Bandwidth Optimization for Through Silicon Via (TSV) bundles in 3D Integrated Circuits

Awet Yemane Weldezion*, Roshan Weerasekera+, Dinesh Pamunuwa+,
Li-rong Zheng*, and Hannu Tenhunen*

*School of Information and Communication Technologies,
Department of Electronics, Computer, and Software Systems,
KTH Royal Institute of Technology, Forum 120, Kista SE 16440, Sweden
{aywe, lirong, hannu}@kth.se

+Centre for Microsystems Engineering, Department of Engineering
Lancaster University, Lancaster LA1 4YW, UK
{r.weerasekera, d.pamunuwa}@lancaster.ac.uk

Abstract

Through silicon vias (TSVs) are the backbone of 3D integration technology connecting vertically stacked ICs. Parallel TSVs in the form of bundles are used for vertical signaling. In this paper, we present the ways of maximizing the total bandwidth of a TSV bundle placed in a fixed area by varying the density and the geometries. The ways of optimizing the total bandwidth using analytical methods for a bundle of TSVs placed in a structure with a fixed area and length are examined. The result shows that for uniformly distributed TSVs, maximum bandwidth by proportional placement of fewer number of TSV in the bundle can be achieved.

1. Introduction

TSVs are primarily used as wires for vertical signaling while at the same time as a mechanical support strengthening the 3D structure. In interconnect analysis; a local wire is characterized by short wire length as compared to a global wire that possibly has its length increased with scaling [1]. TSVs can be treated as local wires reasonably because like local wires, TSV length gets shorter with scaling. Moreover a TSV driven by a device, the TSV RC effect is insignificant and the overall device resistance is the dominant factor. According to the ITRS, the diameter of high density - HDTSV decreases by a factor of half per year [2] making TSVs to be densely placed into a small area and in consequence the coupling effect increases sharply.

These changes in density and TSV geometries define the overall signaling characteristics such as cross-talk, propagation delay and ultimately the total bandwidth. Thus the challenge is how to effectively distribute few TSVs in a given area and attain maximum total bandwidth with less coupling. There are other relevant works such as [6] with focus on bandwidth optimization method for planar wires in 2D integrated circuits. In this paper, we extend the bandwidth optimization method for 3D integrated circuits, for the first time. We examine the ways of optimizing the total bandwidth based on the compact models for a bundle of TSVs placed in a structure with a fixed area and length. Given the assumptions that the TSV behaves like local wire, a lumped model is accurate enough to estimate the delay and hence the bandwidth.

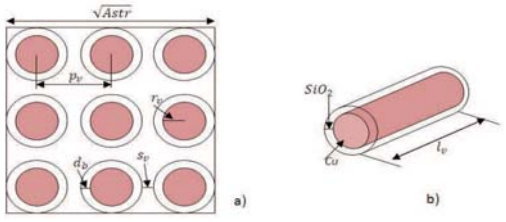


Figure 1. 3×3 **TSV structure (a) top view for** $N_{tsv} = 9$ **(b) single TSV**

2. TSV delay and bandwidth modelling

The bundle of TSVs shown in Figure 1, with number, $N_{tsv} = 3 \times 3$, is made out of copper vias embedded in silicon substrate. Each via is surrounded by SiO_2 dielectric barrier d_b. In the compact models described in [3], it has been shown that for such TSV structure the RLC extraction equivalent values can be deduced through analytical method for a specific range of radius, $10\mu m < r_v < 45\mu m$ and spacing, $40\mu m < s_v < 180\mu m$. The total capacitance C_t and inductance L_t is the sum of the self capacitance C_s and inductance L_s and mutual capacitance C_m and inductance L_m, respectively. The effective TSV resistance R_t remains the same in all conditions.

$$R_t = \rho_{cu}\frac{l_v}{\pi r_v^2}\Omega \tag{1}$$

The via pitch, p_v, between two TSV for a given N_{tsv} can be varied by changing the area of the TSV structure A_{str} of Figure 1a. The p_v, is uniformly distributed and measured as the distance between two lateral TSVs from center to center.

$$p_v = \frac{\sqrt{A_{str}} - 2(r_v + d_b)}{\sqrt{N_{tsv}} - 1} for N_{tsv} > 2 \tag{2}$$

In the compact model [3], it has been shown that the total capacitance C_t and total inductance L_t of any TSV in a uniformly distributed bundle can be expressed in a closed form equation as a function of the radius, r_v, the distance between two TSVs, i.e., pitch,p_v, and TSV length, l_v

$$C_t = \frac{63.34\epsilon_o l_v}{ln(1+5.26\frac{l_v}{r_v})} + \frac{\epsilon_o l_v}{\ln(\frac{k_5 p_v}{r_v})}[k_1(\frac{p_v}{r_v})^{k_2} + k_3(\frac{p_v}{l_v})^{k_4}] \tag{3}$$

$$L_t = \frac{\mu l_v}{2\pi}\ln(1+\frac{2.84 l_v}{\pi r_v}) + \mu l_v \ln(1+0.34\frac{l_v}{p_v}) + \mu l_v \ln(1+0.24\frac{l_v}{p_v}) \tag{4}$$

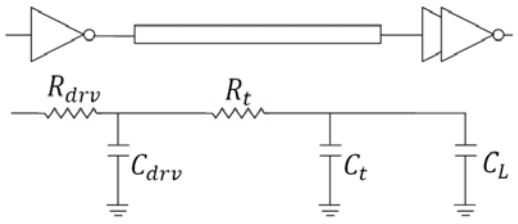


Figure 2. RC model of TSV with driver and load

where k_i, is a constant extracted through simulation based on the geometric position of each TSV in the structure. The compact model equations are used to derive the Elmore delay, $T_{d,RC}$, for the lumped model in Figure 2 [4]. This model represents a single TSV in the bundle. For simple series RC dominant wire the Elmore delay can accurately estimate the delay. The Elmore delay for rise time between 0%-50% is the sum of individual node delays of the circuit can be described as follows.

$$T_{d,RC} = \ln(2)\Sigma RC \tag{5}$$

Now based on these equations, we extract the time delay for the equivalent RC circuit shown in Figure 2 as follows.

$$T_{d,RC} = (R_{drv}(C_{drv}+C_t+C_L)+R_t(C_t+C_L))\ln(2) \tag{6}$$

Equation (6) contains two terms, delay from the driver throughout the wire and delay due to the total wire impedance. With further rearrangement

$$T_{d,RC} = \ln(2)(R_{drv}C_{drv}+(R_{drv}+R_t)C_L+(R_t+R_{drv})C_t) \tag{7}$$

Now if we substitute R_t by equation (1) and C_t by equation (3), we get the time delay for each TSV in the worst case.

$$T_{d,RC} = \ln(2)(R_{drv}C_{drv} + (R_{drv} + (\rho_{cu}\frac{l_v}{\pi(r_v)^2}))C_L) + \ln(2)((R_{drv} + (\rho_{cu}\frac{l_v}{\pi(r_v)^2}))(\frac{63.34\epsilon_o l_v}{ln(1+5.26\frac{l_v}{r_v})} + \frac{\epsilon_o l_v}{\ln(\frac{k_5 p_v}{r_v})}[k_1(\frac{p_v}{r_v})^{k_2} + k_3(\frac{p_v}{l_v})^{k_4}])) \tag{8}$$

Given the assumptions that the TSV behaves like local wire, the lumped model is accurate enough to estimate the delay and hence the bandwidth. The bandwidth (BW) of N numbers of TSV (N_{tsv}) in any given structure can be estimated by measuring the rate of bit streaming of each TSV multiplied by the total number of TSVs in the bundle.

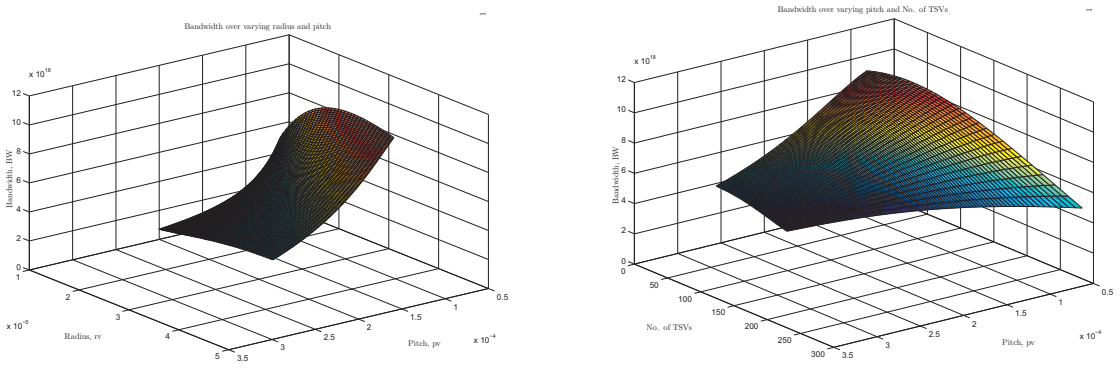
$$BW_{max} = \frac{N_{tsv}}{T_{d,RC}} \tag{9}$$

This is the maximum BW that can be achieved considering the worst case delay found in the mid-TSV of Figure 1a. To be in the safe region, it is sufficient approximation to assume that the average BW is only one-third less than that of the maximum.

$$BW_{avg} = \frac{2}{3}BW_{max} \tag{10}$$

Substituting N_{tsv} by re-arranging equation (2) and $T_{d,RC}$ by equation (8) and with some simplification gives

$$BW(r_v, p_v) = \left(\frac{\left(\frac{(\sqrt{A_{str}} - 2(r_v + d_v) + p_v)^2}{(p_v)^2}\right)}{(R_{drv}C_{drv} + (R_{drv} + (\rho_{cu}\frac{l_v}{\pi(r_v)^2}))C_L + (R_{drv} + (\rho_{cu}\frac{l_v}{\pi(r_v)^2}))(\frac{63.34\epsilon_o l_v}{ln(1+5.26\frac{l_v}{r_v})} + \frac{\epsilon_o l_v}{\ln(\frac{k_5 p_v}{r_v})}[k_1(\frac{p_v}{r_v})^{k_2} + k_3(\frac{p_v}{l_v})^{k_4}])}\right) \tag{11}$$



(a) (b)

Figure 3. Maximum BW is attained when N_{tsv} = 64, p_v = $130.40\mu m$, r_v = 45μm, $BW_{max} = 10.05 \times 10^{18}$ bits per second

us the BW expressed in terms of BW = $F(r_v, p_v)$. The model is the RC equivalent of a TSV found in the middle of the bundle which is worst affected by the coupling effect. Thus we derive the worst case expression for the total bandwidth from the Elmore delay, $T_{d,RC}$, and the number of TSVs in the bundle, N as a function of r_v, and p_v which is shown in equation 11. To get the maximum bandwidth, we use differential equation to derive the local maxima of this expression or use computational tools such as Matlab as shown in Figure 3(a) and b as described in [6]. From these Figures the BW within range can be evaluated and the maximum BW can be easily extracted.

3. Discussion and Conclusions

In practice, compared to the long horizontal metal wires, TSVs behave more like local wires. They are short and fat wires with radius in micro meter range.It is conclusive from the analysis that the method of BW optimization is effective for ideal drivers to attain maximum bandwidth with number of TSVs fewer than the maximum possible $N_{tsv} < N_{max}$. For example, the plot in Figure 3(a) and b. shows that with ideal drivers in a fixed area of A = 1 mm^2, though it is possible to distribute maximum of 296 TSVs, the maximum bandwidth is found only when the number of TSVs = 64. The coupling effect increases when more than 64 TSVs are stacked affecting the BW negatively, where as less than 64 TSVs do not give the maximum possible BW. When both ends of the TSVs are wired to a driver, the maximum bandwidth is achieved only by stacking as many TSVs as possible. This is because for real drivers, the delay is dependent on the driver impedance and the TSV impedance change has small significance on the total delay. The maximum number of TSV is achieved only when the smallest radius TSV with the minimum pitch distance is used. With scaling, when a radius r_v in nanometer range is possible to process and when the TSV impedance is significant enough to drop the delay ($R_t >> R_{drv}$), this optimization method can be used to design and fabricate a circuit with fewer TSVs and higher bandwidth.

4. Acknowledgments

Funding support from the European Commission under grant $FP7 - ICT - 215030$ (ELITE) of the 7th framework program is gratefully acknowledged.

References

[1] M. Celik, Lawrence Pileggi and Altan Odabasioglu. *IC book:IC interconnect Analysis*. Kluwer Academic

Publishers, May 2002.

[2] ITRS. *The International Technology Roadmap for Semiconductors*. 2007 Edition, www.itrs.net/reports.html, 2007.

[3] R. Weerasekera. *System interconnection design tradeoffs in three dimensional integrated circuits*. Ph.D. dissertation, The Royal Institute of Technology (KTH), Stockholm, Sweden,pp. 67-78, 2008.

[4] W. C. Elmore. The transient response of damped linear networks with particular regard to wideband amplifiers. *Journal of Applied Physics, vol. 19, no.1, pp. 55-63, 1948*.

[5] Y. I. Ismail, E. G. Friedman and J. L. Neves. Equivalent Elmore delay for RLC trees. *IEEE Trans. Computer-Aided Design*, vol. 19, pp. 83-97, Jan. 2000.

[6] Dinesh Pamunuwa, Li-Rong Zheng, and Hannu Tenhunen. Maximizing throughput over parallel wire structures in the deep submicrometer regime. *IEEE Trans. on VLSI systems, vol. 11, No. 2, pp 224-243, Apr. 2003*.