

Performance 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

Received 13 March 2019; received in revised form 11 April 2019; accepted 10 May 2019

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 average data access time of a data bus to the timing periods dominates the program performance. In this paper, we proposed an evaluated approach to reconstruct a 3D data bus with inserted unit-size repeaters to motivate that the average data access time of the bus on a complete timing period can speed up at least 10%. The approach 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 at each timing period to reduce their access time. The above process is repeated until no any improvement for each access time. Each inserted repeater with just one unit size due to the limited space of a chip area and the minor reconstruction of a data bus in practical. The approach has the advantages of uniform repeater insertion, less extra area occupation, and simplified time-to-space tradeoff. Experimental results show that our approach has the rapid capable evaluation for a stacked-layer data bus within one millisecond and the saving in average access time is up to 50.81% with the inserted repeater sizes of 70 on average.

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 (Through Silicon Vias) is used to vertically connect these local data buses to integrate them to be a 3D 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 as the propagation delay 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 number of timing periods. The average access time dominates the program performance.

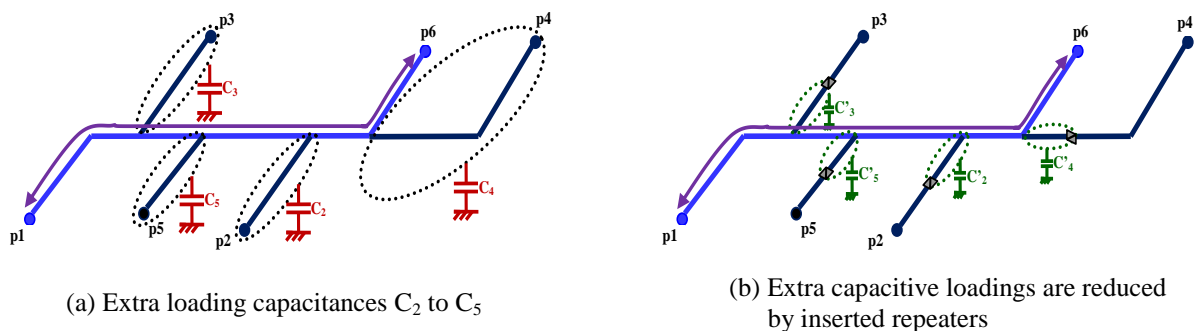


Fig. 1 Data access on the source-sink pair p1-p6 of a 2D local data bus

* Corresponding author. E-mail address: chun@nhu.edu.tw
Tel.: +886-5-2721001#5030

In nanotechnology, a longer interconnection wire always dominates the propagation delay more than a gate delay because of their incremental wire resistances and capacitances. 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 cause to increase the data access time of the source-sink pair of p1-p6. Each extra loading capacitance comes from their branch wire capacitance and terminal capacitive loading along the path of p1-p6 or p6-p1. As shown in Fig. 1(b), most of these extra capacitive loadings can be isolated by inserting a bidirectional repeater into each branch wire for the data bus reconstruction. That is, these extra loading capacitances will be 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 $C'_5 < C_5$. The data bus reconstruction will result that the data access time between two source-sink pairs with terminals p1 and p6 can be clearly reduced.

The above concept of reconstructing a data bus can be expanded to other source-sink pairs for isolating unnecessary capacitive loadings by inserting repeaters into their branch wires to reduce their access times. For a program ran on a data bus with a number of timing periods, its average access time can thus be reduced and its performance can also be upgraded. However, all the inserted repeaters will also cause extra area occupation. This is the time-to-space tradeoff of a data bus reconstruction, such as the saving in average access time is at least 10% with paying a number of repeater sizes.

The repeater insertion was widely applied to a one-way signal path that can effectively reduce their propagation delay, but a few papers discussed the repeater insertion to apply a bidirectional data bus. Ismail [2] proposed repeater insertion for the path delay of an RLC-based wire to estimate the delay and their inductive effects of on-chip interconnects. Lin [3] presented buffer insertion to construct a clock tree on multimode multivoltage islands. They used adjustable delay buffers (ADB) for controlling the clock delay under a boundary skew. Ghoneima [4] introduced the optimal positioning of interleaved repeaters in bidirectional buses. His solution was in focus to reduce noise interference between buses. Acton [5] summarized some studies of signal repeater insertion in multi-source multi-sink data bus. Daneshtalab et al. [6] 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. [7] 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. [8] presented the analysis of system bus considering the interconnection of TSVs on a stacked-layer SoC (System-on a Chip). They found the maximum throughput of the system bus of a 3D stacked-layer chip depending on the data bandwidth. Mohamed [9] introduced a master-slave data access by adding NoCs (Network-on-Chips) among multiple processors and there was a number of data interchange rules that would limited the access time between processors. Khan et al. [10] analyzed the performance for current NoC simulation tools in terms of latency, throughput, and energy consumption, but this comparison was just for 2D NoCs. Tsai [11-12] first conducted repeater insertion and sized the repeaters to minimize the propagation delay for a 3D data bus based on RC delay model, but they do not to consider the capacitive loading effect of un-accessed local data buses. Tsai [13] created an effective method associated with embedded isolated switches [14-15] and inserted repeaters for a 3D data bus to reduce their critical access time, but no any considerations about the pre-evaluation in average access time for a data bus reconstruction.

Most of the above reconstructed data bus methods were based on the repeater insertion and sized them as possible for maximally reducing the data access time. These approaches can decrease the access time effectively, but their data bus would be required to have extra areas for inserting different-size repeaters. This causes the incremental difficulty for reconstructing a data bus at the post refinement step in physical design. The above problem for the optimal solution in the time-to-space tradeoff (data access time minimization vs. repeaters' locations and sizes) by inserting repeaters into a data bus had been approved to be intractable [16]. How to evaluate the data bus of a stacked-layer chip to run well for reducing the average access time? A few papers conduct to this topic and it is the valuable problem for investigation in advance.

In this work, we proposed an approach to evaluate the bus performance by reconstructing a 3D data bus. With inserting unit-size repeaters into a data bus at each timing period, the average data access time of a bus on a complete timing period can be motivated to speed up at least 10% (here, we call it as *basic performance ratio*). The approach is trying to insert a number of unit-size repeaters to isolate most of extra capacitive loadings to reduce the access time of each source-sink pair at different timing periods. This process is repeatedly done until no any improvement for each access time. Then we can estimate the new average access time of a data bus on a complete timing period. If the saving in average access time with inserted unit-size repeaters is larger than the basic performance ratio, the reconstructed data bus can be accepted for reducing the average access time and applied for most of multiple programs ran on the bus. Here, we emphasize the inserted repeater with just a unit size due to the limited space of a chip area and the minor reconstruction of a data bus. The evaluated approach has advantages: uniform repeaters, less extra occupied area, and simplified the time-to-space tradeoff with a basic performance ratio. The demonstrated results show that most of 3D stacked-layer data buses with inserted unit-size repeaters their average access time for any program can be dramatically reduced.

2. Problem Formulation

2.1. Symbols and definitions

Table 1 shows all the symbols and their definitions that are used to go through the whole article for accordance.

Table 1 Symbols and their definitions

Symbol	Definition	Symbol	Definition
n	The total number of terminals of a 3D data bus	q	The total number of bus wires of a 3D data bus
$n(n-1)$	The complete timing period of a 3D data bus	pk	The kth terminal on a 3D data bus
T_{ij}	The access time from source i to sink j without any inserted unit-size repeaters	T'_{ij}	The access time from source i to sink j with inserted unit-size repeaters
T_{av}	The average access time of a 3D data bus without any inserted unit-size repeaters	$U-T_{av}$	The average access time of a 3D data bus with inserted unit-size repeaters
r_w	The resistance of a unit-length wire	$U-size$	The number of inserted unit-size repeaters
c_w	The capacitance of a unit-length wire	RP_k	The kth unit-size repeater
r_{TSV}	The resistance of a TSV	r_B	The resistance of a unit-size repeater
c_{TSV}	The capacitance of a TSV	c_B	The capacitance of a unit-size repeater
r_{fR}	The resistance of a segment wire (f,g)	t_B	The intrinsic delay of a unit-size repeater
c_{fR}	The capacitance of a segment wire (f,g)	R_{di}	The output driving resistance of a source i
r_1	The resistance of a bus wire l1	C_{Lj}	The input loading capacitance of a sink j
c_1	The capacitance of a segment wire l1	$c(T_g)$	The lumped capacitance at node g
C_A	The total capacitance of a wiring area A	C_S	The extra capacitance at a node
C'_A	The reduced total capacitance of a wiring area A with inserted unit-size repeaters	C'_S	The reduced extra capacitance at a node with inserted unit-size repeaters
C_{BUSj}	The total capacitance of the jth-layer local bus, e.g., CBUS2	m	The multiple times of wire capacitance c_1 e.g., $CS = mc_1, m \geq 0$

2.2. Gartner's hype cycle phases

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 $n \times (n-1)$ timing periods as well as a number of $n \times (n-1)$ data access times. The number of $n \times (n-1)$ timing periods is called the *complete timing period* of a data bus. 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 a complete timing period. If most of data access times for a program at different timing periods can be reduced a little, then its average access time will be decreased, that is, the program performance can thus be promoted.

Fig. 2(a) shows the bidirectional data access between two terminals p4 located on layer1 and p16 located on layer3 at the different timing periods of a 3D stacked-layer data bus. Obviously, their data access times, T_{p4-p16} and T_{p16-p4} , 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 data access. As shown in Fig. 2(b), if we insert a number of unit-size repeaters to some bus wires, then most of extra loading capacitances can be reduced to be $C'_A, C'_B, C'_C, C'_D, C'_E, C'_F$ and C'_{BUS2} , respectively. Thus, their data access times, T_{p4-p16} and T_{p16-p4} , can be effectively reduced.

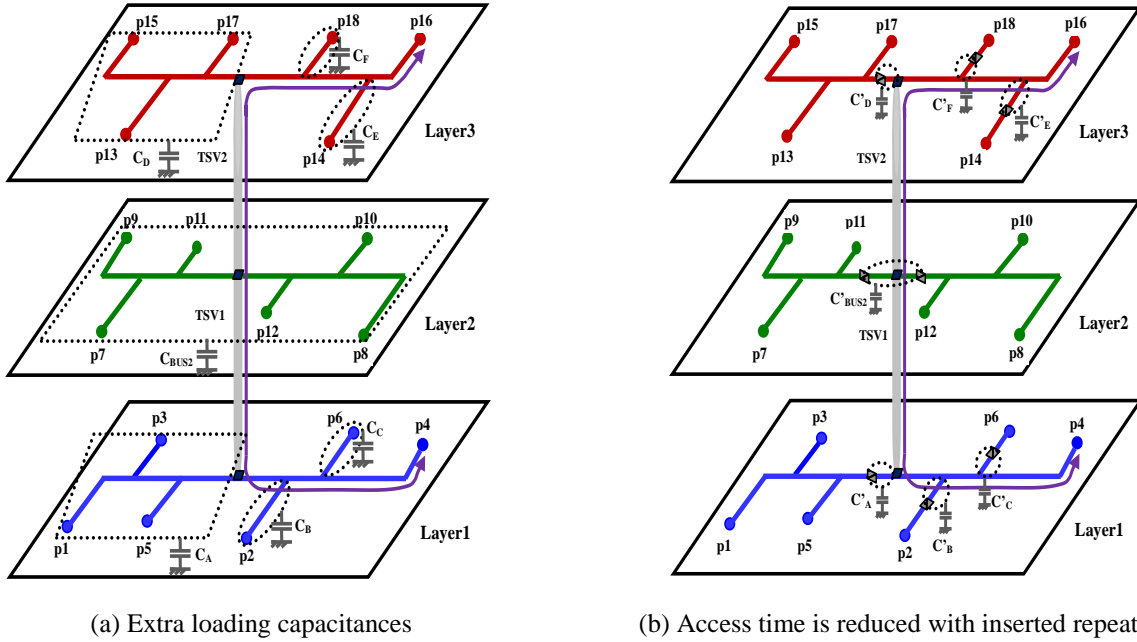


Fig. 2 Data access on the source-sink pair p4-p16 of a 3D stacked layer

In Fig. 2(a), the access time T_{ij} (T_{ji}) from source i (j) to sink j (i) along the path of the source-sink pair of p4-p16 based on Elmore Π -RC delay model [17] is represented as below.

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

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 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 inserted unit-size repeaters RP_{11}, RP_{12} , and RP_{13} ; extra capacitances C_{21} and C_{22} on layer2 are isolated from the 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 inserted unit-size repeaters RP_{31}, RP_{32} , and RP_{33} . A unit-size bidirectional repeater has two equivalent 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 [18] with the resistance r_{TSV} and two half capacitances of $c_{TSV}/2$. The access time T'_{ij} (T'_{ji}) from source i (j) to sink j (i) along the path of a source-sink pair of p4-p16 with isolated unit-size repeaters is represented as below.

$$T'_{ij} = \sum_{(f,g) \in path(i,j), RP_{(k)} \notin path(i,j)} (R_{di} + r_{fg}) \left(\frac{c_{fg}}{2} + C(T_g) \right) \tag{2}$$

where $c(T_g)$ is the lumped capacitance of branch rooted at node g including the capacitances within those isolated unit-size repeaters $RP_{(k)}$.

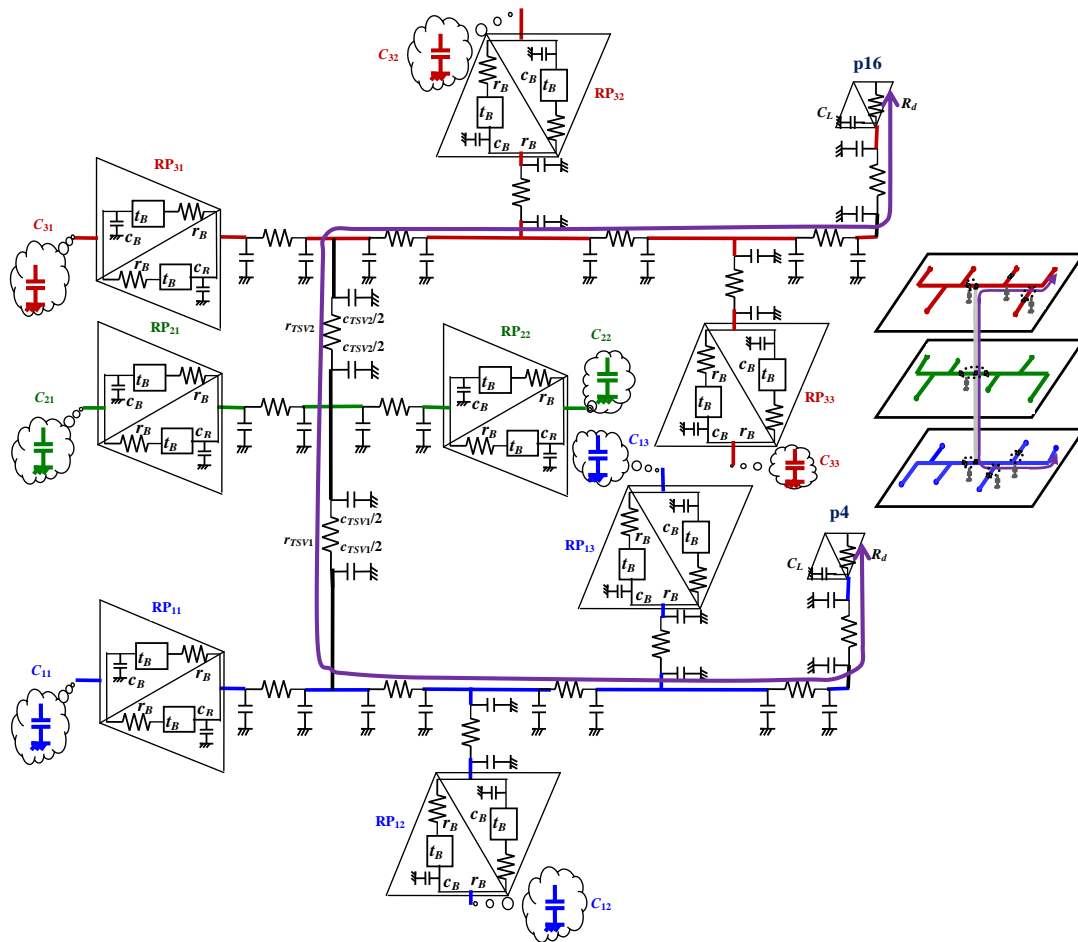


Fig. 3 The equivalent Π-RC circuit of a source-sink pair p4-p16 shown in Fig. 2(b)

For a reconstructed data bus with inserted unit-size repeaters, the average access time on a complete timing period is represented as the bus 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 pairs with inserted repeaters will be affected with each other. Thus, we need to estimate a data bus with inserted unit-size repeaters whether its average access time on a complete timing period is decreased or not. That is, we can calculate the saving percentage in the average access time on a complete timing period for the data bus without/with inserted unit-size repeaters. If the saving is over the basic performance ratio, the data bus can be reconstructed with inserted a number of unit-size repeaters that has good performance improvement in average access time.

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

Given the topology of a stacked-layer data bus that has a number of n terminals and a number of q bus wires on a complete timing period (i.e., the number of $n \times (n-1)$ timing periods), the objective is to evaluate the possible reconstruction of a data bus by inserting unit-size repeaters into the bus wires such that the saving in average access time with inserted repeaters is at least the basic performance ratio than that of without any inserted repeaters, where the basic performance ratio depending on the user's definition, such as 10 %.

3. Performance Evaluation of a Stacked-Layer Data Bus

3.1. The estimation of a unit-size repeater insertion

To understand the effects in data access time of a source-sink pair, it is required to make the estimation of a data access time 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 obtained. If a sink connects the wire segments of a subtree, then the sink has the extra loading capacitance C_S and the access time T_{ij} will be increased, and T_{ij} is represented as below.

$$T_{ij} = r_1(c_1/2 + C_{Lj} + C_S) + R_{di}(C_{Li} + C_1 + C_{Lj} + C_S) \tag{3}$$

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

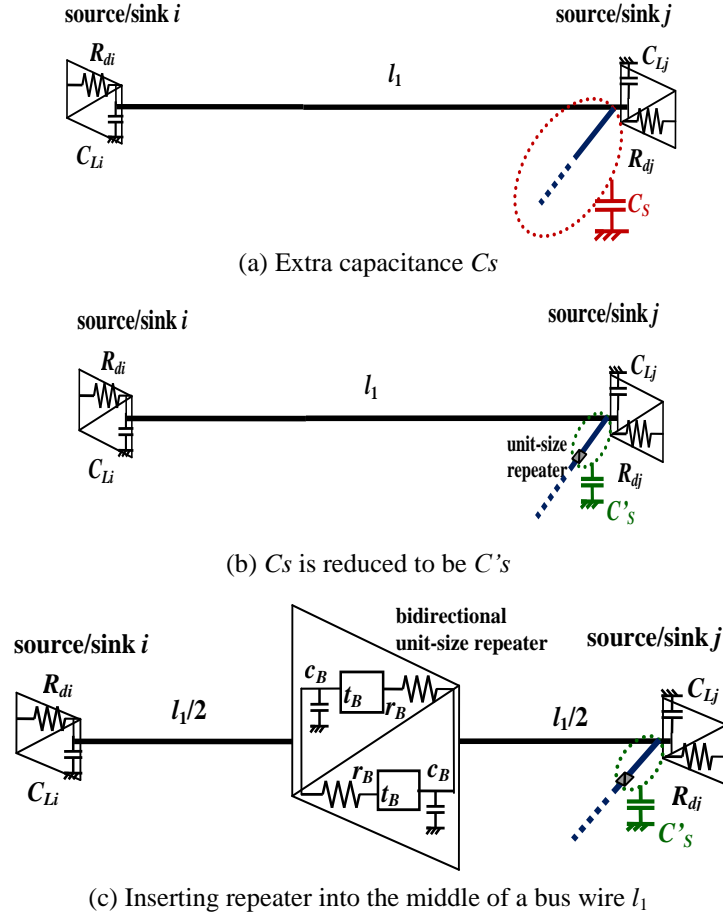


Fig. 4 The bus wire l_1 is inserted into a bidirectional unit-size repeater

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, Eq. (3) is updated to be T'_{ij} and T'_{ij} is denoted as below.

$$T'_{ij} = r_1(c_1/2 + C_{Lj} + C'_S) + R_{di}(C_{Li} + C_1 + C_{Lj} + C'_S) \tag{4}$$

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 a bus wire l_1 if it was enough longer, that is, Eq. (4) is updated as

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

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

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

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 . With inserting a unit-size repeater into the wire l_1 , if its access time T'_{ij} is always less than T_{ij} and then the reduced quantity in 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.

$$T'_{ij} = r_1/2(c_1/4+C_{Lj}+C'_S)+r_B(c_B+c_1/2+C_{Lj}+C'_S)+t_B+r_1/2(c_1/4+c_B)+R_{di}(C_{Li}+c_1/2+c_B) \quad (7)$$

Case 1: $m = 0$,

$$T_{ij} - T'_{ij} = r_1c_1/4 - 2r_Bc_B - t_B = (r_w c_w) l_1^2 / 4 - (2r_Bc_B + t_B) > 0 \quad (8)$$

where the unit of r_w and r_B is Ω , the unit of c_w and c_B is 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

$$l_1 > 2\sqrt{\frac{2r_Bc_B + t_B}{r_w c_w}} \quad (9)$$

Case 2: $m > 0$,

$$\begin{aligned} T_{ij} - T'_{ij} &= mr_1c_1 - r_1c_B/2 - 3r_Bc_B + (m-1/2)r_Bc_1 - t_B \\ &= mr_w c_w l_1^2 - (r_w c_B/2 + (1/2 - m)r_B c_w) l_1 - (3r_Bc_B + t_B) > 0 \end{aligned} \quad (10)$$

The wire length l_1 (μm) can be formulated as

$$l_1 > \frac{0.5r_w c_B + (0.5 - m)r_B c_w + \sqrt{(0.5r_w c_B + (0.5 - m)r_B c_w)^2 + 4mr_w c_w (3r_Bc_B + t_B)}}{2mr_w c_w} \quad (11)$$

3.2. The effects of data access time with inserted unit-size repeaters

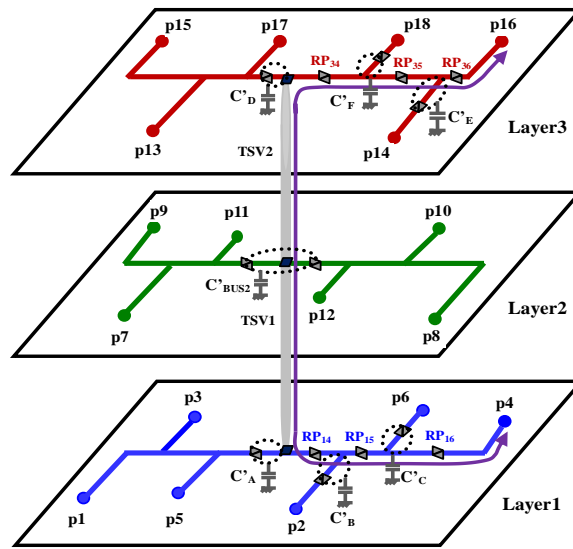
Due to the strategy of extra capacitive loading isolation is adopted by inserting unit-size repeaters, the access time of a source-sink pair for the shorter path has larger reduction in extra capacitances than the longer path. For a data bus on a complete timing period, all the bus wires are almost inserted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

$$T'_{ij} = \sum_{(f,g), RP_{(x)} \in path(i,j), RP_{(k)} \notin path(i,j)} (R_{di} + r_{fg} + r_{Bx}) \left(\frac{C_{fg}}{2} + C_{Bx} C(T_g) \right) + t_{Bx} \quad (12)$$

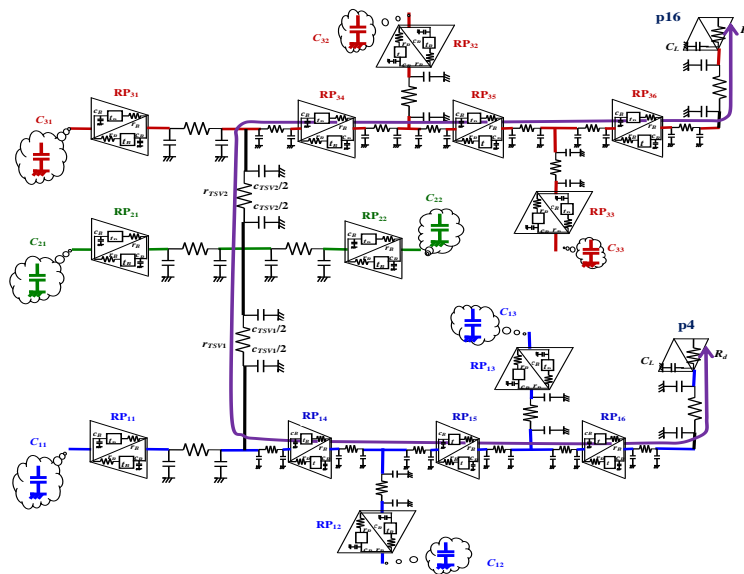
where $RP_{(k)}$ is the number of isolated unit-size repeaters that are not located on the path of p4-p16 and $RP_{(x)}$ is the number of inserted repeaters that are located on the path of p4-p16.

3.3. The evaluation for reconstructing stacked-layer data bus with inserted unit-size repeaters

The evaluated algorithm, *Evaluate_Stacked-layer_DataBus_Reconstruction()*, for the bus performance by reconstructing a stacked-layer data bus with inserted unit-size repeaters is introduced in Fig. 6 to solve the above problem defined in Section 2. 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 on a complete timing period using the function, *Find_AverageAccessTime()*, where T_{av} is defined as the total access times divided by the number of $n \times (n-1)$ timing periods. The *for* loop in step3 is for each timing of a complete timing period and 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. (9) and (11) for isolating the branch capacitive loadings, but at most one 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 on a complete timing period is obtained using the same function *Find_AverageAccessTime()* in step4. Finally, if the saving *U-saving* in average access time defined as $(T_{av} - U-T_{av}) / T_{av} * 100\%$ is larger than the *basic performance 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.



(a) Six inserted unit-size repeaters to the path of p4-p16



(b) The equivalent circuit of the bus structure

Fig. 5 A source-sink pair p4-p16 has six inserted unit-size repeaters and its equivalent circuit


```

Evaluate_Stacked-layer_DataBus_Reconstruction()
{
  /* A 3D data bus topology with the number of  $n$  terminals and
   $q$  bus wires on 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  $(U-saving = (Tav - U-Tav) / Tav * 100%$ 
   $> basic\ performance\ ratio\ 10\%)$ 
  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 The algorithm is used for the evaluation of a data bus performance

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

4. Experimental Results

We have implemented the proposed evaluated algorithm in C language on an i7 CPU@2.7GHz, dual cores with 8GB RAM, running MS-Windows 10. Table 2 shows the parameters of 45nm technology [19] based on Elmore RC delay model. Terms r_w and c_w represent the resistance and capacitance of a unit-length wire, respectively. r_{TSV} and c_{TSV} 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 2 Parameters based on 45nm technology

a unit-length wire		a TSV		a unit-size repeater		
r_w	c_w	r_{TSV}	c_{TSV}	r_B	c_B	t_B
0.1 Ω	0.2fF	0.035 Ω	15.48fF	122 Ω	24fF	17ps

We refer six 3D data bus topologies with 3 stacked layers from [11-12] and reduce them in total length by five times for testing our proposed algorithm. For a data bus, the driving resistances of all the sources and the loading capacitances of all the sinks are assumed to be those parameters of a unit-size repeater. The inserted repeaters into bus wires are also fixed by a unit size due to the limited space of chip area and the minor reconstruction of a data bus.

Table 3 shows the evaluation in average access time for six 3D data bus topologies that their total lengths are reduced by 5 (marked with r5) on their complete timing periods (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 inserted the number of U -size unit-size repeaters, respectively. $U-Saving$ is the saving ratio defined as $(Tav - U-Tav) / Tav * 100%$. Since we always try to insert a bidirectional unit-size repeater into each bus wire for conducting the complete timing periods, thus their number of

U-size unit-size repeaters is near double to the bus wires *#Loc*. For all the cases on their complete timing periods (marked with $\$r5-nxn$) and 2000 timing periods (marked with $\$r5-2k$), their corresponded *U-Tav* and *U-savings* are almost equivalent with each other, for example, the *U-savings* of Test0r5-*nxn* and Test0r5-2k are 37.09% versus 37.27%. These average access times, *U-Tavs*, have better savings, *U-savings*, in the range of 37.09% to 60.88% and they are always larger than the basic performance 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 ran on the bus.

Table 3 The evaluation in average access times *Tav* and *U-Tav* for data buses (their total length is reduced by 5, r5) without/with inserted unit-size repeaters on their complete timing periods and 2000 timing periods, respectively

Example	#Term	#Loc	Tlength	Complete timing period ($\$r5-nxn$)					2k timing periods ($\$r5-2k$)				
				#Peri	Tav (ns)	U-Tav (ns)	U-size	U-savings	#Peri	Tav(ns)	U-Tav(ns)	U-size	U-savings
Test0r5-*	18	38	7510 μ m	306	0.3625	0.2281	76	37.09%	2000	0.3627	0.2275	76	37.27%
CaseFr5-*	15	29	12069 μ m	210	0.6078	0.2378	58	60.88%	2000	0.6099	0.2386	58	60.88%
CaseGr5-*	10	21	8797 μ m	90	0.4419	0.2270	42	48.62%	2000	0.4405	0.2240	42	49.16%
CaseHr5-*	9	20	8538 μ m	72	0.4393	0.2380	40	45.73%	2000	0.4392	0.2398	40	45.40%
CaseJr5-*	21	44	11166 μ m	420	0.6117	0.2917	88	52.31%	2000	0.6104	0.2903	88	52.43%
CaseKr5-*	30	58	13776 μ m	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 4 shows their corresponded *U-savings* of the cases on their complete timing periods (marked with $\$r10-nxn$) and 2000 timing periods (marked with $\$r10-2k$). Like the evaluation in Table 3, their *U-savings* 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) on their complete timing periods (2000 timing periods) are failed because their corresponded *U-savings* have -8.58% (-9.79%), 7.24% (7.27%), and 0.33% (-0.1%) under the basic performance ratio 10%. Obviously, these three-case data buses are not suitable for inserting a number of unit-size repeaters.

Table 4 The evaluation in average access times *Tav* and *U-Tav* for data buses (their total length is reduced by 10, r10) without/with inserted unit-size repeaters on their complete timing periods and 2000 timing periods, respectively

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

*: *nxn* or 2k

Table 5 The evaluation in average access times *Tav* and *U-Tav* for CaseK (the total length is reduced by 5 or 10, r5 or r10) without/with inserted unit-size repeaters on their complete timing periods and different timing periods, respectively

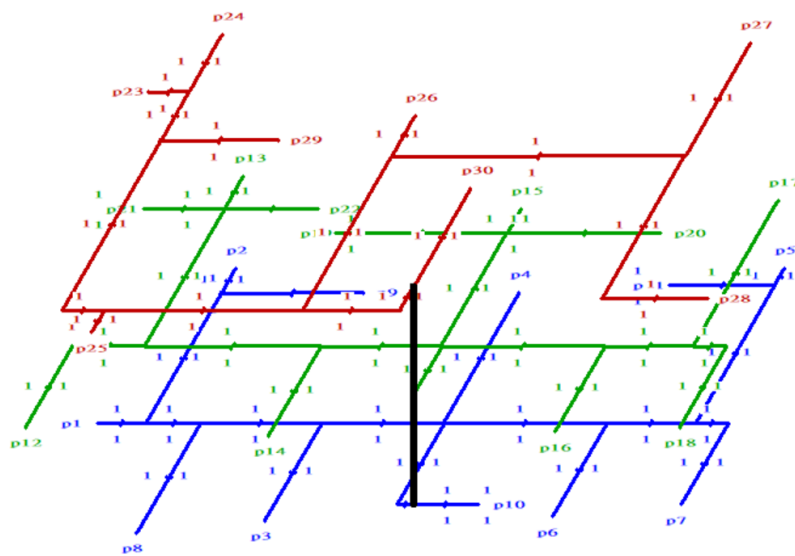
Example	#Term	#Loc	#Peri	Total length reduced by 5 (CaseKr5-*)					Total length reduced by 10 (CaseKr10-*)				
				Tlength	Tav(ns)	U-Tav(ns)	U-size	U-savings	Tlength	Tav(ns)	U-Tav(ns)	U-size	U-savings
CaseKr?- <i>nxn</i>	30	58	870	13776 μ m	0.7686	0.3059	116	60.20%	6859 μ m	0.3622	0.2601	116	28.18%
CaseKr?-1k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

*: by 5 or 10-3k: 300 timing periods

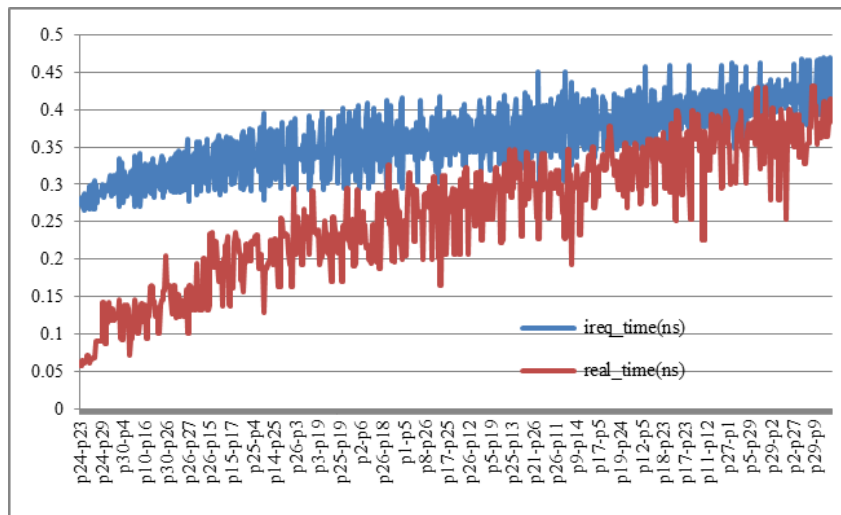
Table 5 presents the evaluation in average access time for the data bus topology CaseK that the total length is reduced by 5 or 10 (marked with r5 or r10) on their complete timing periods (marked with -*nxn*) and different number of timing periods (marked with -0.1k to -5k). From the table, the *U-savings* of CaseKr5-*nxn* and CaseKr10-*nxn* on their complete timing periods

are 60.20% and 28.18%, respectively. For the cases CaseKr5-.1k to CaseKr5-5k of a 3D data bus at different number of 100-5000 timing periods, their average access times have good U-savings in range of 52.44% to 60.42% and they are suitable for inserting a number of unit-size repeaters for the access time reduction. For the cases CaseKr10-.1k to CaseKr10-5k of a 3D data bus at different number of 100-5000 timing periods, the U-savings of three cases CaseKr10-.1k, CaseKr10-.2k, and CaseKr10-.3k are -1.82%, 4.53%, and 8.04%, respectively are less than the basic performance ratio 10% and they are not suitable for their data bus reconstruction.

Fig. 7(a) shows the 3D data bus topology of CaseKr10- $n \times n$ with inserting a number of 116 unit-size repeaters. 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 without/with inserted repeaters to each source-to-sink pair on a complete timing period. The real access time (marked with real_time) of each source-sink pair with inserted repeaters is always less than that the required access time (marked with ireq_time) without any inserted repeaters. The average access times without and with inserted repeaters are 0.3622 ns and 0.2601 ns, respectively. The saving in average access time is up to 28.18%.



(a) The 3D data bus topology with inserting the number of 116 unit-size repeaters



(b) Real access times (real_times) with inserted repeaters are less than the required access times (ireq_times) without inserted repeaters

Fig. 7 The bus topology and all the required and real access times of case CaseKr10- $n \times n$

5. Conclusions

The proposed evaluated approach for the bus performance by reconstructing a stacked-layer data bus based on inserted unit-size repeaters on a complete timing period has been successfully applied for estimating whether the average data access time is reduced more or not. Inserting a number of unit-size repeaters for a data bus reconstruction can reduce the impact in the requirements of each repeater area. Conducting the complete timing period can cover all the possible data accesses for any programs executed on the bus at different timing periods. Evaluating the average access time of a data bus can respond the performance of an executed program in practical. Therefore, our evaluated approach is simple but very fast and effective. Extending work is to investigate different diverse evaluated approaches such that can suit for the various data bus topologies of emerging 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, November 27, 2013.
- [2] Y. I. Ismail, E. G. Friedman, and J. L. Neves, "Repeater insertion in tree structured inductive interconnect," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 48, no. 5, pp.471-481, May 2001.
- [3] K. Y. Lin, H. T. Lin, T. Y. Ho, and C. C. Tsai, "Load-balanced clock tree synthesis with adjustable delay buffer insertion for clock skew reduction in multiple dynamic supply Voltage designs," *ACM Trans. on Design Automation of Electronic Systems*, vol. 17, no. 3, Article 34, 2012.
- [4] M. Ghoneima and Y. Ismail, "Optimum positioning of interleaved repeaters in bidirectional buses," *IEEE Trans. on CAD of Integrated Circuits and Systems*, vol. 24, no. 3, pp. 461-669, March 2005.
- [5] Q. Ashton Acton, *Issues in Electronic Circuits, Devices, and Materials: 2011 Edition*, Scholarly Editions, January 2012.
- [6] M. Daneshtalab, M. Ebrahimi, and J. Plosila, "HIBS- Novel inter-layer bus structure for stacked architectures," *Proc. IEEE International Conference on 3D System Integration (3DIC 12)*, February 2012, pp. 1-7.
- [7] I. G. Thakkar and S. Pasricha, "3D-Wiz: A novel high bandwidth, optically interfaced 3D DRAM architecture with reduced random access time," *Proc. IEEE International Conference on Computer Design (ICCAD 14)*, November 2014, pp. 1-7.
- [8] K. Cho, H. S. Na, T. W. Cho, and Y. You, "Analysis of system bus on SoC platform using TSV interconnection," *Proc. IEEE Asia Symposium on Quality Electronic Design (ASQED 12)*, August 2012, pp. 255-259.
- [9] K. S. Mohamed, *IP cores design from specifications to production, Chap-4 SoC buses and peripherals*, 1st ed. Switzerland: Springer International Publishing, 2016.
- [10] S. Khan, S. Anjum, U. A. Gulzari, and F. S. Torres, "Comparative analysis of network-on-chip simulation tools," *IET Computers & Digital Techniques*, vol. 12, no. 1, pp. 30-38, January 2018.
- [11] C. C. Tsai, "Repeater insertion for 3D data bus with TSVs for reducing critical propagation delay," *Proc. International Conference on Computer Science and Information Engineering (CSIE 15)*, June 2015, pp. 203-208.
- [12] C. C. Tsai, "An effective algorithm for minimizing the critical access time of a 3D-chip data bus," *International Journal of Electronics Communication and Computer Engineering*, vol. 9, no. 4, pp. 117-123, July 2018.
- [13] C. C. Tsai, "Embedded bus switches on 3D data bus for critical access time reduction," *Proc. IEEE Latin American Symposium on Circuits and Systems (LASCAS 18)*, February 2018, pp. 1-4.
- [14] IDTQS3245 data sheet, IDT Co., November 2014.
- [15] 74VHCT126AFT data sheet, Toshiba Co., November 2014.
- [16] C. C. Tsai, D. Y. Kao, and C. K. Cheng, "Performance driven bus buffer insertion," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, no. 4, pp. 429-437, April 1996.

- [17] W. C. Elmore, "The transient response of damped linear networks," *Journal of Applied Physics*, vol. 19, no. 1, pp. 55-63, January 1948.
- [18] 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, June 2011.
- [19] Y. Cao, W. Zhao, E. Wang, W. Wang, J. Velamala, A. Balijepali, and S. Sinha, "Predictive Technology Model (PTM)," <http://ptm.asu.edu>, June 1, 2012.



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).