The Y Architecture for On-Chip Interconnect: Analysis and Methodology

Hongyu Chen, *Student Member, IEEE*, Chung-Kuan Cheng, *Fellow, IEEE*, Andrew B. Kahng, *Member, IEEE*, Ion I. Măndoiu, Qinke Wang, *Student Member, IEEE*, and Bo Yao, *Student Member, IEEE*

***Abstract*—The Y architecture for on-chip interconnect is based on pervasive use of 0°, 120°, and 240° oriented semiglobal and global wiring. Its use of three uniform directions exploits on-chip routing resources more efficiently than traditional Manhattan wiring architecture. This paper gives in-depth analysis of deployment issues associated with the Y architecture. Our contributions are as follows. 1) We analyze communication capability (throughput of meshes) for different interconnect architectures using a multicommodity flow approach and a Rentian communication model. Throughput of the Y architecture is largely improved compared to the Manhattan architecture, and is close to the throughput of the X architecture. 2) We improve existing estimates for the wirelength reduction of various interconnect architectures by taking into account the effect of routing-geometry-aware placement. 3) We propose a symmetrical Y clock tree structure with better total wire length compared to both H and X clock tree structures, and better path length compared to the H tree. 4) We discuss power distribution under the Y architecture, and give analytical and SPICE simulation results showing that the power network in Y architecture can achieve (8.5%) less IR drop than an equally resourced power network in Manhattan architecture. 5) We propose the use of *via tunnels* and *banks of via tunnels* as a technique for improving routability for Manhattan and Y architectures.**

***Index Terms*—Interconnect architectures, very large scale integration (VLSI), Y architecture.**

I. Introduction

THE Y architecture refers to the use of 0°, 120°, and 240° oriented wires for on-chip interconnect, along with supporting methodologies including hexagonal die shapes, hexagonal power and clock distribution, etc. This name is first used in [8] in the same spirit as the "X architecture" for pervasive use of 45° and 135° angles [32].

Compared to the traditional Manhattan (M) architecture, the Y architecture offers many potential advantages, such as substantially reduced wirelength and power consumption, and increased communication bandwidth for a wide range of demand topologies. Combined with the M architecture, the Y architecture can be applied to the upper two layers to improve global interconnects, such as clock and power distribution networks. Moreover, unlike the X architecture, the Y architecture supports a regular routing grid and novel means of avoiding via blockage effects.

Two previous series of works examine the potential use of Y architecture for integrated circuits: 1) a series of LSI Logic patents by Rostoker *et al.* [24]–[26] and 2) a series of works by Cheng *et al.* [7], [8]. Together, these works set out a number of ideas for device architecture, floorplanning, and place-and-route. However, a number of technical gaps still exist, ranging from clock and power distribution methodology to wireability and throughput analysis. In this paper, we provide a more complete, technically in-depth analysis of key deployment and methodology issues associated with the Y architecture. Our main contributions are as follows.

1) We give a more realistic throughput analysis using a communication model based on Rent's rule. Our results show that the Y architecture provides a throughput improvement of about 20% over the M architecture for a square chip, very close to the throughput of the X architecture.
2) We improve existing estimates for the wirelength reduction of various interconnect architectures by taking into account the effect of routing-geometry-aware placement. Our estimate is based on a simulated annealing placer, driven by wirelength in different routing geometries. We also discuss and analyze a "virtuous cycle" effect: reduction of overall wirelength results in decreased routing area, which in turn leads to further wirelength reduction.
3) We discuss clock and power distribution under the Y architecture. For clock distribution, we propose a symmetrical Y clock tree structure with better total wire length compared to both H and X clock tree structures, and better path length compared to the H tree. For power distribution, we give analytical and SPICE simulation results showing that a mesh power network in the Y architecture can achieve 8.6% less IR drop

Manuscript received August 26, 2003; revised February 4, 2004. The work was supported in part by Cadence Design Systems, Inc., in part by the California MICRO program, in part by the MARCO Gigascale Silicon Research Center, in part by the National Science Foundation under Grant MIP-9987678, and in part by the Semiconductor Research Corporation. A preliminary version of this work appeared in the *Proceedings of the Fifth International Workshop on System-Level Interconnect Prediction*, Monterey, CA, 2003, pp. 71–76, and the *Proceedings of the International Conference on Computer-Aided Design*, San Jose, CA, 2003, pp. 13–19. This paper was recommended by Associate Editor T. Yoshimura.

H. Chen, Q. Wang, and B. Yao are with the Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093-0114 USA (e-mail: hchen@cs.ucsd.edu; qiwang@cs.ucsd.edu; byao@cs.ucsd.edu).

C.-K. Cheng and A. B. Kahng are with the Departments of Computer Science and Engineering, and of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0114 USA (e-mail: kuancs.ucsd.edu; abk@cs.ucsd.edu).

I. I. Măndoiu was with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA. He is now with the Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: ion@engr.uconn.edu).

Digital Object Identifier 10.1109/TCAD.2005.844096

0278-0070/$20.00 © 2005 IEEE

than an equally resourced mesh power network in M architecture.

4) To fully utilize the uniform routing grid available in M and Y architectures, and to deal with future increases in via demand due to repeaters [27], we propose the use of *via tunnels* and *banks of via tunnels* to improve routability in these architectures. Such techniques are not obvious with the X architecture.

5) We discuss lithography and manufacturing infrastructure needs, particularly in mask write, related to possible adoption of the Y architecture.

The remainder of the paper is organized as follows. Section II presents throughput analysis for square-shaped chips. Section III discusses wirelength reduction with hexagonal routing. Sections IV and V examine clock and power distribution, and Section VI discusses routability issues. The paper concludes in Section VII. Discussion about the "virtuous cycle" wirelength reduction effect, manufacturing issues and a supporting approximation of the IR-drop are given in the Appendices.

II. Communication Throughput in Meshes

A multicommodity flow (MCF) approach was developed by Chen et al. [8], [9] to evaluate communication efficiency of different interconnect architectures. Communication resources are decomposed into a two-dimensional (2-D) array of slots. A uniform communication requirement is assumed, i.e., every pair of nodes communicates with equal demand and all communications occur at the same time. The throughput, defined as the maximum amount of communication flow simultaneously achievable between every pair of nodes, is computed by a provably good MCF algorithm [13] and is used to measure communication capabilities of different interconnect architectures.

A. Rentian Communication Demand

The uniform pairwise communication used in [8] is simple and general. However, it is not very realistic, since in a well-designed layout, the probability of communication decreases with increasing distance between nodes. Stroobandt and Campenhout [28] derive from Rent's rule an expression for occupation probability, i.e., the probability that a given pair of points will be connected by a wire in an optimal physical placement of the circuit. For a hierarchical placement of a circuit with Rent exponent p in a 2-D Manhattan grid, the occupation probability of a pair of points with Manhattan distance D between them can be approximated by CD^{2p-4}, where C is a normalization constant.[1] When only 2-pin nets are considered, the occupation probability indicates the probability of communication between pairs of nodes. In the following, to ensure a fair comparison of the communication throughput capabilities of different interconnect architectures, we assume a Rentian communication demand, i.e., we set the communication demand between any two unit-area slots to be proportional to D^{2p-4}, where D is the Euclidean distance D between them. A widely quoted survey of Bakoglu [3] indicates that the Rent exponent at the chip and module level of high-speed computers is approximately 0.63.

[1] C depends on the routing architectures and the underlying distance metric.

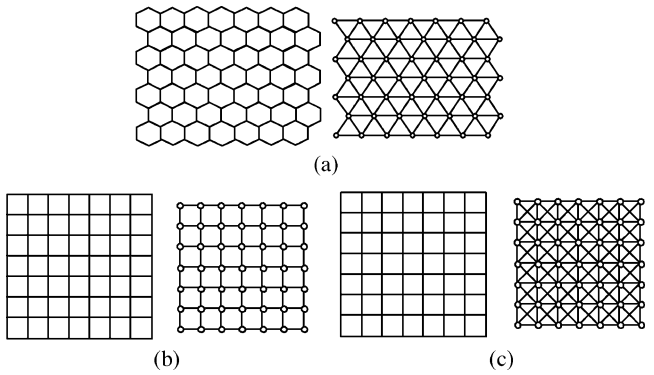


Fig. 1. 7×7 meshes with different interconnect architectures. (a) A 7×7 mesh using Y architecture. (b) A 7×7 mesh using M architecture. (c) A 7×7 mesh using X architecture.

B. Communication Throughput

Definition 1: The throughput is defined to be the maximum fraction of communication demand simultaneously satisfied between every pair of nodes in $n \times n$ square meshes.

We compute the throughput using the MCF algorithm. The throughput is tightly correlated to routability, and describes communication capabilities of different interconnect architectures. Fig. 1 illustrates three 7×7 meshes using different interconnect architectures. For Y architecture, the shape of each slot is hexagonal, and the enclosing box of the slots is close to square. Although Y architecture meshes are different from M and X architecture meshes, this does not significantly affect the communication demand. For the 17×17 Y mesh, total communication demand is only 1.8% different from that for other architectures.

In the experiments, total routing area is set to be the same for all meshes. We normalize the computed throughput so that it is independent of the dimension of meshes and total communication demand.[2] Table I lists the results for $n \times n$ meshes with n ranging between 9 and 17.[3] Compared to the M architecture, the Y architecture provides an average throughput improvement of 19.8% for these meshes, which is comparable to the 21.9% improvement achieved by the X architecture. For a 17×17 mesh, Y architecture provides a throughput improvement of 20.6% while X architecture achieves an improvement of 22.1%.

For Manhattan architecture and Y architecture, equally distributed edge capacities produces maximum throughput on n by n meshes. For X architecture, we show the optimal ratio of the routing area of diagonal routing edges to that of Manhattan edges in the last column. That ratio approaches 1.25 when n increases.

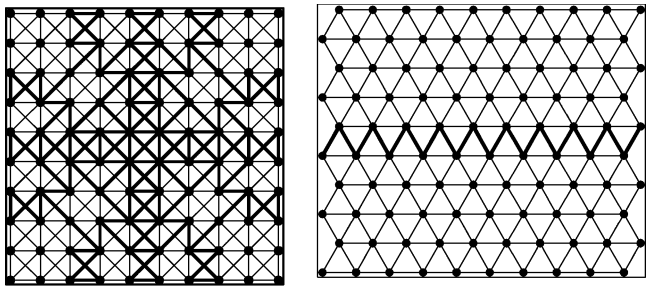
Fig. 2 show bottlenecks of communication flows for 10 by 10 meshes using different interconnect architectures. The fully saturated edges are highlighted with bold lines. For Y architecture, there is only one central horizontal cut line, instead of horizontal

[2] For example, the computed throughput on a $n \times n$ mesh using Y architecture is normalized by $(TD_M/TD_Y) \cdot D_c/n$, where TD_M and TD_Y are total demand for M and Y architectures, respectively, and D_c is the communication demand crossing the horizontal middle cut line on the Manhattan mesh.

[3] The experiments end up with size 17 is mainly because of CPU limit. However, the current improvement results roughly show convergence and are close to theoretical improvement bounds.

TABLE I
NORMALIZED THROUGHPUT (AND IMPROVEMENT VERSUS M ARCHITECTURE IN PERCENTAGE) IN SQUARE MESHES WITH RENTIAN DEMAND. LAST COLUMN GIVES THE OPTIMAL RATIO OF DIAGONAL TO AXIS-PARALLEL ROUTING EDGES

n	#Mesh Nodes	M-arch. Thrpt	Y-architecture Thrpt	Y-architecture Impr.	X-architecture Thrpt	X-architecture Impr.	X-architecture RA Ratio
9	81	1.989	2.354	18.30	2.412	21.25	1.18
10	100	1.989	2.366	18.92	2.419	21.59	1.20
11	121	1.987	2.374	19.47	2.420	21.78	1.23
12	144	1.986	2.382	19.94	2.423	22.00	1.25
13	169	1.991	2.386	19.84	2.425	21.76	1.25
14	196	1.990	2.392	20.19	2.429	22.02	1.25
15	225	1.988	2.395	20.47	2.429	22.14	1.25
16	256	1.992	2.400	20.44	2.430	21.98	1.25
17	289	1.992	2.402	20.58	2.433	22.11	1.25



(a) (b)

Fig. 2. Congestion patterns of 10 by 10 mesh. (a) X architecture. (b) Y architecture.

and vertical cut lines for Manhattan architecture. For X architecture, there are two types of cut lines: horizontal and vertical cut lines, and diagonal cut lines. The throughput of meshes using X architecture depends on both types of cut lines.

Summing up the capacities of the edges passing across the cut lines, we can derive a throughput upper bound for n by n meshes with different interconnect architectures.

For the Manhattan architecture, there are n edges crossing each cut line. The total edge capacity is n. For Y architecture, there are $2n-1$ edges passing across each cut line and each edge has capacity 0.6205, the total edge capacity crossing the cut line is $1.241n-0.6205$. When n approaches infinity, an n by n mesh using Y architecture can have 24.1% more flow crossing the cut line. Thus, Y architecture can achieve at most 24.1% throughput improvement over Manhattan architecture on a squared mesh.

For the X architecture, there are $2(n-1)$ diagonal edges and n Manhattan edges crossing each of the horizontal and vertical cut lines. To achieve maximum throughput, the ratio of routing area for diagonal edges and that for Manhattan edges is 1.25. Under this ratio, the edge capacities are 0.444 and 0.393 for Manhattan edges and diagonal edges, respectively. The total flow amount that go across the cut line is at most $1.230n-0.393$. When n approaches infinity, the throughput improvement bound is 23.0%.

A rectangular chip has communication bottlenecks on two (horizontal and vertical) middle cut lines. The physical dimension of the middle part of the chip restricts the communication flow and thus prevents us from achieving larger throughput. For M and Y architectures, convex-shaped chips (diamond chip for the M architecture and hexagonal chip for the Y architecture) produce better throughput by allowing more wires to cross the original middle cut lines [8].[4] Note that the use of octagonal chips for the X architecture is undesirable since the wafer cannot be tiled by octagons without waste.

III. WIRELENGTH REDUCTION

Because of its restrictions on routing directions, the M architecture entails significant added wirelength beyond the Euclidean optimum. In the Y architecture, routing is allowed along three uniform orientations, and total wirelength is expected to be reduced. An accurate cost-benefit analysis of Non-Manhattan routing is impossible without good estimation of the expected wirelength reduction when switching from Manhattan to non-Manhattan routing. However, the literature contains only simplistic (and seemingly conflicting) estimates.

1) For nets with ten or more pins, experiments with both exact [21] and heuristic Steiner algorithms [16], [17], [22] suggest an average improvement between Manhattan and octilinear Steiner trees of approximately 10% when the nets are randomly generated, and even smaller improvements when nets are extracted from real very large scale integration (VLSI) designs.
2) For 2-pin nets, the octilinear over Manhattan improvement is estimated to be 17.17% in [24], respectively, 14.6% in [31]. [24] and [31] assume different probability distributions over 2-pin nets: in [24], one pin is assumed to be chosen uniformly at random from an Euclidean circle centered at the other pin, and in [31], the Euclidean circle is replaced by a Manhattan circle.
3) For full commercial designs placed and routed with octilinear-aware tools, [15] and [31] report wirelength improvements of 20% or more.

The previous estimates do not adequately address the effect of routing-geometry-aware placement on the overall wirelength improvement. Previous studies of the routing demand using different traditional placers [19] show that Manhattan placers tend to align circuit elements either vertically or horizontally, leaving few opportunities to exploit additional routing directions. A Y-aware or X-aware placer factors in hexagonal or octilinear wiring during placement, and results in better placements of nets when such wiring is used to route the nets. Therefore, total wirelength can be greatly reduced.

To estimate the wirelength improvement achieved by the Y- or X-aware placement and routing versus Manhattan placement and routing, we have built a simplified placer which uses simulated annealing driven by hexagonal or octilinear wirelength estimation. The input of the placer is a simplified netlist extracted from Microelectronics Center of North Carolina (MCNC) instances, in which a list of cells is specified for each net. After a random initial placement of cells, two cells are randomly selected, and we decide whether to swap these two cells based on the current annealing temperature and the new SMT cost with hexagonal or octilinear routing, which is computed using an exact SMT algorithm, GeoSteiner [21]. The initial temperature of the simulated annealing algorithm is specified so that it is far larger than the standard deviation

[4]Note that it is not necessary to use a regular hexagon for the Y architecture; either horizontal or vertical symmetry suffices.

TABLE II
AVERAGE WIRELENGTH IMPROVEMENTS FOR NON-MANHATTAN PLACEMENT AND ROUTING VERSUS. MANHATTAN PLACEMENT AND ROUTING (%)

Instance	#nets	Y-Arch	X-Arch	Euclidean
C2	601	4.81	8.92	11.04
BALU	658	7.13	9.29	11.07
PRIMARY1	695	7.32	10.31	13.03
C5	1438	8.34	11.48	12.73

TABLE III
TOTAL WIRELENGTH OF INSTANCE "C2" WITH DIFFERENT COMBINATIONS OF PLACEMENT AND ROUTING

routing-geometry driven placement	routing-geometry			
	Rect	**Hex**	**Oct**	**Euclidean**
Rectilinear	1805	1841.8	1719.2	1683.9
Hexagonal	1896	1743.0	1665.9	1621.5
Octilinear	1908	1799.6	1644.0	1617.4
Euclidean	1865	1772.1	1646.9	1605.7

TABLE IV
TOTAL WIRELENGTH OF INSTANCE "BALU" WITH DIFFERENT COMBINATIONS OF PLACEMENT AND ROUTING

routing-geometry driven placement	routing-geometry			
	Rect	**Hex**	**Oct**	**Euclidean**
Rectilinear	1820	1856.0	1728.4	1694.7
Hexagonal	1878	1748	1660.8	1623.6
Octilinear	1886	1785.6	1650.9	1621.3
Euclidean	1898	1769.8	1654.6	1616.5

TABLE V
TOTAL WIRELENGTH OF INSTANCE "PRIMARY1" WITH DIFFERENT COMBINATIONS OF PLACEMENT AND ROUTING

routing-geometry driven placement	routing-geometry			
	Rect	**Hex**	**Oct**	**Euclidean**
Rectilinear	2058	2080.8	1942.5	1903.7
Hexagonal	2126	1958.3	1882.7	1833.9
Octilinear	2136	2004.8	1862.0	1828.0
Euclidean	2124	1976.8	1854.1	1805.5

of total wirelength distribution [14]. For each temperature, the number of swaps is on the order of 100 times the number of cells [20]. The new temperature is generated by multiplying the current temperature by $\alpha = 0.95$, which is a relatively large α for simulated annealing [20].

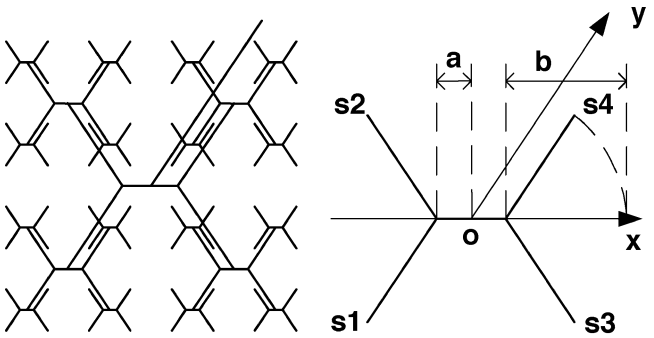
For each instance and each routing geometry (rectilinear, hexagonal, octilinear, and Euclidean), we run the placer five times, and get the best wirelength with routing-geometry-aware placement and routing. The wirelength improvements achieved by non-Manhattan placement and routing are summarized in Table II. In Tables III–VI, we give total wirelengths obtained with different combinations of placement and routing for each of the four testcases.

According to the results, the Y architecture achieves a wirelength improvement up to about 8.3%. The X architecture further reduces total wirelength to be up to about 11.4% over M architecture and it produces about 3.3% wirelength reduction over Y architecture with the cost of one more routing direction.

We note that in the above experiments, the placer uses a fixed-area die. However, reduction of overall wirelength results in decreased routing area, which in turn leads to further wirelength reduction, creating a "virtuous cycle" effect. An analysis of this effect is given in Appendix I.

TABLE VI
TOTAL WIRELENGTH OF INSTANCE "C5" WITH DIFFERENT COMBINATIONS OF PLACEMENT AND ROUTING

routing-geometry driven placement	routing-geometry			
	Rect	**Hex**	**Oct**	**Euclidean**
Rectilinear	3557	3625	3340	3272
Hexagonal	3619	3334.99	3182.48	3107.45
Octilinear	3569	3390	3149	3097
Euclidean	3628	3397	3183	3104



(a) (b)

Fig. 3. Y clock tree. (a) Y clock tree. (b) A one-level Y clock tree.

IV. Y CLOCK TREE

Clock distribution networks synchronize the flow of data signals among synchronous data paths. The design of these networks can dramatically affect system-wide performance and reliability. The H clock tree [4] is widely used in the IC industry. In the H tree, clock terminals are arranged in a symmetric fashion, and are connected by a planar hierarchy of symmetric "H" structures. When octilinear routing is allowed, the "H" structure can be replaced with an "X" structure, so that source-sink path (i.e., insertion) delay and total wirelength are decreased. However, significant undesirable overlapping (superposition) will occur between parallel interconnect wires in the X tree.

With three uniform routing directions, a Y clock tree can be built as depicted in Fig. 3(a), essentially giving a "distorted X tree" with reduced wirelength and no superposed parallel wires. Let the distance between two adjacent clock terminals be 1. Path length from the clock source to clock terminal, as well as total wirelength, are compared with the H and X trees in Table VII. In the table, n is the level of clock trees. Path length and total wirelength of n-level clock trees can be easily derived. E.g., source-to-sink path length of H clock tree is derived according to the following formula.

$$\mathrm{PathLength}_H(n) = \mathrm{PathLength}_H(n-1) + 2^{n-1}. \quad (1)$$

The Y clock tree has a path length of $0.7887 \cdot (2^n - 1)$, 21.1% less than the H tree. Its total wirelength is $1.366 \cdot 2^n(2^n - 1)$, 8.9% less than the H tree, and 3.4% less than the X tree. Actually, the one-level Y tree shown in Fig. 3(b) is the optimal Euclidean Steiner minimum Tree to connect four adjacent clock

TABLE VII
PATH LENGTH AND TOTAL WIRELENGTH OF H TREE, X TREE AND Y TREE

	Path Length	Total Wirelength
H-tree	(2^n-1)	$\frac{3}{2}\cdot 2^n(2^n-1)$
X-tree	$\frac{\sqrt{2}}{2}\cdot(2^n-1)$	$\sqrt{2}\cdot 2^n(2^n-1)$
Y-tree	$\frac{1}{2}(1+\frac{\sqrt{3}}{3})\cdot(2^n-1)$	$\frac{1+\sqrt{3}}{2}\cdot 2^n(2^n-1)$

terminals $s1$, $s2$, $s3$, $s4$, and the clock source o. Thus, the Y clock tree provides minimal total wirelength among all clock trees with similar symmetric structure. The further advantage of Y clock tree is that there is no overlapping of parallel interconnect wires. The following can be shown.

Theorem 1: Let the distance between two adjacent clock terminals be D. The minimum distance between two parallel interconnect wires is $(\sqrt{3}-1/4)D$.

Proof: Suppose there is a coordinate system with a 0° x axis, a 60° y axis, and the origin (0,0) at the center of the main Y tree structure [see Fig. 3(b)]. Then, in a one-level Y tree, the two interconnect wires that are parallel to the y axis in the figure have x coordinates of $\pm a$. In a two-level Y tree, the lowest-level y axis-parallel interconnect wires have x coordinates of $\pm a \pm 2a$ and $\pm a \pm 2(a+b)$. Generally, in an n-level Y tree, x coordinates of the lowest-level y axis-parallel interconnect wires are $\pm a \pm (2a \text{ or } 2(a+b)) \pm \cdots \pm (2^{n-1}a \text{ or } 2^{n-1}(a+b))$.

Since $a=(D/2)(1-(\sqrt{3}/3))$, and $(a+b)=(D/2)(1+(\sqrt{3}/3))$, the y coordinates can be written as $(\pm 2^0 \pm 2^1 \pm \cdots \pm 2^{n-1})\cdot(1/2)D + (\pm 2^0 \pm 2^1 \pm \cdots \pm 2^{n-1})\cdot(\sqrt{3}/6)D$. These values cannot be zero because the values of $\pm 2^0 \pm 2^1 \pm \cdots \pm 2^{n-1}$ must not be zero, and the minimum absolute value among them is $a=(D/2)(1-(\sqrt{3}/3))$. Thus, the minimal distance between two parallel interconnect wires in the Y clock tree is $(\sqrt{3}/2)a = (\sqrt{3}-1/4)D$. ■

V. Y POWER DISTRIBUTION

Excessive voltage drop in the power grid can slow device switching speed and reduce noise margin. Robust power distribution within the available area resource is critical to chip performance and reliability. Hierarchical mesh structures are widely used for power distribution in high-performance chips because of their robustness [5]. In this section, we show that power distribution in the Y architecture is not only natural, but achieves less IR drop than equally resourced mesh distribution in the M architecture.

Our comparison is based on the following model of the power distribution network.

1) The power distribution network is constructed by a hierarchy of mesh structures connected by vias at crossing points of wires. Each mesh has equal wire spacing and wire width. Ignoring the resistance of vias,[5] we assume perfect contact at each crossing point.

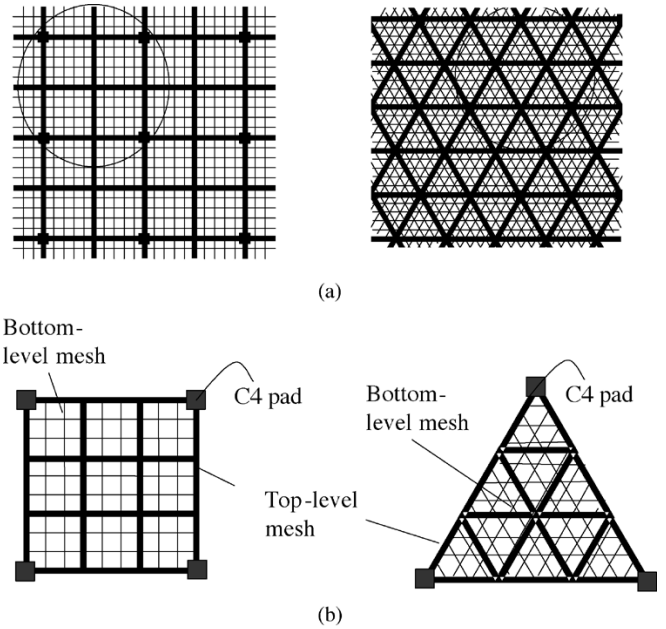


Fig. 4. Power distribution networks and representative areas for M and Y architectures. (a) Two-level power mesh. (b) Representative areas.

2) On top of metal layers, there are arrays of C4 power pads evenly distributed on the surface of the power mesh.
3) Under the bottom-level mesh, there are devices connected to the wires of the bottom-level mesh. The devices are modeled as uniform current sinks and placed at crossing points of the bottom-level mesh.

In state-of-art designs, there is a fairly large number (>100) of power pads evenly distributed on the surface of the top-level power mesh [34]. It is reasonable to assume that the whole power mesh is an infinite resistive grid constructed by replicating the area surrounded by adjacent power pads. Fig. 4 illustrates two-level power meshes and the *representative areas* in the M and Y architectures. Our analysis and circuit simulations consider only the worst-case IR-drop on the representative area. This method is also used in [12].

A. IR-Drop on Single-Level Power Mesh

Static IR-drop on a hierarchical power mesh depends largely on the top-level mesh since usually the top-level mesh is wider and coarser and most current flows along the top-level mesh. Here, we analyze and compare the worst-case static IR-drop on a single-level power mesh in the M and Y architectures.

1) IR-Drop on Single-Level Power Mesh in the Y architecture: A single-level power mesh in the Y architecture is abstracted as an infinite triangular resistive lattice with edge resistance R_Y.[6] We examine IR-drop in the triangular area with N_Y rows surrounded by three adjacent power pads.[7] In this

[5]In practice, high current density on vias often causes reliability problems. In the Y architecture, assuming the same wire width, the area of intersection (overlap) between two adjacent-layer wires is larger than in the M architecture. Hence, we can place a bigger via between adjacent layers or place more vias in the via array between adjacent layers to reduce resistance and current density for vias. Let A_Y, A_X, and A_M represent this area for Y, X, and M architectures, respectively. We have $A_Y = 1.1547A_M$ and $A_X = 1.414A_M$.

[6]Note that for a uniform mesh with fixed total routing area, the edge resistance is independent of the number of metal lines on the mesh. When the number lines increases, wire pitch and wire width decreases with the same ratio, and the edge resistance remains the same.

[7]E.g., for the top-level Y architecture mesh shown in Fig. 4(b), N_Y is equal to 3.

case, the worst-case IR-drop appears at the center of this representative area. Each power pad supplies a current $I_Y = N_Y^2 i$ to the power mesh, where i is the current drain at each intersection on the mesh.

Assume there is a coordinate system with the origin at the center of the power mesh, and 0° and 120° lines used as m and n axes, respectively. We analyze the voltage drop between the node (0,0) and the power pad at $(N_Y/3, -N_Y/3)$ by considering currents from power pads and evenly distributed current sinks separately.

IR-drop caused by currents from power pads: Suppose that a current I_Y enters the lattice at the node (m_s, n_s) and leaves at infinity. The voltage drop for any node on the lattice is analyzed in [2]. The voltage drop between (m_s, n_s) and (m, n), denoted as $V_{(m_s,n_s)}(m, n)$, is given by the integral

$$\frac{I_Y R_Y}{2\pi} \int_0^{\frac{\pi}{2}} \left(1 - e^{-|(m-m_s)-(n-n_s)|x}\right) \times \frac{\cos\left(((m-m_s)+(n-n_s))\,y\right)}{(\sinh x \cos y)} \mathrm{d}y \quad (2)$$

where $2\cosh x \cos y + \cos 2y = 3$. When $|(m-m_s)-(n-n_s)|$ is large, the voltage drop $V_{(m_s,n_s)}(m, n)$ can be approximated as

$$\frac{I_Y R_Y}{4\sqrt{3}\pi} \left[\ln\left((m-m_s)^2 + (n-n_s)^2 - (m-m_s)(n-n_s)\right) + c_1\right] \quad (3)$$

where $c_1 = 3.6393$ is a constant.[8]

Let $V_{(m_s,n_s)}$ denote the voltage drop between (0,0) and the power pad at $(N_Y/3, -N_Y/3)$ caused by the current source at (m_s, n_s). According to the above approximation:

1) when $(m_s, n_s) = (N_Y/3, -N_Y/3)$, $V_{(m_s,n_s)} \approx (I_Y R_Y/4\sqrt{3}\pi)(2\ln N_Y - \ln 3 + c_1)$.
2) when $(m_s, n_s) \neq (N_Y/3, -N_Y/3)$, $V_{(m_s,n_s)} = V_{(m_s,n_s)}(0,0) - V_{(m_s,n_s)}(N_Y/3, -N_Y/3) \approx (I_Y R_Y/2\sqrt{3}\pi)\ln(D_0/D_s)$, where D_s is the Euclidean distance between (m_s, n_s) and $(N_Y/3, -N_Y/3)$, and D_0 is the Euclidean distance between (m_s, n_s) and (0,0). The constant $c_2 = \sum_{(m_s,n_s)\neq(N_Y/3,-N_Y/3)} \ln(D_0/D_s)$ can be computed by a simple algorithm, which calculates the summation for all the current sources within a circle around the origin. As the radius of the circle increases, the summation converges to a value of $c_2 = -1.173\,679$.

Therefore, if only currents from power pads are considered, the voltage drop between (0,0) and the power pad at $(N_Y/3, -N_Y/3)$ is

$$V_{\text{source}} = \sum_{(m_s,n_s)} V_{(m_s,n_s)} = \frac{I_Y R_Y}{2\sqrt{3}\pi}(\ln N_Y + C_Y) \quad (4)$$

where $C_Y = c_1/2 - \ln 3/2 + c_2 = 0.096\,66$.

[8]See Appendix II for details of this approximation.

TABLE VIII
SIMULATION RESULTS FOR WORST-CASE IR-DROP ON THE SINGLE-LEVEL POWER MESH IN THE Y ARCHITECTURE, COMPARED TO ESTIMATED VALUES (mV)

N_Y	IR-Drop	Estimated IR-Drop	Error
3	166.67	165.39	1.28
6	229.17	229.08	0.09
9	266.36	266.34	0.02
12	292.78	292.77	0.01
15	313.28	313.27	0.01
18	330.03	330.03	0.00
21	344.20	344.19	0.00

IR-drop caused by evenly distributed current sinks: Next, we consider the voltage drop caused by current sinks at the intersections of the power mesh. If the voltage between (0,0) and (m, n) is denoted by $V_{\text{sink}}(m, n)$, by a combination of Ohm's and Kirchhoff's Laws we have

$$\begin{aligned} &V_{\text{sink}}(m-1,n) + V_{\text{sink}}(m+1,n) + V_{\text{sink}}(m,n+1) \\ &+ V_{\text{sink}}(m,n-1) + V_{\text{sink}}(m-1,n-1) \\ &+ V_{\text{sink}}(m+1,n+1) - 6V_{\text{sink}}(m,n) = iR_Y. \end{aligned} \quad (5)$$

If the resistive lattice is regarded as a discrete approximation to a continuous resistive medium, we will obtain a potential function proportional to D^2, where D is the Euclidean distance from the origin. Therefore, we assume the following representation for the voltage between (0,0) and (m, n):

$$V_{\text{sink}}(m,n) = k(m^2 + n^2 - mn) \quad (6)$$

where k is a constant. Equation (5) then yields

$$V_{\text{sink}}(m,n) = \frac{iR_Y}{6}(m^2 + n^2 - mn). \quad (7)$$

When only current sinks are considered, the voltage drop between (0,0) and the power pad at $(N_Y/3, -N_Y/3)$ is

$$V_{\text{sink}} = V_{\text{sink}}\left(\frac{N_Y}{3}, -\frac{N_Y}{3}\right) = \frac{I_Y R_Y}{18}. \quad (8)$$

Verification of worst-case IR-drop: From the above analysis, we obtain the voltage drop at the center

$$V_Y = V_{\text{source}} + V_{\text{sink}} \approx \frac{I_Y R_Y}{18} + \frac{I_Y R_Y}{2\sqrt{3}\pi}(\ln N_Y + C_Y) \quad (9)$$

where $C_Y = 0.096\,66$.

To verify the above formula for worst-case IR-drop on the single-level power mesh, we use HSpice to simulate various power meshes with different values of N_Y's. Since the problem is linear in nature, in our experiments the resistance of each wire segment R_Y is simply set to be 1 KΩ, and the total current drain in the area I_Y is set to be 1 mA. We list simulation results for N_Y from 3 to 21 in Table VIII, and compare them with the estimated values from the formula. The results show that the formula is accurate, with error less than 1%.

2) Comparing IR-Drop on Single-Level Power Mesh: For a single-level power mesh in the M architecture, worst-case IR-drop is analyzed and verified in [11]. Suppose the power mesh has edge resistance R_M, the number of rows within the

TABLE IX
IR-DROP IMPROVEMENTS IN SINGLE-LEVEL Y MESH VERSUS M MESH

N_M	Estimated IR-Drop (mV) in M-mesh	IR-Drop Impr. (%) with Y-mesh
2	214.25	10.78
3	278.78	8.28
4	324.56	7.11
5	360.08	6.41
6	389.09	5.93
7	413.63	5.58
8	434.88	5.31
9	453.63	5.09

representative area N_M, and the current supplied by each power pad I_M, the worst-case IR-drop on the single-level Manhattan (M) mesh is

$$V_M \approx \frac{I_M R_M}{8} + \frac{I_M R_M}{2\pi}(\ln N_M + C_M) \tag{10}$$

where $C_M = -0.1324$.

To fairly compare the Y and M meshes, we constrain the two meshes to have the same wire material and thickness, cover the same area (same total current drain) with the same wiring resource, and have the same number of crossing points and power pads. Therefore, we have $R_Y = \sqrt{3}R_M$, $I_Y = I_M$, and $N_Y = N_M$. According to (9) and (10), the worst-case IR-drop on the single-level Y mesh is less than that on the M mesh by

$$\Delta V = V_M - V_Y = cI_M R_M \tag{11}$$

where $c = 0.023\,09$. We list IR-drop improvements with Y mesh for different values of N_M in Table IX. The number of wire lines between two adjacent power pads on the top-level power mesh is usually small [11]. When $N_M = 4$, static IR-drop improvement of the Y mesh over M mesh is 7.1%.

B. IR-Drop on Hierarchical Power Mesh

In practice, power is distributed through a hierarchy of six or more metal layers. In this section, we simulate hierarchical power networks for the X, Y, and M architectures using HSpice, explore different configurations of power networks, and compare the best solutions. We assume an equal sum of routing resources (i.e., total routing area) for X, Y, and M architecture power distribution across layers M6, M5, and M4. In our experiment below, we set the total wiring area of M6, M5, and M4 to be 52% of the total representative area. The representative area for the Manhattan and X mesh is set to be a 1.2×1.2 mm^2. To achieve the same power pad density, the representative area for the Y power grid is an equilateral triangle with edge length 1.289 mm. Further details of our comparison are as follows.

1) Layer thickness and resistivity parameters of a six-layer process are taken from TSMC 0.13-μm copper process information [30]. Layer thicknesses are 0.33 μm for M1, 0.36 μm for M2–5, and 1.02 μm for M6.
2) M1–M3 power distribution is native to library cells and blocks, requiring a common interface (0°) at M4. Power routing in M1–3 has the same pitch in both the Y and Manhattan solutions. M1 has a pitch of 8 μm and wire width of 2 μm, M2 has a pitch of 60 μm and wire width of 4 μm, and M3 has pitch of 60 μm and wire width of 4 μm. M4 pitch is fixed at 75 μm to enable matchup with M1–3 macros and an apples-to-apples comparison.
3) Allowed values of wiring separations (= pitches) on M5 and M6, denoted by S5 and S6, are 600, 300, 150, and 75 μm. Allowed percentages of total wiring area used on M4 and M5, denoted as P4 and P5, are 10%, 20%, 30%, 40%, . . ., 80%.
4) There is a via at each intersection of wires on adjacent metal layers. The vias are regarded as perfect electrical contact.
5) One V voltage sources are placed at the corners of representative areas. Each current sink on M1 (between two adjacent vias) is 5.21×10^{-7} A.

TABLE X
STATIC IR-DROP ON BEST POWER NETWORKS IN X, Y, AND MANHATTAN ARCHITECTURES

Arch.	M6 pitch (um)	M6 area (%)	M5 pitch (um)	M5 area (%)	M4 pitch (um)	IR-drop (mv)
M	300	70	75	20	75	38.5
Y	600	70	150	20	75	35.2
X	300	40	300	40	150	34.8

All combinations of wire pitch and wire width of M4–M6 are exhaustively searched. Table X shows the best configurations in M, Y, and X architecture and the corresponding static IR-drop. In the best M architecture configuration, M6 has a wire pitch of 300 μm and uses 70% of the power routing resource and M5 has a wire pitch of 75 μm and uses 20% of the resource. The IR-drop produced by this configuration is 38.5 mV. In the best Y architecture configuration, M6 has a pitch 600 μm and uses 70% of the power routing resource, while M5 has a pitch 150 μm and again uses 20% of the wiring area. The IR-drop is 35.2 mV, which is 8.6% smaller than that of the best M architecture solution. In the best X architecture configuration, M6 and M5 both have a 300-μm wiring pitch and uses 40% of wiring area, while the M4 has pitch 150 μm. The IR-drop is 34.8 mV, which is 1.2% smaller than that of the best Y architecture configuration. Ongoing research seeks a more general and formal comparison.

VI. ROUTABILITY IN THE Y ARCHITECTURE

A. Uniform Routing Grid

A nice property of the Y architecture is that there is a natural, uniform routing grid. Fig. 5(a) and (b) illustrates the routing grid in the M and Y architectures, wherein each routing layer has exactly the same wiring pitch. Fig. 5(c) shows the X architecture grid, where identical layer pitches imply that wire intersection points are not coincident. It is therefore difficult to find a natural, resource-efficient, uniform wiring grid in the X architecture.

A uniform routing grid is expected to benefit large VLSI designs for two main reasons. 1) It enables continued use of today's dominating gridded routing algorithms. 2) The uniform routing grid can permit integral coordinates (even if absolute positions have irrational coordinates!), significantly simplifying detailed routing and design rule checking algorithms.

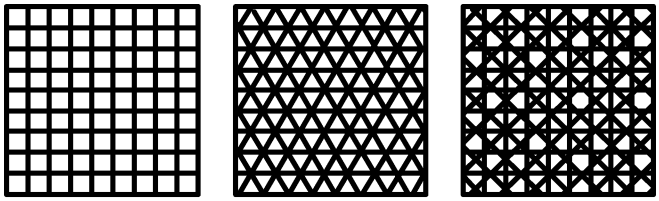


Fig. 5. Routing grids in M, Y, and X architectures.

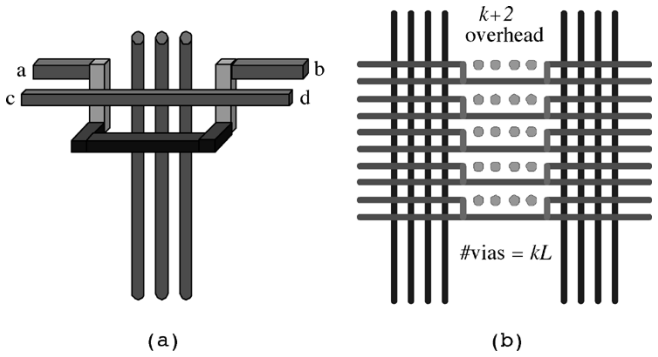


Fig. 6. Via tunnel and bank of via tunnels in the M architecture. (a) Avia tunnel. (b) Bank of via tunnels.

B. Via Tunnels and via Tunnel Banks

Another advantage of the uniform global routing grid is that we can utilize *via tunnels* and *via tunnel banks* to avoid the fragmentation of routing resources caused by vias; this improves overall chip routability. In multilayer routing, wire tracks are blocked on the layers that a via passes through. Traditional routing schemes scatter vias all over the chip, and this fragmentation of routing resources may cause serious wireability problems; this is called "via blockage effect." As we approach the 65-nm technology node, this effect becomes more serious, since buffering of global wires introduces many via chains that go through all the way from the top-level metal down to the gate layer. We believe that the proposed use of via tunnels and via tunnel banks will reduce the via blockage effect and thus improve routability and wiring density.

Fig. 6(a) shows an example of a *via tunnel* in the Manhattan architecture. There are two routing layers shown in the figure: the upper layer is for horizontal routing and the lower layer is for vertical routing. Terminals a and b are connected by detouring the horizontal wires around the via using the space on the vertical layer. Because the detour happens on the lower layer, it will not affect the wire between terminals c and d on the upper layer.

By aligning a number of *via tunnels* in vertical direction, we obtain a *bank of via tunnels*, which is shown in Fig. 6(b). Suppose each *via tunnel* have k vias arranged in a horizontal line [in Fig. 6(b), $k = 3$], and we align L *via tunnels* into a bank. In the resulting bank, all the horizontal tracks are free to route, and only $k + 2$ vertical tracks are blocked. Note that there are a total of kL vias in the bank; without the bank of via tunnels, up to kL tracks could be blocked on each layer that the vias pass through. The use of *via tunnel banks* can thus significantly reduce the "via blocking effect."

We have designed similar *via tunnel* and *bank of via tunnels* for the Y architecture.

1) Fig. 7(a) shows the bird's eye view of a *via tunnel* design in the Y architecture. In this example, we have three layers. From top to bottom, the routing direction is 60°, 120°, and 0° in each layer, respectively. The circle in the center represents a through via. The space in the middle layer is used to detour wires around the via. We can achieve blockage-free routing on the top and bottom layers, and have four tracks blocked on the middle layer.
2) Similar to the construction of *banks of via tunnels* in M architecture, we align the *via tunnels* together to obtain a *bank of via tunnels* in the Y architecture. Fig. 7(b) illustrates how two *via tunnels* shown in Fig. 7(a) are aligned along the 120-degree direction.
3) In order to reduce the average track overhead, each via tunnel can have more than one vias in a line. Fig. 8 illustrates the construction of via tunnels with k ($k = 2, 3, 4$) vias aligned in a line. The figures show detour routing patterns on the middle layer for $k = 2, 3, 4$, respectively. Fig. 9 is an example of three via tunnels with $k = 2$ aligned along the 120° direction. From these examples, we can see that for via tunnel with k vias aligned in a line, the track overhead on the middle layer is $2k + 1$.
4) Fig. 10 depicts a *bank of via tunnels* in the Y architecture. Suppose the bottom m layers are used to perform intracell routing, and the top $n - m$ layers are used for distributing signals to the banks. Assume each *via tunnel* has k vias in a line, and there are L *via tunnels* in the bank. All the kL vias introduce only $2k + 1$ tracks of routing blockage on the 120° routing layers.

VII. Conclusion

In this paper, we have examined key issues concerning the potential use of Y architecture for semiconductor ICs, including throughput analysis, estimates of wirelength savings, clock and power distribution methodology, wireability, and manufacturing. We have not discussed such issues as graphics engine changes, computational-geometric data structures, number and coordinate systems, calibration of parasitic extraction (especially capacitance extraction) models, etc. Such "mundane" issues are part of the necessary groundwork for the eventual deployment of the Y architecture, and the subject of ongoing work in our group, but are beyond the scope of this paper.

Compared to the X architecture, the Y architecture supports a regular routing grid, which is important for simplifying manufacturing process and routing and design rule checking algorithms. Besides, novel means of via tunnel and via tunnel bank can be used to avoid via blockage effects; such techniques are not obvious with the X architecture. Moreover, a Y clock tree has a better total wire length compared to X clock tree structure without overlapping of parallel interconnect wires. Finally, the Y architecture provides a throughput and wirelength improvement close to the X architecture with one less routing direction;

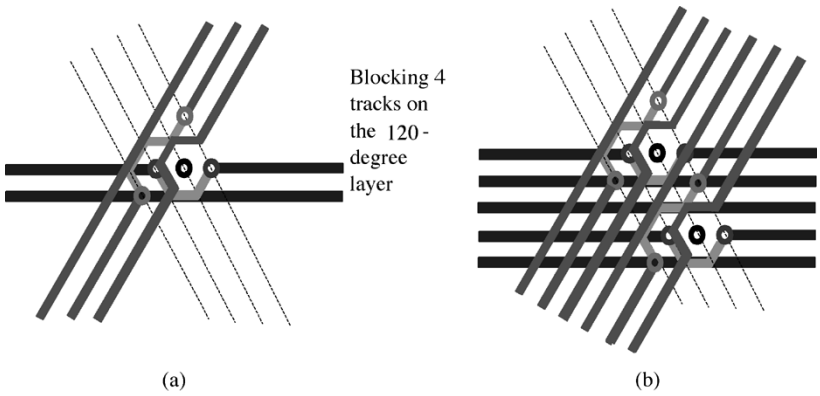


Fig. 7. Via tunnels and bank of via tunnels in Y architecture. (a) A via tunnel in Y architecture. (b) Two tunnels aligned together.

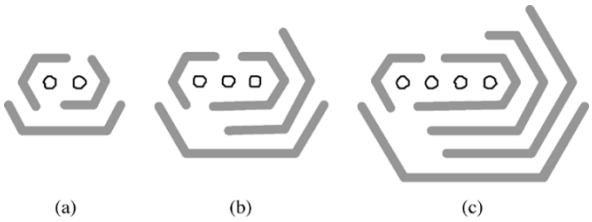


Fig. 8. Via tunnels with vias aligned in a line in Y architecture. (a) $k = 2$ overhead $= 5$; (b) $k = 3$ overhead $= 7$; (c) $k = 4$ overhead $= 9$.

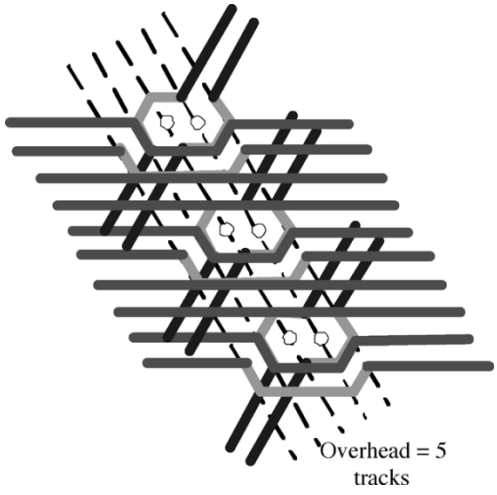


Fig. 9. Three tunnels with $k = 2$ aligned together.

convex-shaped chips can produce further improvement for the Y architecture without wafer waste.

Further research directions include: 1) theoretical analysis and high-impact designs or codes to demonstrate Y architecture advantages; 2) more accurate estimations of expected wirelength improvement which formalizes interactions between nets; and 3) interfaces to current library cells and new Y specific library cells. Many parts of a commercially successful Y architecture methodology remain open. The Y architecture also has applications beyond the die, e.g., it may be valuable on laminates used for multidie integration, and on the buildup layers (e.g., BBUL [33]) that will replace traditional packages.

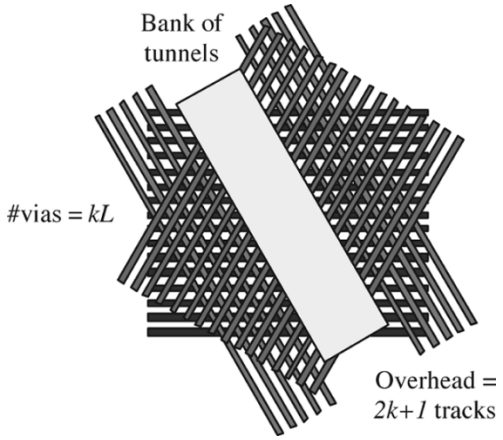


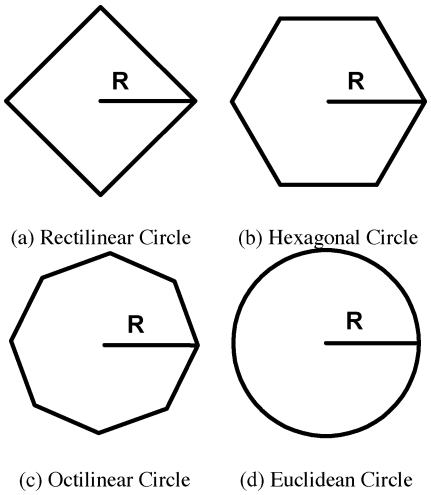
Fig. 10. Bank of via tunnels in Y architecture.

APPENDIX I
ANALYSIS OF THE "VIRTUOUS CYCLE" WIRELENGTH REDUCTION EFFECT

The simulated annealing placer in Section III places cells within a chip that has fixed area. However, reduction of overall wirelength results in decreased routing area, which in turn leads to further wirelength reduction, creating a "virtuous cycle" effect.

Consider a cluster of two-pin nets which are connected to one pin A. All other pins are uniformly located in a circle by a routing-geometry-aware placer. Circles for different routing geometries are shown in Fig. 11. Based on the "virtuous cycle" effect, the circle will have an area proportional to the total routing area. For Manhattan placement and routing, suppose the pins are placed in a rectilinear circle with radius R. We have the area of the rectilinear circle $A = 2R^2$ and the total routing area, $A_{\text{routing}} = 4\int_0^R (x \cdot (xdx/A)N) = (4/3)(R^3/A)N = (2/3)RN$, where N is the number of two-pin nets and $(xdx/A)N$ is the number of pins located between unit circles with radii x and $x + dx$. Let $A \sim A_{\text{routing}}$. We have $R \sim N/3$ and $A_{\text{routing}} \sim (2/9)N^2$. Similar analysis can be done for other routing geometries, with the results summarized as follows:

1) **Rectilinear**: $A_{\text{routing}} \sim (2/9)N^2$.
2) **Hexagonal**: $A_{\text{routing}} \sim (8\sqrt{3}/81)N^2$, 23.0% less compared to Manhattan placement and routing.
3) **Octilinear**: $A_{\text{routing}} \sim (\sqrt{2}/9)N^2$, 29.3% less compared to Manhattan placement and routing.



(a) Rectilinear Circle (b) Hexagonal Circle

(c) Octilinear Circle (d) Euclidean Circle

Fig. 11. Circles with radius R for each routing geometry (rectilinear, hexagonal, octilinear, and Euclidean).

4) **Euclidean**: $A_{\text{routing}} \sim (4/9\pi)N^2$, 36.3% less compared to Manhattan placement and routing.

This simple analysis shows that the wirelength reduction caused by the "virtuous cycle" effect is significant, and can partly explain the large wirelength reductions reported in [15] and [31].

Appendix II
Approximation of Equation (2)

Suppose a current I enters a uniform infinite triangular resistive lattice with edge resistance R at the origin and leaves at infinity. The voltage drop for any node on the lattice is analyzed in [2]. The final result for the voltage drop is expressed as an integral representation. The voltage between (0,0) and (m, n), $V(m, n)$, is

$$V(m,n) = \frac{IR}{2\pi} \int_0^{\frac{\pi}{2}} \frac{\left(1 - e^{-|(m-n)|x} \cos(m+n)y\right)}{\sinh x \cos y} dy \quad (12)$$

where $2\cosh x \cos y + \cos 2y = 3$. When $|m - n|$ is large, the exponential term in the above expression become negligible except when x is very small. When x is very small, we have:

1) $\cosh x \approx 1 + x^2/2$;
2) $\sinh x \approx x$;
3) $\cos y = [(8 + (\cosh x)^2)^{1/2} - \cosh x]/2 \approx 1 - x^2/6$;
4) $y \approx x/\sqrt{3}$.

The above expression can be rewritten as the sum of three integrals $V_{m,n}/IR = I_1 + I_2 + I_3$, where

$$I_1 = \left(\frac{1}{2\pi}\right) \int_0^{\frac{\pi}{2}} \frac{\left(1 - e^{-|m-n|\sqrt{3}y} \cos(m+n)y\right)}{\sqrt{3}y} dy$$

$$I_2 = \left(\frac{1}{2\pi}\right) \int_0^{\frac{\pi}{2}} \left(\frac{1}{\sinh x \cos y} - \frac{1}{\sqrt{3}y}\right) dy$$

$$I_3 = \left(\frac{1}{2\pi}\right) \left[\int_0^{\frac{\pi}{2}} \frac{e^{-|m-n|\sqrt{3}y} \cos(m+n)y}{\sqrt{3}y} dy \right.$$

$$\left. - \int_0^{\frac{\pi}{2}} \frac{e^{-|m-n|x} \cos(m+n)y}{\sinh x \cos y} dy \right]. \quad (13)$$

The first integral can be expressed in terms of the exponential integral $Ein(z)$

$$Ein(z) = \int_0^z \left[\frac{(1 - e^{-t})}{t}\right] dt = \int_0^{\frac{\pi}{2}} \left[\frac{\left(1 - e^{\frac{-2yz}{\pi}}\right)}{y}\right] dy, \quad (14)$$

so that

$$I_1 = \frac{1}{2\sqrt{3}\pi} \text{Re}\left\{Ein\left(\frac{\pi}{2}\left[|m-n|\sqrt{3} - i(m+n)\right]\right)\right\}. \quad (15)$$

For large values of its argument, $Ein(z) \approx \ln z + c_1$, where $c_1 = 0.577\,21$. So, we have

$$I_1 \approx \frac{1}{4\sqrt{3}\pi} \left[\ln(m^2 + n^2 - mn) + 2(\ln \pi + c_1)\right]. \quad (16)$$

The second integral can be integrated numerically. Let $I_2 = (1/2\sqrt{3}\pi)\, c_2$, where $c_2 = \int_0^{\pi/2} (\sqrt{3}/\sinh x \cos y - 1/y) dy = 0.09772$.

The exponentials in **the third integral** are negligible, except for small values of x and y, and for those values, $\sinh x \cos y \approx x \approx \sqrt{3}y$, so the third integral can be neglected.

Finally, we have

$$V_{m,n} \approx \frac{IR}{4\sqrt{3}\pi} \left[\ln(m^2 + n^2 - mn) + c\right] \quad (17)$$

where $c = 2(\ln \pi + c_1 + c_2) = 3.6393$.

Appendix III
Manufacturing and Other Issues

As is well known from the example of the X initiative [32], any new back end of the line architecture requires engagement throughout the mask and process infrastructure. According to our discussions with domain experts [6], [23], the Y architecture presents a number of generic challenges to manufacturing; there are no show-stoppers, but engineering efforts will be required across several domains. Space limits preclude detailed discussion here, but we sketch several main points.

With respect to mask making, vector shaped beam ebeam lithography tools [1] create "shots" of varying shape and size by imaging the overlap of two apertures, typically both square. This allows a range of rectangular shots to be created and exposed on the mask. Existing Toshiba ebeam lithography systems can produce 45° pattern at high speed through the combination of one rectangular aperture and one with 45° and 135° edges [32]. The new JEOL JBX3030 tool [18] also has apertures to produce 45° and 135° edges. These new tools mitigate the write time implications of angled data since they provide an alternative to approximating an angled line with a series of small rectangles; Fig. 12 illustrates mask fracturing using both rectangle and triangle shots versus mask fracturing using only rectangle shots [32]. With successful experiences with 45° edges in mind, it is possible that 60° and 120° edges can be printed assuming future availability of 30° and 60° angles in apertures.

Current support for angular edges is really focused on small edge segments rather than long lines. To produce long lines efficiently, it is necessary to have a pair of rectangular apertures

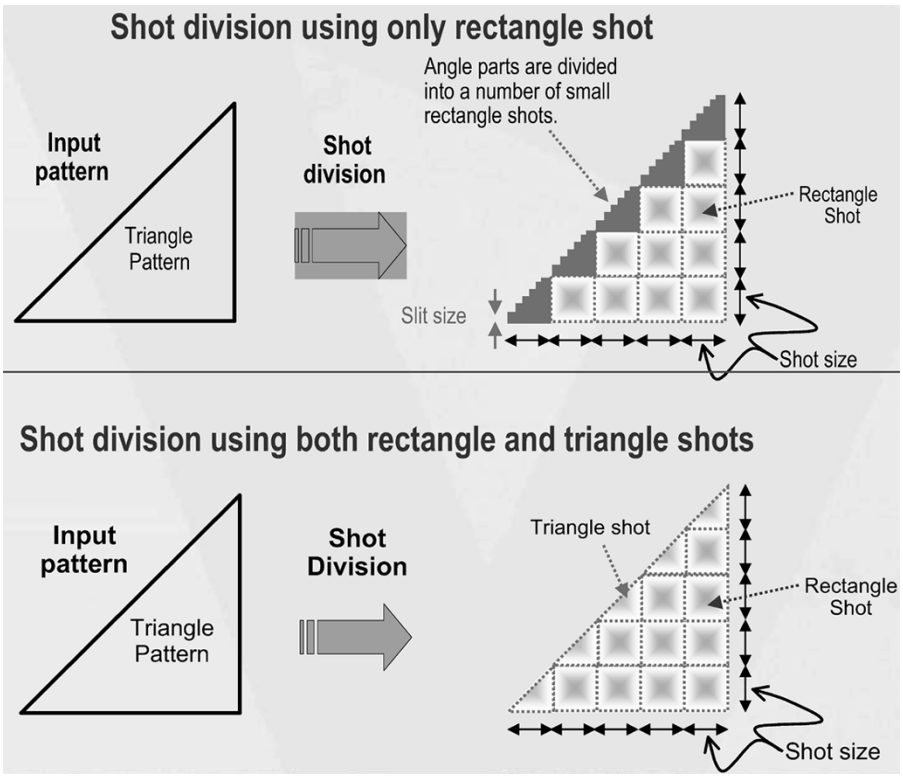


Fig. 12. Toshiba machine triangle shots.

rotated to still produce rectangular shots, but rotated to the desired angle. On the other hand, if the Y architecture is applied only to the upper and lower resolution metal layers—as we have proposed—the write time issue could be solved if the masks could be made with optical (laser) lithography (e.g., ETEC Alta writers), where throughput is independent of angular edges.

Of course, algorithms for optical proximity correction (OPC) and mask data preparation (layout data fracturing for vector shaped e-beam writers) will need to be updated for Y architecture geometries. We believe that all forms of rule- and model-based OPC can be adapted and calibrated to handle Y architecture geometries. Current fracturing algorithms already handle (45°) "slant" edges when partitioning layout into trapezoids; introduction of 30° and 60° slant edges should not pose a significant challenge.

The potential of nonrectangular die also presents challenges to package input/output (I/O) design and dicing. Current side-to-side die sawing cannot cut hexagonal dies due to the silicon lattice structure. New technologies, such as waterjet-guided laser [35], are emerging to confront the challenges.

There are other challenges related to inspection, exposure, repair, metrology and pattern compensation. Ultimately, the deployment of the Y architecture will depend on careful engineering, and provable cost reductions vis-a-vis achievable design quality with pervasive 60° and 120° wiring.

REFERENCES

[1] F. Abboud, S. V. Babin, V. Charkarian, A. Ghanbari, R. Innes, F. Raymond, III, and C. A. Sauer, "Design considerations for an electron-beam pattern generator for the 130-nm generation of masks," *SPIE Symp. Photomask X-Ray Mask Technol. VI*, vol. 3748, pp. 385–399, 1999.

[2] D. Atkinson and F. J. van Steenwijk, "Infinite resistive lattices," *Amer. J. Phys.*, vol. 67, pp. 486–492, 1999.

[3] H. B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.

[4] H. B. Bakoglu, J. T. Walker, and J. D. Meindl, "A symmetric clock distribution tree and optimized high-speed interconnections for reduced clock skew in ULSI and WSI circuits," in *Proc. IEEE Int. Conf. Computer Design*, Oct. 1986, pp. 118–122.

[5] S. Boyd, L. Vandenberghe, and A. El Gamal, "Design of robust global power and ground networks," in *Proc. ACM/SIGDA Int. Symp. Phys. Design*, 2001, pp. 283–288.

[6] *Dupont Photomasks*

[7] H. Chen, B. Yao, F. Zhou, and C. K. Cheng, "Physical planning of on-chip interconnect architectures," in *Proc. IEEE Int. Conf. Comput. Design*, Sep. 2002, pp. 30–35.

[8] ——, "The Y architecture: Yet another on-chip interconnect solution," in *Proc. Asia South Pacific Design Automation Conf.*, 2003, pp. 840–846.

[9] H. Chen, C.-K. Cheng, A. B. Kahng, I. Măndoiu, and Q. Wang, "Estimation of wirelength reduction for λ-geometry vs. manhattan placement and routing," in *Proc. ACM/IEEE Workshop System Level Interconnect Prediction*, 2003, pp. 71–76.

[10] H. Chen, C.-K. Cheng, A. B. Kahng, I. Măndoiu, Q. Wang, and B. Yao, "The Y architecture for on-chip interconnect: analysis and methodology," in *Proc. Int. Conf. Computer-Aided Design*, 2003, pp. 13–19.

[11] H. Chen, C.-K. Cheng, A. B. Kahng, Q. Wang, and B. Yao, Optimal Sizing Analyzes for Mesh-Based Power Plans, 2003.

[12] A. Dharchoudhury and R. Panda, "Design and analysis of power distribution networks in POwerPC microprocessors," in *Proc. Design Automation Conf.*, 1998, pp. 738–743.

[13] N. Garg and J. Konemann, "Faster and simpler algorithms for multicommodity flow and other fractional packing problems," in *Proc. 39th Annu. Symp. Foundations Comput. Sci.*, 1998, pp. 300–309.

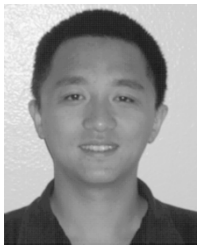
[14] T. Hildebrandt, "An annotated placement bibliography," *ACM SIGDA Newsletter*, pp. 12–21, 1985.

[15] M. Igarashi, T. Mitsuhashi, and A. Lee *et al.*, "A diagonal-interconnect architecture and its application to RISC core design," in *Proc. Int. Solid-State Circuits Conf.*, 2002, pp. 166–167.

[16] A. B. Kahng, I. I. Măndoiu, and A. Z. Zelikovsky, "Highly scalable algorithms for rectilinear and octilinear Steiner trees," in *Proc. Asia South Pacific Design Automation Conf.*, 2003, pp. 827–833.

[17] C.-K. Koh and P. H. Madden, "Manhattan or non-Manhattan? A study of alternative VLSI routing architectures," in *Proc. Great Lakes Symp. VLSI*, 2000, pp. 47–52.

[18] M. Lemke, J. Gramss, and H. J. Doering *et al.*, "Advanced writing strategies for high-end mask making," *Proc. SPIE*, vol. 3996, pp. 166–172, 2000.
[19] P. H. Madden, Congestion Reduction in Traditional and New Routing Architectures.
[20] C. Sechen, "Placement and global routing of integrated circuits using simulated annealing," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, 1987.
[21] B. K. Nielsen, P. Winter, and M. Zachariasen, "An exact algorithm for the uniformly-oriented steiner tree problem," in *Proc. 10th Eur. Symp. Algorithms*, vol. 2461, 2002, pp. 760–772.
[22] M. Hayase and S. Meki, "An algorithm for Steiner trees in λ-geometry," *IPSJ J.*, vol. 38, no. 4, 1997.
[23] C. Progler, private communication, Nov. 2002.
[24] M. D. Rostoker, J. S. Koford, R. Scepanovic, E. R. Jones, G. R. Padmanahben, A. K. Kapoor, V. B. Kudryavtsev, A. E. Andreev, S. V. Aleshin, and A. S. Podkolzin, "Hexagonal Architecture," U.S. Pat. US6407434B1, Jun. 2002.
[25] ——, "CAD for Hexagonal Architecture," U.S. Pat. US5822214, Oct. 1998.
[26] R. Scepanovic, J. S. Koford, V. B. Kudryavtsev, A. E. Andreev, S. V. Aleshin, and A. S. Podkolzin, "Microelectronic Integrated Circuit Structure and Method Using Three Directional Interconnect Routing Based on Hexagonal Geometry," U.S. Pat. US5578840, Nov. 1996.
[27] P. Saxena, N. Menezes, P. Cocchini, and D. A. Kirkpatrick, "The scaling chanllenge: can correct-by-construction design help?," in *Proc. Int. Symp. Physical Design*, 2003, pp. 51–58.
[28] D. Stroobandt and J. V. Campenhout, "Accurate interconnection length estimations for predictions early in the design cycle," *VLSI Design*, vol. 10, no. 1, pp. 1–20, 1999.
[29] S. Teig and J. L. Ganley, "Method and Apparatus for Considering Diagonal Wiring in Placement," Int. Patent Application, no. WO 02/47 165 A2, June 2002.
[30] TSMC 0.13 μm Design Rules. [Online] Available: http://www.tsmc.com
[31] S. Teig, "The X architecture," in *Proc. ACM/IEEE Workshop on Syst. Level Interconnect Prediction*, 2002, pp. 33–37.
[32] X Initiative. [Online] Available: http://www.xinitiative.org.
[33] Intel Research Webpage on Packaging. [Online] Available: http://www.intel.com/research/silicon/packaging.htm
[34] The ITRS Assembly and Packaging Roadmap. [Online] Available: http://public.itrs.net
[35] WaterJet-Guided Laser in Wafer Cutting—Synova SA. [Online] Available: http://www.gemcity.com/downloads/synova01.pdf



Hongyu Chen (S'03) received the B.S. degree in computer science and technology from Tsinghua University, Beijing, China, and the M.S. degree in computer science in 2002 from University of California at San Diego, La Jolla, where he is currently pursuing the Ph.D. degree.

His research interests include interconnect architectures, on-chip networks, power and clock distribution, and high-speed signal propagation.

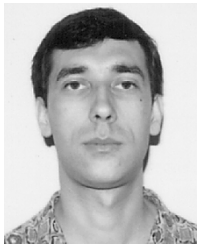
Chung-Kuan Cheng (S'82–M'84–SM'95–F'00) received the B.S. and M.S. degrees in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1976 and 1978 and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1984.

From 1984 to 1986, he was a Senior Computer-Aided Design Engineer with Advanced Micro Devices Inc. In 1986, he joined the University of California at San Diego (UCSD), La Jolla, where he is a Professor in the Computer Science and Engineering Department and an Adjunct Professor in the Electrical and Computer Engineering Department. He served as a Chief Scientist with Mentor Graphics in 1999. His research interests include network optimization and design automation of microelectronic circuits.

Prof. Cheng was an Associate Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS from 1994 to 2003. He is a recipient of the IEEE Transactions on Computer-Aided Design Best Paper Awards in 1997 and 2002 and the NCR Excellence in Teaching Award, School of Engineering, UCSD, 1991.

Andrew B. Kahng (A'89–M'03) received the A.B. degree in applied mathematics is from Harvard College and the M.S. and Ph.D. degrees in computer science are from University of California (UC) at San Diego, La Jolla.

From 1989 to 2000, he was a member of the Computer Science Department, UC, Los Angeles. He is a Professor of Computer Science Engineering and Electrical and Computer Engineering at UC, San Diego. He has published over 200 papers in the very large scale integration computer-aided design literature, receiving three Best Paper awards and an NSF Young Investigator award. His research is mainly in physical design and performance analysis of VLSI, as well as the VLSI design-manufacturing interface. Other research interests include combinatorial and graph algorithms, and large-scale heuristic global optimization. Since 1997, he has defined the physical design roadmap for the International Technology Roadmap for Semiconductors (ITRS), and since 2001 has chaired the U.S. and international working groups for Design technology for the ITRS. He has been active in the MARCO Gigascale Silicon Research Center since its inception. He was also the founding General Chair of the ACM/IEEE International Symposium on Physical Design, and cofounded the ACM Workshop on System-Level Interconnect Planning.



Ion I. Măndoiu received the M.S. degree from Bucharest University, Bucharest, Romania, in 1992 and the Ph.D. degree from Georgia Institute of Technology, Atlanta, in 2000, both in computer science.

He worked as a Research Assistant at Bucharest University (1992–1995), Instructor at Georgia Institute of Technology (2000–2001), and Postgraduate Researcher at University of California, Los Angeles (2000–2001). Since 2001 he has been a Research Scientist at University of California, San Diego. He is the author of over 40 refereed scientific publications, including an Asia-South Pacific Design Automation Conference Best Paper. His research interests include approximation algorithms, VLSI physical layout design, and combinatorial optimization.

Qinke Wang (S'03) received the B.S. degree in computer science and technology and the M.S. degree in computer software from Tsinghua University, Beijing, China, in 1998 and 2001, respectively. He is currently pursuing the Ph.D. degree at the University of California at San Diego, La Jolla.

His current research interests include analytical placement, multiproject wafer and reticle cost optimization, and interconnect architectures.

Bo Yao (S'01) received the B.S. degree in electrical engineering from Zhejiang University, Zhejiang, China, in 1997 and the M.S. degree in computer science from Tsinghua University, Beijing, China, in 2000. He is currently pursuing the Ph.D. in computer science at the University of California at San Diego, La Jolla.

His research interests include floorplanning, placement, interconnect optimization, and circuit simulation.