Efficient VLSI Layout of Grid Pyramid Networks

Saeedeh Bakhshi1,2, Marzieh Bakhshi1 and Hamid Sarbazi-Azad1,2

1Sharif University of Technology, Tehran, Iran
2IPM School of Computer Science, Tehran, Iran
E-mail: bakhshi@ipm.ir, m_bakhshi@ce.sharif.edu, azad@ipm.ir

ABSTRACT
Reducing the VLSI layout area of on-chip networks can result in lower costs and better performance. Those layouts that are more compact can result in shorter wires and therefore the signal propagation through the wires will take place in less time. The grid-pyramid network is a generalized pyramid network based on a general 2D Grid structure (such as mesh, torus, hypermesh or WK-recursive mesh). Such pyramid networks form a wide class of interconnection networks that possess rich topological properties. In this paper, we study these topologies from the VLSI-layout efficiency point of view. Also, we investigated on the layout of RTCC-pyramid networks that we believe can be considered in the class of Grid-pyramid networks.

Keywords: VLSI layout, Grid Pyramid networks

1. INTRODUCTION
The number of transistors per chip has vastly increased in recent years and will likely increase by another order of magnitude in the next two decades [8]. These changes made it necessary for chip-multiprocessors and SoCs to use smaller layouts. More compact layouts can lead to shorter wire lengths and therefore reducing signal propagation delay as well as lower cost in the implementation process. The collinear layout that we discuss in this paper can result in better packagability of the network, not only in one-layer chips, but also in today’s technology for multi-level chip designs.

When the same physical chip area is used and the network is wire-bounded, a network layout with smaller layout area (when unit link width is assumed) will lead to an implementation with faster communication links. More precisely, a layout with smaller area by a factor of \( f \) (in terms of unit of area where a unit of length is taken as the physical width of a link) will lead to communication links whose bandwidth is higher than that of another layout with larger area by a factor of \( \sqrt{f} \), since more wires can be used to implement a single communication link using the former layout. Thus, the impact of efficient VLSI layout on cost and performance become more crucial [10].

The grid-pyramid networks were first introduced in [2] as a generalized pyramid topology. Various topological properties of these networks were studied in [2]. Here, we study the packagibility and implementation properties of this family of networks. As we see in the following sections this class of networks are so appropriate in VLSI implementation according to their cost-effective layout area and the regular topologies they have. We selected these class of networks to show how efficient they can be placed on chips.

In section 2, we formally state some definitions, which are used throughout the paper; collinear layouts...
2. DEFINITIONS AND PRELIMINARIES

Definition 1. An $a \times b$ mesh network, $M_{a,b}$, is a set of nodes $V(M_{a,b}) = \{(x, y) | 1 \leq x \leq a, 1 \leq y \leq b\}$, where nodes $(x_1, y_1)$ and $(x_2, y_2)$ are connected by an edge iff $|x_1 - x_2| + |y_1 - y_2| = 1$. See Fig. 1 (a).

Definition 2. The $a \times b$ torus network, denoted as $T_{a,b}$, consists of a set of nodes $V(T_{a,b}) = \{(x, y) | 1 \leq x \leq a, 1 \leq y \leq b\}$, where each node $(x_1, y_1)$ is connected to its four neighboring nodes $(x_1 \pm 1 \text{ mod } a, y_1)$ and $(x_1, y_1 \pm 1 \text{ mod } b)$. See Fig. 1 (b).

Definition 3. An $a \times b$ hypermesh, $H_{a,b}$, is a set of nodes $V(H_{a,b}) = \{(x, y) | 1 \leq x \leq a, 1 \leq y \leq b\}$ where nodes $(x_1, y_1)$ and $(x_2, y_2)$ are connected by an edge iff $x_1 = x_2$ or $y_1 = y_2$. See Fig. 1 (d).

Definition 4. A $t$-level WK-recursive mesh network $WK(d, t)$ with amplitude $d$ and expansion level $t$, consists node set $V(WK(d, t)) = \{a_1 a_2 \ldots a_t | 0 \leq a_1 < t\}$. The node with address $A = (a_1 a_2 \ldots a_t)$ is connected to:

1. all the nodes with addresses $(a_1 a_2 \ldots a_k)$, $0 \leq k < t$, $k \neq a_i$, as sister nodes, and
2. node $(a_1 a_2 \ldots a_j, a_{j+1})$, if for one $j$, $1 \leq j < t; a_{j+1} = a_{j-1} = \ldots = a_1$ and $a_j \neq a_{j+1}$, as a cousin node. See Fig. 1.

Definition 5. A grid-pyramid of $n$ levels, denoted by $P_{G,n}$, consists of a set of nodes $V(P_{G,n}) = \{(k, x, y) | 0 \leq k \leq n, 1 \leq x, y \leq 2k\}$. A node $(k, x, y) \in V(P_{G,n})$ is said to be node at level $k$. All the nodes in level $k$ form a $2^k \times 2^k$ grid network $G$ which can be one of the grid-based networks: mesh, torus, hypermesh or WK-recursive mesh, i.e. $(G \in M, T, HM, WK)$; the resulted pyramid can then be then denoted as $P_{G,n}$, $P_{T,n}$, $P_{HM,n}$, and $P_{WK,n}$, respectively.

Fig. 2 shows all grid pyramid networks.

Definition 6. A VLSI layout is called collinear if all the nodes are placed along a straight line [11].

Definition 7. A C-node cycle consists of a set of nodes $\{0, 1, 2, 3, \ldots, C-1\}$ and set of edges $\{e_0, e_1, e_2, \ldots, e_{C-1}\}$ such that $e_i = (i, i+1 \text{ mod } C)$. 

![Fig. 1: The topologies of (a)M4x4 (b)T4x4 (c)WK4x2 (d)HM4x4](image)

![Fig. 2: The topologies of (a)PM2 (b)PT2 (c)PWK2 (d)PHM4](image)
Definition 8. The definition of the RTCC(C, L) is based on a C-node cycle. We name all the nodes in this cycle, 'Extern nodes' or 'Open nodes'. An RTCC(C, 2), consists of a number of C discrete RTCC(C, 1) networks, or C-node cycles, numbered 0 to C-1. Each external node i of each C-node cycle j is connected to node j of C-node cycle i. It is obvious that a node whose number is equal to the number of the RTCC(C, 2) in which it resides, is not directly connected to any other cycle, and is of node degree one less than other nodes in the network. There is one such node in each RTCC(C, 1) used to construct an RTCC(C, 2), and thus a total of C such nodes. We name these nodes as the external nodes of the RTCC(C, 2), the number of each one being equal to the number of the RTCC network at level k > 0, to node \((k, a, a, \ldots, a)\), as the neighbouring brother nodes, and connected to a node with address schema \((k, a, a, \ldots, a)\), 1 ≤ j ≤ L-1 if there exists one j such that 1 ≤ j ≤ L-1, ai = aj = ... = aj, and aj ≠ aj; as a cousin node (nodes at the same level). This node is also connected to nodes \((k+1, a, a, \ldots, a, b)\), for 1 ≤ b ≤ C, in level k+1, as a child node, and connected to node \((k-1, a, a, \ldots, a)\), in level k-1, as a father node. Fig. 4 illustrates an RTCC-pyramid network, P-RTCC(C, 2) [19].

All the nodes in level k form an RTCC(C, L) network. Hence, there exist a total of \(N = \sum_{k=0}^{L} C^k = (C^{L-1}+1)/C-1\) nodes in a P-RTCC(C, L). A node with address \((k, a, a, \ldots, a)\) is connected, within the RTCC network at level k > 0, to node \((k, a, a, \ldots, a)\), as the neighbouring brother nodes, and connected to a node with address schema \((k, a, a, \ldots, a)\), 1 ≤ j ≤ L-1 if there exists one j such that 1 ≤ j ≤ L-1, ai = aj = ... = aj, and aj ≠ aj;

Definition 9. An RTCC-pyramid network, denoted as P-RTCC(C, L), consists of a set of nodes \(V(P-RTCC(C, L)) = \{ (k, a, a, \ldots, a) | 0 ≤ k ≤ L, 0 ≤ a ≤ C-1, 1 ≤ i ≤ k or k = 0 and a_i = 1 \} \). A node with addressing scheme \((k, a, a, \ldots, a)\) is said to be a node at level k, e.g. the apex is at level 0. The part \(a, a, \ldots, a\) of the address determines the address of a node within the RTCC network at layer k.

3. THOMPSON GRID MODEL

Thompson proposed a mathematical model for VLSI computations which is widely accepted and is known as the Thompson grid model [7]. In this model, he presumed the chip consists of some vertical and horizontal tracks which are spaced apart at unit intervals. Two layers of interconnect are used to route the wires. Vertical wires are routed in one layer while horizontal wires are routed in the other. The circuit is viewed as a graph G in which vertices correspond to processing elements and edges to wires.

Wires can cross each other but cannot overlap with each other. To change direction, wires may turn into the other layer by vias. The graph is then
embedded in a two-dimensional grid. An embedding of a graph $G$ in a Thompson grid is an assignment of nodes of $G$ to intersection points in the grid and the edges of $G$ to paths along the grid tracks. The area of a layout is the area of the smallest upright rectangle that contains all the nodes and wires. When there are two layers of wires, it is guaranteed that we can lay out the network within the area. The maximum wire length is the length of the longest wire in the layout.

4. LAYOUT OF GRID-PYRAMID

In this section, we present a method for laying out $P_{G,n}$ network, on chip. We use a recursive method based on layout of the base network for constructing the layout.

First, we assume that we have an efficient layout scheme for the base network.

To find the collinear layout complexity of $P_{G,n}$ we use a recursive process. Let the number of tracks which are needed by $P_{G,n}$ be $T(P_{G,n})$. Consider the case where there is just one node in $P_{G,n}$ and no links. This is the case of $P_{G,0}$ for which we have $T(P_{G,0}) = 1$, because it is an isolated node that can be put in one square.

For $P_{G,n}$, we need to lay $P_{G,n-1}$ out and add a $G_{2^n}$ with $4N(G_{2^n-1},2^n-1)$ edges connected the base network to the last layer of $P_{G,n-1}$. Hence, the layout of $P_{G,n}$ requires $T(P_{G,n}) = 4N(G_{2^n-1},2^n-1) + T(G_{2^n},2^n)$ tracks.

4.1 Mesh-Pyramid layout

Now, let us investigate the layout of the mesh-pyramid. For this purpose it is needed to find the number of tracks required by mesh networks in their layout. We can put each row of the mesh in a track, then all links connecting those nodes in the same track. Then, we should put $n$ copies of such a row in a grid area to obtain an efficient layout of $n^2$-area for the $M_{n,n}$.

Fig.5 shows the collinear layout of $M_{4,4}$. For building Mesh-Pyramid network we can use the same layout mechanism.

4.2 Torus-Pyramid layout

Now that we have obtained the number of tracks and the area needed by mesh-pyramid to lay out on a chip, we use the same approach to describe layout of the torus-pyramid.

Yeh and Parhami in [4] described a method for laying out k-ary n-cubes. They used a bottom-up approach, starting with a k-node ring. Using their method, we need $2n + 2$ tracks and $n^2 + O(n^2)$-area for the layout of $T_{n,n}$. In Fig. 6 we can see a collinear layout of $T_{4,4}$.
Efficient VLSI Layout of Grid Pyramid Networks

Fig. 7: Collinear layout of T_{4,4}

Similar to that of Mesh-Pyramid, for the PT,n, we need to lay PT,n–1 out and add a T_{2n,2n} with 4N (T_{2n–1,2n–1}) edges connecting the mesh to the last layer of PT,n. Hence, the layout of PT,n needs T(PT,n) = 4N (T_{2n–1,2n–1}) + T(T_{2n,2n}) = 2^{2n} + O(2^{2n}) tracks. Since N = N(PT,n) = (4^n + 1 – 1)/3, we have T(PT,n) = 1.5N + o(N).

In a hypermesh-pyramid, each node is connected to all of nodes in the same row and the same column. Also, each node is adjacent to its father (except for the apex) and four children (except for the last layer nodes), hence the maximum degree of a node in a hypermesh-pyramid equals 2^{n+1} – 1 for n ≥ 2. So, the hypermesh-pyramid can be laid out in the area of 2^{3n+1} + o (2^{3n+1}) which is 6N√N + o (N√N)-area. You can see collinear layout of PHM,2 in Fig. 10.

Fig. 10: Collinear layout of PHM,2

4.3 Hypermesh-Pyramid layout

Layout of the hypermesh and its pyramid network is slightly different to that of the mesh-pyramid and torus-pyramid.

In [5], it is shown that an N-node complete graph can be optimally laid out using \( \left\lceil N^2/4 \right\rceil \) tracks for a collinear layout, and can be laid out in \( N^3/16 + o(N^3) \) area for a 2D layout. Fig. 9 shows the collinear layout of HM_{4x4}.

Fig. 9: Collinear layout of HM_{4x4}

For the layout of a hypermesh-pyramid, we have: T(P_{HM,a}) = 4N (P_{HM,0}) + T(HM_{2^a,2^a}). Since T(HM_{2^a}) = 4, P_{HM,1} needs 8 tracks for its layout. For a P_{HM,n}, we have:

\[
T(P_{HM,n}) = 4N (HM_{2^{n-1},2^{n-1}}) + T (HM_{2^n,2^n}) = 2^{2n} + o (2^{2n}) \text{ tracks.}
\]

Since N = N(P_{HM,n}) = (4^n + 1 – 1)/3, we have T(P_{HM,n}) = 1.5N + o(N).

4.4 WK-Pyramid layout

Finally, we work on the VLSI layout of WK-pyramid networks. In [6], an efficient layout for WK-recursive networks was proposed. The WK(d, t) network requires \( \frac{d^{t+1} - d^t + 2d^{t-1} - 2}{d - 1} \) tracks to lay out collinearly. It means it needs an area of \( N^3/16 + o (N^3) \). The collinear layout of WK_{4,2} is shown in Fig. 11.

Fig. 11: Collinear layout of WK_{4,2}

Here, we use the same scheme to obtain the layout of a WK-pyramid network. To this end, it is enough to lay out WK (4, n). So, the layout of PWK_n needs T(P_{WK,n}) = 4N (WK (4, n – 1)) + T (WK (4,
n) tracks. Since \( N = W K (4, n - 1) = 4^n - 1 \), we have
\[
T(P_{WK,n}) = \frac{13}{6} \times 4^n - \frac{2}{3} .
\]
That means \( P_{WK,n} \) can be laid out in \( 39/2N^2 + O(N) \) area. Collinear layout of \( P_{WK,2} \) is shown in Fig. 12.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig12.png}
\caption{Collinear layout of \( P_{WK,2} \)}
\end{figure}

5. RTCC AND RTCC-PYRAMID LAYOUTS

In this section we investigate on another grid network, which is not one of the networks introduced in [2] for using as a base network while building pyramids, but we think that it has the important characteristics of a grid network. We follow with the two dimensional and multilayer layout of RTCC, and then use them for constructing the RTCC-Pyramid layout.

5.1 Two dimensional layout of RTCC

An RTCC\((C, L)\) contains \( C \) subgraphs isomorphic to RTCC\((C, L-1)\), each pair of which are connected by 1 link. If we let RTCC\((C, L-1)\) as a supernode, then the RTCC\((C, L)\) becomes a C-supernode complete graph.

To lay out an RTCC\((C, L)\), we first place nodes belonging to each RTCC\((C, L-1)\) into a block which we call \((L-1)\)-block, and lay out the RTCC\((C, L)\) using that of complete graph, stated in [5]. Then we should continue this process recursively until the remainder graph is the C-ring and then we can simply lay out the ring.

To lay out RTCC\((C, L)\), base on the layout of the complete graph [5], the area of \( N^2 / 16 + o(N^2) \) will be required. For level \( h \)-block we first connect the wires outside the block to appropriate the \((h-1)\)-blocks within it, and then follow the technique of laying out the \( K_n \) graph. Hence, the layout area will be obtained from:

\[
\frac{C^{2L}}{16} + o(C^{2L}) + \frac{C^{2(L-1)}}{16} + o(C^{2(L-1)}) + \ldots + 1 = \frac{C^{2L}}{16} + o(C^{2L})
\]

and so an RTCC\((C, L)\) can be laid out in \( N^2/16 + o(N^2) \), where \( N = C^L \). The collinear layout of RTCC\((4, 2)\) is shown in Fig. 13.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig13.png}
\caption{Collinear layout of RTCC\((4,2)\)}
\end{figure}

5.2 Multilayer layout of RTCC

In order to obtain the multilayer layout of RTCC\((C, L)\) we use the same bottom-up approach, used in [4] for k-ary n-cube and several other networks.

A collinear layout of a ring can be obtained by placing the \( C \) nodes along a row and connecting neighboring nodes in one track and then connecting node 0 to \( C-1 \), in another track [4]. Let the number of tracks which are needed by RTCC\((C, L-1)\) be \( f_C(L-1) \). To obtain the collinear layout of RTCC\((C, L)\) we need \( C \) copies of RTCC\((C, L-1)\) which are placed in a space, horizontally \( C \) times of the space of RTCC\((C, L-1)\) and two extra tracks to connect the level \( L \) inter-block links. Therefore, the number of tracks needed to layout the RTCC\((C, L)\) is:

\[
f_C(L) = Cf_C(L-1) + 2 = C^L - 1 + \ldots + C + 1
\]

Now, we use the approach of orthogonal multilayer layout [4] scheme to obtain an S-layer layout.

The number of tracks per layer above a row is

\[
\left[ \frac{2\left(C^{L/2}-1\right)}{C-1} \right]
\]

and the number of tracks per layer to the right of the column is

\[
\left[ \frac{2\left(C^{L/2}-1\right)}{S/2\left(C-1\right)} \right]
\]

So the area of the S-layer RTCC\((C, L)\) is:

\[
\left[ \frac{2\left(C^{L/2}-1\right)}{S/2\left(C-1\right)} \right]
\]
6. CONCLUSION

In this paper, we have investigated on the area of layouts for grid-pyramid networks, including the mesh-pyramid, torus-pyramid, hypermesh-pyramid and WK-pyramid. Also we added RTCC as base network for grid pyramid networks and studied RTCC-Pyramid’s layout on chip. We have obtained the number of tracks required for collinear layout of these networks as well as the area required for them on a chip. We showed that a mesh-pyramid (or torus-pyramid) requires $13.5N^2 + O(N\sqrt{N})$ area for its layout, while a hypermesh-pyramid consumes an area of $6N\sqrt{6N} + o(N\sqrt{N})$ and a WK-pyramid network can be laid out on an area of $39/2N^2 + O(N)$. We also showed that an RTCC network can be laid out in $N^2/16 + o(N^2)$ area and $8N^2/(S/2)^2 + o(N^2)$ volume (when the chip is multilayer). By using the RTCC layout results, we obtained $3/2 \cdot C^2N^2 + O(CN^2)$ as the area boundary needed for RTCC-Pyramid network to sit on the chip.

REFERENCES