# Power Trends and Performance Characterization of 3-Dimensional Integration

Rongtian Zhang, Kaushik Roy, Cheng-Kok Koh, and David B. Janes
ECE, Purdue University, 1285 EE Bldg., West Lafayette, IN 47907-1285
{rzhang, kaushik, chengkoh, janes}@ecn.purdue.edu

*Abstract*—**3-D technology promises higher integration density and lower interconnection complexity and delay. At present, however, not much work on circuit applications has been done due to lack of insight into 3-D circuit architecture and performance. One of the purposes of realizing 3-D integration is to reduce the interconnect complexity and delay of 2-D, which is widely avowed as the barrier to the continued performance gains in the future technology generations. Therefore, in this paper, we present a stochastic 3-D interconnect model, study the impact of 3-D integration on circuit performance and power consumption. We show that 3-D structures effectively reduce the number of long delay nets, significantly reduce the number of repeaters, and dramatically improve the circuit performance. With 3-D integration, a circuit can be clocked at frequencies much higher (double, even triple) than with 2-D. However, we also show that the impacts of vertical wires on chip area and interconnect delay can be limiting factors on the vertical integration of device layers; and that 3-D integration offers limited relief of power consumption.**

*Keywords*— **3-Dimensional Circuits, Repeaters, Performance Characterization, Power Consumption, Technology Generations**

## I. INTRODUCTION

Scaling has been the primary approach in the past few decades to meet circuit performance and power consumption requirements in VLSI circuits. However, as device dimensions shrink to submicron and below, short channel effects, quantum effects, interconnect complexity and delay, pose serious challenges to the conventional 2-D integration and can lead to the emergence of 3-D integration [1]. The main purpose of realizing 3-Dimensional integration is to reduce the interconnect complexity and delay. Therefore, it is imperative to gain thorough understanding of wiring requirements and its impact on the architecture and performance of 3-D circuits. In this paper, the impacts on power and performance of 3-D integration for future technology generations are studied. We investigate repeater insertion, circuit speed, and power consumption of 3-D integration based on a stochastic 3-D interconnect model.

Various approaches have been proposed and developed in realizing multi-device-layer integration [2][3][4][5]. A schematic of 3-D circuit structure with 3 device layers is shown in Fig. 1. Intuitively, 3-D technology offers compact logic gates and, in general, compact functional blocks and circuit structures. In 3-D circuits, transistors are stacked in several device layers, rather than spanned in a vast single device layer. Significant area saving and performance improvement can be achieved due to the increased integration density and reduced wire length of 3-D structures. In [6], for example, we studied the 3-D integration of array multiplier under 10 different schemes (Fig. 2 is the layout of one such scheme with 4 device layers and 2 metals per layer). Nearly 40% area reduction with 18% interconnect delay gain for 2-device-layer structure over standard 2-D structure, and 70% area reduction with 40% interconnect delay gain for 4-device-layer structures were obtained.

However, while 3-D circuits overcome the problem of long and complicated wiring in 2-D layout and increase the integration density, we foresee a new problem arising from the connections of gates in different device layers. Typically, we ignore the effects of vias in multi-metal layers of 2-D circuits. Unfortunately, the same can not be said for the vertical channel that connects one device layer to another in 3-D structures (see Fig. 1). The process technology, materials, and structure of the 3-D vertical channels are completely different from those of 2-D metal contacts, leading to completely different effects on chip area and circuit performance. A large number of vertical channels in a 3-D structure can be detrimental to the integration density due to the additional area required for such channels. The combined parasitic effects of several vertical channels may significantly impact the circuit performance, especially for circuits with large logic depth. Our study shows that more than 90% of wires go through vertical channels in a 8-device-layer structure. Without careful planning, the impact of the vertical channels may negate the advantages that 3-D structures offer.

## II. STOCHASTIC 3-D INTERCONNECT MODELING

A stochastic 3-D interconnect mode is developed for investigating the impact of 3-D structures on circuit performance and power consumption [8]. In order to best understand the wiring requirements, we divide the overall wires into two parts: horizontal wires and vertical wires. For a wire connecting one gate to another, we define the portions that are parallel to the device layers as a *horizontal wire* and the portions that are perpendicular to the device layers as a *vertical wire* (Fig. 3). Horizontal wires determine the overall routing resources on top of each device layer, and are generally realized by metal layers. Vertical wires are realized by vertical channels and have impact on the area of device layer. Both horizontal and vertical wires contribute to the overall interconnection delay. This section presents some important results of the modeling.

For a system with $N$ gates distributed in $m$ device layers, The closed form expressions of the horizontal and vertical wire-length distributions are obtained by extending Rent's Rule [9] and are listed as follows:

**Horizontal Wire-Length Distribution:**

$$h(\ell) = \begin{cases} \Theta(\frac{\ell^3}{3} - 2\ell^2\sqrt{\frac{N}{m}} + 2\ell\frac{N}{m})l^{2p-4}, & 1 \le \ell < \sqrt{\frac{N}{m}}; \\ \frac{\Theta}{3}(2\sqrt{\frac{N}{m}} - \ell)^3\ell^{2p-4}, & \sqrt{\frac{N}{m}} \le \ell \le 2\sqrt{\frac{N}{m}}; \end{cases}$$

where $h(\ell)$ is the number of connections with horizontal distance of $\ell$ gate pitches; and

$$\Theta = \frac{\alpha A m^p \frac{N}{m}(1-(\frac{N}{m})^{p-1})}{-(\frac{N}{m})^p\frac{1+2p-2^{2p-1}}{p(2p-1)(p-1)(2p-3)} - \frac{1}{6p} + \frac{2\sqrt{\frac{N}{m}}}{2p-1} - \frac{\frac{N}{m}}{p-1}}.$$

**Vertical Wire-Length Distribution:**

$$V(k) = \frac{\alpha A N(1-N^{p-1}-m^{p-2}+m^{-1}N^{p-1})}{m(m-1)}(2m-2k),$$

where $V(k)$ is the number of connections with vertical distance of $k$ device layers. $k = 1, 2, ..., m-1$.

$A$, $\alpha$, $p$ are Rent's parameters for interconnect complexity [9][10].

When $m = 1$, the expressions give the wire-length distributions of 2-D circuits, which have no vertical wires.

Figs. 4 and 5 show the horizontal and vertical wire-length distributions, respectively. Significant reduction of horizontal wires with 3-D integration is obtained. However, the reduction is at the expense of a large number of vertical wires.

## III. PERFORMANCE CHARACTERIZATION OF 3-DIMENSIONAL CIRCUITS

With the knowledge of wire-length distributions, we are ready to study the impact of 3-D integration on circuit performance and power dissipation. The evaluation is based on 1999 *International Technology*

Acknowledgment: This research is supported in part by DARPA (N66001-97-1-8903) and SRC (98-HJ-638).

0-7803-6685-9/01/$10.00©2001 IEEE



Authorized licensed use limited to: Purdue University. Downloaded on July 9, 2009 at 14:38 from IEEE Xplore. Restrictions apply.

*Roadmap for Semiconductors* (ITRS) [11]. Elmore delay model [9] is used in accessing interconnect delay and repeater insertion. To simplify the calculation, we assume that each gate is a two-input NAND gate, and the repeater is the two cascaded inverters, as shown in Fig. 6. Based on transistor density predicted by ITRS and on performance requirements, the sizes of the gate transistors are estimated to be $W_n/L_n = 10$ and $W_p/L_p = 10$. Also, for a net with both horizontal wire and vertical wire, we assume that the vertical wire is in the middle of the net regardless of the length of the net (Fig. 7). We denote the resistance and capacitance of a horizontal wire with one gate pitch as $R$ and $C$, respectively; and denote the resistance and capacitance as $Rv * R$ and $Cv * C$ respectively for a vertical wire with one device layer depth, where $Rv$ and $Cv$ are the coefficients that relate the resistances and capacitances of vertical wires and horizontal wires. Under the assumption that tungsten plugs are used as vertical channels; $Rv$ and $Cv$ are estimated to be 10 and 0.85, respectively.

Table I lists the parameters of high performance MPU's for the six technology generations. We assume that the wire width is two times the minimum gate length. And the normalization coefficient $Rv$ and $Cv$ remain the same across the six technology generations ($Rv = 10.0$ and $Cv = 0.85$). Fan-out of a gate is assumed to be 3, which gives us $A = 3.8$ and $\alpha = 0.74$ [10]. Rent's constant $p$ is arbitrarily set as 0.45 based on the facts that $p = 0.45$ roughly reflects the interconnect complexity of modern MPU's and that our discussions and conclusions of this section are independent of the value of $p$.

Fig. 8 plots the number of repeaters needed over technology generations for various number of device layers. We observe the significant reduction of repeaters with 3-D structures. The reduction of the repeaters would relieve the difficulty of placement and routing due to repeater insertion. Moreover, reducing the number of repeaters will further reduce the chip area and increase the integration density. Fig. 8 also shows that the demand of repeaters increases rapidly with the migration from one technology generation to another. This implies that 3-D integration will become important for future technology generations in reducing repeater requirement.

While the number of repeaters is reduced, the clock rate is increased by 3-D structure. Fig. 9 plots the worst case clock rates over technology generations for different number of device layers. The worst case clock period is assumed to be proportional to the longest net delay [11][9]. We also assume that the corresponding 2-D circuit performance at each technology node meets its ITRS target so that we can normalize the clock rates of all generations to $180nm$ node. We observe that 3-D structures have two or three technology generations advantage over 2-D circuits in terms of performance.

## IV. Power Trend of 3-Dimensional Circuits

Following the study of 3-D circuit performance, we continue to investigate the power trend. Figs. 10 and 11 plot the average power consumption per switching for a single gate and for the whole chip, respectively. In estimating the power consumption, we assume that the supply voltage remains the same within the technology generation regardless of the number of device layers. Therefore, the power due to gate switching capacitance is the same for any number of device layers in the same technology generation. Based on our estimation, power consumption is still dominated by the gate switching capacitance. However, power consumed by interconnects (including repeaters) is becoming substantial, taking about 15% to 25% of the total power. This is not an insignificant portion and should call for our attention.

Despite the continuing reduction of power consumption in each gate, the total power for the whole system continually increases over technology generations due to the continuous increase of the transistor count. Nevertheless, 3-D integration can provide a relief, as shown in Figs. 10 and 11. This mainly comes from the reduction of the interconnect capacitance, since we do not alter the device or circuit structures.

However, we observe that there is a limit on how much the reduction can be achieved by 3-D integration. In fact, integrating more device layers may increase the power consumption. To understand this phenomenon, we take $180nm$ technology node as an example and plot the interconnect capacitance distribution for different number of device layers (Fig. 12). The plot shows that while 2-D larger capacitance nets are converted into lower capacitance nets by 3-D integration, the 2-D lower capacitance nets are also converted into higher capacitance nets. Therefore, there is a tradeoff between reducing higher capacitance nets and increasing lower capacitance nets. In another words, there exists an optimum number of device layers that gives the lowest total interconnect capacitance.

The existence of an optimum number of device layers that gives the lowest total interconnect capacitance can be more clearly seen in Fig. 13, where the horizontal wire capacitance, vertical wire capacitance, and the total interconnect capacitance versus the number of device layers are plotted. The plot shows that while the horizontal wire capacitance is reduced by 3-D integration, the vertical wire capacitance is increased. For the small number of device layers, the decrease of horizontal wire capacitance overrides the increase of the vertical wire capacitance. Therefore, the total interconnect capacitance decreases, so does the power consumption. However, with large number of device layers, the trend of decreasing horizontal wire capacitance slows down while the rate at which the vertical wire capacitance increases almost remains the same. Thus, the total wire capacitance increases. This in turn increases the power consumption.

The increase of interconnect complexity outpaces the increase of the circuit complexity [11]. The percentage of total power consumption due to the interconnect increases with each new technology generation. Reducing interconnect power consumption through 3-D integration, together with circuit low-power techniques, is important for future VLSI technology.

## V. Conclusions

Base on the 3-D wire-length distributions, 3-D circuit applications for future technology generations are investigated. Our analyses show that 3-D circuits have two or three technology generations advantage over 2-D in terms of performance. Moreover, 3-D integration can naturally provide a relief for power consumption due to the reduction of the interconnect capacitance, even though the relief is only limited. We conclude that 3-D integration can be a viable solution for low-power high-performance designs in the future technology generations.

## References


[1] S. J. Souri, K. Banerjee, A. Mehrotra, and K. C. Saraswat, :Multiple Si Layer ICs: Motivation, Performance Analysis, and Design Implications" *Design Automation Conference*, pp.213-220, 2000.

[2] R. Zingg, J. A. Friedrich, G. W. Neudeck, and B. Hofflinger, "Three-Dimensional Stacked MOS Transistors by Localized Silicon Epitaxial Overgrowth", *IEEE Transactions on Electron Devices*, Vol. 37, No. 6, pp.1452-1461, 1990.

[3] G. W. Neudeck, S. Pae, J. P. Denton, and T. Sue, "Multiple Layers of Silicon-on-Insulator for Nanostructure Devices", *Journal of Vacuum Science & Technology B*, Vol. 17, No. 3, pp.994-998, 1999.

[4] K. Sarawat, S. J. Souri, V. Subramanian, A. R. Joshi, and A. W. Wang, "Novel 3D Structures", *Proceedings of 1999 International SOI Conference*, pp.54-55, 1999.

[5] P. Ramm, D. Bollman, R. Braun, R. Buchner, et al, "Three Dimensional Metalization for Vertically Integrated Circuits", *Microelectronic Engineering*, 37/38, pp.39-47, 1997.

[6] R. Zhang, K. Roy, and D. B. Janes, "Architecture and Performance of 3-Dimensional Circuits", *IEEE International SOI Conference*, pp.44-45, 1999.

[7] S. J. Abou-Samra, P. A. Aisa, A. Guyot, and B. Coutois, "3D CMOS SOI for High Performance Computing", *Proceedings of 1998 International Symposium on Low Power Electronics and Design*, pp.54-58, California, 1998.

[8] R. Zhang, K. Roy, C.-K. Koh, and D. B. Janes, "Stochastic Wire-Length and Delay Distributions of 3-Dimensional Circuits", *International Conference on Computer-Adide Design*, pp.208-213, 2000.

[9] H. B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley, 1990.

[10] J. A. Davis, V. K. De, and J. Meindl, "A Stochastic Wire-Length Distribution for Gigascale Integration (GSI) – Part I: Derivation and Validation", *IEEE Trans. Electron Devices*, Vol. 45, No. 3, pp.580-589, Mar. 1998.

[11] Semiconductor Industry Association, *The International Technology Roadmap for Semiconductors*, 1999 Edition.




Authorized licensed use limited to: Purdue University. Downloaded on July 9, 2009 at 14:38 from IEEE Xplore. Restrictions apply.

TABLE I
PARAMETERS FOR HIGH PERFORMANCE MPU'S OVER FUTURE TECHNOLOGY GENERATIONS

| Year<br>Technology Nodes | 1999<br>180nm | 2002<br>130nm | 2005<br>100nm | 2008<br>70nm | 2011<br>50nm | 2014<br>35nm |
|---|---|---|---|---|---|---|
| Gate Length $(L_n, L_p)$ $(nm)$ | 140 | 85 | 65 | 45 | 32 | 22 |
| Oxide Thickness $(t_{ox})$ $(nm)$ | 1.9 | 1.5 | 1.0 | 0.8 | 0.6 | 0.5 |
| Supply Voltage $Vdd$ $(V)$ | 1.8 | 1.5 | 1.0 | 0.8 | 0.6 | 0.5 |
| Performance $(MHz)$ | 1200 | 1600 | 2000 | 2500 | 3000 | 3600 |
| Chip Size $(mm^2)$ | 450 | 509 | 622 | 713 | 817 | 937 |
| Transistor Count | 22M | 67M | 180M | 546M | 1560M | 4320M |
| Switch Resistance $R_s w W_n$ $(\Omega - cm)$ | 2400 | 2000 | 1600 | 1200 | 800 | 667 |
| Input Capacitance $C_{in}/(W_n + W_p)$ $(fF/\mu m)$ | 2.54 | 1.96 | 2.24 | 1.94 | 1.84 | 1.52 |
| Output Capacitance $C_{in}/(W_n + W_p)$ $(fF/\mu m)$ | 2.04 | 1.69 | 1.32 | 1.36 | 1.39 | 2.08 |
| Gate Pitch $(\mu m)$ | 9.0 | 5.5 | 3.7 | 2.3 | 1.4 | 0.9 |
| Wiring Aspect Ratio | 2.2 | | | | | |
| Wiring Resistivity $(\mu\Omega - cm)$ | 3.3 | | | | | |
| Wiring Capacitivity $(pF/cm)$ | 2 | | | | | |

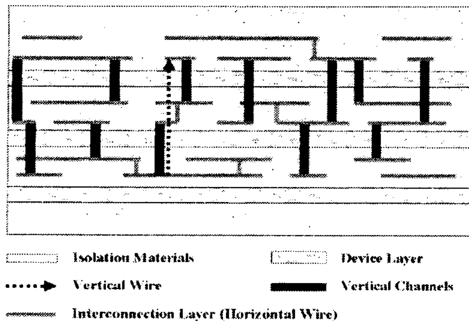


Fig. 1. Schematic of 3-D circuit structure.

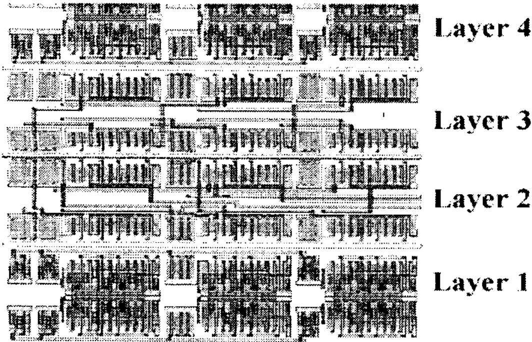


Fig. 2. Layout of a 3-D array multiplier with 4 device layers and 2 metal per device layer.

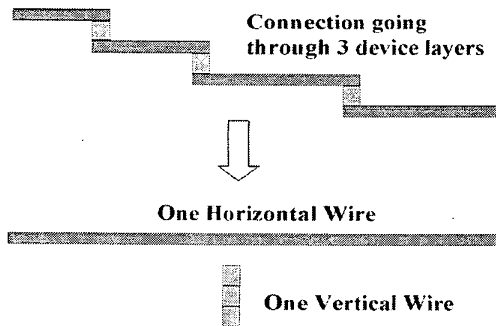


Fig. 3. Definition of horizontal and vertical wire.

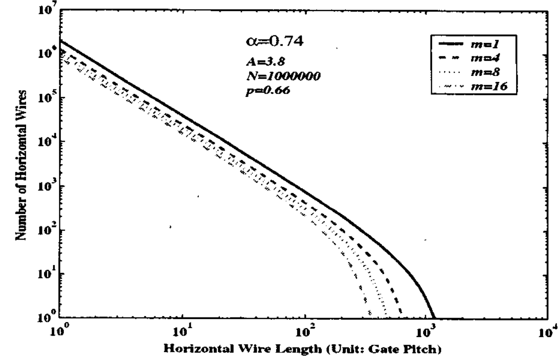


Fig. 4. Horizontal wire-length distribution.

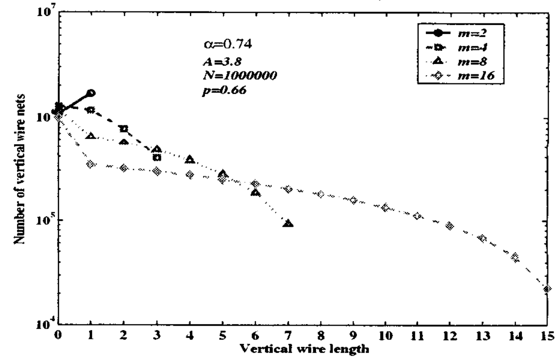


Fig. 5. Vertical wire-length distribution.

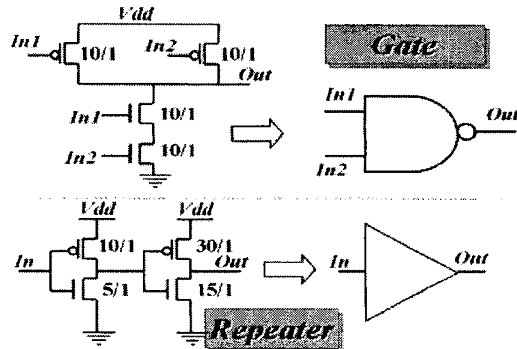


Fig. 6. 2-input NAND gate, repeater and their logic symbols.



Authorized licensed use limited to: Purdue University. Downloaded on July 9, 2009 at 14:38 from IEEE Xplore. Restrictions apply.

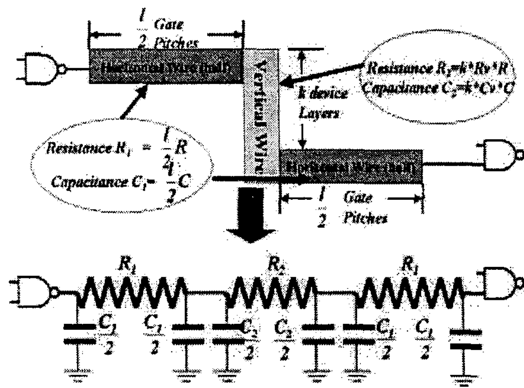

Fig. 7. 3-D wire and its delay model. $R$ and $C$ are resistance and capacitance of horizontal wire with length of one gate pitch. $Rv$ and $Cv$ normalize the vertical wire resistance and capacitance with respect to that of horizontal wire.

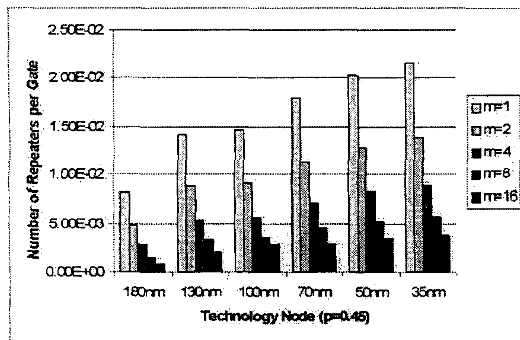


Fig. 8. Number of repeater over technology generations for various device layers.

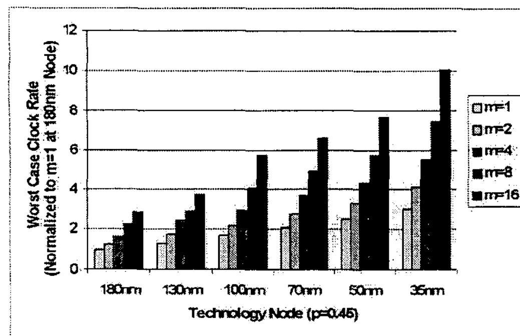


Fig. 9. Clock rate over technology generations for various device layers.

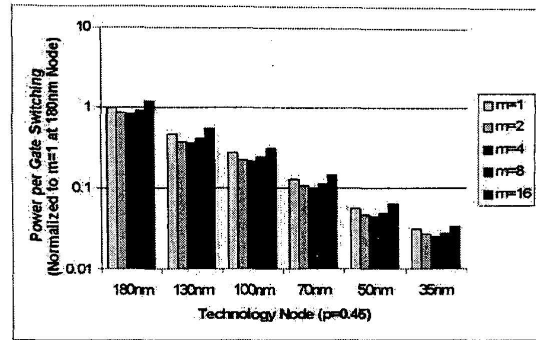


Fig. 10. Power consumption per gate per switching over technology generations for various device layers.

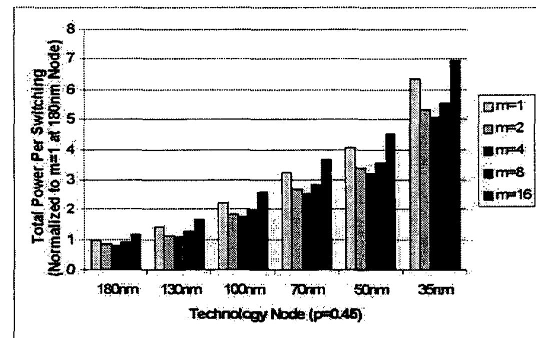


Fig. 11. Total power consumption per switching over technology generations for various device layers.

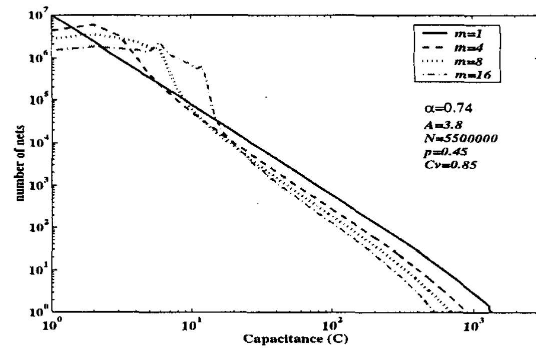


Fig. 12. Interconnect capacitance distribution for various device layers at $180nm$ node (capacitance is in unit of one gate pitch long horizontal wire capacitance, $C$).

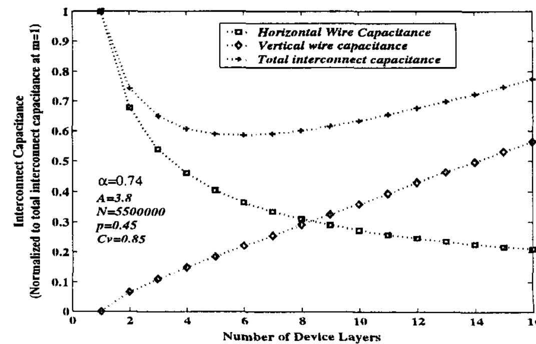


Fig. 13. Interconnect capacitance for various device layers at $180nm$ node.



Authorized licensed use limited to: Purdue University. Downloaded on July 9, 2009 at 14:38 from IEEE Xplore. Restrictions apply.