Interconnect Delay Minimization through Interlayer Via Placement in 3-D ICs

Vasilis F. Pavlidis, Eby G. Friedman
Department of Electrical and Computer Engineering
University of Rochester
Rochester, New York 14627, USA
[pavlidis, friedman] @ ece.rochester.edu

ABSTRACT

The dependence of the propagation delay of the interlayer 3-D interconnects on the vertical through via location and length is investigated. For a variable vertical through via location, with fixed vertical length, the optimum vertical through via location that minimizes the propagation delay of an interconnect line connecting two circuits on different planes is determined. The optimum vertical through via location and length or, equivalently, the number of physical planes traversed by the vertical through via, are determined for varying the placement of the connected circuits. Design expressions for the optimal via locations and lengths have been developed to support placement and routing algorithms for 3-D ICs.

Categories and Subject Descriptors

B.7.2 [**Design Aids**]

General Terms: Performance, Design.

Keywords: 3-D ICs, Elmore delay, *RC* Interconnects.

1. INTRODUCTION

Technology scaling has enabled an increase in integration density and a considerable decrease in the intrinsic gate delay, through smaller and faster devices. Higher integration densities require both a greater number of interconnects and longer interconnects. Therefore, as the device delay is reduced, the performance of the integrated circuits is now dominated by the interconnect delay. In addition, other interconnect related issues, such as power consumption and signal integrity, have become more pronounced with technology scaling. To manage these issues, a variety of techniques have been developed, such as tapered buffers, repeater insertion, wire sizing, and shielding, to name a few. Nonetheless, these techniques increase silicon area and power consumption. As a result, innovative design processes are sought to satisfy the ever increasing demand for greater performance.

Three-dimensional integration is an effective design paradigm for interconnect-centric circuits that offers significant reductions in interconnection length. This advantage is achieved by using the third dimension, as shown in Figure 1. In Figure 1, multiple physical planes are stacked to create a 3-D system. A variety of bonding techniques to accomplish this task have been described in the literature [1-3]. Most of these techniques involve bonding forces with elevated temperatures and the bonding materials include adhesive polymers or eutectic metal pads [4]. Each physical plane of the stack is similar to a conventional 2-D circuit, in that a plane includes a device layer and multiple metal layers are used to connect individual circuits on the same physical plane (the intralayer interconnects). Each of the bonded planes can utilize completely different processes or design disciplines. For instance, non-silicon analog, digital, and RF circuits such as GaAs and SiGe can be stacked within a single 3-D multilayer system. Such technology diversity extends the capabilities of 3-D systems over a conventional CMOS platform, greatly facilitating the Systems-on-Chip design concept. Communication among circuits on different physical planes (the interlayer interconnects) is implemented by vertical through vias, which are called vias here for brevity.

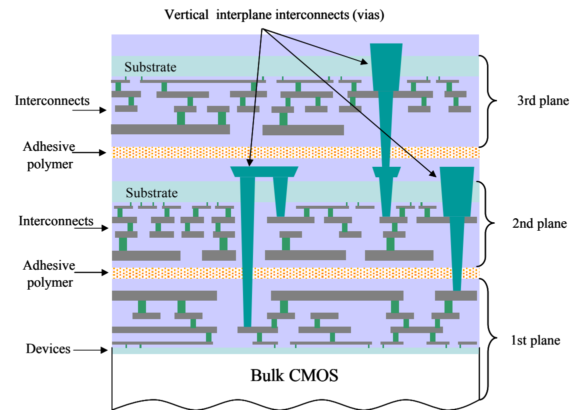

Figure 1. Schematic of a three-dimensional circuit [1].

These short interlayer vias replace the long interconnects along the edges of the die, yielding a considerable reduction in wirelength. This behavior is verified by a number of wirelength distribution models described in the literature which predict significant savings in the number of long interconnects. The fundamental assumption of these models is that Rent's rule [5] can be applied to three-dimensional partitions. More specifically, Joyner [6] extended a stochastic model for 2-D circuits introduced by Davis [7] to three-dimensions. Rahman [8] also used the Davis model as the basis for his 3-D interconnect distribution model.

Although wirelength reduction is an attractive feature of 3-D ICs, volumetric design poses a number of challenges. Placement and routing algorithms are examples of such a challenge, since another degree of freedom is added to the design process [9]. Certain

This research was supported in part by the Semiconductor Research Corporation under Contract No. 2003-TJ-1068 and 2004-TJ-1207, the National Science Foundation under Contract No. CCR-0304574, the Fullbright Program under Grant No. 87481764, a grant from the New York State Office of Science, Technology & Academic Research to the Center for Advanced Technology in Electronic Imaging Systems, and by grants from Xerox Corporation, IBM Corporation, Intel Corporation, Lucent Technologies Corporation, and Eastman Kodak Company.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
GLSVLSI'05, April 17–19, 2005, Chicago, Illinois, USA.
Copyright ACM 1-59593-057-4/05/0004...$5.00.

transformations can be applied to convert a 3-D routing problem to a 2-D routing task [10]. In addition, novel CAD tools are required for three-dimensional circuits. Recently, such CAD tools have been published [11] that validate the interconnect prediction model described in [8].

In all of these algorithms, however, the particular nature of the interlayer interconnects is not considered. Additionally, delay expressions used in the aforementioned interconnect prediction models for 3-D circuits are similar to traditional CMOS models, neglecting the impact of the vias and the non-uniform impedance characteristics of these lines. Zhang *et al.* [12] consider the effect of the vertical vias on the interlayer interconnects in their delay expression by modeling the line with different impedances; however, they apply two restrictive assumptions. First, that the via is always placed in the middle of the line, independent of the line length and, second, each horizontal segment of the interconnect has the same impedance characteristics. As shown in this paper, the former assumption leads to severe performance inaccuracy, while the latter assumption does not accurately depict the physical nature of the interlayer interconnects. The optimum via location that yields the minimum propagation delay of the line is determined in this work. The effect of the via length, or equivalently, the number of physical planes that the via spans, on the propagation delay of the line is also investigated. The inclusion of variable via locations into routing algorithms and a variable number of physical planes in placement algorithms will considerably enhance the efficiency of the 3-D design process.

The rest of the paper is organized as follows. In the following section, the problems are formulated, and various traits of the interlayer 3-D interconnects are outlined. For a fixed via length, the optimum via location to minimize the propagation delay of these interconnects is described in Section 3. Additionally, conditions for the minimum propagation delay, where the via length and location are varied, are described. Expressions for the via length and location that yield the minimum delay are also provided. In Section 4, simulation results are presented, verifying the theoretical results presented in Section 3. Finally, in Section 5, some conclusions are offered.

2. PROBLEM FORMULATION

The problems explored in this paper can be better explained with the aid of Figure 2, where a cross section of a three-dimensional manhattan grid is illustrated. The circles and the bold solid, dashed, and dotted lines represent possible placement locations and routing paths, respectively. Since the overall performance is typically based on the delay of the long interconnects, it is assumed that these interconnects are routed first, thus multiple routing paths are possible. In Figure 2a, two circuits are placed on two different device layers. The total and interlayer distances are fixed. The via location (the length of the horizontal segments of the line), however, is allowed to change, as shown by the dashed and dotted routing paths. In Figure 2b, not only does the length of the horizontal segments change, but also the via is allowed to span a variable number of planes. In the first case, the optimum via location that minimizes the propagation delay is determined, while in the second case, both the optimum via location and number of planes that the via traverses are determined. Because the total geometric length for all of the possible routing paths is the same, in most placement and routing algorithms a typical criterion to determine the most appropriate via location and/or number of planes is to select the routing path and/or placement location that yields the minimum routing blockage. This characteristic occurs because in 2-D circuits, interconnects are usually considered to occupy a single metal layer and to be of uniform impedance throughout the length, and the via location and impedance are assumed to not affect the delay of the line.

This assumption is no longer valid, however, for interlayer interconnects in three-dimensional circuits, as shown in Figure 3. To model interlayer interconnects as an assemblage of non-uniform segments is justified by the physical nature of 3-D ICs. For example, the circuit can be placed on more than one die and therefore process variations not only exist within a single die (intradie) but also die-to-die (interdie) variations should be considered. Consequently, interconnect impedances in different physical planes can differ from each other by a non-negligible amount. In addition, 3-D ICs may combine dies from totally disparate technologies, resulting in different interconnect parameters for each physical plane. Furthermore, each segment of the interlayer interconnect can be laid out on different metal layers, which exhibit different impedance characteristics. Also, the coupling capacitance between each segment will vary, as the interconnect structure that surrounds each segment is typically different. Consider the case where a 3-D circuit consists of only two physical planes and the bonding process is similar to that illustrated in Figure 1. The metal layers of the bottom physical plane will exhibit a greater line-to-ground capacitance because the layers are sandwiched between two substrates, while for those interconnects within the upper plane, the same capacitance component is smaller due to the absence of a nearby second substrate.

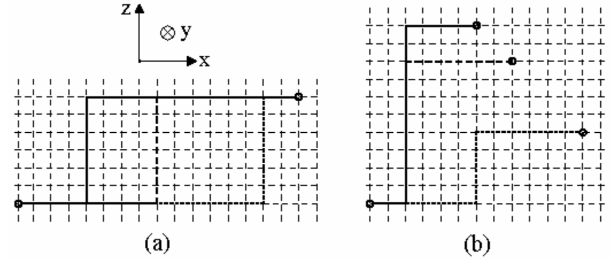

Figure 2. Interlayer interconnect (a) with multiple routing paths, (b) with multiple placement locations and routing paths.

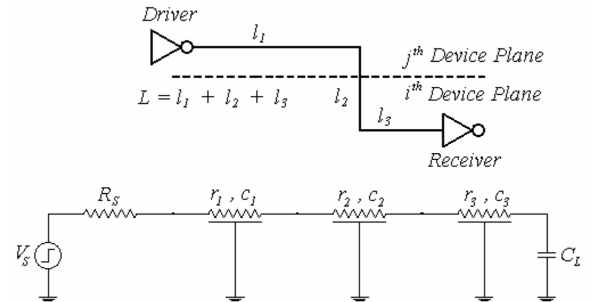

Figure 3. Interlayer interconnect and corresponding model, composed of a set of non-uniformly distributed *RC* segments.

The Elmore delay model has been adopted to analyze the propagation delay of these interconnects. If accuracy is an issue, a fitted Elmore delay model can be used [13]. However, unlike a single plane, more than one set of fitting coefficients is required in a 3-D system. In the following section, the optimum via location problem is further explored.

3. OPTIMUM VIA LOCATION

In this section, two variants of the optimum via location problem, which enhance the task of routing and placement for interlayer interconnects in 3-D circuits, are considered. In the first subsection, the optimum via location of an interconnect, for a fixed via length, is provided. The case where the via length constraint is removed is investigated in the second subsection.

3.1 Fixed number of physical planes

An interlayer interconnect that connects two circuits located on two different physical planes is depicted in Figure 3. As mentioned previously, due to the non-uniformity of the interconnects, each segment is modeled as a distributed *RC* line with different impedance characteristics (see Figure 3). Inductance is not considered in this work. The driver is modeled as a step input voltage and a linear resistance R_S, and the interconnect is terminated with a capacitive load C_L. The total resistance and capacitance of segment i are $R_i = r_i l_i$ and $C_i = c_i l_i$, where r_i and c_i denote the resistance and capacitance, respectively, per unit length and l_i is the length of the segment. The length of the horizontal segments is l_1 and l_3 and the via length is l_2. Since the via may span more than one physical plane, l_2 can be expressed as

$$l_2 = (n-1)l_v\,, \qquad (1)$$

where n is the number of physical planes making up the connected circuits and l_v is the length of the via that interconnects two metal layers located in two adjacent physical planes. This value is determined by the fabrication process and can range from 15 µm to 70 µm [3], [14]. The total length of the line can be expressed as

$$L = l_1 + l_2 + l_3\,. \qquad (2)$$

The Elmore delay for the system is

$$T_{el} = \frac{R_1C_1}{2} + C_1R_S + \frac{R_2C_2}{2} + C_2(R_S+R_1) + \frac{R_3C_3}{2} + C_3(R_S+R_1+R_2) + C_L(R_S+R_1+R_2+R_3)\,. \qquad (3)$$

Substituting the total resistance and capacitance with the per unit length parameters, and using (1) and (2), the Elmore delay described in (3) can be written as a function of the length of the first segment l_1,

$$T_{el}(l_1) = A_1 l_1^2 + A_2 l_1 + A_3\,, \qquad (4)$$

where

$$A_1 = (1/2)\,(r_1c_1 - 2r_1c_3 + r_3c_3)\,, \qquad (5)$$

$$A_2 = R_s(c_1-c_3) + (n-1)l_v(r_1c_2 - r_2c_3 + r_3c_3 - r_1c_3) + L(r_1c_3 - r_3c_3) + C_L(r_1-r_3)\,, \qquad (6)$$

$$A_3 = R_S((n-1)l_v(c_2-c_3) + Lc_3) + ((n-1)l_v)^2\left(\frac{r_2c_2}{2} - r_2c_3 + \frac{r_3c_3}{2}\right) + L(n-1)l_vc_3(r_2-r_3) + C_L((n-1)l_v(r_2-r_3) + R_S + Lr_3) + \frac{L^2r_3c_3}{2}. \qquad (7)$$

Equation (4) describes a parabola, but the existence of a minimum is not guaranteed. The second derivative of (4) with respect to l_1 is

$$\frac{d^2T_{el}}{dl_1^2} = 2A_1\,. \qquad (8)$$

Depending upon the sign of A_1, the propagation delay of the line will exhibit either a minimum or a maximum as l_1 varies or, alternatively, as the location of the via along the line changes. The following notations are introduced to facilitate the analysis,

$$a = \frac{r_3}{r_1}\,, \qquad b = \frac{c_1}{c_3}\,, \qquad m = \frac{r_2}{r_1}\,, \quad \text{and} \qquad v = \frac{c_2}{c_3}\,. \qquad (9)$$

From (9), the second derivative is

$$\frac{d^2T_{el}}{dl_1^2} = r_1c_3(a+b-2)\,. \qquad (10)$$

Since r_1c_3 is always positive, the sign of (10) and, consequently, the timing behavior of the line only depends upon the sign of the term in the parentheses. For the propagation delay to be minimum, the following inequalities should be satisfied,

$$\frac{d^2T_{el}}{dl_1^2} > 0 \Rightarrow \begin{cases} a > 2, or\ b > 2 \\ b \in (1,2],\ a \in (1,2] \\ b \in (0,1],\ a \in (1,2], and\ a > 2-b \\ a \in (0,1],\ b \in (1,2], and\ b > 2-a\,. \end{cases} \qquad (11)$$

If the inequalities in (11) are not satisfied, the delay of the line exhibits a maximum. The existence of either a minimum or maximum delay with the via location depending on the values of a and b can be roughly explained as follows. Neglecting the via, the line comprises two segments with different impedance characteristics. Alternatively, a non-uniform line can be seen as a uniform tapered line that only consists of two segments. Setting $r_1 = r_0/w_1$, $r_3 = r_0/w_3$, $c_1 = c_0w_1$, and $c_3 = c_0w_3$, a and b describe the tapering factor of the line. It has been shown that an optimum tapering factor exists in terms of the delay, where the width of the line decreases towards the receiver [15]. If a is greater than one, the tapering decreases.

The value of l_1 for which the delay exhibits an extremum, either minimum or maximum, is

$$l_1 = -\frac{A_2}{2A_1} = -\left[\frac{(n-1)l_v(r_1c_2 - r_2c_3 + r_3c_3 - r_1c_3)}{r_1c_1 - 2r_1c_3 + r_3c_3} + \frac{R_s(c_1-c_3) + L(r_1c_3 - r_3c_3) + C_L(r_1-r_3)}{r_1c_1 - 2r_1c_3 + r_3c_3}\right]. \qquad (12)$$

Since no restrictions have been applied on the value of l_1, the extreme point can occur for values other than within the physical domain of l_1, *i.e.*, $l_1 \in [0, l_0]$ where $l_0 = L - l_2 - l_{min}$. l_{min} is the minimum distance between a via and a cell, determined by the design rules of the fabrication process. The following Lemma is used to determine the optimum via location for various values of a, b, and l_1. The proof is omitted due to space limitations.

Lemma 1: *If $f(x) = Ax^2+Bx+C$ and $\frac{d^2 f(x)}{dx^2} < 0$ then*

(a) for $x_{max} \in [0,x_0]$

(i) if $x_{\max} > \frac{x_0}{2}$, $f(0) < f(x_0)$, (ii) if $x_{\max} < \frac{x_0}{2}$, $f(0) > f(x_0)$,

(b) for $x_{max} < 0$, $f(0) > f(x_0)$,

(c) for $x_{max} > x_0$, $f(0) < f(x_0)$.

Depending upon the sign of (10) and the value of l_1 in (12), the optimum via location is determined for each possible case:

A) $\frac{d^2T_{el}}{dl_1^2} > 0$. If $l_1 \in [0, l_0]$, the propagation delay is minimized when l_1 is the value described in (12). Consequently, the via should be placed at a distance l_1 from the driver. The Elmore delay for a 5 mm line is illustrated in Figure 4 versus the via location l_1. Two

observations can be made. First, that the delay exhibits a minimum and that the minima position shifts to the right as a increases. If $l_1 < 0$, the via should be placed closest to the driver, while if $l_1 > l_0$, the via should be placed closest to the receiver.

B) $\frac{d^2T_{el}}{dl_1^2} < 0$. In this case, the delay of the line reaches a maximum for the value of l_1 described in (12). If $l_1 \in [0, l_0]$, according to Lemma 1, for $l_1 < l_0/2$, the via should be placed closest to the receiver, while for $l_1 > l_0/2$, the via should be placed closest to the driver. In Figure 5, the Elmore delay of a 5 mm line is shown as a function of the via location l_1. Note that the delay reaches a maximum and that the maximum shifts to the right as b increases. If $l_1 < 0$, the via should be placed closest to the receiver, while if $l_1 > l_0$, the via should be placed closest to the driver.

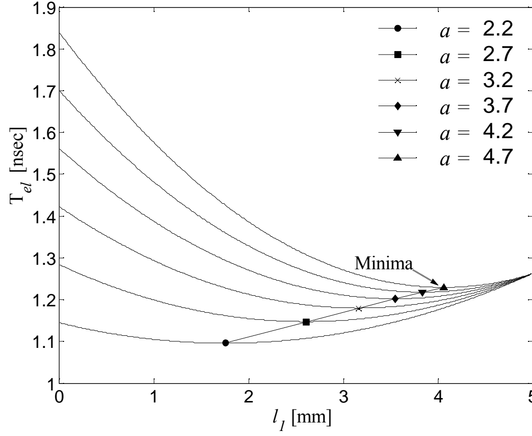

Figure 4. Propagation delay of a 5 mm line versus via location l_1 for various values of a. The interconnect parameters are r_1 = 76 Wmm, r_2 = 53 Wmm, c_2 = 223 fF/mm, c_3 = 279 fF/mm, b = 1.674, l_v = 20 µm, and n = 2. The driver resistance and load capacitance are R_S = 410 Wand C_L = 180 fF, respectively.

C) $\frac{d^2T_{el}}{dl_1^2} = 0$. If the second derivative equals zero, (4) becomes

$$T_{el}(l_1) = A_2 l_1 + A_3 , \tag{13}$$

which is a linear function of l_1. The first derivative of (4) is equal to A_2. For $A_2 < 0$, (4) is strictly decreasing, and the delay is a minimum by placing the via closest to the receiver. For $A_2 > 0$, (4) is strictly increasing, and the delay is a minimum by placing the via closest to the driver. Note that the fundamental minimum distance of a via from a cell is technology dependent. In the special case where $a = b = 1$, from (6) and (7), $A_2 << A_3$ and the delay is independent of l_1. However, as n increases, A_2 also increases and the choice of n affects the rate of change in the delay, as discussed in the following section.

As illustrated in Figure 4, the optimum via location shifts to the right (left) when a increases (decreases). The same applies to Figure 5 in terms of b. To explain this behavior, consider the definitions of a and b in (9), where a (b) describes the resistance (capacitance) ratio of the horizontal segments. Referring to Figure 4, both a and b are greater than one, which means that segment 1 is less (more) resistive (capacitive) than segment 3. Assuming for the moment that a = 1, the delay of the line decreases as the length of the more capacitive segment l_1 (*i.e.*, C_1) decreases. However, l_1 does not vanish because C_3 increases as l_1 decreases, approaching C_1. Consequently, as l_1 is decreased beyond a certain distance, described in (12), the delay starts to increase. As a increases, the optimum point occurs at values of $l_1 > l_0/2$, although segment 1 is more capacitive than segment 3 ($b > 1$). This behavior occurs because the delay depends not only on the capacitance, but also on the current, which is controlled by the resistance of each segment. In the case where $b >> 1$ and $a << 1$, where segment 1 is both more capacitive and resistive than segment 3, l_1 becomes small to reduce the overall delay and, in this case, (12) yields negative values for l_1. Similar arguments apply to the dependence of the optimum via location on b shown in Figure 5. In the following subsection, the optimum via location and number of planes in terms of the propagation delay are determined for a variable number of planes n.

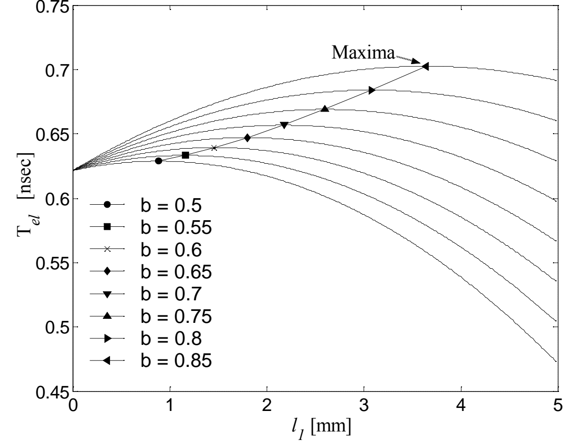

Figure 5. Propagation delay of a 5 mm line versus via location l_1 for various values of b. The interconnect parameters are r_1 = 92 Wmm, r_2 = 43 Wmm, c_2 = 196 fF/mm, c_3 = 286 fF/mm, a = 0.471, l_v = 20 µm, and n = 2. The driver resistance and load capacitance are R_S = 410 Wand C_L = 180 fF, respectively.

3.2 Variable number of physical planes

If the number of physical planes is varied, multiple placement locations can exist, and (4) becomes

$$T_{el}(l_1, l_2) = r_1 c_3 \left(F_1 l_1^2 + F_2 l_1 l_2 + F_3 l_2^2 + F_4 l_1 + F_5 l_2 + F_6 \right), \tag{14}$$

where

$$F_1 = (1/2)(a + b - 1), \; F_2 = (v - m - 1 + a) , \tag{15}$$

$$F_3 = \frac{mv}{2} - m + \frac{a}{2} , \; F_4 = \frac{R_S}{r_1}(b-1) + \left(L + \frac{C_L}{c_3} \right)(1-a) , \tag{16}$$

$$F_5 = \frac{R_S}{r_1}(v-1) + \left(L + \frac{C_L}{c_3} \right)(m-a) , \tag{17}$$

$$F_6 = \left(\frac{C_L}{c_3} + L \right)\frac{R_S}{r_1} + \frac{1}{2}L^2 a + L\frac{C_L}{c_3} a . \tag{18}$$

Equation (14) is a quadratic function with respect to l_1 and l_2, and describes a paraboloid. As a polynomial function is both continuous and differentiable, global extrema exist which can occur either on the boundary, or the interior of the domain of l_1 and l_2. The solutions of the gradient $\nabla T_{el}(l_1, l_2)$, which are called critical points, describe possible interior extrema. The interior point at which the delay can exhibit an extremum is

$$l_1 = \frac{-R_s c_3 \left(ab - va + m - 2mb + 2v - v^2 + vmb - 1\right)}{r_1 c_3 (2(v+m) + mv(b+a) + ab - 2mb - 2va - v^2 - m^2 - 1)} - \frac{r_1 \left(C_L + c_3 L\right)\left(va - m\right)\left(1 - m\right)}{r_1 c_3 (2(v+m) + mv(b+a) + ab - 2mb - 2va - v^2 - m^2 - 1)}, \quad (19)$$

$$l_2 = \frac{\left(R_S c_3 + r_1 \left(C_L + c_3 L\right)\right)\left(m - mb + ab - av + v - 1\right)}{r_1 c_3 (2(v+m) + mv(b+a) + ab - 2mb - 2va - v^2 - m^2 - 1)}. \quad (20)$$

From multivariable calculus, the interior critical point is a minimum for (14) when (19) and (20) satisfy the following conditions,

$$H_1 = \frac{\partial^2 T_{el}\left(l_1, l_2\right)}{\partial l_1^2} = 2 r_1 c_3 F_1 > 0, \quad (21)$$

$$H_2 = \frac{\partial^2 T_{el}\left(l_1, l_2\right)}{\partial l_1^2}\frac{\partial^2 T_{el}\left(l_1, l_2\right)}{\partial l_2^2} - \left[\frac{\partial^2 T_{el}\left(l_1, l_2\right)}{\partial l_1 \partial l_2}\right]^2 = = r_1^2 c_3^2 \left(4 F_1 F_3 - F_2^2\right) > 0. \quad (22)$$

Since (14) is a quadratic function, both (21) and (22) are independent of the critical points of (14) and the values depend only on the ratios described in (9). Depending upon these values, the following cases are distinguished:

A) $H_1 > 0$, $H_2 > 0$. The delay is a minimum in terms of both the via location and length. The optimum via location is given by (19), while the optimum number of planes n_{opt} is obtained after the discreteness constraints for variable l_2, described in (20), are applied according to (1),

$$n_{opt} = \begin{cases} \left[\left[\dfrac{l_2}{l_v}\right]\right] + 1, \ if\ l_2 < \left[\dfrac{l_2}{l_v}\right] + \dfrac{1}{2} \\ \left[\left[\dfrac{l_2}{l_v}\right]\right] + 2, \ if\ l_2 > \left[\dfrac{l_2}{l_v}\right] + \dfrac{1}{2} \end{cases}, \quad (23)$$

where, in (23), $[x]$ denotes the integer part of number x. In Figure 6, the Elmore delay of a 5 mm line is plotted versus the via location l_1 and the via length l_2. The global minimum is depicted in Figure 6 by the dot.

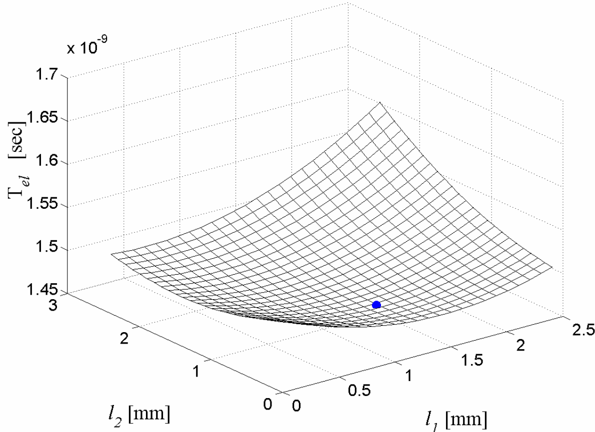

Figure 6. Propagation delay of a 5 mm line versus via location l_1 and via length l_2. The interconnect parameters are r_1 = 51 Ωmm, c_3 = 288 fF/mm, a = 3.1, b = 1.85, m = 2.1, and v = 1.35. The driver resistance and load capacitance are R_S = 476 Ω and C_L = 288 fF, respectively.

B) $H_1 < 0$, $H_2 > 0$. The delay is a maximum for the values of l_1 and l_2 described in (19) and (20), respectively. In this case, the optimal point both for the via location and length is on the boundary of the domain of l_1 and l_2, which defines a rectangle. Searching in the subset of boundary points that consists of the vertices of the rectangle, namely $T_{el}(l_{min}, l_v)$, $T_{el}(l_{min}, (n_{max}-1)l_v)$, $T_{el}(L-l_v-l_{min}, l_v)$, and $T_{el}(L-(n_{max}-1)l_v-l_{min}, (n_{max}-1)l_v)$ suffices for the minimum delay to be determined. n_{max} is the maximum number of stacked planes permitted by the target technology. The point that yields the smallest delay provides the optimal values for l_1 and l_2.

C) $H_2 < 0$. If (22) is negative, (19) and (20) do not represent an extremum but rather a saddle point. The optimal point is again located on the boundary. Note that in this case, depending on the signs of F_1 and F_2, a different subset of boundary points should be examined to determine the optimal values of l_1 and l_2. In addition, selecting the number of planes can be based on other criteria as discussed in the following section, where simulation results are presented.

4. RESULTS AND DISCUSSION

In this section, experimental results are presented and a discussion of the proposed approach is offered, particularly for those cases where the minimum occurs at the boundary. The analysis has been applied to interconnects with lengths ranging from 1 mm to 10 mm, for different interconnects parameters. The ratios in (9) range from 0.1 to 10, and cover all practical cases of interest. The maximum variation in delay with the length of the segment l_1 (or equivalently the via location) for different line lengths and interconnect parameters is listed in Table 1. This variation expresses the difference between the delay of the line when the via is placed at the optimum point and the delay of the line when the via location coincides with a point close to either the receiver or the driver. It is assumed that the physical design rules impose a minimum 10 µm distance between a via and a cell. The number of planes is equal to two and the via length is assumed to be $l_2 = l_v$ = 20 µm [14]. As listed in Table 1, the optimum via location shifts to the right as a increases, in agreement with the theoretical results depicted in Figure 4.

In Table 2, the optimum via location and number of planes for different line lengths are listed. In this case, the minimum via length l_v = 50 µm. As mentioned in Section 3, a minimum does not exist if (22) is negative and the values of l_1 and l_2, from (19) and (20), respectively, correspond to a saddle point. In addition, (20) may result in an infeasible solution, since the maximum number of planes that can be stacked n_{max} is technology limited. In these cases, the optima occur at the boundary of the domain of l_1 and l_2. Thus, either n = 2 or $n = n_{max}$ should be selected depending upon which value produces the smaller delay.

The choice of the number of planes n can also be based on a different criterion than the propagation delay. For example, if the impedance parameters of the via are much greater than those of the horizontal segments (m, v >> 1), n can be increased such that the delay of the line does not vary significantly as the via location changes. Alternatively, if the impedance parameters of the line segments are of similar value, the number of planes n can be selected such that the delay with respect to l_1 is a minimum when the via is placed at the center of the line. In this case, the variation of the line delay with via location is smallest, as shown in Figure 4, for a = 2.7. Thus, if an increase in the overall delay is acceptable, the number of layers n can be chosen such that the skew between interconnects (*e.g.*, clock skew) varies slowly with via location.

5. CONCLUSIONS

The interlayer via location and length to minimize the signal delay in 3-D systems is described in this paper. The Elmore delay model is adopted to investigate the propagation delay of interlayer 3-D interconnects. The non-uniform impedance characteristics of the line are also considered. The propagation delay of the interconnect is shown to depend both on the via location and the number of planes that the via vertically traverses. Expressions for the optimum via location and number of planes are provided. Simulation results verifying these expressions are also presented. The proposed design expressions can be used to enhance placement and routing algorithms targeting 3-D ICs.

6. REFERENCES

[1] R. J. Gutmann *et al*., "Three-dimensional (3D) ICs: A Technology Platform for Integrated Systems and Opportunities for New Polymeric Adhesives," *Proceedings of the Conference on Polymers and Adhesives in Microelectronics and Photonics*, pp. 173-180, October 2001.

[2] L. Xue, C. Liu, and S. Tiwari, "Multi-layers With Buried Structures (MLBS): An Approach to Three-dimensional Integration," *Proceedings of the IEEE International Conference on Silicon On Insulator*, pp. 117-118, October 2001.

[3] M. Koyanagi *et al*., "Future System-on-silicon LSI Chips," *IEEE Micro*, Vol. 18, No. 4, pp 17-21, July/August 1998.

[4] A. Fan, A. Rahman, and R. Reif, "Copper Wafer Bonding," *Electrochemical and Solid-State Letters*, Vol. 10, No. 2, pp. 534-536, October 1999.

[5] B. Landman and R. Russo, "On a Pin vs. Block Relationship for Partitions of Logic Graphs," *IEEE Transactions on Computers*, Vol. 20, No. 12, pp. 1469-1479, December 1971.

[6] J. W. Joyner *et al*., "Impact of Three-dimensional Architectures on Interconnects in Gigascale Integration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 9, No. 6, pp. 922-927, December 2000.

[7] J. Davis and J. Meindl, "A Stochastic Wire-length Distribution for Gigascale - Part I: Derivation and Validation," *IEEE Transactions on Electron Devices*, Vol. 45, No. 3, pp. 580-589, March 1999.

[8] A. Rahman and R. Reif, "System-level Performance Evaluation of Three-dimensional Integrated Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 8, No. 6, pp. 671-678, December 2000.

[9] A. Cohoon *et al*., "Physical Layout for Three-Dimensional FPGAs," *Proceedings of the ACM/SIGDA Physical Design Workshop*, pp. 142-149, April 1996.

[10] C. C. Tong and C.-L. Wu, "Routing in a Three-dimensional Chip," *IEEE Transactions on Computers*, Vol. 44, No.1, pp. 106-117, January 1995.

[11] S. Das *et al*., "Calibration of Rent's Rule for Three-Dimensional Integrated Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 12, No. 4, April 2004.

[12] R. Zhang *et al*., "Stochastic Modeling, Power Trends, and Performance Characterization of 3-D Circuits," *IEEE Transactions on Electron Devices,* Vol. 48, No. 4, April 2001.

[13] A.I. Abou-Seido, B. Nowak, and C. Chu, "Fitted Elmore Delay: A Simple and Accurate Interconnect Delay Model," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 12, No. 7, pp. 691-696, July 2004.

[14] P. Ramn *et al.*, "InterChip Via Technology for Vertical System Integration," *Proceedings of the IEEE International Interconnect Technology Conference*, pp. 160-162, June 2001.

[15] C.-P. Chen *et al.*, "Optimal Sizing Formula Under the Elmore Delay Model," *Proceedings of the International ACM/IEEE Design Automation Conference*, pp. 487-490, June 1996.

Table 1. Maximum delay variation and optimum via location for different interconnect lengths and impedance parameters

	r_1 = 86 Ω/mm, r_2 = 53 Ω/mm, c_2 = 279 fF/mm, c_3 = 396 fF/mm, b = 1.232, C_L = 180 fF, and l_v = 20 µm											
	L = 1 mm, R_S = 1240 Ω			L = 2 mm, R_S = 1240 Ω								
a	4.0	4.5	5.0	3.0	3.5	4.0	4.5	5.0				
Del. Variation (%)	2.1	2.0	3.5	2.9	3.7	6.9	9.5	11.7				
l_{1opt} [mm]	0.35	0.43	0.86	1.02	1.39	1.51	1.58	1.63				
	L = 3 mm, R_S = 940 Ω						L = 4 mm, R_S = 940 Ω					
a	2.0	2.5	3.0	3.5	4.0	4.5	2.0	2.5	3.0	3.5	4.0	4.5
Del. Variation (%)	4.8	4.7	8.6	13.4	17.5	21.2	4.0	7.6	13.0	18.0	22.5	26.7
l_{1opt} [mm]	0.64	1.78	2.22	2.43	2.57	2.67	1.87	2.79	3.24	3.45	3.62	3.73
	L = 5 mm, R_S = 560 Ω						L = 6 mm, R_S = 560 Ω					
a	1.5	2.0	2.5	3.0	3.5	4.0	1.5	2.0	2.5	3.0	3.5	4.0
Del. Variation (%)	5.2	10.3	18.9	26.2	32.5	37.7	5.1	12.6	21.7	29.4	35.7	41.0
l_{1opt} [mm]	1.92	3.62	4.21	4.50	4.68	4.80	2.71	4.49	5.10	5.42	5.59	5.74
	L = 7 mm, R_S = 310 Ω						L = 8 mm, R_S = 310 Ω					
a	1.25	1.75	2.25	2.75	3.25	3.75	1.25	1.75	2.25	2.75	3.25	3.75
Del. Variation (%)	7.6	15.3	27.7	37.1	44.5	50.0	7.6	16.5	29.1	38.5	46.1	51.8
l_{1opt} [mm]	1.86	5.26	6.07	6.43	6.64	6.78	2.55	6.13	6.96	7.36	7.63	7.71
	L = 9 mm, R_S = 210 Ω						L = 10 mm, R_S = 210 Ω					
a	1.25	1.75	2.25	2.75	3.25	3.75	1.25	1.75	2.25	2.75	3.25	3.75
Del. Variation (%)	7.1	21.0	34.5	44.3	51.3	57.1	7.1	21.8	35.5	45.2	52.4	58.0
l_{1opt} [mm]	3.74	7.16	7.99	8.36	8.59	8.72	4.27	7.98	8.89	9.28	9.51	9.66

Table 2. Optimum via location l_1 and number of physical planes n for different interconnect lengths and impedance parameters

Inter. Length	l_{1opt} [mm]	n_{opt}	a	b	m	v	R_S [Ω]	C_L [fF]	r_1 [Ω/mm]	c_3 [fF/mm]
L = 2 mm	0.207	2	3.35	1.35	2.42	1.13	1042	103	81	189
L = 3 mm	0.674	6	3.10	2.10	1.80	1.60	410	213	97	278
L = 5 mm	2.493	10	3.10	1.60	1.65	1.38	410	100	46	329
L = 5 mm	2.643	9	3.10	1.60	2.35	1.18	410	100	46	329
L = 10 mm	1.073	18	2.10	2.35	1.51	1.65	210	90	61	278