Evaluation of 3DICs and Fabrication of Monolithic Interlayer Vias

William Wahby, Ashish Dembla, Muhannad Bakir
Georgia Institute of Technology

***Abstract*—A compact model for interconnect length in homogeneous 3DICs is presented. The new model accounts for lateral TSV size, which is often much larger than the gate pitch, leading to TSV-induced gate blockage and potentially affecting the wirelength distribution. The impact of TSV diameter on maximum wirelength and wiring power is investigated, and systems with smaller vias are found to have better properties. Accordingly, fabrication results for nanoscale monolithically integrated copper vias are presented to demonstrate the feasibility of developing 3D systems with dense vertical integration which do not suffer from TSV-induced gate blockage.**

I. Introduction

Accurate wirelength distributions are required for rapid and accurate prediction of 3DIC system parameters. By combining a wirelength distribution with wire layer assignment and repeater insertion algorithms, the electrical and thermal behavior of a design can be predicted in advance of its actual development.

Complicating projections of 3D system performance are the additional degrees of freedom and complexity inherent to 3D design. Stacked device layers must be interconnected vertically, and the method of interconnection strongly influences other aspects of the design. Conventional 3D designs rely on through-silicon vias (TSVs) for vertical interconnection, but TSV aspect ratios are currently limited by etch technology to roughly 20:1 [1–3]. Additionally, large keep-out zones are required around TSVs to isolate adjacent logic from thermo-mechanical stresses. Consequently, TSVs have a much larger footprint than individual logic gates, significantly limiting the number of vertical interconnects which can be incorporated into a design without consuming excessive die area. Additionally, TSVs must be allocated to power delivery as well as signal routing, further reducing the resources available for wirelength reduction. Monolithic 3D systems skirt the issue of TSV size by integrating device layers directly atop one another and interconnecting them with monolithic interlayer vias (MIVs) of comparable size to the gate pitch [4–6]. The differences between these types of 3D interconnects impact the realizable performance benefits from 3D integration and affect the sensitive tradeoff between die size and interlayer connectivity, but are not considered in existing 3D wirelength models.

Stochastic wirelength distributions have been shown to be effective tools for system prediction in conventional 2D chips [7, 8]. By determining both the number of gates separated by a given length and the probability that gates separated by that length will be connected, an estimate for the distribution of wirelengths in a given system can be obtained. Much work has been done to extend conventional 2D wirelength prediction to 3DICs [9, 10], but the impact of finite TSV area on the wirelength distribution has received little consideration.

Recently, [11] adapted the method of [10] to account for TSV area, but the corrected distribution depends on terms which must be calculated via brute force methods, which become cumbersome when considering large systems. Significant simplifications can be made by considering regular TSV geometries. Using the method detailed in [10, 12] as a starting point, we present a compact model for the 3D wirelength distribution which accounts for the displacement of logic gates by TSVs. This new wirelength distribution can be applied to any vertical interconnect technology, as the corrections depend only on interconnect surface area and placement.

The new wirelength distribution is combined with wire layer assignment and repeater insertion algorithms and a GUI frontend to create a CAD tool capable of 3D system property prediction. We investigate the impact to maximum interconnect length and power consumption as the die area allocated to TSVs is varied. The design implications of conventional TSVs and MIVs are compared by considering a design with a fixed gate count and sweeping the TSV diameter from tens of microns down to dozens of nanometers. To justify the consideration of MIVs and nanoscale TSVs, experimental results are presented for copper-filled vias with diameters of 117 nm and aspect ratios of roughly 15:1.

II. Methodology

For a homogeneous chip, [10] defines the probability distribution function (PDF) for a gate existing at position x,y to be

$$f[x,y] = \frac{1}{N_g} r(x,0,N_x) r(x,0,N_y) \tag{1}$$

where $r(x,a,b)$ is the unit rectangle function beginning at $x = a$ and ending at $x = b$, N_g is the number of gates on the chip, and N_x and N_y are the lengths of the chip in the x and y directions, measured in gate pitches. To simplify the analysis, we assume a periodic array of TSVs. Following [10] and [12] we assume that each gate has the same length and width, and that the chip is square. We further assume that each TSV is symmetric. In order to account for locations without TSVs, all locations in which a gate is forbidden due to the presence of a TSV must be subtracted from Eq. (1). The modified 2D PDF then becomes

$$\begin{aligned} f[x,y] =& \frac{1}{N} \left(r(x,0,N_x) r(y,0,N_y) \right) \\ & - \frac{1}{N} \sum_{n,m} r(x, nT+t, w) r(y, mT+t, w) \end{aligned} \tag{2}$$

978-1-4673-6484-3/13/$31.00

where T is the TSV pitch, t is the TSV offset (distance from edge of the chip to the first TSV), w is the TSV width, N_x is the length of the chip in the x direction, N_y is the length of the chip in the y direction, n is the TSV index in the x direction, and m is the TSV index in the y direction. T, t, w, N_x, and N_y are all measured in gate pitches. For simplicity, let q_{xy} designate the TSV correction term, and let f_o to be the nonnormalized two-dimensional PDF in the absence of TSVs. Then Eq. (2) becomes

$$f[x,y] = \frac{1}{N}\left(f_o - q_{xy}\right) \tag{3}$$

In order to determine the total number of interconnects of length l, we must first determine the number of gate pairs separated by l gate lengths. If the starting point of the interconnect is (x_1, y_1), and the ending point is (x_2, y_2), then both $f[x_1, y_1]$ and $f[x_2, y_2]$ must be nonzero for the interconnect to be considered valid. In order to determine the number of gate pairs which satisfy this criterion, M_t^*, all possible combinations of x_1, x_2, y_1, and y_2 must be considered.

$$M_t^*[l] = \sum_{l_x=0}^{l} \sum_{x_1=0}^{N_x-1} \sum_{y_1=0}^{N_y-1} f[x_1,y_1] f[x_2,y_2] \tag{4}$$

The various coordinates can be constrained as follows

$$l_x = x_2 - x_1 \tag{5}$$
$$l_y = y_2 - y_1 \tag{6}$$
$$l = l_x + l_y \tag{7}$$

Expanding Eq. (4) we find

$$M_t^*[l] = \sum_{l_x=0}^{l} \sum_{x_1=0}^{N_x-1} \sum_{y_1=0}^{N_y-1} (f_{o_1} f_{o_2} - f_{o_1} q_{xy_2} - f_{o_2} q_{xy_1} + q_{xy_1} q_{xy_2}) \tag{8}$$

The first term in Eq. (8) yields the number of gate pairs separated by distance l in the case where TSV width is ignored. This is simply the original result from [10]

The second term in Eq. (8) determines the number of interconnects which start on an allowed gate location, but end on a forbidden location. Since the x_2 and y_2 portions of the summation in term 2 are completely independent of one another, they may be separated and treated as independent quantities. Interchanging the order of the summations yields expression in the form of $f(x,y) = g(x)h(y)$.

$$M_{t2}^*(l) = \sum_{l_x} M_{t2_x}^* M_{t2_y}^* \tag{9}$$

$$M_{t2_x}^*(l_x) = \sum_{n,x_1} r(x_1, 0, N_x) r(x_1 - l_x, nT + t, w) \tag{10}$$

$$M_{t2_y}^*(l_y) = \sum_{m,y_1} r(y_1, 0, N_y) r(y_1 - l_y, mT + t, w) \tag{11}$$

For a periodic rectangular TSV array, we can construct a function $g_x(l_x)$ which reproduces the behavior of the complete summation.

$$g_x(l_x) = \begin{cases} n_t w & l_x' < T - t - w \\ (n_t - 1)w & l_x' > T - t \\ n_t w - l_x' + T - t - w & \text{else} \end{cases} \tag{12}$$

$$l_x \in [0, N_x] \tag{13}$$
$$l_x' = l_x \bmod T \tag{14}$$
$$n_t = (n_{max} + 1) - \left\lfloor \frac{l_x}{T} \right\rfloor \tag{15}$$

A similar analysis for $M_{t2_y}^*$ yields an equivalent function, $g(l_y)$, which is identical to Eq. (12) with $x \to y$, $n \to m$, and $l_y = l - l_x$. For a square chip, $N_x = N_y$, $g_x = g_y = g$, and M_{t2}^* becomes

$$M_{t2}^*(l) = \sum_{l_x} g_x(l_x) g_y(l - l_x) \tag{16}$$

This is simply the formula for the discrete convolution of g with respect to itself. We can therefore write

$$M_{t2}^*(l) = g_x * g_y \tag{17}$$

The third term in Eq. (8) counts the number of possible interconnects which start in a forbidden zone, but end in an allowed zone. Its similarity to term 2 can be used to to quickly determine a simplified form.

$$M_{t3}^*(l) = h_x * h_y \tag{18}$$

where h_x is equivalent to g_x with $T - t - w \to t$ and $T - t \to t + w$

The final term in Eq. (8) corrects for the fact that terms 2 and 3 double-count forbidden interconnections. Term 2 counts interconnects which end in forbidden zones, whereas term 3 counts interconnects which start at forbidden zones. Potential interconnects which both start and end in forbidden locations are counted in both terms 2 and 3, and the final term corrects for this issue. This term is not as analytically tractable, but it is straightforward to compute numerically. It can be safely ignored in many cases, as the number of doubly-forbidden interconnections is much smaller than the number of singly-forbidden paths.

Now Eq. (8) can be rewritten in terms of known quantities

$$M_t^*[l] = M_t^o - g * g - h * h + \sum_{l_x=0}^{l} \sum_{x_1=0}^{N_x-1} \sum_{y_1=0}^{N_y-1} q_{xy_1} q_{xy_2} \tag{19}$$

where M_t^o is the original 2D gate separation function given as M_s in [12].

In order to make quantitative predictions about system behavior, a wire layer assignment algorithm (WLA) was implemented, following the method of [12]. An optimal repeater insertion (RI) scheme is also used, following [13].

In conjunction with these WLA and RI algorithms, the new wirelength model can be used to estimate power consumption in a system of interest for different 3D integration scenarios. Once power draw is known, an estimate for the number of TSVs required for power delivery can be determined.

III. Results

A. Model evaluation

The corrected wirelength distribution is compared to the original model in Figs. 1 and 2. In order to demonstrate the operation of the new model, a hypothetical high performance logic design at the $32nm$ node is considered. The design under

consideration has 10^9 gates, an average fanout of 4, a TSV aspect ratio of 20:1, a rent exponent $p = 0.6$ corresponding to a logic-heavy design, a rent constant $k = 3.75$, a logic activity factor of 0.1, a gate length of $32nm$, and a gate pitch of $100nm$. In this work, the number of gates is fixed, and the tier area is allowed to grow as TSVs are inserted.

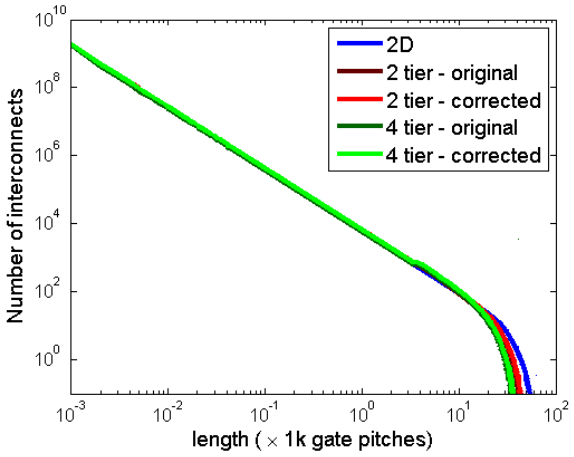

Fig. 1. Comparison between the original distribution and the corrected distribution The distributions are compared for 3D designs folded across 2 and 4 layers.

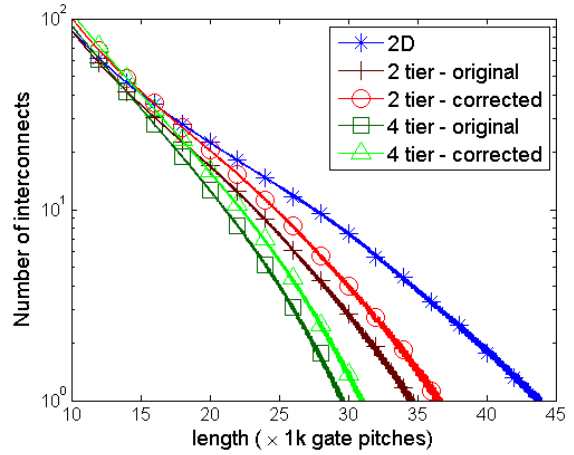

Fig. 2. Comparison between the original distribution and the corrected distribution zoomed in on long interconnects. The corrected distribution predicts an increase in maximum interconnect length as well as a greater number of long interconnects than the original model. In two dimensions, the corrected distribution collapses to the original result, and is omitted here for clarity.

The same test case is used to investigate the impact of silicon wafer thickness on wirelength distribution. In Fig. 3 the test case is folded across four logic tiers. The TSV aspect ratio is fixed at 20:1, and the interlayer separation is swept from $300\mu m$ down to $10\mu m$. For designs with small separations between active tiers, the wirelength distribution closely agrees with the ideal result, but as layer thickness increases beyond $10\mu m$ significant deviation is observed due to the displacement of logic gates by TSVs.

Wiring and repeater power predictions for 2D and four tier systems operating at 1 GHz are compared in Fig. 4. The 3D

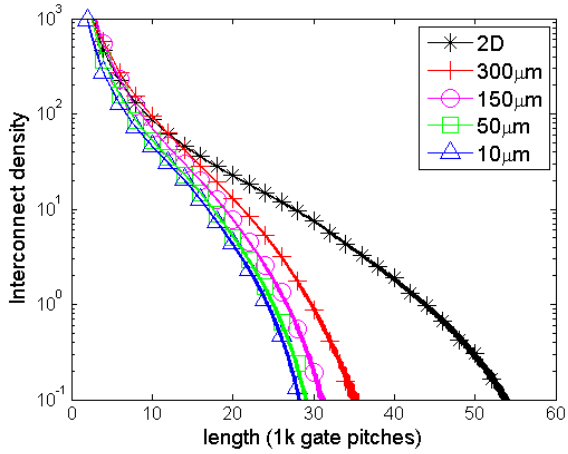

Fig. 3. Wirelength distribution for a 4 tier design with 10^9 logic gates implemented with different separations between the active layers. As the interlayer separation is increased, both the length and number of long interconnects are increased, impacting wiring capacitance, signal delay, and wiring power.

design exhibits significant power savings over the 2D case due to its reduced wirelength and wiring capacitance.

Complicating 3DIC design is the fact that 3D systems may also require 3D power delivery, in which case only a fraction of the TSVs in a design will be available for signal routing. In this case the impact to the wirelength distribution may be much greater, as power TSVs displace logic gates without providing the corresponding reduction in overall wirelength. Additionally, since the total area required for both signal and power TSVs will increase, TSVs may consume unacceptably high portions of silicon area. In such cases increased TSV density, achievable either by increasing TSV aspect ratios or by thinning substrates, may be necessary to meet design goals.

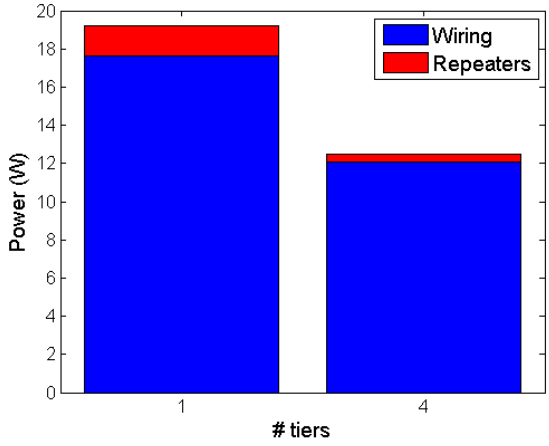

Fig. 4. Wiring and repeater power dissipation estimates for a system with 10^9 logic gates, a 20:1 TSV aspect ratio, and an interlayer separation of $10\mu m$ operating at a frequency of 1GHz. The 4-tier design exhibits significant reductions in both wiring and repeater power due to the wirelength reduction obtained over the 2D design.

To simplify the use of the new model a simple GUI was developed, as shown in Fig. 5. The user may enter the relevant system parameters, run the simulation, and view the results

without the use of a command line. Data can be exported either as complete plots or as comma-delimited text files.

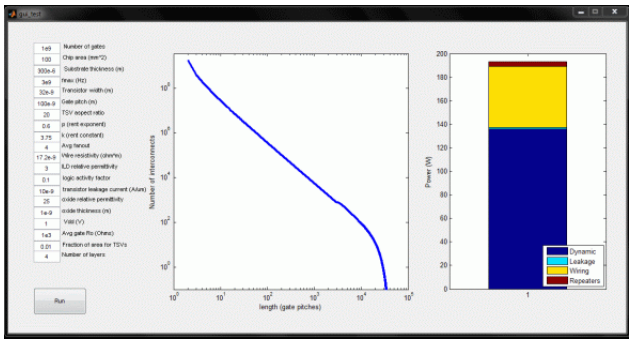

Fig. 5. GUI frontend for the new model. Users input the relevant design parameters and the interconnect distribution is automatically generated and displayed. A breakdown of the power dissipation in the design under consideration is also displayed.

B. Nanoscale Copper Vias

Small interlayer vias are a key enabler of high performance 3DICs. As shown in Fig. 6, decreasing interlayer via size decreases the capacitance of each via (and hence decreases delay and power dissipation), and increases the number of vias which can be incorporated into the same area, increasing interlayer connectivity. Shorter vias also reduce wirelength and wiring power dissipation, as shown in Figs. 3 and 4. Accordingly, we investigated the fabrication of nanoscale copper vias with diameters ranging from $500nm$ down to $100nm$.

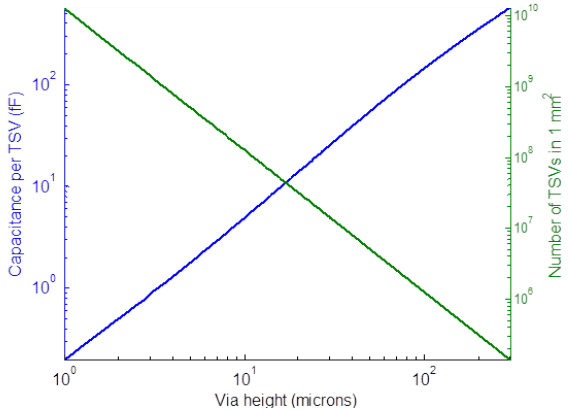

Fig. 6. Via capacitance (blue) and number of vias (green) as via height is varied from $1\mu m$ to $300\mu m$ with a fixed aspect ratio of 20:1. Smaller vias have lower capacitance (and hence lower delay and power dissipation) and can be integrated at much higher densities than large vias.

Nanoscale copper vias of various geometries and sizes were fabricated on conventional silicon wafers, as well as SOI wafers with a $2\mu m$ device layer, $1.5\mu m$ buried oxide, and a $300\mu m$ silicon handle layer. Rectangular vias nominally measuring $250nm \times 1\mu m$ and $350nm \times 700nm$, as well as circular vias with $500nm$ radius were fabricated to characterize the process. Ultimately, circular nanoscale copper vias with diameters of $117nm$ and heights of $1.8\mu m$ were fabricated. The fabrication process is shown in Fig. 7.

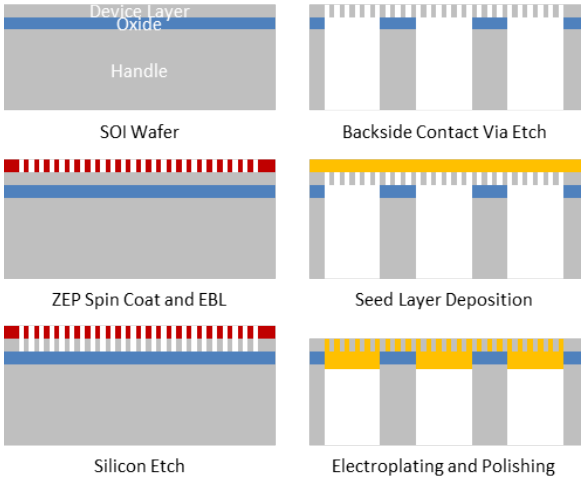

Fig. 7. Nanoscale via fabrication flow. Large backside vias are etched in order to simplify the copper-fill process, enabling the use of bottom-up electroplating to fill the vias.

First, ZEP520A, a positive electron beam resist, was spun on the device side of the wafer. The nanoscale vias were defined using a JEOL JBX 9300FS electron beam lithography system at $100kV$, $2nA$, and a shot pitch of $6nm$, and the resist was developed in amyl-acetate for 2 minutes. Patterned photoresist after development is shown in Fig. 8. The pattern was transferred into the silicon via a Bosch process etch, using the buried oxide as an etch stop. Etched rectangular and circular vias on conventional silicon wafers are shown in Figs. 9 and 10. Via openings after etching are shown in Fig. 11. Vias etched on an SOI wafer are shown in Fig. 12.

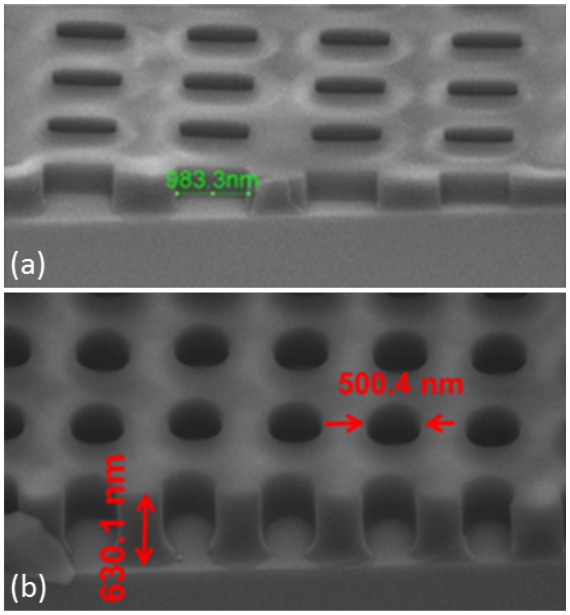

Fig. 8. Photoresist after exposure and development. (a) Rectangular via measuring roughly $200nm \times 1000nm$. (b) Circular via with $500nm$ radius.

To simplify fabrication, large backside contact vias are etched through the handle layer using a similar process, and a 5:1 BHF etch is used to remove the buried oxide between

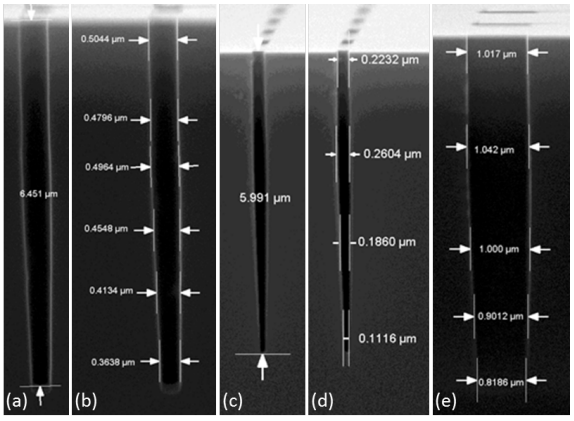

Fig. 9. Cross sections of etched vias. (a) and (b) $6.451\mu m$ deep circular via with diameter $500nm$. (c), (d), and (e) $5.991\mu m$ deep rectangular via with $223nm \times 1017nm$ opening.

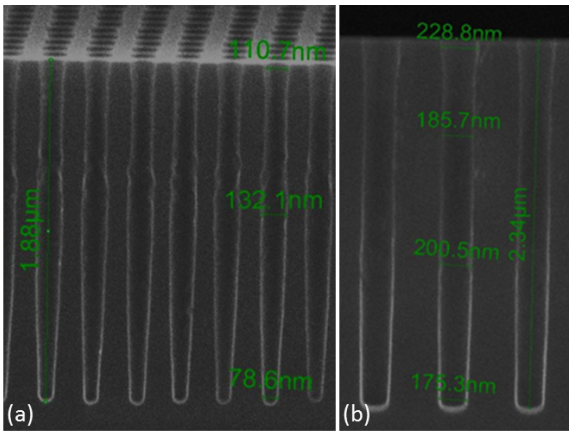

Fig. 10. FIB cross section showing etched nanoscale circular vias. (a) Vias with $100nm$ openings and aspect ratios of roughly 20:1. (b) Vias with $200nm$ openings and aspect ratios of roughly 10:1.

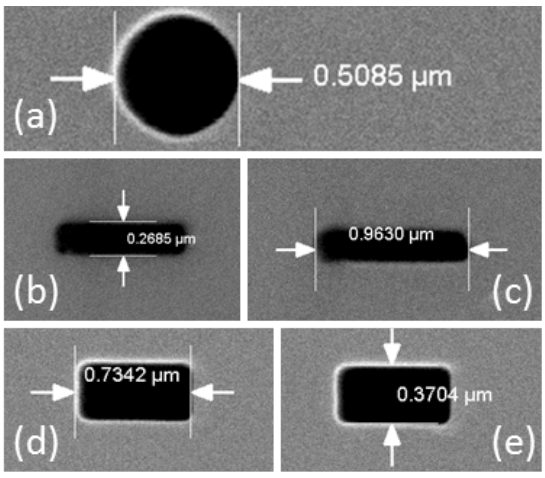

Fig. 11. Nanoscale via openings after etch. (a) Circular via with $508nm$ diameter. (b) and (c) Rectangular via measuring $269nm \times 963nm$. (d) and (e) Rectangular via measuring $734nm \times 370nm$.

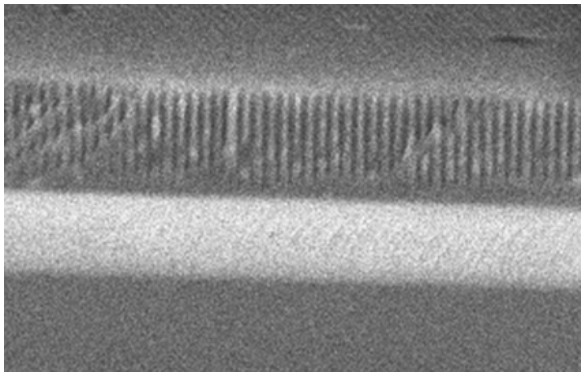

Fig. 12. Densely packed sea of etched nanoscale vias. The oxide layer is clearly visible as the white band below the etched vias.

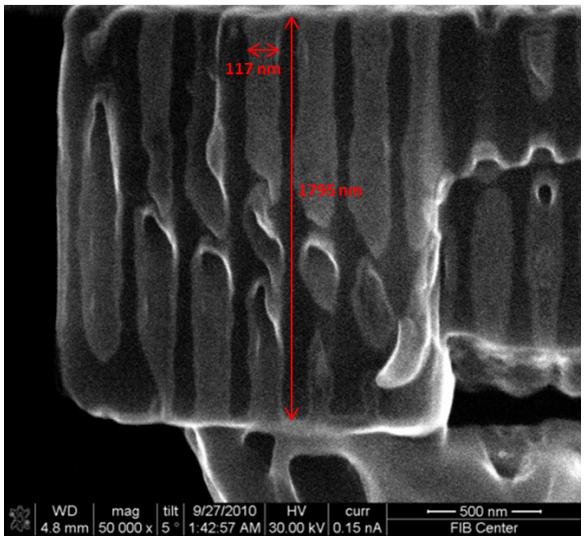

Fig. 13. Fully plated copper vias with widths of $117nm$ and heights of $1.8\mu m$ (aspect ratio of 15). Voids are introduced during the FIB process

the backside vias and the nanoscale vias. A thin ($20nm$) oxide liner is grown using thermal oxidation to isolate the vias from the surrounding silicon. A thin electroplating seed layer (Ti/Cu/Ti) is deposited on the device side of the wafer via electron beam evaporation. The small dimensions of the nanoscale vias allow them to act as a superfine mesh [14]. Finally, bottom up copper electroplating is used to fill the vias. The fully-plated nanoscale interlayer vias are shown in Fig. 13.

IV. Conclusions

A modified wirelength distribution suitable for TSV-based 3DICs was developed and used to examine the impact of substrate thickness and via size on wirelength distribution and power consumption. The wirelength distribution was paired with simple wire layer assignment and repeater insertion algorithms in order to generate estimates for total system performance, and a GUI interface was developed to simplify operation. Fabrication results for nanoscale copper vias were presented in order to justify the consideration of ultrathin substrates.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Semiconductor Research Corporation Global Research Collaboration (task ID 2254.001).

REFERENCES

[1] A. Dembla, Y. Zhang, and M. Bakir, "High aspect ratio TSVs in micropin-fin heat sinks for 3D ICs," in *Nanotechnology (IEEE-NANO), 2012 12th IEEE Conference on*, 2012, pp. 1–6.

[2] F. Inoue, H. Philipsen, A. Radisic, S. Armini, Y. Civale, P. Leunissen, and S. Shingubara, "Novel seed layer formation using direct electroless copper deposition on ALD-Ru layer for high aspect ratio TSV," in *Interconnect Technology Conference (IITC), 2012 IEEE International*, 2012, pp. 1–3.

[3] A. C. Fischer, S. J. Bleiker, T. Haraldsson, N. Roxhed, G. Stemme, and F. Niklaus, "Very high aspect ratio through-silicon vias (TSVs) fabricated using automated magnetic assembly of nickel wires," *Journal of Micromechanics and Microengineering*, vol. 22, no. 10, p. 105001, 2012.

[4] S. Wong, A. El-Gamal, P. Griffin, Y. Nishi, F. Pease, and J. Plummer, "Monolithic 3D Integrated Circuits," in *VLSI Technology, Systems and Applications, 2007. VLSI-TSA 2007. International Symposium on*, 2007, pp. 1–4.

[5] D. Sekar and Z. Or-Bach, "Monolithic 3D-ICs with single crystal silicon layers," in *3D Systems Integration Conference (3DIC), 2011 IEEE International*, 2012, pp. 1–2.

[6] Y.-J. Lee, P. Morrow, and S. K. Lim, "Ultra high density logic designs using transistor-level monolithic 3D integration," in *Proceedings of the International Conference on Computer-Aided Design*, ser. ICCAD '12. New York, NY, USA: ACM, 2012, pp. 539–546.

[7] J. Davis, V. De, and J. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI). I. Derivation and validation," *IEEE Transactions on Electron Devices*, vol. 45, no. 3, pp. 580–589, 1998.

[8] J. Davis, V. De, and J. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI). II. Applications to clock frequency, power dissipation, and chip size estimation," *IEEE Transactions on Electron Devices*, vol. 45, no. 3, pp. 590–597, 1998.

[9] A. Rahman, A. Fan, J. Chung, and R. Reif, "Wire-length distribution of three-dimensional integrated circuits," in *Interconnect Technology, 1999. IEEE International Conference*, 1999, pp. 233–235.

[10] J. Joyner, P. Zarkesh-Ha, J. Davis, and J. Meindl, "A three-dimensional stochastic wire-length distribution for variable separation of strata," in *Interconnect Technology Conference, 2000. Proceedings of the IEEE 2000 International*, 2000, pp. 126–128.

[11] D. H. Kim, S. Mukhopadhyay, and S.-K. Lim, "TSV-aware interconnect length and power prediction for 3D stacked ICs," in *Interconnect Technology Conference, 2009. IITC 2009. IEEE International*, 2009, pp. 26–28.

[12] J. Joyner, R. Venkatesan, P. Zarkesh-Ha, J. Davis, and J. Meindl, "Impact of three-dimensional architectures on interconnects in gigascale integration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 9, no. 6, pp. 922–928, 2001.

[13] H. Bakoglu, *Circuits, interconnections, and packaging for VLSI*, ser. VLSI systems series. Addison-Wesley Pub. Co., 1990.

[14] J.-H. Lai, H. S. Yang, H. Chen, C. R. King, J. Zaveri, R. Ravindran, and M. S. Bakir, "A 'mesh' seed layer for improved through-silicon-via fabrication," *Journal of Micromechanics and Microengineering*, vol. 20, no. 2, p. 025016, 2010.