PROCEEDINGS
NINETEENTH ANNUAL ALLERTON CONFERENCE
ON COMMUNICATION, CONTROL, AND COMPUTING

W. K. Jenkins
D. V. Sarwate
Conference Co-Chairmen

Conference held
September 30-October 2, 1981
Allerton House
Monticello, Illinois

Sponsored by
The Department of Electrical Engineering
and
The Coordinated Science Laboratory
of the
UNIVERSITY OF ILLINOIS
at
Urbana-Champaign
I/O LIMITATIONS ON MULTI-CHIP VLSI SYSTEMS
Extended Abstract
by
Marc Snir
Computer Science Department
New York University
251 Mercer Street, New York, N. Y. 10012

0. ABSTRACT

Decompositions of regular interconnection networks used for multiprocessing are investigated. Lower bounds on the number of components as a function of the number of connections between components are given.

1. INTRODUCTION

The advances in VLSI technology have fostered interest in new computer architectures suited to this technology. In particular, much research has been devoted to the design of multiprocessor systems consisting of large, regular networks connecting identical processors. The same approach has been followed in the design of special-purpose VLSI systems that implement parallel algorithms in hardware. Among the interconnection networks that have been considered are mesh-connected, hexagonally-connected, tree and shuffle exchange networks [2,4]. These are single-stage interconnection networks, with processors at the nodes being directly connected by lines. Alternatively, one may use multistage networks consisting of interconnected switches. A ubiquitous network of this type is the Ω-network [3].

The complexity of VLSI systems is usually measured in area. This measure takes into account the area required for the connecting wires as well as the area required for the computation nodes. For each of the networks mentioned above area-optimal layouts are known. These layouts are obtained under the assumption that the entire network is embedded on one chip. In practice, large multiprocessor systems are built today from tens of thousands of discrete components, and taking into account the human inclination to stretch any technology to its limits, we may safely forecast that large multiprocessor machines will be multi-chip systems in the foreseeable future.

When a network is partitioned into several chips we have to take into account the additional restrictions due to pin limitations. The limited bandwidth available for off-chip communication is a major constraint in VLSI environment, which seems to get worse with increased miniaturization. It is therefore important to find strategies for partitioning networks such that the number of connections between
distinct components is minimized. We shall study in this paper the existence of such partitions.

2. DEFINITIONS

Let \( G = (V,E) \) be an (undirected) graph and let \( \Pi = \{V_1, \ldots, V_k\} \) be a partition of the nodes of \( G \). We define \( b(V_i) \), the boundary of \( V_i \), to be the set of edges connecting nodes from \( V_i \) to nodes outside \( V_i \). The size of the partition \( \Pi \) is defined to be the number \( s(\Pi) = k \) of subsets. The weight of the partition \( \Pi \) is defined to be the maximal number \( w(\Pi) \) of nodes in a subset \( V_i \) of the partition; the girth of the partition \( \Pi \) is defined as \( g(\Pi) = \max \#b(V_i) \).

An implementation of a network on several chips defines a partition of the corresponding graph. The chip area restricts the maximal number of nodes that can be implemented on one chip, that is the weight of the partition; the pin limitation restricts the maximum number of lines connecting nodes on one chip with nodes from other chips, that is the girth of the partition. We attempt to minimize the number of distinct components, that is the size of the partition, under these two constraints. We shall investigate the non-trivial cases where the size of the partition is restricted by the partition girth rather than the partition weight.

3. MESH-CONNECTED NETWORKS

A \( d \)-dimensional mesh-connected network of size \( N^d \) consists of the nodes \( \langle n_1, \ldots, n_d \rangle, 1 < n_i < N \). The node \( \langle n_1, \ldots, n_d \rangle \) is connected to the nodes \( \langle n_1, \ldots, n_i+1, \ldots, n_d \rangle, 1 < j < d \), provided such nodes exist. Each node has at most \( 2d \) neighbours. Of particular interest are \( 2 \)-dimensional grids which have simple planar layouts. With such layout the girth of a set of nodes is related to the perimeter of a surface, whereas the number of nodes in the set is related to its area. We can therefore bound the the maximal size of a set as a function of its girth.

We recall the following facts from planar geometry.

**FACT 1.** The maximal area of a surface circumscribed by a curve of fixed length is achieved by a circle.

**FACT 2.** The maximal area of a surface circumscribed by a straight segment and a curve of fixed length is achieved by a half-circle.

**FACT 3.** The maximal area of a surface circumscribed by two orthogonal segments and a curve of fixed length is achieved by a quadrant.

**THEOREM 3.1.** Let \( M \) be a \( 2 \)-dimensional mesh-connected network of size \( N^2 \). Then for any subset \( V \) of nodes from \( M \) with \( \#V < N^2/2 \) we have \( \#V = O(\#b(V)^2) \).
PROOF: Let \( b = \#b(V) \) and \( v = \#v \). We can assume w.l.o.g. that \( b < N \), the theorem being otherwise trivial. We represent the network by a square grid, with nodes at unit distances. We associate to each node \( P \) in \( V \) a unit square centered at \( P \). The union of these squares forms a surface \( S \) with area \( v \) and perimeter \( b \) (not counting that part of the perimeter provided by the boundary of the grid—see Fig. 1). Assume that the surface \( S \) is the union of connected components \( S_i \), each with area \( v_i \) and perimeter \( b_i \), so that \( v = \sum v_i \) and \( b = \sum b_i \).

As \( b < N \) and \( \#v < N^2/2 \) no connected component of \( S \) can touch more than two of the sides of the surrounding square. It follows therefore that the area of \( S_i \) is no larger than the area of a quadrant bounded by an arc of length \( b \), that is \( v_i < b_i^2/\pi \). It follows that

\[
v = \sum v_i < \frac{1}{\pi} \sum b_i^2 < \frac{1}{\pi} (\sum b_i^2)^2 = b^2/\pi.
QED
\]

**COROLLARY 3.2.** For any partition \( \Pi \) of a 2-dimensional mesh-connected network of size \( N^2 \) we have \( s(\Pi) = O([N^2/g(\Pi)]^2) \).

The same argument can be carried through for \( d \)-dimensional mesh-connected networks, \( d > 2 \), by considering regular grids in \( d \)-dimensional space. We obtain:

**THEOREM 2.3.** Let \( M \) be a \( d \)-dimensional mesh-connected network of size \( N^d \). Then

1. For any set \( V \) of nodes from \( M \) we have \( \#v = O(\#b(V)^d/(d-1)) \).
2. For any partition \( \Pi \) of \( M \) we have \( s(\Pi) = \Omega(N^d/g(\Pi)^d/(d-1)) \).

Note that the constants implicit in the \( O \) notation depend on \( d \). The same lower bound argument is valid for hexagonally connected networks, and for any network that can be embedded in \( d \)-dimensional space such that the distance between two nodes is at least 1, and the length of each edge is at most \( c \), for some constant \( c \) which is independent of the network size. These lower bounds can be matched in each case by similar upper bounds, by taking the obvious decomposition into subsquares (sub-hypercubes).

4. REGULAR TREES

A regular tree of degree \( k \) and depth \( d \) consists of the nodes \( \langle a_1 \ldots a_i \rangle \), where \( 0 < i < d \) and \( 1 < a_i < k \). The node \( \langle a_1 \ldots a_i \rangle \) is connected to the node \( \langle a_1 \ldots a_{i-1} \rangle \) (its parent) and to the nodes \( \langle a_{i+1}, a_{i}, a_{i-1} \rangle \) (its children), if such exist. The tree contains \((k^{i-1}-1)/(k-1)\) nodes and each node has at most \( k+1 \) neighbours.

**LEMMA 4.1.** A regular tree with \( N \) nodes can be partitioned into a partition of weight \( \langle w, g \rangle \) and size \( O(N/w) \), for any \( g > 2 \).

**PROOF.** Let \( k \) be the degree of the tree. We assume w.l.o.g. that \( w = (k^{r+1})/(k-1) \). The \( r \) bottom levels of the tree can be partitioned into \((k^{r+1})/(k-1)\) subtrees of weight \( w \). The remaining \((N-w)/(k-1)w+1)\) nodes can be partitioned into \( O(N/kgw) \) subtrees with \( g-1 \) outgoing edges at the leaves and one outgoing edge at the root (see Fig. 2).

QED
The last result is asymptotically optimal. It implies that girth restrictions do not affect significantly the number of components in a partition of a tree. This is true since most of the nodes of a tree are found at the bottom levels and these levels can be efficiently packed into subtrees with only one outgoing connection. The situation is completely different if we are looking for a partition of the internal nodes of the tree, leaves not included. Such a situation occurs in practice, where the tree network is a connection network with simple switches at the internal nodes, whereas the leaves are more complex processors or memory banks. We have:

**LEMMA 4.2.** Let \( V \) be a set of internal nodes in a tree of degree \( k \). Then \( \#b(V) > (k-1)\cdot\#V \).

**PROOF.** The number \( e \) of leaves in a tree of degree \( k \) is related to the number \( n \) of internal nodes by the relation \( e = n(k-1)+1 \). Each connected component \( C \) of \( V \) is a subtree of degree \( k \). If \( C \) has \( c \) nodes then it has at least \( c(k-1)+1 \) outgoing edges. The claim now follows. QED

**COROLLARY 4.3.** Let \( T \) be a tree of degree \( k \) with \( N \) internal nodes. For any partition \( \Pi \) of the internal nodes of \( T \) we have \( s(\Pi) > N(k-1)/(g(\Pi)-1) \).

Thus, for a partition of the internal nodes of a tree, the limiting factor is the girth rather than the weight.

5. \( \Omega \)-NETWORKS

\( \Omega \)-networks have been introduced by Lawrie [3] as interconnection networks for parallel processing. Many other networks considered in the literature are topologically equivalent to \( \Omega \)-networks. This includes the Flip (Staran) network, the indirect binary cube network, the baseline network, and the SW-banyan network with 2×2 switches [4]. Our results are valid for each of them. They can be also applied to Benes networks which essentially consist of two \( \Omega \)-networks joined end to end.

We represent for convenience \( \Omega \)-networks as directed graphs. The network \( \Omega_n \) has \( 2^n \) inputs \( I[a_1\ldots a_n] \), \( 2^n \) outputs \( 0[a_1\ldots a_n] \), and \( n \) columns of \( 2^n-1 \) 2×2 switches \( S[a_1\ldots a_{n-1};j] \), where \( a_i=0,1 \) and \( 1\leq j\leq n \). The edges of \( \Omega_n \) connect the input \( I[a_1\ldots a_n] \) to the switch \( S[a_1\ldots a_{n-1};1] \); the switch \( S[a_1\ldots a_{n-1};j] \) to the switches \( S[b,a_1\ldots a_{n-2};j+1], b=0,1 \); and the switch \( S[a_1\ldots a_{n-1};n] \) to the outputs \( 0[a_1\ldots a_{n-1};b], b=0,1 \). The network \( \Omega_n \) is illustrated in Fig. 3. Note that the switches in two adjacent columns of \( \Omega_n \) are connected by the unshuffle connection.

**THEOREM 5.1.** Let \( V \) be a set of switches in \( \Omega_n \) with \( v \) nodes and girth \( g \). Then \( v < g \cdot l_g(g)/4 \).

**PROOF:** Since each switch has the same number of incoming edges as of outgoing edges, the number \( p \) of edges entering \( V \) is equal to the
number of edges leaving $V$. We shall prove by induction on $n$ that $v < p \log(p)/2$.

The claim is true for $\Omega_1$, which contains one switch. Assume it is true for $\Omega_{n-1}$. Note that the graph of $\Omega_n$ contains two isomorphic copies $\Omega^0$ and $\Omega^1$ of $\Omega_{n-1}$; $\Omega^\varepsilon$ consists of the subgraph spanned by the inputs $I[\varepsilon,a_2\ldots a_n]$, the switches $S[a_1\ldots a_{i-1},\varepsilon,a_{i+1}\ldots a_{n-1}; j]$, for $1 \leq j \leq n$, and the switches $S[a_1\ldots a_n;n]$ (see Fig. 3).

Let $V^\varepsilon$ be the subset of nodes of $V$ belonging to $\Omega^\varepsilon$, with the exclusion of switches in the last column of $\Omega_{n}$. Let $V^2$ be the subset of nodes of $V$ in the last column of $\Omega_n$, and let $v^2 = \#V_2$. Let $p^\varepsilon$ be the number of edges entering $V^\varepsilon$, and let $q^\varepsilon$, $\varepsilon = 0,1$, be the number of edges connecting nodes from $\Omega^\varepsilon$ to nodes in $V^2$. The number of edges leaving $V^\varepsilon$, $\varepsilon = 0,1$, and therefore the number of edges connecting nodes from $V^\varepsilon$ to nodes in $V^2$ is at most $p^\varepsilon$. We have therefore

$q^\varepsilon > p^\varepsilon$, for $\varepsilon = 0,1$, and

$q^0 + q^1 < p^0 + p^1 + p^2 = p$.

Each node of $V^2$ is incident with one edge coming from $\Omega^0$ and one edge coming from $\Omega^1$. It follows that

$v^2 < q^\varepsilon$, for $\varepsilon = 0,1$.

By the inductive assumption, we have

$v^\varepsilon < \frac{1}{2} p^\varepsilon \log p^\varepsilon$, for $\varepsilon = 0,1$.

Assume w.l.g that $q^0 < q^1$. Then

$v = v^0 + v^1 + v^2$

$< \frac{1}{2} p^0 \log p^0 + \frac{1}{2} p^1 \log p^1 + q^0$

$< \frac{1}{2} q^0 \log q^0 + \frac{1}{2} q^1 \log q^1 + q^0$

$< \frac{1}{2} (q^0 + q^1) \log (q^0 + q^1) < \frac{1}{2} p \log p$.

QED

A similar result is proven is [1].

**Corollary 5.2.** For any partition $\Pi$ of the switches of the network $\Omega_n$ we have $s(\Pi) = \Omega(n 2^n/(g(\Pi) \log g(\Pi)))$.

This lower bound can be matched by a corresponding upper bound. Indeed, let $k$ divide $n$. We can partition the switches of $\Omega_n$ into $(n/k) 2^{n-k}$ subsets containing $k 2^{k-1}$ switches, each set with $2^k$ incoming lines and $2^k$ outgoing lines. Such set will contain $2^{k-1}$ consecutive switches in one column, and their successors in the next $k-1$ columns — see Fig. 4.

Note that the connections between components will follow the same pattern as in an $\Omega$-network of degree $2^k$: An $\Omega$-network of degree $d$ has
d×d switches; It has the same formal definition as the usual Ω-network, with addresses being given to base d rather than in binary notation.

In the general case (k does not divide n) it is still possible to partition Ωn into no more than \( \lceil n/k \rceil 2^{n-k} \) subsets, each with girth at most \( 2^{k+1} \).

6. SHUFFLE-EXCHANGE NETWORKS

The shuffle-exchange network \( SE_n \) consists of \( 2^n \) nodes \( <a_1...a_n> \) and \( 3 \cdot 2^{n-1} \) edges. Node \( <a_1...a_n> \) is connected to node \( <a_2...a_n,a_1> \) (shuffle connection), node \( <a_n,a_1...a_{n-1}> \) (unshuffle connection), and node \( <a_1...a_n> \) (exchange connection). We have drawn \( SE_3 \) in Fig. 5.

For convenience we define two related networks. The doubly linked shuffle-exchange network \( DSE_n \) is obtained from the shuffle-exchange network \( SE_n \) by replacing each exchange edge with two edges. The four pin shuffle-exchange network \( 4SE_n \) is obtained by merging in \( SE_{n+1} \) pairs of nodes differing in the last address bit only. It consists of \( 2^n \) nodes, each incident to four edges. Node \( <a_1...a_n> \) is connected to the nodes \( <b,a_1...a_{n-1}> \) and \( <a_2...a_n,b> \), for \( b = 0,1 \). We shall assume that the edges of the four pin shuffle exchange-network are directed in the unshuffle direction, i.e. connect \( <a_1...a_n> \) to \( <b,a_1...a_{n-1}> \). This last network is closely related to the Ω-network. It is obtained from an Ω-network by coalescing together switches in successive columns that have the same address within the column. We can therefore apply the results obtained for Ω-networks to four pin shuffle-exchange networks.

THEOREM 6.1. Let \( α, β \) and \( γ \) be constants such that \( α > 0, β < 1 \) and \( γ < 1/2e \). Let \( V \) be a set of nodes in \( 4SE_n \) of size \( w < γ 2^n \) and girth \( g \), where \( n^α < g < 2^m \). Then \( w = O(g lg g) \).

PROOF: Since each node has two incoming edges and two outgoing edges, there are \( g/2 \) edges entering \( V \) and \( g/2 \) edges leaving \( V \). Let \( V' \) be the set of switches in the Ω-network \( Ω_n \) defined by \( V' = \{ S[a_1...a_n,j] : <a_1...a_n> \in V \} \). The set \( V' \) contains \( nw \) nodes and has at most \( 2w+(n-1)g/2 \) incoming edges. It follows, by Th. 5.1, that

\[
\text{nw} < \frac{1}{2} (w+(n-1)g/2) \cdot \text{lg} \left( w+(n-1)g/2 \right). \tag{5.1}
\]

Let

\[
P(w) = \frac{1}{2} (w+(n-1)g/2) \cdot \text{lg} \left( w+(n-1)g/2 \right) - nw. \tag{5.2}
\]

\( P(w) \) is decreasing for

\[
w < w_{\text{min}} = \left( 2^{n-1}/e - (n-1)g/4, \right. \tag{5.3}
\]

increasing afterward, and it has at \( w_{\text{min}} \) a minimum value of

\[
P(w_{\text{min}}) = n(n-1)g/2 - 2^n/e 2n. \tag{5.4}
\]

229
From the choice of $\beta$ and $\gamma$ it follows that for $n$ large enough $w_{\text{min}} > \gamma^{2^n}$ and $F(w_{\text{min}}) < 0$. It follows that $F(w)$ has a unique root $w_0$ in the interval $[0, w_{\text{min}}]$, and that inequality (5.1) is satisfied only if $w < w_0$. We shall now estimate the root $w_0$.

Let

$$w_0 = \frac{1}{4}(n-1)g(n/\gamma - 1). \quad (5.5)$$

Substituting for $w_0$ in the equation $F(w_0) = 0$, we obtain

$$\gamma - 1g \gamma = n - 1g \left(n(n-1)/2\right). \quad (5.6)$$

Let $C = n - 1g \left(n(n-1)/2\right)$. For $n$ large enough $C > 1$.

It is easy to see now that the iteration

$$y_0 = C,$$

$$y_{n+1} = C + 1g y_n,$$

produces a sequence that converges to the root $\gamma$ of the equation (5.6). Also, since $C > 1$, it follows by induction that $y_n > C$, and therefore, that $\gamma > C$. Substituting back, we obtain that

$$w_0 < \frac{1}{4}(n-1)g(n/C - 1)$$

$$= \frac{(n-1)(n-C)g}{4C}$$

$$= \frac{(n-1)(1g n + 1g(n-1) + 1g (n-1)}{4(n - 1g n - 1g(n-1) - 1g (n-1) + 1)}$$

$$< \frac{(n-1)(1g n + 1g(n-1) + 1g (n-1)}{4((1-\beta)n - 1g n - 1g(n-1) + 1)}$$

$$= 0(g 1g n)$$

QED

CORROLARY 6.2. The claim of theorem 6.1 is valid for subsets of (i) the doubly linked shuffle-exchange network, and (ii) of the shuffle exchange network.

PROOF: (i) Let $V$ be a set of nodes in $DSE_n$. Assume that the node $<a_1...a_n>$ is in $V$ whereas the node $<a_1...a_n>$ is not in $V$. Then the set $V' = V + \{<a_1...a_n>\}$ contains one more node and no more outgoing edges than $V$. It follows that we can assume w.l.g. that $<a_1...a_{n-1}0> \in V$ iff $<a_1...a_{n-1}1> \in V$. The claim now follows immediately from Th. 6.1.

(ii) If $V$ is a set of weight $w$ and girth $g$ in $SE_n$, then the corresponding set of nodes in $DSE_n$ has weight $w$ and girth $\lesssim 2g$. The claim follows therefore from (i).

QED

As any fixed permutation can be performed on the shuffle exchange network $SE_n$ in $O(n)$ routing steps [5], we must have for any set $V$ of nodes in $SE_n$ $n#(V) = \Omega(#V, 2^n - #V)$. This yields a new proof of the inequality in Th. 6.1, for large $g$ and $v$. Some of the restrictions in this theorem can be relaxed, yielding
COROLARY 6.3. Let \( \alpha \) and \( \beta \) be constants such that \( \alpha > 0 \) and \( \beta < 1 \). Let \( V \) be a set of nodes in \( SE_n \) (DSE\(_n\), 4SE\(_n\)) of size \( w < \beta 2^n \) and girth \( g \geq n^\alpha \). Then \( w = O(g \lg g) \).

COROLARY 6.4. Let \( \Pi \) be a partition of the shuffle-exchange network \( SE_n \) (DSE\(_n\), 4SE\(_n\)) of weight \( w \) and girth \( g \), where \( w < \beta 2^n \) and \( n^\alpha < g \), for \( \alpha > 0 \), and \( \beta < 1 \). Then \( s(\Pi) = \Omega(2^n / g \lg g) \).

The lower bounds given in Cor. 6.3 are the best possible of this form: For any \( k < n \) it is possible to build in \( SE_n \) a set of nodes of size \( O(2^k) \) and girth \( \Theta(k 2^k) \). Nevertheless, it is not obvious that the lower bound of Cor. 6.4 can be matched by a corresponding upper bound.

7. CONCLUSION

Our results show that in order to take full advantage of the increasing gate count \( G \) of chips the pin count \( P \) must be increased at least as \( \sqrt{G} \) for architectures using 2D mesh-connected networks, as \( G \) for tree-like communication networks, as \( G / \lg G \) for \( \Omega \)-networks and shuffle exchange networks. The past trend seems to indicate a relation of the form \( P = \alpha G^\beta \), with \( \beta < 1/2 \). If this can be extrapolated, then the performance of any of the above structures will be restricted by the pin count of its components rather than by their gate count.

ACKNOWLEDGMENTS

I wish to thank Larry Rudolph and Chee Yap for their helpful comments.

REFERENCES

Figure 1
Subset of 2D grid

Figure 2
Partition of a binary tree
Figure 3
Ω₃

Figure 4
Partition of Ω₄

Figure 5
SE₃