Balancing the Interconnect Topology for Arrays of Processors between Cost and Power

Esther Y. Cheng, Feng Zhou, Bo Yao, Chung-Kuan Cheng, Ronald Graham

Department of Computer Science and Engineering
University of California San Diego, La Jolla, CA 92093-0114

ABSTRACT

High performance SoC requires nonblocking interconnections between an array of processors built on one chip. With the advent of deep sub-micron technologies, switches are becoming much cheaper while wires are still expensive. Therefore, optimization efforts should focus on the wire resources. In this paper, we devise an objective function to balance the interconnect topology between routing area and power dissipation. Based on the objective function, we find the best one-dimensional and two-dimensional nonblocking interconnect architectures. Furthermore, we define a derivative benefit and devise a strategy for improving the performance of hierarchical nonblocking interconnect architectures and derive optimized results.

1. INTRODUCTION

Interconnect architectures [1, 2, 3, 4, 5, 6, 9, 10, 11] play a key role for arrays of processors on a single chip. Currently, bus-based architectures offer standards for communication interfaces. However, the physical size of long bus wires limits the scalability of the architecture. Furthermore, the contention for the buses adds the latency of the communication.

Nonblocking interconnects [3] give the best solution to high speed systems. Whenever a processor needs to communicate with another one, a route always exists for setting up the connection. Designing a nonblocking interconnect has been a research topic for several decades. Initially, it was used to optimize telephone switching systems; research [3] focused on minimizing the number of switches because as individual components, switches are more expensive than wires.

In [4], D. Knuth discussed the problem taking into consideration both the number of switches and signal delays, in terms of switch stages. However, today's devices have shrunk to very small sizes, while wires, especially buses, have lengthened with the enlargement of the chip size. Wires occupy most of the chip area and dominate the power and signal delays, and therefore have become the main consideration when we optimize our nonblocking interconnect architectures.

In [10], Moritz et al. adopted a packet communication strategy where the cost of the architecture is proportional to the area of the network and the memory for the packets. Then they tried to reduce the latency. Mai et al. utilized crossbars for local communication [9]. They used buses to link the crossbars for a second-level hierarchy and a mesh for global communication. Their goal was to minimize the delay of communication. Dally and Towles [11] also proposed a packet communication strategy. They used a mesh to link an array of processors. They also adopted area and latency as the metric for the network.

We use the total wire length to measure the cost of the resources since the wire length is proportional to the amount of area taken on the routing layers. In deep submicron technologies, the number of routing layers remains limited. Even as the number of layers increases, the coupling capacitance due to congestion and the required vias that connect the signal to the layers high above make routing area a precious resource.

We also try to reduce the power dissipation of the interconnect since power consumption has become one of the main concerns in many applications such as telecommunication, portable appliances, and high performance computation. In CMOS technology, the power dissipation is proportional to the wire capacitance, which is again proportional to the distance the signal travels. Thus, we try to reduce the total traveled distance of the signal communication. Let us assume that each processor has to communicate with the rest of the processors with equal demand. Then the total power dissipation is measured by the total pairwise distance between the processors. We adopt this equal demand model because the demand is symmetrical and thus independent of the placement implementation.

It is conceivable that by adding wires for the communication, the traveling distance can be reduced. On the other hand, the wire resources are limited by the physical space. Furthermore, the same resources are precious for other purposes such as making internal connections in each processor, or testing. Thus, the product of the total wire length and the total power dissipation is chosen as the metric to balance the design. Moreover, the derivative of the product can be used to improve the interconnect architecture.

Proceedings of the 2002 IEEE International Conference on Computer Design: VLSI in Computers and Processors (ICCD'02)
1063-6404/02 $17.00 © 2002 IEEE

In [7], Igarashi et. al. demonstrated that diagonal interconnect architectures are feasible for integrated circuits. They showed that the diagonal interconnect approach can reduce wire length by 19.2% and area by 10%. Moreover, the critical path delay is 19.8% shorter. Therefore, we also try non Manhattan interconnect as part of our proposed architecture.

Leiserson [1, 2] proposed hierachical architectures which are suitable for large scale processors arrays. In our paper, we also adopt the same strategy. However, we will use a proposed objective function to justify the choice of the architecture. We explore the hierarchical structures of H trees, X trees and Y trees. Furthermore, we use the derivative of the objective function to search for the best architecture improvement.

Overall, our contributions to this research topic are:

1. An objective function is proposed to measure the quality of an interconnect architecture. The total wire length is used to measure the cost of the resources since the wire length is proportional to the amount of area taken on the routing layers. The total pairwise distance between the processors is used to measure the power dissipation as well as the time necessary for communication.

2. The best basic blocks for one-dimensional and two-dimensional nonblocking interconnect architectures are found, with diagonal interconnect architectures introduced [7].

3. One of the large scale nonblocking interconnect architectures is based on the architecture of [1] due to its hierarchical structure and regularity. To further improve the performance, we propose to set up direct connections at some levels to avoid detouring and derive the most efficient solution with respect to the performance/cost.

The paper is organized as follows. Section 2 describes the objective function. Section 3 and Section 4 discuss in detail the basic nonblocking architectures in one and two dimensions . In Section 4, we extend the best basic two-dimensional architectures to a hierarchical structure and derive the optimal incremental improvement. Finally, we give our conclusions in Section 5.

2. STATEMENT OF THE PROBLEM

Here we define an objective function with an emphasis on the wires instead of the switches. The bus width of individual processors is assumed to be uniform and is thus considered to be a basic unit.

Definition 1: The objective of interconnect architecture design is to minimize:

$$M = L * D$$

where,

$$\begin{aligned} L &= \sum Length\ of\ each\ wire \\ D &= \sum_{1 \le i < j < \le P} d_{i,j} \end{aligned}$$

$d_{i,j}$ is the shortest route length between processor(i) and processor(j). P is the total number of processors.

The derivative form of the objective function will be used to further improve the interconnect architecture.

$$\begin{aligned} \frac{\Delta M}{\Delta L} &= \frac{(L+\Delta L)(D+\Delta D) - L*D}{\Delta L} \\ &= \frac{\Delta D}{\Delta L} L + \Delta D + D \\ &\approx \frac{\Delta D}{\Delta L} L + D \end{aligned}$$

The last equation is based on the assumption that $\frac{L}{\Delta L}$ is much larger than 1. To identify the most cost effective incremental improvement due to the change of L, we introduce the derivative benefit:

Definition 2: The derivative benefit is defined to be:

$$I = -\frac{\Delta D}{\Delta L}$$

Note that a negative sign is used because D is expected to decrease when L increases.

3. ONE-DIMENSIONAL ARCHITECTURES

For the cases in which the processors are placed along a straight line, we design a one-dimensional interconnect architecture. Fig.1 shows an example of six processors. In Fig.1(a), every x between two crossing segments represents a switch between the 2 crossing buses. When the switch is on the vertical bus connects the horizontal bus. Otherwise the two buses are not connected. In Fig.1(b), a non-filled circle represents a group of six switches connecting six possible pairs of the four ports. Thus, the architecture of Fig.1(b) requires more switches but less wire resources than the architecture of Fig.1(a).

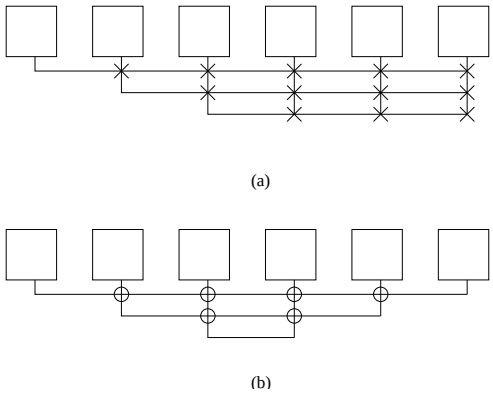

Figure 1: Structures in one dimension

Table 1 shows the objective function values of Fig.1, assuming that the distance between adjacent processors is 1.

Table1

Model	L	D	M
a	$\frac{3n^2-2n}{8}$	$\frac{n(n+1)(n-1)}{6}$	$\frac{n^2(3n-2)(n-1)(n+1)}{48}$
b	$\frac{n^2}{4}$	$\frac{n(n+1)(n-1)}{6}$	$\frac{n^3(n-1)(n+1)}{24}$

Model (b) has the following features: 1. for every vertical cutline, we have the minimum number of wires necessary to connect the two parts separated by the cutline; 2. for every pair of processors, we have the shortest signal route. So model (b) has the smallest M, and we have:

Lemma l: Model (b) is the best solution of one-dimensional nonblocking architectures.

4. TWO-DIMENSIONAL ARCHITECTURES

For the two-dimensional cases, we first study the possible nonblocking interconnect architectures for 2x2 blocks. Various configurations are enumerated and measured according to the proposed objective function. Two tree structures, the X model and the H model, are compared. In addition, a completely new tree structure, Y model, which uses hex-shaped cells, is presented. Then, these three types of tree structures are extended for hierarchical interconnect architectures. We identify the optimal improvement of the three hierarchical structures in terms of the derivative benefit.

4.1 Basic interconnect architectures of 2x2 blocks

In Fig.2((a)-(f)), we list six nonblocking interconnect architectures of 2x2 blocks. Models ((a)-(c)) are of mesh structure, models ((e)-(f)) are of tree structure, and model (d) is the mixture of mesh and tree structures.

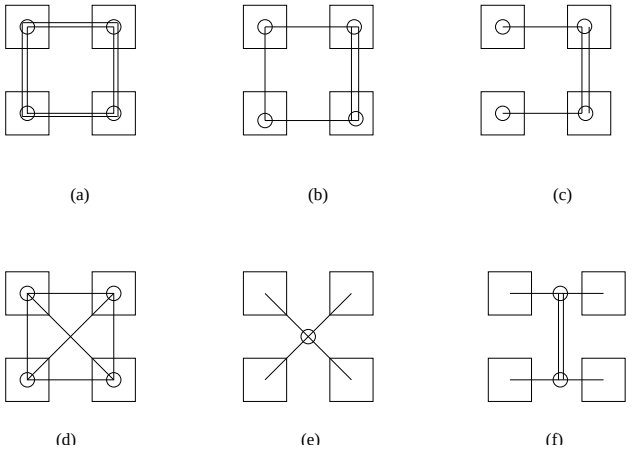

Figure 2: Nonblocking interconnect architectures of 2x2 blocks

Table 2

Model	L	D	M
a	8	8	64
b	5	8	40
c	4	10	40
d	$4+2\sqrt{2}$	$4+2\sqrt{2}$	$24+16\sqrt{2}$
e	$2\sqrt{2}$	$6\sqrt{2}$	24
f	4	10	40

Table 2 gives their objective function values. From the data, we have the following remarks:

1. The set of wires of model (b) is a subset of the wire set of model (a), but the total pairwise distance D is the same. Thus model (b) is superior to model (a);

2. The set of wires of model (c) is a subset of model (b). However, model (c) has the total pairwise distance ($D = 10$) longer than that of model (b) ($D = 8$). The quality of the two models are equal in terms of the objective function;

3. Models (d) and (e) adopt 45-degree wires. Model (d) has less total pairwise distance ($D = 4 + 2\sqrt{2}$) than model (e). However, model (d) requires much longer length of wires. Thus, measured by the objective function, model (d) is worse than model (e);

4. Model (f) uses an H tree topology. Since the wires are forced to follow a Manhattan pattern, the quality of model (f) is worse than that of model (e);

5. Model (d) has the minimum D, for it provides the shortest signal route for any pair of processors. Model (e) consumes the fewest wire resources.

Using the objective function evaluation, we conclude:

Lemma 2: Model (e) is the best among the proposed 2x2 nonblocking architectures.

4.2 Hex-shaped blocks

If the physical layout of processors has a shape of hexagon, as illustrated in Fig.3, the connection between cells will be completely different. We compared two interconnect architectures, Y-type and Δ-type. The results are shown in Table 3. We assume that the distance between the centers of neighbouring cells is 1. It is obvious that Y-type is better than Δ-type.

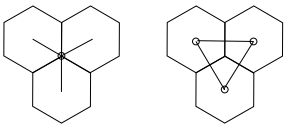

Figure 3: Hex-shaped blocks

Table 3

Model	L	D	M
Y	$\sqrt{3}$	$2\sqrt{3}$	6
Δ	3	3	9

4.3 H tree architecture

We first study model (f) because this H tree structure has been adopted by many researchers [1, 2, 5].

Fig.4(a) illustrates the H tree interconnect architecture using model (f). Here non-filled circles are replaced with filled circles. Apart from the switches, two buses of the same level are bundled together to form a new bus connecting to the level above (Fig.5). The bus width doubles for every expansion to a higher level. The expansion continues until the root of the tree is reached. Fig.4(b) describes the hierarchy of the tree structure and the definition of levels for the tree. To make up a square array, the number denoting the tree's top level, n, has to be even. The number of processors covered by a tree with n levels is equal to $P_n = 2^n$.

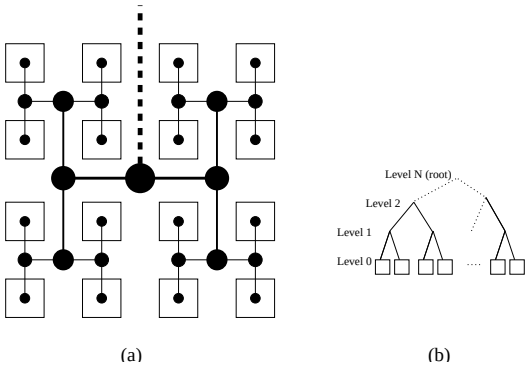

Figure 4: Construction from basic model (f)

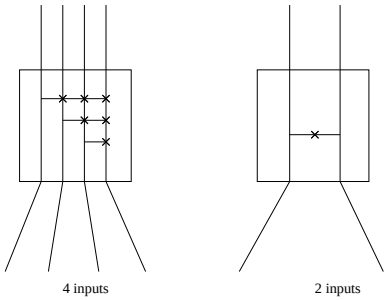

Figure 5: Switch group and bus bundling

The main shortcoming of this structure is the extra detouring problem. An extreme example is depicted in Fig.6. Two processors may be close in geometric distance, but their actual route can be much longer if their lowest common ancestor of the hierarchical tree structure is the root.

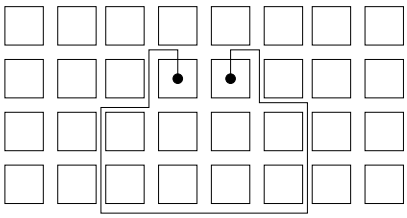

Figure 6: Detouring problem

Our strategy is to add wires bridging nodes of the same level. The communication between the bridged nodes can thus bypasses the detour going toward upper levels by taking advantage of the bridge. Fig.7 shows some possible additional interconnects.

Given an n-level H-tree structure, for each integer $m(0 \leq m < n)$, we state the incremental improvement of level-m nodes as follows. We define two level-m nodes (the T joints of the H tree) are physically adjacent if the Euclidean distance between the pair are the closest among all level-m nodes. We connect a pair of level-m nodes if the pair are physically adjacent and their lowest common ancestor of the tree structure is the root. Level-m nodes are linked with 2^m buses.

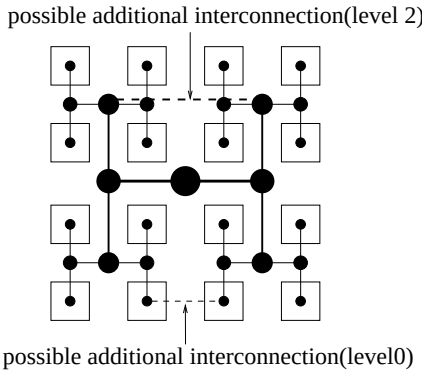

Figure 7: Possible additional interconnects

Fig.8 illustrates the alternative bridges at five levels using an array of 8 X 8 processors as an example. Only the upper half array is shown. We eliminate the tree structure as depicted in Fig.4 to clarify the illustrations. The additional wires are symmetrical with respect to the central vertical line which divides the whole processor array into two halves.

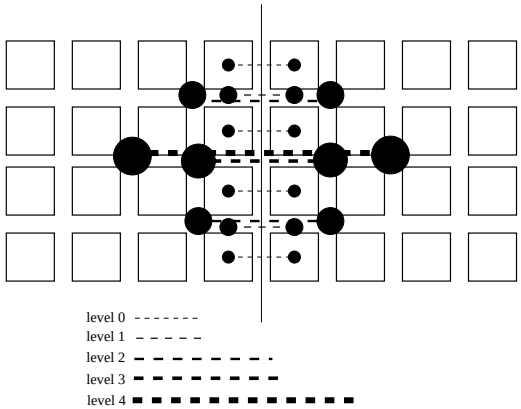

Figure 8: Alternative additional interconnects

Now we are faced with an optimization problem: on which level should we establish our additional interconnects to get the largest benefit? Let us derive our derivative benefit function according to Definition 2:

Given a tree of level n, and the level we are investigating, m, we have:

$$\Delta D(n,m) \quad = \quad A(n,m) * B(n,m) \qquad (1)$$

where A(n,m) represents the number of processors pairs which will benefit from the addition of interconnects; B(n,m) represents their route length saved due to the additional interconnects.

If m is odd:

$$\begin{aligned} A(n,m) &= 2^{\frac{n+3m-1}{2}} \\ B(n,m) &= -(2^{\frac{n+2}{2}} - 2^{\frac{m+3}{2}}) \\ \Delta L(n,m) &= 2^{\frac{n+2m-2}{2}} \\ I(n,m) &= 2^{\frac{n+m+3}{2}} - 2^{m+2} \end{aligned}$$

Refer to Fig.4, if $m = n-1$, $I = 0$ because $B(n, n-1) = 0$.

If m is even:

$$\begin{aligned} A(n,m) &= 2^{\frac{n+3m}{2}} \\ B(n,m) &= -(2^{\frac{n+2}{2}} - 3 * 2^{\frac{m}{2}}) \\ \Delta L(n,m) &= 2^{\frac{n+2m}{2}} \\ I(n,m) &= 2^{\frac{n+m+2}{2}} - 3 * 2^m \end{aligned}$$

For any even m($0 < m < n$), we have:

$$\begin{aligned} I(n,m) - I(n,m-1) &= 2^{\frac{n+m+2}{2}} - 3 * 2^m - 2^{\frac{n+m+2}{2}} + 2^{m+1} \\ &= -2^m < 0 \end{aligned}$$

From the above inequality, we conclude that $I(n,m) < I(n,m-1)$. Hence, we inspect only the odd levels for the maximal derivative benefit. For a continuous variable function: $I(n,x) = 2^{\frac{n+x+2}{2}} - 3 * 2^x$, we calculate that when $x = n-3$, the derivative benefit is the maximal, $I(n, n-3) = 2^{n-1}$.

Lemma 3: For an H tree architecture, level $m = n-3$ gives the best derivative benefit for additional interconnects: $I_{max} = 2^{n-1}$.

One example is given in Fig.9 with optimized additional interconnects(dashed lines). The best solution is neither at the highest level nor at the lowest level.

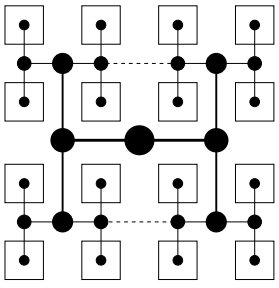

Figure 9: An example for best additional interconnects

Table 4 gives a group of values of $I(n,m)$ to make the above conclusions more clear. We take $n = 10$ for an example.

Table 4

m	ΔL	ΔD	I
0	32	-1952	61
1	32	-3840	120
2	128	-14848	116
3	128	-28672	224
4	512	-106496	208
5	512	-196608	384
6	2048	-655360	320
7	2048	-1048576	512
8	8192	-2097152	256

From this table, we can see clearly the derivative benefits for different levels. The best solution lies neither on the top level nor on the bottom level.

4.4 X tree architecture

Fig.10 shows the hierarchical extension of model (e). The bus width expands 4 times for every migration to a higher level. For the case that the tree's root lies at the mth level, the number of processors is $P_n = 4^n$. Again we add interconnects. Fig.11 illustrates alternatives for additional interconnects at different levels. The additionally connected nodes are always symmetrical with respect to the large cross which divides the whole processor array into 4 parts.

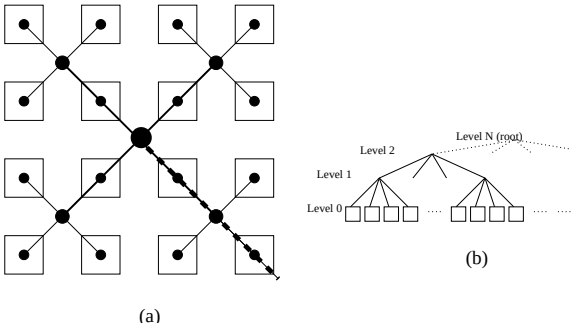

Figure 10: Construction from basic model (e)

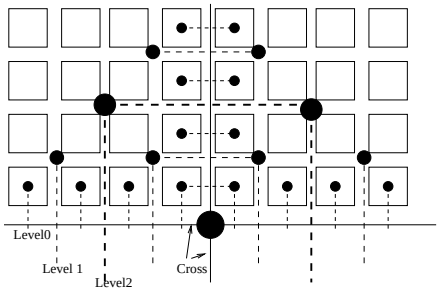

Figure 11: Alternative additional interconnects

Given an n-level X-tree structure, we explore incremental improvements by linking nodes at different levels. For each level m : $0 \le m < n$, we connect pairs of level-m nodes if the pairs are physically adjacent and their lowest common ancestor in the X-tree is the root. Level-m nodes are linked with 4^m buses. We derive the derivative benefit as follows:

$$
\begin{aligned}
A(n,m) &= 4 * 2^{n-m-1} * 4^m * 4^m = 2^{n+3m+1} \\
B(n,m) &= -(\sqrt{2} * \sum_{i=m}^{n-1} 2^i - 2^m) \\
\Delta L(n,m) &= 4 * 2^{n-m-1} * 2^m * 4^m = 2^{n+2m+1} \\
I(n,m) &= \sqrt{2} * 2^{n+m} - 2^{2m} * (\sqrt{2}+1)
\end{aligned}
$$

For the continuous variable function: $I(n,x) = \sqrt{2} * 2^{n+x} - 2^{2x} * (\sqrt{2}+1)$, there exists an x_0: $1 < n - x_0 < 2$, such that $I(n,x_0)$ has a maximum value. Further calculation shows that $I(n,n-2) > I(n,n-1)$. So we conclude that:

Lemma 4: For a quad tree architecture, level $m = n-2$ gives the best derivative benefit for additional interconnects: $I_{max} = 2^{2(n-2)}[2^{\frac{5}{2}} - (\sqrt{2}+1)]$

4.5 Y tree architecture

Fig.12 shows one type of Y tree architecture. There are dead cells (shaded cells), which means some cells are excluded from the interconnect covered by the Y tree; Fig.13 and Fig.14 show another type of Y tree architecture, in which there are no dead cells, but the directions of Y-type connections rotate with the increase of tree levels. Table 5 gives the values of L and D for a Y trees with n levels. It can be easily proved that, for a large n, M_{no_empt} has a smaller value than M_{with_empt}. Although the architecture without dead cells is better, the rotation of Y connections brings additional difficulty for adding extra interconnects.

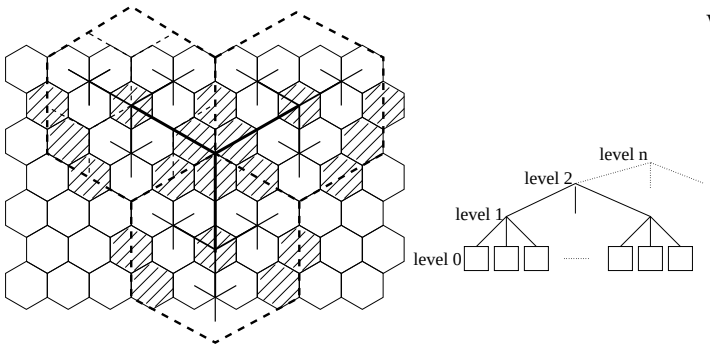

Figure 12: Construction from hex-cells, with empty cells

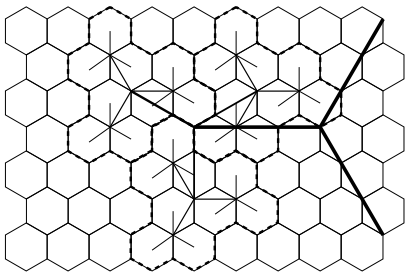

Figure 13: Construction from hex-cells, without empty cells (level 0-3)

Table 5

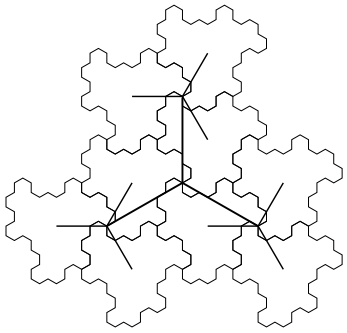

Figure 14: Construction from hex-cells, without dead cells (level 4-5)

Type	L
With dead cells	$L_1 = \sqrt{3}$ $L_n = 3L_{n-1} + 6^{n-1}\sqrt{3}$
Without dead cells	$L_1 = \sqrt{3}$ $L_n = 3L_{n-1} + 3^{n-1}\sqrt{3}^n$

Type	D
With dead cells	$D_1 = 2\sqrt{3}$ $D_n = 3D_{n-1} + 2\sqrt{3}(2^{n-1}-1)9^{n-1}$
Without dead cells	$D_1 = 2\sqrt{3}$ $D_n = 3D_{n-1} + (\sqrt{3}+3)(\sqrt{3}^n - 1)9^{n-1}$

Fig.15 (a) gives the possible addition of extra interconnects on a Y architecture with dead cells. Interestingly, we found that, the numbers of extra buses are just the same no matter which level we are considering (all equal 3).

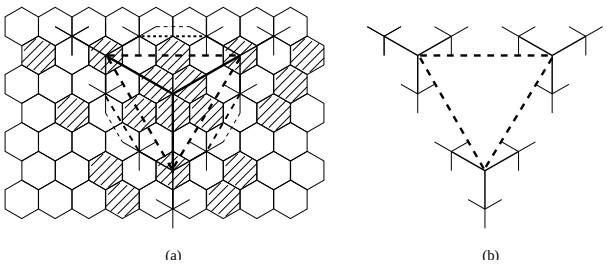

Figure 15: Alternative additional interconnects for Y architecture with dead cells

With the similar methods mentioned above, we derived the derivative benefit.

$$
\begin{aligned}
A(n,m) &= 3^m * 3^m * 3 = 3^{2m+1} \\
B(n,m) &= -(\frac{2}{\sqrt{3}} * \sum_{i=m}^{n-1} 2^i - 2^m) \\
\Delta L(n,m) &= 3^{m+1} * 2^m \\
I(n,m) &= \sqrt{3} * 3^{m-1} * (2^{n-m+1} - 2) - 3^m
\end{aligned}
$$

The optimal level to put the additional interconnects on is level $n-2$, with the maxium incremental benefit :$I(n,n-2) = (2\sqrt{3}-1) * 3^{n-2}$. We noticed that, if we put the

additional interconnects on level $n-1$, the top level Y connection can be removed, as illustrated in Fig.15 (b). Then ΔL becomes $3^n * 2^{n-1} - \sqrt{3} * 6^{n-1}$, resulting in the incremental benefit of $\frac{1}{2}3^{n-1}(\sqrt{3}-1)$. However, it is still less than $I(n, n-2)$.

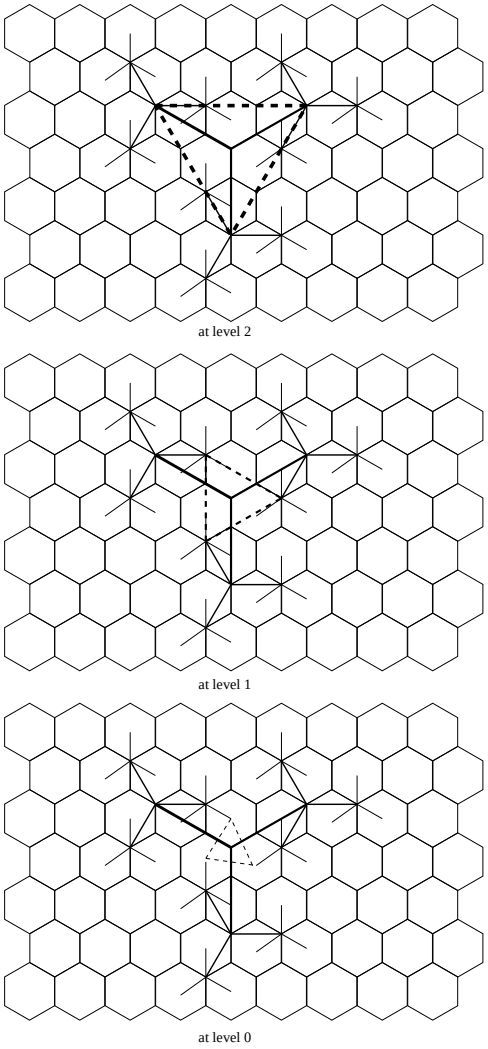

Figure 16: Alternative additional interconnects for Y architecture without dead cells

Fig.16 gives the possible addition of extra interconnects on a Y architecture without dead cells, and we have:

$$
\begin{aligned}
A(n,m) &= 3^m * 3^m * 3 = 3^{2m+1} \\
B(n,m) &= -(\frac{2}{\sqrt{3}} * \sum_{i=m}^{n-1} \sqrt{3}^i - \sqrt{3}^m) \\
\Delta L(n,m) &= 3^{m+1} * \sqrt{3}^m \\
I(n,m) &= 3^m(\frac{2}{3-\sqrt{3}}(\sqrt{3}^{n-m} - 1) - 1)
\end{aligned}
$$

Again we have the maxium incremental benefit at level $n-2$.

Lemma 5: For a Y tree architecture, level $m = n-2$ gives the best derivative benefit for additional interconnects. $I_{max} = 3^{n-2}(2\sqrt{3}-1)$ for the architecture with dead cells and $I_{max} = 3^{n-3}(2\sqrt{3}+3)$ for the architecture without dead cells.

5. CONCLUSION

In this paper, we proposed an objective function and derivative benefit to measure the quality of nonblocking interconnect architectures and their improvements. We have enumerated the basic interconnect structures for one-dimensional and two-dimensional nonblocking architectures, and identified the best solutions. The three best tree structures were extended for hierarchical interconnect. We perturbed the three hierarchical tree structures for further improvement. The incremental improvement connecting nodes at 2, 3 and 2 levels below the root are optimal for the X, H and Y trees, respectively.

6. ACKNOWLEDGEMENT

Esther Y. Cheng would like to thank Mrs. V. Coordt for her encouragement and support in starting this research as a Science Fair project.

This work was supported in part under grants from NSF project number MIP-9987678, the California MICRO program and SRC support. The authors would like to thank all the reviewers for their comments.

7. REFERENCES

[1] C.E. Leiserson, Fat Tree: Universal Networks for Hardware – Efficient Supercomputing, IEEE Trans on Computers, Oct 1985. pp. 892-901.

[2] C.E. Leiserson, VLSI Theory and Parallel Supercomputing, pp. 5-16. Advanced Research in VLSI, Proceeding of teh Decennial Caltech Conference on VLSI, March 1989, edited by C.L. Seitz.

[3] V.E. Benes, Mathematical Theory of Connecting Networks and Telephone Traffic, Academic Press, 1965.

[4] D. Knuth, The Art of Computer Programming, vol. 3, Sorting and Searching

[5] F.T. Leighton, Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes, Kaufmann, 1992.

[6] C. Clos, A Study of Non-Blocking Switching Networks, Bell System Tech. vol. 32, pp. 406-424, 1953.

[7] M. Igarashi, A diagnoal-interconnect architecture and its applications to RISC core design, ISSCC Digest of Technical Papers, pp.210-211, San Francisco, 2002.

[8] R. Carden and C.K. Cheng, A Global Router Using An Efficient Approximate Multicommodity Multiterminal Flow Algorithm, ACM/IEEE Design Automation Conf., pp.316-321, July 1991.

[9] K. Mai, Smart Memories, a modular reconfigurable architecture, Proceedings of 27th International Symposium on Computer Architecture, pp.161-171, Vancouver, BC Canada, 2000.

[10] C.A. Moritz, Exploring optimal cost-performance designs for Raw microprocessors, Proceedings. IEEE Symposium on FPGAs for Custom Computing Machines, pp.12-27, Los Alamitos, CA, 1998.

[11] W.J. Dally, Route Packets Not Wires: On-chip Interconnection Networks, IEEE Proceedings of 38th Design Automation Conference, pp.684-689, 2001.