From the table, three examples, CaseKr5-*nxn* and CaseKr10-*nxn*, with complete timing periods and total length reduced by 5 and 10 that their *U-saving*s are 60.20% and 28.18%, respectively. For the cases CaseKr5-.1k to CaseKr5-5k of a 3D data bus with different timing periods, their average access times have good *U-saving*s in range of 52.44% to 60.42% and they are suitable for inserting a number of unit-size repeaters for access time reduction. For the cases CaseKr10-.1k to CaseKr10-5k of a 3D data bus with different timing periods, three cases CaseKr10-.1k, CaseKr10-.2k, and CaseKr10-.3k are -1.82%, 4.53%, and 8.04%, respectively that their *U-saving*s are not larger than the basic time-space ratio 10% and they are not suitable for inserting a number of unit-size repeaters for reducing access time.

Table 5 shows the comparison of our approach and the minimized critical access-time method [5] for all the cases of 3D data bus with complete timing periods and total length reduced by 5. The method is always to minimize the critical access time for all the timing periods until no any improvement. From the table, although the saving in average access time of our approach has 9.35% (i.e., 60.16%-50.81%) less than that of the method, but the running time has speed up to 21,252 times (i.e., 21.252s/0.001s) and the sized repeaters is reduced by 4 (i.e., 74-70) sizes. Therefore, our approach still has capable for rapidly evaluating a data bus of stacked-layer chip.

Table 4 Evaluation in average access times *Tav* and *U-Tav* for CaseK without/with inserted unit-sized repeaters for their total length reduced by 5 and 10 with a complete timing period and different timing periods.

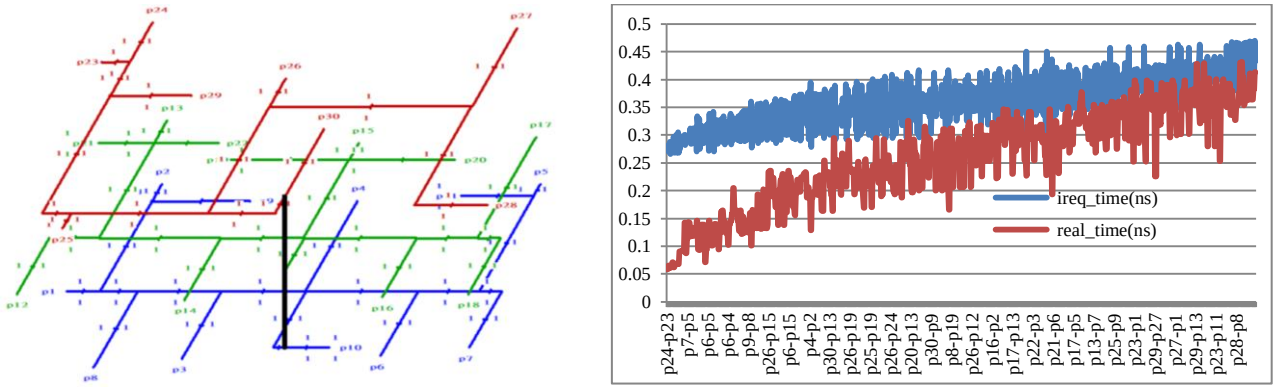
Example	*#Term*	*#Loc*	*#Peri*	Total length reduced by 5 (CaseKr5-*k)					Total length reduced by 10 (CaseKr10-*k)				
				Tlength	*Tav*(ns)	*U-Tav*(ns)	*U-size*	*U-saving*	*Tlength*	*Tav*(ns)	*U-Tav*(ns)	*U-size*	*U-saving*
CaseKr?-nxn	30	58	870	13776µm	**0.7686**	**0.3059**	116	**60.20%**	6859µm	**0.3622**	**0.2601**	116	**28.18%**
CaseKr?-.1k	30	58	100	13776µm	0.6324	0.3008	116	52.44%	6859µm	0.2581	0.2628	116	**-1.82%**
CaseKr?-.2k	30	58	100	13776µm	0.6446	0.3019	116	53.16%	6859µm	0.2710	0.2588	116	**4.53%**
CaseKr?-.3k	30	58	300	13776µm	0.6757	0.3103	116	54.07%	6859µm	0.2798	0.2573	116	**8.04%**
CaseKr?-.5k	30	58	500	13776µm	0.6875	0.3053	116	55.60%	6859µm	0.2964	0.2594	116	12.48%
CaseKr?-.7k	30	58	700	13776µm	0.6960	0.3067	116	55.94%	6859µm	0.3100	0.2610	116	15.82%
CaseKr?-1k	30	58	1000	13776µm	0.7288	0.3086	116	57.66%	6859µm	0.3252	0.2599	116	20.08%
CaseKr?-2k	30	58	2000	13776µm	0.7556	0.3075	116	59.30%	6859µm	0.3521	0.2610	116	25.89%
CaseKr?-3k	30	58	3000	13776µm	0.7619	0.3038	116	60.13%	6859µm	0.3602	0.2613	116	27.45%
CaseKr?-4k	30	58	4000	13776µm	0.7636	0.3022	116	60.42%	6859µm	0.3603	0.2596	116	27.95%
CaseKr?-5k	30	58	5000	13776µm	0.7670	0.3059	116	60.12%	6859µm	0.3614	0.2588	116	28.38%

?: by 5 or 10 -.3k: 3000 timing periods

Table 5 Comparison of all the cases in average access time with inserted unit-sized repeaters and critical access-time reduced method [5]. These cases have complete timing periods and their total lengths are reduced by 5.

Example	*#Term*	*#Loc*	*Tlength*	*#Peri*	*Tav*(ns)	Our approach				Minimized critical access-time method [5]			
						U-Tav(ns)	*U-size*	*U-time*	*U-saving*	*W-Tav*(ns)	*W-size*	*W-time*	*W-saving*
Test0r5-*nxn*	18	38	7510µm	306	0.3625	0.2281	76	0.001s	37.09%	0.1746	74	8.163s	51.84%
CaseFr5-*nxn*	15	29	12069µm	210	0.6078	0.2378	58	0.001s	60.88%	0.2185	82	2.010s	64.06%
CaseGr5-*nxn*	10	21	8797µm	90	0.4419	0.2270	42	0.001s	48.62%	0.2001	45	0.420s	54.73%
CaseHr5-*nxn*	9	20	8538µm	72	0.4393	0.2380	40	0.001s	45.73%	0.1840	53	0.281s	58.12%
CaseJr5-*nxn*	21	44	11166µm	420	0.6117	0.2917	88	0.001s	52.31%	0.2126	94	25.281s	65.24%
CaseKr5-*nxn*	30	58	13776µm	870	0.7686	0.3059	116	0.002s	60.20%	0.2538	103	91.361s	66.98%
Average	-					-	**70**	**0.001s**	**50.81%**	-	**74**	**21.252s**	**60.16%**

Fig. 7(a) shows the 3D data bus topology of CaseKr10-*nxn* with inserted 116 unit-size repeaters. The average access time is 0.2601 ns and the access time for the shorter length (118 µm) is 0.0585 ns from source p23 to sink p24 and the access time for the longer length (1812µm) is 0.4132 ns from source p28 to sink p9. In the figure, two numbers located on the middle of a bus wire are the sizes of an inserted unit-size bidirectional repeater. Fig. 7(b) presents all the access times to each source-to-sink of a complete timing period without/with inserted repeaters. Each pair of source-sink access time (marked with real_time) with inserted repeaters is always less than that the access time (marked with ireq_time) without inserted repeaters. The average access times without and with inserted repeaters are 0.3622 ns and 0.2601 ns, respectively.



(a) (b)

Fig. 7 For the case CaseKr10-nxn, (a) its 3D data bus topologies with inserted unit-sized repeaters and (b) access times for each source-to-sink of a complete timing period without/with inserted repeaters.

5. Conclusions

The proposed approach based on isolated unit-size repeater insertion for a stacked-layer data bus has been successfully applied for the evaluation of a reconstructed stacked-layer data bus topology in average access time. The approach is very simple and effective and is also used for evaluating the average access time of a program ran on the bus with the number of hundreds or thousands timing periods. Extending work is to update the evaluated approach such that can suit for the various

data bus topologies of modern stacked-layer chips.

Acknowledgment

This work was partially supported by NHU-107 research project subsidy of Nanhua University.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] EE Times, The state of the art in 3D IC technologies, Nov. 27, 2013.
[2] M. Daneshtalab, M. Ebrahimi, and J. Plosila, “HIBS- Novel Inter-layer Bus Structure for Stacked Architectures,” in Proc. IEEE International Conf. on 3D System Integration, 2012, pp. 1-7.
[3] Ishan G Thakkar and Sudeep Pasricha, “3D-Wiz: A novel high bandwidth, optically interfaced 3D DRAM architecture with reduced random access time,” in Proc. IEEE International Conf. on Computer Design, 2014, pp. 1-7.
[4] Kyoungrok Cho, Hyeon-Seok Na, Tae Won Cho, and Younggap You, “Analysis of system bus on SoC platform using TSV interconnection,” in Proc. The 4th ASQED, 2012, pp. 255-259.
[5] Chia-Chun Tsai, “Repeater Insertion for 3D Data Bus with TSVs for Reducing Critical Propagation Delay,” in Proc. International Conference on Computer Science and Information Engineering, 2015, pp. 203-208.
[6] Chia-Chun Tsai, “Minimizing Critical Access Time for 3D Data Bus Based on Inserted Bus Switches and Repeaters,” in Proc. IEEE Computer Society Annual Symposium on VLSI, 2017, pp. 140-145.
[7] Chia-Chun Tsai, “Embedded Bus Switches on 3D Data Bus for Critical Access Time Reduction,” in Proc. The 9th IEEE Latin American Symposium on Circuits and Systems (LASCAS), 2018, Paper ID-16.
[8] IDTQS3245 data sheet, IDT Co., Nov. 2014.
[9] 74VHCT126AFT data sheet, Toshiba Co., Nov. 2014.
[10] W. C. Elmore, “The transient response of damped linear networks,” Journal of Applied Physics, Vol. 19, pp. 55-63, 1948.
[11] T. Bandyopadhyay, K-J Han, D. Chung, R. Chatterjee, M. Swaminathan, and R. Tummala, “Rigorous electrical modeling of through silicon vias with MOS capacitance effects,“ IEEE Trans. Components, Packaging, and Manufacturing Technology, Vol. 1, No. 6, pp. 893-903, Jun 2011.
[12] http://ptm.asu.edu/.