COORDINATED SCIENCE LABORATORY
College of Engineering

COMPLEXITIES OF LAYOUTS IN THREE-DIMENSIONAL VLSI CIRCUITS

Mokhtar A. Aboelaze
Benjamin W. Wah

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS None	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UILU-ENG-87-2212 CSG-63		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Coordinated Science Lab University of Illinois	6b. OFFICE SYMBOL (If applicable) N/A	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research/NSF	
6c. ADDRESS (City, State, and ZIP Code) 1101 W. Springfield Ave. Urbana, IL 61801		7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy St. 1800 G. Street, NW Arlington, VA 22217 Washington, DC	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Joint Services Electronics Program/NSF	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-84-C-0149 NSF DMR 85-19649	

8c. ADDRESS (City, State, and ZIP Code) 800 N. Quincy St. 1800 G. Street, NW Arlington, VA 22217 Washington, DC	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Complexities of Layouts in Three-Dimensional VLSI Circuits

12. PERSONAL AUTHOR(S)   Aboelaze, Mokhtar A. and Wah, Benjamin W.

13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) February 1987	15. PAGE COUNT 24

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	cost; graph embedding, one-active-layer layout; separator; three-dimensional layout; undirected graph; unrestricted layout; VLSI complexity; volume; wire length

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Recent advances in Very Large-Scale Integration (VLSI) fabrication technologies have demonstrated the feasibility of three-dimensional (3-D) circuits in a single chip. Due to the ability and flexibility to connect non-adjacent circuits using the third dimension, the cost of mapping non-planar circuits to two-dimensional (2-D) systems can be reduced. In this report, we examine the complexities in volume and maximum wire length of mapping circuits represented as undirected graphs to 3-D systems. Tighter bounds than those previously known are shown for various families of graphs, in both the one-active-layer and the unrestricted layouts. Finally, we develop a cost model to reflect the cost implementation in the third dimension and present an optimization model on the number of layers to minimize the overall cost.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.   SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete.

COMPLEXITIES OF LAYOUTS
IN THREE-DIMENSIONAL VLSI CIRCUITS

Mokhtar A. Aboelaze and Benjamin W. Wah
Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana–Champaign
Urbana Illinois 61801

ABSTRACT

Recent advances in Very-Large-Scale-Integration (VLSI) fabrication technologies have demonstrated the feasibility of three-dimensional (3-D) circuits in a single chip. Due to the ability and flexibility to connect non-adjacent circuits using the third dimension, the cost of mapping non-planar circuits to two-dimensional (2-D) systems can be reduced. In this paper, we examine the complexities in volume and maximum wire length of mapping circuits represented as undirected graphs to 3-D systems. Tighter bounds than those previously known are shown for various families of graphs, in both the one-active-layer and the unrestricted layouts. Finally, we develop a cost model to reflect the cost of implementation in the third dimension and present an optimization model on the number of layers to minimize the overall cost.

Index terms: Cost, graph embedding, one-active-layer layout, separator, three-dimensional layout, undirected graph, unrestricted layout, VLSI complexity, volume, wire length.

1. INTRODUCTION

The increasing demands for faster processors in scientific as well as commercial computations indicate the need for tremendous computing capacity, in terms of speed and volume. One way to achieve this is to build chips with more active devices. To increase the number of devices in a single chip while maintaining a reasonable yield, the transistor size should be decreased. There exist problems with decreasing transistor size, such as the short-channel effect and the nonstatistical behavior of transistors that span only a few hundred or a few thousand silicon atoms [15]. However, long before these problems become important, the problem of reduced driving capability of smaller transistors will have an equally profound impact on the layout of VLSI chips, since the average wire length grows linearly with the number of transistors. As a result, a limit will be reached at which the size of a transistor cannot be decreased any more without affecting its ability to correctly transmit a signal to another transistor.

Recently, 3-D VLSI circuits have been shown to be feasible. 3-D VLSI circuits are more flexible than their corresponding 2-D counterparts because wire routing is easier and more systematic, the runs of wires are shorter, and the volume of a 3-D realization may be less [19]. Wise has demonstrated this phenomenon in a two-layer layout of the Banyan/FFT networks; however, his work was directed towards the printed-circuit level rather than the VLSI-chip level [22]. With increased flexibility in placing devices in a 3-D circuit, the complexity of the resulting circuit can be reduced, hence, the driving capability of a transistor and the overall power requirement can be reduced. Dr. Gibbons, the president of Texas Instruments, predicted the feasibility of such chips in the earlier 1990s [3]. Examples of current implementations include IBM's "modestly" three-dimensional Thermal Conduction Module (TCM) circuit package [2] and Hughes' 32-by-32 3-D cellular computer to be finished in 1987 [15]. Nudd, Etchells, and Grinberg has proposed a cellular machine employing 3-D technology to perform image understanding operations [4, 15].

The feasibility of 3-D VLSI technology is still plagued by four major problems. One problem is the alignment of the corresponding positions in successive layers of a chip. Another problem is the creation of truly cylindrical holes. Due to effects like diffraction, scattering, and nonuniform exposure to solvents, the holes tend to be accentuated at the top or at the bottom [5]. Recent work on X-ray beam and refined optical lithography [19] suggested that this issue will be less of a problem in the future. The third problem is that placing active devices in MOS technology deep inside a 3-D volume would require multiple layers of monocrystalline silicon to be deposited, and subsequent processing of the chip would destroy the crystal structure of the monocrystalline silicon. Recent work at Texas Instruments [6] and IBM [21] suggested that full layers of the monocrystalline silicon are not needed, and transistors can be fabricated on islands of monocrystalline silicon that reside on a sea of oxide. The fourth problem is the heat generation and the cooling of such chips. However, due to shorter wire lengths, the heat generated will be less than

the corresponding 2-D circuits. Moreover, the problem is less severe in a one-active-layer chip in which active devices exist in one layer, and the rest of the volume is used for wire routing [19, 18].

In this paper, we show improved bounds on volume and maximum wire length of 3-D layouts, in both the one-active-layer and unrestricted models. In Section 2, we present a model for 3-D layouts. In Sections 3 and 4, we develop tighter lower and upper bounds on volume and maximum wire length and propose layouts for the various families of undirected graphs. Finally, we present an optimization model to minimize the overall cost of the design.

2. A MODEL OF 3-D VLSI CIRCUITS

In this section, we describe the model used to obtain the lower and upper bounds of mapping various families of undirected graphs in 3-D circuits. We will also briefly describe other attempts in 3-D layouts.

The model used here is an extension of Thompson's 2-D model into three dimensions [20]. The model consists of a 3-D grid of width W, length L, and height H, respectively (Figure 1). A vertex in this grid, (x, y, z), where $0 \leqslant x \leqslant W$, $0 \leqslant y \leqslant L$, and $0 \leqslant z \leqslant H$, denotes the location where devices can reside. An edge in the grid represents a wire in the circuit. It is assumed that three mutually perpendicular lines in the grid can pass through one point without physically touching each other. As a special case, the traditional 2-D circuit with two levels of metalization can be considered as a one-layer 3-D circuit because the two levels of metallic conductor can cross without touching each other. It is further assumed that any active device will require a unit volume, that the cross section of a wire is a unit area, and that the separation between the wires in any direction is of unit length. These assumptions are not over-restrictive as we are evaluating the order-of-magnitude asymptotic complexities.

To find the upper bounds of mapping an undirected graph $G = (V, E)$, where V is a set of vertices and E is a set of edges connecting the vertices, it is necessary to find a one-to-one mapping between the set of vertices of the graph and the set of nodes of the grid and, at the same time, a one-to-one mapping between the set of edges in the graph and the set of disjoint paths of the grid. The volume of the layout is the minimum volume of a parallelepiped containing the layout, while the maximum wire length is the maximum run of a wire without encountering any active device.

Rosenberg proposed two models of 3-D layouts [18, 19]. The first model is the one-active-layer model in which active devices are allowed to reside in either the top or the bottom layer, and the other layers will be used for the routing of wires. The second model is the unrestricted model in which active devices can be placed anywhere in the volume. In general, the one-active-layer model requires more volume and longer wires than its unrestricted counterpart.
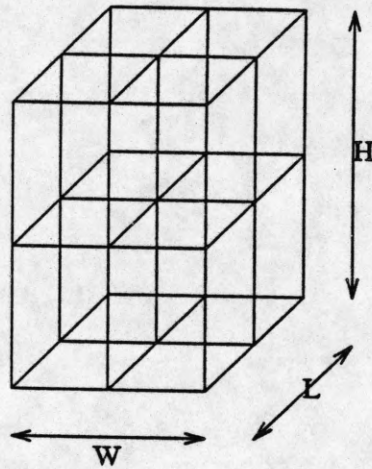
Figure 1. A 3-D grid.

Rosenberg proved that there is an unrestricted 3-D realization of the n-input rearrangeable permutation network that consumes $\Theta(n^{3/2})$ volume and that there is a one-active-layer 3-D realization of the same network with $\Theta(n^{3/2}\log n)$ volume* [18, 19]. Preparata proposed a layout for the cube-connected-cycles and developed the upper and lower bounds using the $VT^{3/2}$ and VT measures, where V is the volume, and T is the computation time [17]. He pointed out that the $VT^{3/2}$ measure is suitable for the unrestricted layout, while the VT measure is suitable for the one-active-layer layout. Leighton and Rosenberg have found lower and upper bounds for the layout of various families of undirected graphs [8, 12].

In the next two sections, we will develop improved lower and upper bounds to map an undirected graph to a 3-D grid for the one-active-layer and the unrestricted layouts and compare our bounds to previously known results. The undirected graphs considered are classified into families as characterized by their separators, which define the relationship between the area or volume of layout and the connectivity of the graph. An N-node graph G is said to have $f(N)$ separator if (a) G can be partitioned into two graphs, each with $N/2$ nodes, by cutting no more than $f(N)$ edges; and (b) both of the two $N/2$-node subgraphs have $f(N/2)$ separators. Lipton and Tarjan proved that any N-node planar graph has $O(\sqrt{N})$ separator [14]. Note that the above result is an upper bound, hence, it is possible for planar graphs to have separators less than $\Theta(\sqrt{N})$ and for

---

* Θ indicates the set of functions with the same order-of-magnitude complexity; O indicates the set of functions with the upper-bound order-of-magnitude complexity; Ω indicates the set of functions with the lower-bound order-of-magnitude complexity.

non-planar graphs to have $\Theta(\sqrt{N})$ separators. The relationship between the separator and the corresponding area of layout was first observed by Thompson [20], who showed that the lower-bound area to lay out a graph with separator ω is $\Omega(\omega^2)$. Leighton obtained lower and upper bounds on the area and maximum wire length to lay out the various families of graphs with $\Theta(N^q)$ separator, where $q < 1/2$, $q = 1/2$, and $q > 1/2$, and the family of planar graphs using 2-D technologies [10].

3. LOWER BOUNDS ON VOLUME AND MAXIMUM WIRE LENGTH

In this section, we develop improved lower bounds for the various families of undirected graphs in the one-active-layer model. To prove the lower bounds on embedding the various families of undirected graphs in a 3-D grid, it is necessary to find a representative graph in each family such that this graph will have the greatest lower bound. We did not find any improvement in lower bounds for the unrestricted model because the existing lower bounds on volume will be shown equal to the improved upper bounds in Section 4.2, except for graphs with $\Theta(N^{2/3})$ separator, and, hence, are already tight. Table 1 summarizes the existing lower bounds on volume and maximum wire length for the unrestricted model [19, 12, 11].

Graph Separator $f(N)$	Volume	Maximum Wire Length
$\Theta(N^q)$, $0 \leqslant q \leqslant 2/3$	$\Omega(N)$ [19, 12, 11]	$\Omega(N^{1/3}/\log N)$ [19]
Planar	$\Omega(N)$ [19, 12, 11]	$\Omega(N^{1/3}/\log N)$ [19]
$\Theta(N^q)$, $2/3 < q \leqslant 1$	$\Omega(N^{3q/2})$ [19]	$\Omega(N^{q/2}/\log N)$ [19]

Table 1. Lower bounds on volume and maximum wire length for the unrestricted model. (Note that a lower bound is intended to mean the largest known lower bound for a graph in the given family.)

The following theorem proves the necessary area for a 3-D circuit to be converted into a 2-D circuit. This theorem is an improvement over the one proved by Leighton and Rosenberg [12] and shows that the area required is $4BH^2$ instead of $9BH^2$ and that the degree of the graph can be six instead of four.

Theorem 1: Any 3-D layout of volume V, base area B, and height H can be transformed into a 2-D layout of area $A = 4BH^2$. If the maximum wire length in the 3-D layout is W_{3d}, then the maximum wire length in the 2-D layout is $W_{2d} \leq 2 \cdot max(H,3) \cdot W_{3d}$.

Proof: Without loss of generality, we will transform a 3-D grid of base area $B = WL$ and height H into a 2-D grid of area $A = 4WLH^2$. Consider the 3-D grid in Figure 2a. Assume that the nodes of this grid are located in the Cartesian coordinates (x,y,z), where $0 \leq x < W$, $0 \leq y < L$, $0 \leq z < H$. Point (x,y,z) in the 3-D grid is mapped to point (x',y') in the 2-D grid such that

$$x' = Hx + z; \qquad y' = Hy + z \tag{1}$$

Note that the width and length of the 2-D grid are $W' = HW$ and $L' = HL$, respectively.

Figure 2b shows the mapping of the 3-by-3-by-3 grid into a 9-by-9 grid, where solid lines represent connections in the first plane in Figure 2a, dashed lines represent connections in the second plane, dotted lines represent connections in the third plane, and diagonal lines represent connections across different planes. The effect of this mapping is that two nodes in a straight line in the x or y direction in the 3-D grid are mapped into two nodes in a straight line in the same direction in the 2-D grid, but the distance between them is multiplied by H. As an illustration, $a_{1,1}$ and $a_{1,3}$ are separated by a distance of two units and are in the x direction in Figure 2a. These two points are separated by a distance of six units ($H=3$) and are also in the x direction in Figure 2b. Note that nodes in a straight line in the z direction are mapped to nodes in a straight line in the diagonal direction in the 2-D plane. Since most models in 2-D VLSI layouts do not allow connections in the diagonal direction, the problem can be circumvented by multiplying the area by four and mapping diagonal connections to a sequence of horizontal and vertical connections. Figure 2c shows a generic node connected to its six neighbors, where the distance between adjacent nodes in the x or y direction is unity. Figure 2d shows the same set of nodes after doubling the distance between two nodes and quadrupling the area of each node, hence, multiplying the total area by four. Here, a diagonal connection is altered to be a horizontal segment followed by a vertical segment and finally a horizontal segment. As a result, the area is

$$A = 4(HW) \cdot (HL) = 4(WLH^2) \tag{2}$$

To prove the result on the maximum wire length, note that a wire connecting any two nodes in the 3-D layout is composed of wires running in the x, y, and z directions. The length of wires in the x or y direction is multiplied by $2H$ during the transformation, while the length of wires in the z direction is multiplied by a constant less than six. It is straightforward to show that

$$W_{2d} \leq 2 \cdot max(H,3) \cdot W_{3d} . \tag{3}$$

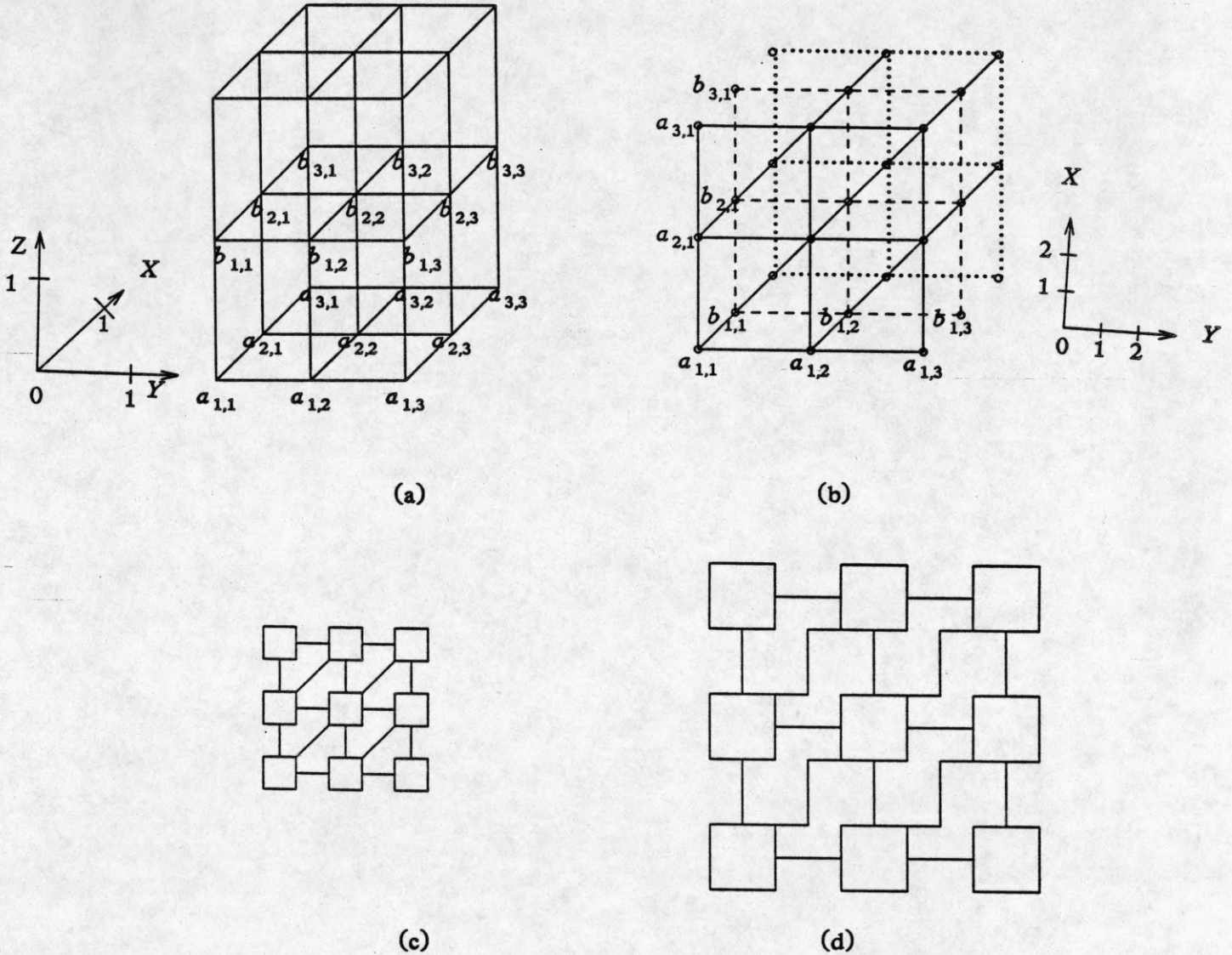which proves the theorem. □

(a)

(b)

(c)

(d)

Figure 2. The mapping of a 3-by-3-by-3 grid into a 9-by-9
grid. (a) a 3-D grid; (b) a mapping of the 3-D grid
in (a) to a 2-D grid; (c) a section of the 2-D grid
with diagonal connections; (d) an expanded sec-
tion of the 2-D grid without diagonal connections.

In the following theorem, the mesh of trees is considered as an example in the family of
graphs with $\Theta(\sqrt{N})$ separator, while the tree of meshes is considered as an example in the family
of planar graphs. These two example graphs were used by Leighton in proving the lower bound of
2-D layouts [7, 10].

The *mesh of trees* is defined as follows [10, 7]. Starting with an n-by-n matrix of nodes (n is assumed to be a power of 2) and adding nodes wherever necessary, a complete binary tree is constructed using nodes in each row and column of the matrix as leaves. Hence, each node in the mesh is a leaf of two orthogonal binary trees, one for the binary tree encompassing nodes in the row containing this node and another for the tree encompassing nodes in the column. (Orthogonal trees is another name for the mesh of trees.) An example of the mesh of trees is shown in Figure 3.
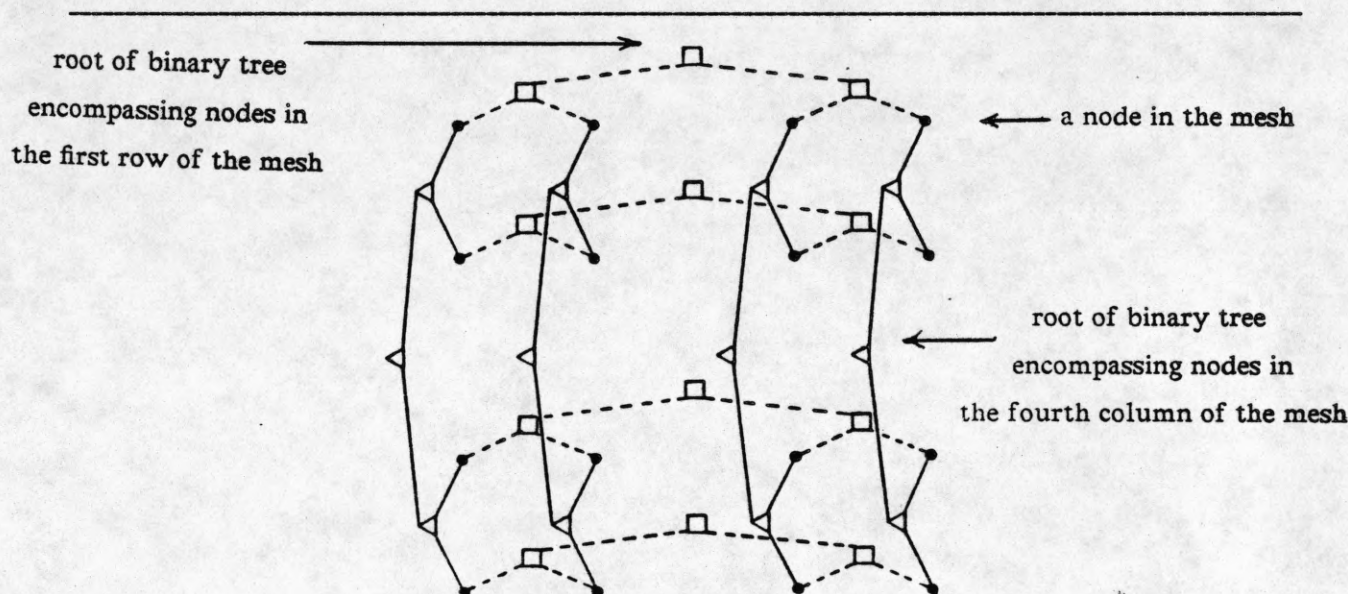
Figure 3. A 4-by-4 mesh of trees.

The *tree of meshes* is defined as follows [7, 10]. In a complete binary tree, each node is replaced by a mesh and each edge by several edges that connect the meshes together. The root is replaced by an n-by-n mesh (n is assumed to be a power of 2), its children are replaced be n/2-by-n meshes, whose children are replaced by n/2-by-n/2 meshes. This continues until the leaves of the original binary tree are replaced by 1-by-1 meshes. Figure 4 shows a 4-by-4 tree of meshes.

Theorem 2: (a) Any 3-D one-active-layer layout of the mesh of trees will require $\Omega(N \log N)$ volume and $\Omega(\sqrt{N} / \log \log N)$ maximum wire length. (b) Any 3-D one-active-layer layout of the tree of meshes will require $\Omega(N \sqrt{\log N})$ volume and $\Omega(\sqrt{N} / \log N)$ maximum wire length.
Proof: We will prove these lower bounds by contradiction. Leighton proved that any 2-D layout of the N-node mesh of trees will require $\Omega(N \log^2 N)$ area, and that this layout must have $\Omega(\sqrt{N} \log N / \log \log N)$ maximum wire length [10]. He also proved that any 2-D layout of the N-node tree of meshes will require $\Omega(N \log N)$ area, and that this layout must have $\Omega(\sqrt{N} / \sqrt{\log N})$ maximum wire length.
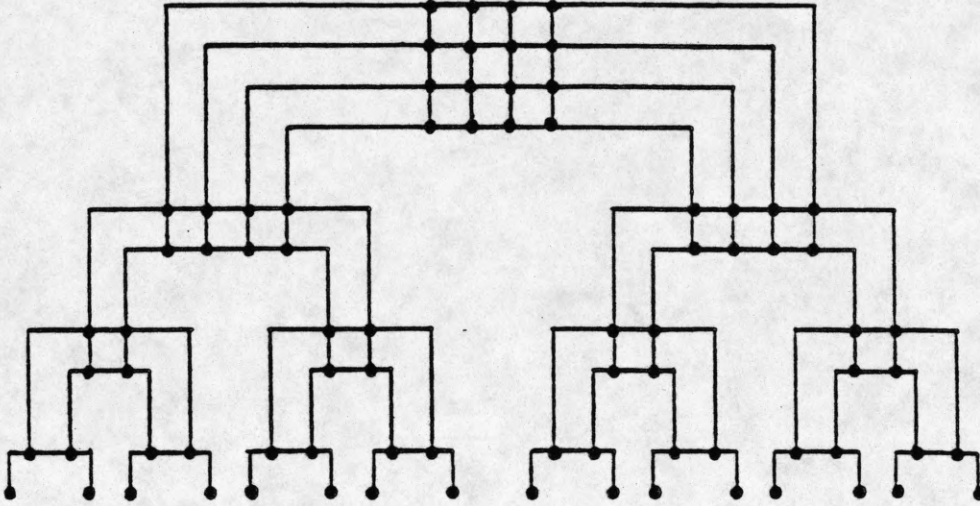
Figure 4. A 4-by-4 tree of meshes.

For the mesh of trees, let us assume the existence of a 3-D one-active-layer layout with a volume $V < \Theta(N \log N)$ and maximum wire length $W < \Theta(\sqrt{N} / \log \log N)$. The base area of this one-active-layer layout should be $\Omega(N)$, as it should be large enough to accommodate the N nodes of the graph. Therefore, the height of this layout is $H < \Theta(\log N)$. According to Theorem 1, this 3-D circuit can be transformed into a 2-D layout with area $A < \Theta(N \log^2 N)$ and maximum wire length $W_{2d} < \Theta(\sqrt{N} \log N / \log \log N)$, which contradict Leighton's results [10]. Thus, any 3-D one-active-layer layout of the mesh of trees will require $\Omega(N \log N)$ volume and $\Omega(\sqrt{N} / \log \log N)$ maximum wire length.

For the tree of meshes, assume the existence of a 3-D one-active-layer layout with volume $V < \Theta(N \sqrt{\log N})$ and maximum wire length $W < \Theta(\sqrt{N} / \log N)$. Since the base of this layout should have $\Omega(N)$ area, the height of this layout is $H < \Theta(\sqrt{\log N})$. Using Theorem 1, we can transform this layout into a 2-D layout with area $A < \Theta(N \log N)$ and maximum wire length $W_{2d} < \Theta(\sqrt{N} / \sqrt{\log N})$, which contradict Leighton's results. Hence, any 3-D one-active-layer layout of the trees of meshes should have $\Omega(N \sqrt{\log N})$ volume and $\Omega(\sqrt{N} / \log N)$ maximum wire length. $\square$

The upper-bound volume of the family of undirected graphs with $\Theta(N^q)$ separator, $1/2 < q \leqslant 1$, will be shown in Section 4.1 to be $O(N^{q+1/2})$. As the base area has $\Omega(N)$ complexity, the height of this layout has $O(N^{q-1/2})$ complexity. Leighton has proved the lower-bound maximum wire length for the family of undirected graphs with $\Theta(N^q)$ separator, $1/2 < q \leqslant 1$, in a 2-D

layout to be $\Omega(N^q)$ [10]. From Theorem 1, the maximum wire length in the 3-D one-active-layer layout should be $W_{2d}/H = \Omega(N^q/N^{q-1/2}) = \Omega(\sqrt{N})$.

For the family of graphs with $\Theta(N^q)$, $0 \leqslant q < 1/2$, separator, Paterson, Ruzzo, and Snyder have proved the lower-bound maximum wire length in a 2-D layout of a binary tree to be $\Omega(\sqrt{N}/\log N)$ [16]. In a one-active-layer 3-D layout, a similar argument can be made such that nodes of a binary tree are in one layer, and that the maximum distance between two nodes separated by $2 \log N$ edges is $\Omega(\sqrt{N})$. Hence, the lower bound in the 3-D case is the same as that of the 2-D case.

Table 2 summarizes the lower bounds obtained for the one-active-layer layout and compares them with previously known results.

Graph Separator $f(N)$	Volume		Maximum Wire Length	
	Previous	New	Previous	New
$\Theta(N^q)$, $0 \leqslant q < 1/2$	$\Omega(N)$ [12]	$\Omega(N)$ [12]	constant [12]	$\Omega(\sqrt{N}/\log N)$ [16]
Planar	$\Omega(N)$ [12]	$\Omega(N\sqrt{\log N})$	constant [12]	$\Omega(\sqrt{N}/\log N)$
$\Theta(\sqrt{N})$	$\Omega(N)$ [12]	$\Omega(N \log N)$	constant [12]	$\Omega(\sqrt{N}/\log\log N)$
$\Theta(N^q)$, $1/2 < q \leqslant 1$	$\Omega(N^{q+1/2})$ [12, 19]	$\Omega(N^{q+1/2})$ [12, 19]	$\Omega(N^{q-1/2})$ [12]	$\Omega(\sqrt{N})$

Table 2. Lower bounds on volume and maximum wire length for the one-active-layer model. (Note that a lower bound is intended to mean the largest known lower bound for a graph in the given family.)

4. UPPER BOUNDS

Before introducing the results on upper bounds, we review the necessary mathematical background behind the theory of layouts. Thompson introduced the idea of the minimum bisection width of an undirected graph and proved a relation between the minimum bisection width and the minimum area required to lay out the given graph [20]. Lipton and Tarjan introduced the idea of separator for a family of undirected graphs and proved that the family of planar graphs has a

$O(\sqrt{N})$ separator [14]. They also proposed a linear time algorithm to compute this separator. Bhatt and Leighton introduced the ideas of bifurcators and decomposition trees [1]. An N-node undirected graph has (F,α) bifurcator if it can be decomposed into two subgraphs, G_1 and G_2, by removing no more than F edges. Both G_1 and G_2 can further be decomposed into two subgraphs by removing no more than F/α edges. In general, any subgraph in level i can be decomposed into two subgraphs by removing no more than F/α^i edges. This decomposition can be represented by the decomposition tree in Figure 5.
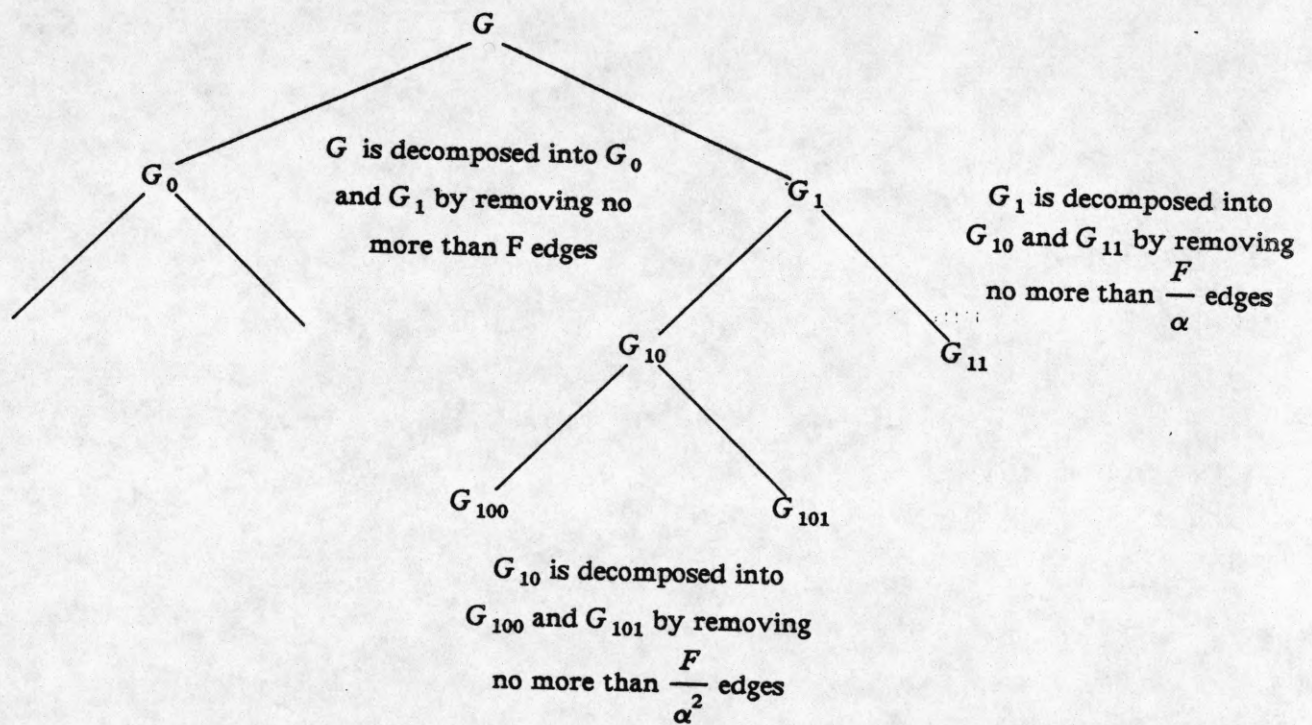
Figure 5. Decomposition tree for a graph with (F,α) bifurcator.

A decomposition tree is said to be a fully balanced decomposition tree if

(1) when decomposing any subgraph into two smaller subgraphs, the number of nodes in the two smaller subgraphs are equal; and

(2) when decomposing any subgraph into two smaller ones, the number of edges connecting this subgraph to the rest of the original graph is divided into two equal sets that are distributed in the two decomposed subgraphs.

Bhatt and Leighton also proved that any graph with (F,α) bifurcator has a fully balanced decomposition tree with (F',α) bifurcator, where F' is related to F by a constant. Leighton showed that if $F=N^q$, then the total number of edges connecting any subgraph with $N/2^i$ nodes in level i of

the decomposition tree to the rest of the original graph is $k(N/2^i)^q$, where k is a constant [9]. As a result, a graph with N^q separator has $(N^q, 2^q)$ bifurcator.

In the rest of this section, we show upper bounds for the various families of undirected graphs in both the one-active-layer and the unrestricted models.

4.1. One-Active-Layer Layouts

The following theorems prove the upper bounds for the various families of undirected graphs in the one-active-layer 3-D layouts. The family of planar graphs is treated in the same way as the family of graphs with $\Theta(\sqrt{N})$ separator here.

In the following theorem, the upper bounds for graphs with $\Theta(\sqrt{N})$ separator are proved. Although Leighton and Rosenberg have proved the same bounds before, they have assumed in their proof the existence of a layout of an n-node subgraph in which the ports of this subgraph "are sufficiently sparse that the routing is guaranteed to be possible" [12]. We will assume in the following proof that the ports of a subgraph are equally spaced along one side of its layout. In connecting two n-node subgraphs into a 2n-node subgraph, a complete crossbar switch will be used to perform the routing, and the ports in the resulting subgraph will also be equally spaced along one side of the resulting layout. The above model allows us to prove a better upper bound on volume for the family of graphs with $\Theta(N^q)$ separator. This will be shown in Theorem 4.

Theorem 3: Any undirected graph with $\Theta(\sqrt{N})$ separator has a 3-D one-active-layer layout with $O(N \log N)$ volume and $O(\sqrt{N})$ maximum wire length.

Proof: We assume that the balanced decomposition tree of the graph is known. The proof is by induction on a graph with n nodes. The case for n=1 is trivial. For the induction hypothesis, assume that an n-node graph can be mapped into a parallelepiped with volume $V(n)$, height $H(n)$, and a square base of side $L(n)=kc\sqrt{n}$, where k is a constant. It is further assumed that the $c\sqrt{n}$ ports to connect any node in this subgraph to another node outside this subgraph are aligned and equally spaced along one side of the top layer of this layout (see Figure 6, where the ports are represented by circles). In the induction step, consider the volume needed to lay out four n-node subgraphs. We will combine these four layouts to produce one 4n-node layout with volume $V(4n)$, height $H(4n)$, a square base of side $L(4n)=kc\sqrt{4n}$, and that the $c\sqrt{4n}$ ports of the 4n-node subgraph are aligned and equally spaced along one side of the top layer. This will be done by first showing that one additional layer is needed to accommodate the necessary interconnections when two n-node subgraph layouts are combined to form one 2n-node subgraph layout.

Consider two n-node layouts placed side by side as shown in Figure 6. Figure 7 shows the additional top layer that is created when the two n-node subgraph layouts are combined. We have (a) to create $c\sqrt{2n}$ ports in the 2n-node subgraph layout; and (b) to connect a maximum of $c\sqrt{n}$ ports in one of the n-node subgraphs to a maximum of $c\sqrt{n}$ ports in the other n-node subgraph.
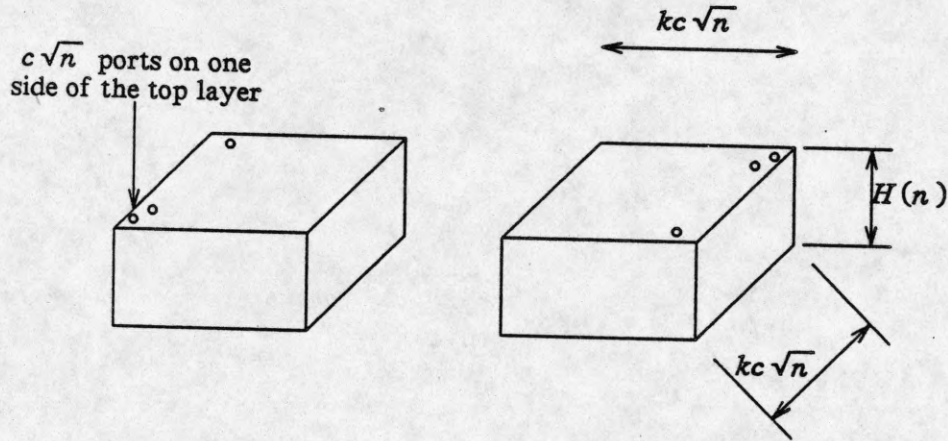
Figure 6. Two n-node subgraph layouts with $\Theta(\sqrt{n})$ separator. (The $c\sqrt{n}$ ports of each layout are represented as circles and are aligned on one side of the top layer.)

Since we have assumed that the subgraph has a balanced decomposition tree, half of the $c\sqrt{2n}$ ports in the combined layout will be connected to ports in the first layout, while the other half will be connected to ports in the second layout.

In Figure 7, the ports of the two n-node subgraphs, each with $c\sqrt{n}$ ports, are represented by $c\sqrt{n}$ circles along the sides. The newly created $c\sqrt{2n}$ ports of the 2n-node subgraph are represented by squares equally spaced along the top side. The objective is to route the $c\sqrt{2n}$ ports in the combined layout to ports in the two original layouts and to connect the $2c\sqrt{n}$ ports in the original layouts together. This can be done by creating a track out of each port in the original layout and extending it across the top layer in the new layout. These tracks are represented by the solid horizontal lines in Figure 7. Tracks are also created for each of the $c\sqrt{2n}$ ports in the combined layout and extended across the top layer. These tracks are shown by the dotted vertical lines in Figure 7. These horizontal and vertical tracks allow us to form a complete crossbar switch that connects any port in the two n-node layouts to any port in the combined layout. To connect the $c\sqrt{n}$ ports in one of the n-node layouts to the $c\sqrt{n}$ ports in the other n-node layout, a maximum of $c\sqrt{n}$ vertical tracks represented as dashed lines in Figure 7 are created to form a complete crossbar switch. The above construction process is feasible for $k \geqslant 2$ in the induction hypothesis because the number of horizontal tracks is $2c\sqrt{n}$, which is less than $kc\sqrt{n}$, and the maximum number of vertical tracks is $(c\sqrt{2n} + c\sqrt{n})$, which is less $2kc\sqrt{n}$.
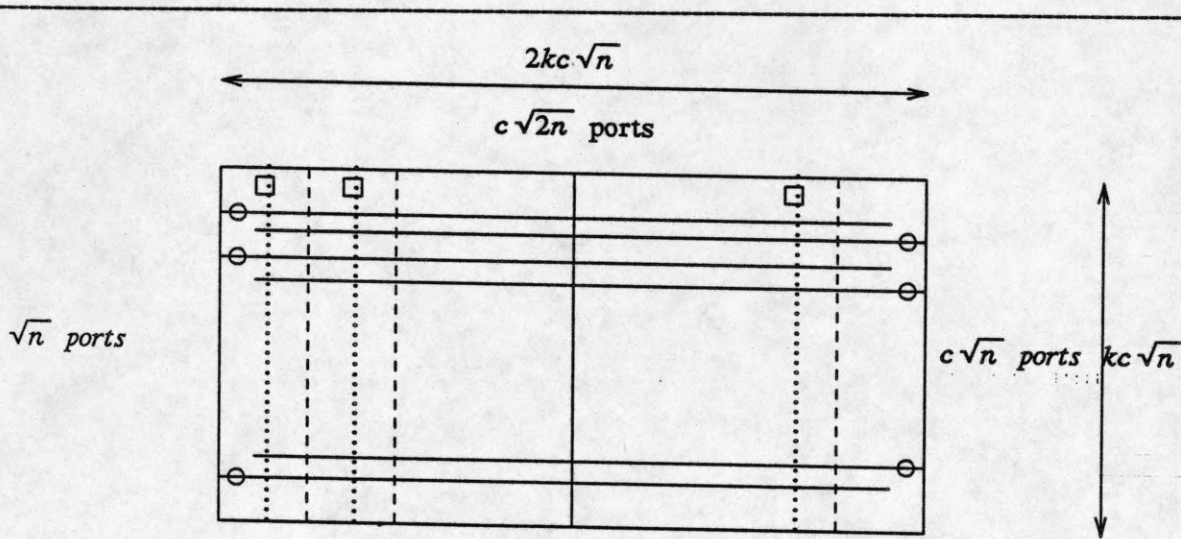
$2kc\sqrt{n}$

$c\sqrt{2n}$ ports

$\sqrt{n}$ ports

$c\sqrt{n}$ ports $kc\sqrt{n}$

Figure 7. New layer on top that is created when two n-node subgraph layouts are combined to form one 2n-node subgraph layout in the one-active-layer model. (Circles represent the $c\sqrt{n}$ ports in each n-node layout. Squares represent the $c\sqrt{2n}$ ports in the combined layout.)

In combining two n-node subgraph layouts to form one 2n-node subgraph layout, an additional layer is needed. The number of layers in the resulting layout can be computed from the following recurrence.

$$H(2n) = \begin{cases} H(n) + 1 & n > 1 \\ 1 & n \leqslant 1 \end{cases} \tag{4}$$

Similarly, we can combine two 2n-node subgraph layouts to form one 4n-node subgraph layout. In general, for an N-node subgraph layout, where N is a power of two,

$$H(N) = \log N \tag{5}$$

Since the base area of an N-node layout is $(kc\sqrt{N})^2$, the total volume will be

$$V_{1AL}(N) = k^2 c^2 N \log N = O(N \log N) \tag{6}$$

In computing the volume, no constraint is put on the routing of wires, hence, a wire can run along the $\log N$ layers in a zig-zag fashion in the worst case. The maximum wire length is

$$W_{1AL}(N) = O\left(2 \sum_{i=0}^{\log_2 N} kc \sqrt{\frac{N}{2^i}}\right) = O(\sqrt{N}) \tag{7}$$

According to the theory of induction, the theorem is proved. □

Theorem 4: Any undirected graph with $\Theta(N^q)$ separator, $1/2 < q \leqslant 1$, has a 3-D one-active-layer layout with $O(N^{q+1/2})$ volume and $O(\sqrt{N})$ maximum wire length.

Proof: The proof is similar to that of Theorem 3 except that cn^q edges connect any n-node subgraph to the rest of the graph. Assume that there are two n-node subgraph layouts, each in the form of a parallelepiped with height $H(n)$, square base with side $kc\sqrt{n}$, and that the cn^q ports are arranged in the form of a rectangle of width $n^{q-1/2}$ and length $c\sqrt{n}$ in the top layer (see Figure 8). In forming a 2n-node layout, we have to create $c(2n)^q$ new ports and route them to a maximum of $2cn^q$ ports of the two n-node layouts, and connect the ports of the two n-node layouts together in a similar way as in the proof of Theorem 3. We will need $n^{q-1/2}$ layers to form a complete 3-D crossbar switch to perform the routing between the $c(2n)^q$ ports in the combined 2n-node layout and the corresponding ports in the two n-node layouts. Another $n^{q-1/2}$ layers are needed to form a complete 3-D crossbar switch to connect the ports in the two n-node layouts.
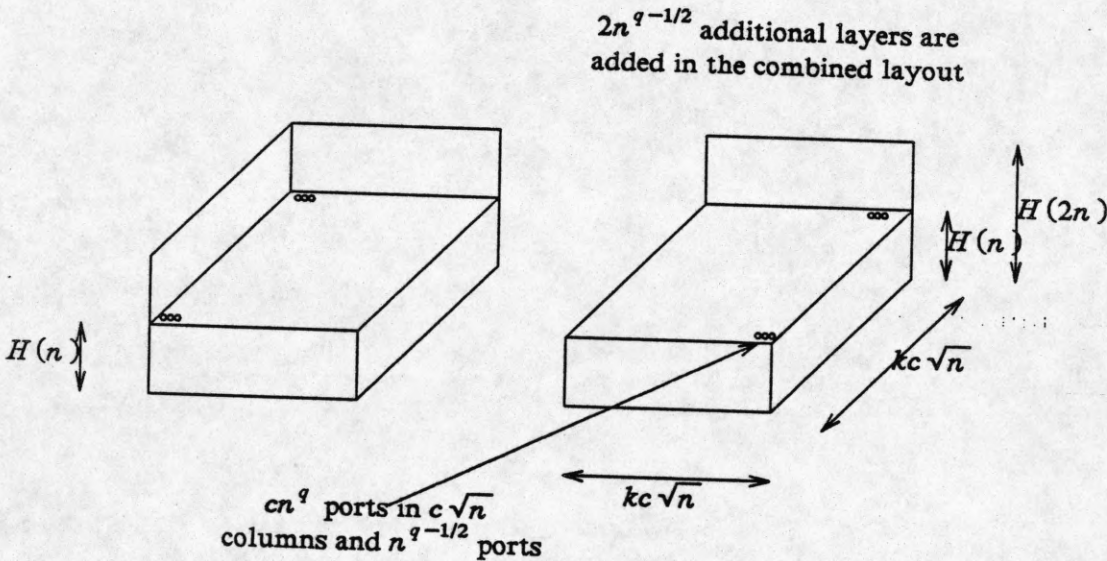
Figure 8. Two n-node-subgraph layouts with $\Theta(N^q)$, $1/2 < q \leqslant 1$, separator.

The height of the layout can be computed from the following recurrence.

$$H(2n) = \begin{cases} H(n) + 2n^{q-1/2} & n > 1 \\ 1 & n \leqslant 1 \end{cases} \tag{8}$$

In general, for an N-node layout, where N is a power of 2,

$$H(N) = \sum_{i=1}^{\log_2 N} 2\left[\frac{N}{2^i}\right]^{q-1/2} = O(N^{q-1/2}) \tag{9}$$

Since the base area of the N-node layout is $(kc \sqrt{N})^2$, the total volume is

$$V_{1AL}(N) = O(N \cdot N^{q-1/2}) = O(N^{q+1/2}) \tag{10}$$

Using the same argument as in the family of graphs with $\Theta(\sqrt{N})$ separator, the maximum wire length is

$$W_{1AL} = O(\sqrt{N}) \tag{11}$$

Note that the upper bounds on volume and maximum wire length are optimal because they are the same as the corresponding lower bounds (see Table 2). □

Table 3 shows the upper bounds in the one-active-layer model and compares them against previously known results [11, 12]. Note that the upper bounds on volume are tight in all cases except for the family of planar graphs.

Graph Separator $f(N)$	Volume		Maximum Wire Length	
	Previous	New	Previous	New
$\Theta(N^q)$, $0 \leqslant q < 1/2$	$O(N)$ [12]	$O(N)$ [12]	$O(\sqrt{N}/\log N)$ [12]	$O(\sqrt{N}/\log N)$ [12]
Planar	$O(N \log N)$ [12]	$O(N \log N)$	$O(\sqrt{N})$ [12]	$O(\sqrt{N})$
$\Theta(\sqrt{N})$	$O(N \log N)$ [12]	$O(N \log N)$	$O(\sqrt{N})$ [12]	$O(\sqrt{N})$
$\Theta(N^q)$, $1/2 < q \leqslant 1$	$O(N^{q+1/2} \log N)$ [12]	$O(N^{q+1/2})$	$O(\sqrt{N})$ [12]	$O(\sqrt{N})$

Table 3. Upper-bound volume and maximum wire length for the 3-D one-active-layer layouts. (The previous and new results may be the same but obtained by different methods.)

4.2. Unrestricted Layouts

The next theorem proves the complexities of layouts in the 3-D unrestricted model, in which devices can be placed anywhere in the 3-D volume.

Theorem 5: Any undirected graph with $\Theta(N^q)$ separator, $0 \leqslant q \leqslant 1$, has a 3-D unrestricted layout with volume

$$V(N) = \begin{cases} O(N) & 0 \leqslant q < 2/3 \\ O(N \log^3 N) & q = 2/3 \\ O(N^{3q/2}) & 2/3 < q \leqslant 1 \end{cases} \tag{12}$$

with maximum wire length

$$W(N) = \begin{cases} O(N^{1/3}) & 0 \leqslant q < 2/3 \\ O(N^{1/3} \log N) & q = 2/3 \\ O(N^{q/2}) & 2/3 < q \leqslant 1 \end{cases} \tag{13}$$

Proof: We assume that the balanced decomposition tree of the graph is known. The proof is by induction on a graph with n nodes. The case for n=1 is trivial. For the induction hypothesis, assume that an n-node layout is in the form of a cube. Further, assume that the cn^q ports of this layout are arranged in the form of a square with side $k\sqrt{cn^q}$ in one of the faces of the cube, where k is a constant. In the induction step, we will show that eight n-node subgraphs can be combined into one 8n-node layout in the form of a cube, and the $c(8n)^q$ ports of this layout are arranged in the form of a square of side $k\sqrt{c(8n)^q}$.

The induction step is proved by first arranging the eight n-node layouts in the corners of a larger cube, such that ports of the four upper cubes are directed downwards, while ports of the four lower cubes are directed upwards (see Figure 9). We will first combine two n-node subgraph layouts to form one 2n-node layout. We have (a) to create $c(2n)^q$ new ports for the 2n-node layout; and (b) to connect the ports in the two n-node layouts. Figure 10 shows the $2(cn^q)$ ports of the upper and lower layouts, each in the form of a square with side $k\sqrt{cn^q}$, where k is a constant. By adding $2^q k\sqrt{cn^q}$ layers between the upper and lower cubes in Figure 10, we can create a complete 3-D crossbar switch to perform the necessary routing between the newly created $c(2n)^q$ ports and any of the $2cn^q$ ports in the two original n-node layouts. The newly created $c(2n)^q$ ports are arranged in the form of a rectangle of length $k\sqrt{cn^q}$ and width $2^q k\sqrt{cn^q}$. To perform the necessary connections between the two n-node subgraph layouts, each with a maximum of cn^q ports, another complete 3-D crossbar switch with $k\sqrt{cn^q}$ layers is created between the upper and lower layouts in Figure 10.
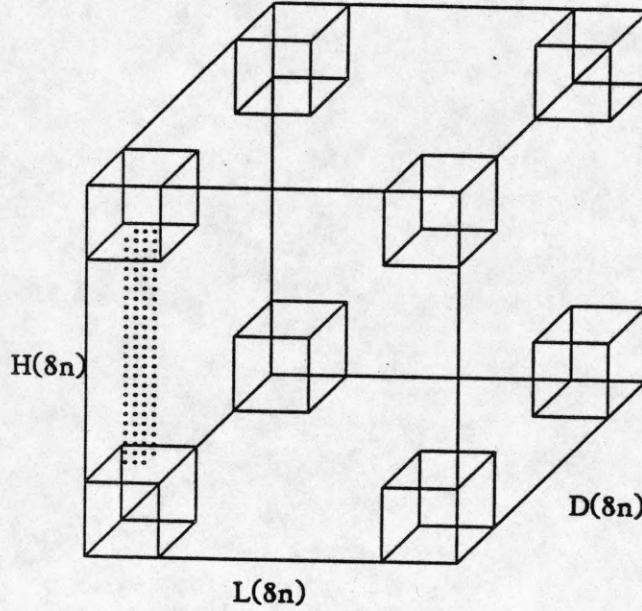
Figure 9. Combining eight n-node layouts to form one 8n-node layout.

In a similar way, we can combine four 2n-node layouts to form two 4n-node layouts and two 4n-node layouts into one 8n-node layout. In each case, we have added $\Theta(\sqrt{cn^q})$ layers between the two layouts concerned. The height, length, and width of the 8n-node layout can be computed from the following recurrences.

$$H(8n) = \begin{cases} 2H(n) + k_1\sqrt{cn^q} & n > 1 \\ 1 & n = 1 \end{cases} \tag{14}$$

$$L(8n) = \begin{cases} 2L(n) + k_2\sqrt{cn^q} & n > 1 \\ 1 & n = 1 \end{cases} \tag{15}$$

$$D(8n) = \begin{cases} 2D(n) + k_3\sqrt{cn^q} & n > 1 \\ 1 & n = 1 \end{cases} \tag{16}$$

where k_1, k_2, and k_3 are constants. Solving the last three equations, we get

$$D(N) = L(N) = H(N) = \begin{cases} O(N^{1/3}) & 0 \leqslant q < 2/3 \\ O(N^{1/3}\log N) & q = 2/3 \\ O(N^{q/2}) & 2/3 < q \leqslant 1. \end{cases} \tag{17}$$

The volume of the layout will be
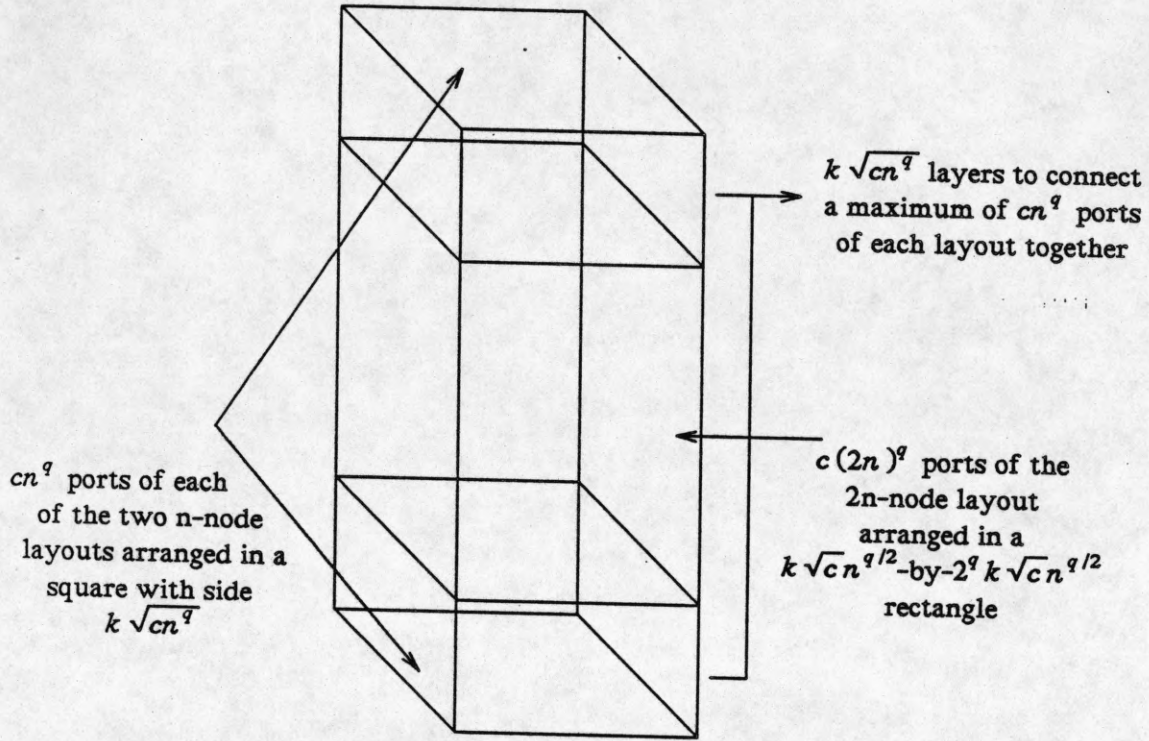
Figure 10. Combining two n-node layouts to form one 2n-node layout in the unrestricted model.

$$V(N) = \begin{cases} O(N) & 0 \leq q < 2/3 \\ O(N \log^3 N) & q = 2/3 \\ O(N^{3q/2}) & 2/3 < q \leq 1 . \end{cases} \tag{18}$$

For the maximum wire length, note that the maximum wire length of the 8n-node layout is equal to the maximum wire length of the n-node layout plus $\alpha L(8n)$, where $L(8n)$ is the length of each side of the 8n-node layout, and α is a constant. (Due to the crossbar connection, we did not extend any port more than a length of $\alpha L(8n)$). The maximum wire length can be computed from the following recurrence.

$$W(8n) = W(n) + \alpha L(8N) . \tag{19}$$

Substituting L from Eq. (17),

$$W(N) = \begin{cases} O(N^{1/3}) & 0 \leq q < 2/3 \\ O(N^{1/3} \log N) & q = 2/3 \\ O(N^{q/2}) & 2/3 < q \leq 1 . \end{cases} \tag{20}$$

According to the theory of induction, the theorem is proved. □

- 20 -

Table 4 shows the upper bounds obtained here for the volume and maximum wire length, respectively, and compares them with previously known results [8, 12, 11]. Comparing the upper bounds in Table 4 and the lower bounds in Table 1, all upper bounds on volume obtained here are tight except for the family of graphs with $\Theta(N^{2/3})$ separator.

Graph Separator $f(N)$	Volume		Maximum Wire Length	
	Previous	New	Previous	New
$\Theta(N^q)$, $0 \leqslant q < 1/2$	$O(N)$ [12]	$O(N)$	$O(N^{1/3})$ [12]	$O(N^{1/3})$
Planar	$O(N\log^{3/2}N)$ [12]	$O(N)$	$O(N^{1/3}\sqrt{\log N})$ [12]	$O(N^{1/3})$
$\Theta(\sqrt{N})$	$O(N\log^{3/2}N)$ [12]	$O(N)$	$O(N^{1/3}\sqrt{\log N})$ [12]	$O(N^{1/3})$
$\Theta(N^q)$, $1/2 < q < 2/3$	$O(N^{q+1/2}\log^{3/2}N)$ [12]	$O(N)$	$O(N^{q/3+1/6}\sqrt{\log N})$ [12]	$O(N^{1/3})$
$\Theta(N^{2/3})$	$O(N^{7/6}\log^{3/2}N)$ [12]	$O(N\log^3 N)$	$O(N^{7/18}\sqrt{\log N})$ [12]	$O(N^{1/3}\log N)$
$\Theta(N^q)$, $2/3 < q < 1$	$O(N^{q+1/2}\log^{3/2}N)$ [12]	$O(N^{3q/2})$	$O(N^{q/3+1/6}\sqrt{\log N})$ [12]	$O(N^{q/2})$

Table 4. Upper-bound volume and maximum wire length for the 3-D unrestricted layouts. (The previous and new results may be the same but obtained by a different method.)

5. OPTIMIZATION OF TOTAL COST OF IMPLEMENTATION

Up to this point, we have considered the volume occupied by the components. In general, the volume is not directly related to the cost of implementation because the cost of running a wire or placing a device in a 3-D volume may depend on its location in the chip. In contrast, in a 2-D implementation, the cost of running a wire or placing a device is independent of its location, and, hence, the area is related to the cost of implementation by a constant. To compare the trade-off between 2-D and 3-D implementations, the criterion used must be based on costs.

In this section, we assume the cost as a function of the layer in the 3-D chip and minimize the total cost instead of volume in a one-active-layer layout. In the following discussion, planar graphs are treated in the same way as graphs with $\Theta(\sqrt{N})$ separator, although planar graphs may have separators less than $\Theta(\sqrt{N})$.

Consider the problem to lay out a graph with $\Theta(N^q)$ separator, $0 \leqslant q < 1/2$. Since the complexities of a 3-D layout is $\Theta(N)$, which is the same as that of a 2-D layout [10], the cost of a 3-D implementation will differ from that of a 2-D implementation by a constant factor.

To lay out a graph with $\Theta(N^q)$, $1/2 \leqslant q \leqslant 1$, separator, Leiserson has proved that the area required for a 2-D implementation is [13]

$$A = \alpha N D^2(N),$$ (21)

where

$$D(N) = \sum_{i=0}^{\log_2 N} \left[\frac{N}{2^i} \right]^{q-1/2}$$ (22)

α is a graph-dependent constant, and N is a power of 2. The i'th term in the summation in Eq. (22), $(N/2^i)^{q-1/2}$, is the increase in area to connect two $(N/2^i)$-node subgraphs together in a 2-D implementation. Recall from Section 4.1 that $2(N/2^i)^{q-1/2}$ layers are needed in a 3-D implementation to connect two $(N/2^i)$-node subgraphs together (one layer will be needed if $q = 0.5$). Hence, if the interconnections of a set of subgraphs are implemented in the third dimension, then the corresponding terms in Eq. (22) should be eliminated in computing the area of the base, and the height of the chip will be increased by the sum of the terms eliminated. Note that this is an upper bound on the number of layers, since we are assuming a crossbar connection to perform the routing. Suppose that h layers are used in the third dimension, then a number of terms will be chosen from Eq. (22) to sum up to h. The terms chosen will depend on the graph concerned and the cost of implementation in the third dimension. The area of the base of the layout using h layers is

$$B_h = \alpha N [D(N) - h + 1]^2 \qquad 1 \leqslant h \leqslant H,$$ (23)

where H is the maximum height. The maximum height can be computed from Eq's (5) and (9).

$$H = \begin{cases} \log N & \text{for graphs with } \Theta(\sqrt{N}) \text{ separator} \\ \beta N^{q-1/2} & \text{for graphs with } \Theta(N^q), 1/2 < q \leqslant 1, \text{ separator} \end{cases}$$ (24)

where β is a graph-dependent constant. The volume of this chip is

$$V_h(N) = B_h h = \alpha N [D(N) - h + 1]^2 h.$$ (25)

Let $c(j)$ to be the cost of implementing layer j, then the total cost of using h layers is

$$cost_h = \alpha N [D(N) - h + 1]^2 \sum_{j=1}^{h} c(j) \qquad 1 \leqslant h \leqslant H. \tag{26}$$

Hence, to minimize the total cost of implementation, it is necessary to solve the following optimization problem.

$$\min_{1 \leqslant h \leqslant H} cost_h = \min_{1 \leqslant h \leqslant H} \left\{ \alpha N [D(N) - h + 1]^2 \sum_{j=1}^{h} c(j) \right\} \tag{27}$$

As an illustration, if $c(j)$ is a constant independent of j, then the cost of the circuit is the same as its volume. h should be set as $D(N)$ to minimize the volume, and the maximum number of layers will be used.

6. CONCLUSIONS

In this paper, we have proved improved lower and upper bounds on volume and maximum wire length in both the one-active-layer and unrestricted layouts. From the previously known bounds and our improved bounds, optimal complexities on the volume of layout have been found except for two cases: the layout of planar graphs in the one-active-layer model and the layout of graphs with $\Theta(N^{2/3})$ separator in the unrestricted model. To compare between 2-D and 3-D implementations, we have proposed a simple model to compute the total cost of layout. Table 5 summarizes the results obtained in this paper.

REFERENCES

[1] S. N. Bhatt and F. T. Leighton, "A Framework for Solving VLSI Graph Layout Problems," *J. of Computer and System Sciences*, vol. 28, pp. 300-343, 1984.

[2] A. J. Blodgett and D. R. Barbour, "Thermal Conduction Module: A High-Performance Multilayer Ceramic Package," *IBM J. of Research and Development*, vol. 26, no. 1, pp. 30-36, Jan. 1982.

[3] J. F. Gibbons, "SOI—A Candidate for VLSI?," *VLSI Design*, vol. 3, pp. 54-55, 1982.

[4] J. Grinberg, G. R. Nudd, and R. D. Etchells, "A Cellular VLSI Architecture," *Computer*, vol. 17, no. 1, pp. 69-81, IEEE, Jan. 1984.

[5] H. T. Kung, "Highly Concurrent Systems," in *Introduction to VLSI Systems*, ed. C. A. Mead and L. A. Conway, Addison-Wesley, 1980.

[6] H. W. Lam, A. F. Tasch Jr., and T. C. Holloway, "Characteristics of MOSFETs Fabricated in Laser-Recrystallized Polysilicon Islands with a Retaining Wall Structure on an Insulating Substrate," *IEEE Electron Dev. Letters*, vol. EDL-1, pp. 206-208, 1980.

[7] F. T. Leighton, "New Lower Bounds Techniques for VLSI," *Proc. 22nd Annual IEEE Symposium on Foundation of Computer Science*, pp. 1-12, Oct. 1981.

[8] F. T. Leighton and A. L. Rosenberg, "Automatic Generation of Three-Dimensional Circuit Layouts," *IEEE Int'l Conf. on Computer Design: VLSI in Computers*, pp. 633-636, 1983.

[9] F. T. Leighton, *Complexity Issues in VLSI*, Foundations in Computing Series, M.I.T. Press, Cambridge, Mass., 1983.

[18] A. L. Rosenberg, "Three-dimensional Integrated Circuits," in *VLSI Systems and Computations*, ed. H. T. Kung, B. Sproull, and G. Steele, pp. 69-80, Computer Science Press, Rockville, MD., 1981.

[19] A. L. Rosenberg, "Three-Dimensional VLSI: A Case Study," *J. of the ACM*, vol. 30, no. 3, pp. 397-416, July 1983.

[20] C. D. Thompson, A Complexity Theory for VLSI, Ph.D. Dissertation, Dept. Computer Science, Carnegie-Mellon University, Pittsburgh, PA, 1980.

[21] Z. A. Weinberg, "Polysilicon Recrystallization by $CO2$ Laser Heating of $SiO2$," Technical Report RC-8835, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1981.

[22] D.S. Wise, "Compact Layouts of Banyan/FFT Networks," in *VLSI Systems and Computations*, ed. H.T. Kung, B. Sproull, and G. Steele, Computer Science Press, 1981.