10-1996

An Empirical Comparison of Networks and Routing Strategies for Parallel Computation

Ronald I. Greenberg
Rgreen@luc.edu

Lee Guan

Author Manuscript

This is a pre-publication author manuscript of the final, published article.

Recommended Citation

Ronald I. Greenberg and Lee Guan. An empirical comparison of networks and routing strategies for parallel computation. In Proceedings of the Eighth IASTED International Conference on Parallel and Distributed Computing and Systems, pages 265--269, Chicago, October 1996.

An Empirical Comparison of Networks and Routing Strategies for Parallel Computation
(preliminary version)

Ronald I. Greenberg
Mathematical and Computer Sciences
Loyola University
6525 North Sheridan Road
Chicago, IL 60626
rig@math.luc.edu

Lee Guan
Electrical Engineering
University of Maryland
College Park, MD 20742
leegaun@eng.umd.edu

Abstract

This paper compares message routing capabilities of important networks proposed for general-purpose parallel computing. All the networks have been proven to have some type of universality property, i.e., an ability to simulate other networks of comparable cost with modest slowdown, using appropriate cost and communication models. But in this paper we seek an empirical comparison of communication capability under typical direct use rather than an analysis of worst-case results for simulating message traffic of another network.

1 Introduction

A significant challenge to massively parallel computing is providing an economical interconnection network that can support general patterns of communication among processors. It has been noted that the hypercube is universal in the sense that it can simulate any network on the same number of processors with logarithmic slowdown (e.g., see [18]). The high pin-out and area requirements of the hypercube are serious detriments, however, which led to development of various types of "fat-tree" networks, which have the property that they can simulate any network of comparable VLSI area with slowdown polylogarithmic in the area under circuit-switched, packet, or wormhole routing models [3, 4, 8, 9, 10, 13, 14]. This research has influenced the design of parallel computers by Thinking Machines Corporation and Meiko [5, 12, 15]. Most area-universality analyses have been performed in the "unit wire delay model", i.e., assuming that unit time suffices to send a bit across a wire regardless of its length. This model has generally proved reasonable for current technology, but may become less appropriate as we build larger systems; an extension of the fat-tree referred to as the "fat-pyramid" [8] has been shown to be area-universal given any reasonable dependence of delay on wire length. We use a simple version of the "butterfly fat-tree" (BFT) and fat-pyramid as illustrated in Figure 1. Finally, the mesh has been a popular network for parallel computing, and it is easy to see that the mesh is area-universal under linear wire delay, which may be the most accurate model in the distant future (e.g., see [6]).

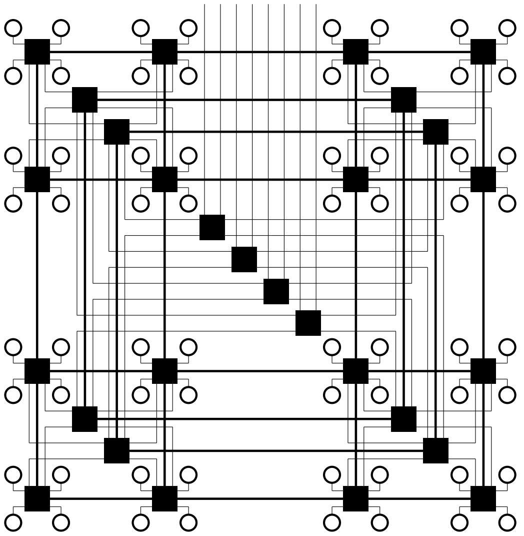

Figure 1: A fat-pyramid. Processors are placed at the leaves, represented by circles; the squares are switches. This network can be viewed as containing 2^h copies of a mesh of size $\sqrt{n/4}/2^h \times \sqrt{n/4}/2^h$ for levels h with $0 \leq h \leq \log_2 \sqrt{n/4}$. The switch denoted (h, c, x, y), where h is the level, c is the copy number of the mesh $(0 \leq c < 2^h)$ at this level that contains the switch, and x and y specify a mesh position in an ordinary Cartesian coordinate system $(0 \leq x, y < \sqrt{n/4}/2^h)$ is connected to $(h+1, 2c, \lfloor x/2 \rfloor, \lfloor y/2 \rfloor)$ and $(h+1, 2c+1, \lfloor x/2 \rfloor, \lfloor y/2 \rfloor)$ by "tree edges" for $h < \log_2 \sqrt{n/4}$. The fat-tree is as above, but with only the tree edges; i.e., the edges within the meshes are removed. (A different layout of the fat-pyramid is used to obtain results independent of wire delay.)

We perform an empirical comparison of message routing on the networks mentioned above, since examination of typical performance in practice may lead to substantially different conclusions than an analysis of worst-case slowdown for simulating another network. This paper focuses on the unit wire delay model; though the universality advantages of the mesh or fat-pyramid come in to play with different models of wire delay, it is interesting to see whether there is significant detriment to using these networks in the unit delay model. Most of the simulations here are performed in the simple (store-and-forward) packet routing model, but we also compare to wormhole routing, where messages (worms) are composed of *flits* or *flow control digits*, and worms snake through the network one flit after another with only a constant number of flits being stored in an intermediate node at any time. In store-and-forward routing, messages are conceptually transferred from node to node as atomic units, but we still count an appropriate number of flit steps to transfer a packet of many bits from one node to another to achieve a fair comparison with wormhole routing.

2 Equalizing Hardware Cost

To make fair comparisons between different networks (with a given number of processors), we adjust the channel width (the number of wires connecting adjacent nodes) in each, to make the cost of interconnections equal. We consider three models of hardware cost that have received substantial recent attention. Bisection width and pin-out constraints have been considered in prior empirical studies of wormhole routing on k-ary n-cube networks [1, 7]. The Thompson model for area [16, 17] has been the focus of theoretical analyses of area-universality [3, 4, 8, 9, 10, 13, 14]. The bisection width of a network is the minimum number of wires cut when the network is divided into two equal halves. The pin-out of a node is the degree times the channel width W, and total pin-out is the sum of the pin-outs over all nodes. VLSI layout area is evaluated based on the assumption that all processors and switches are placed on a 2-D substrate. The substrate has two layer of interconnect for the x-direction and y-direction, respectively, with a minimum wire width and separation. Such a model can serve as a good abstraction for a variety of VLSI packaging technologies, such as wafer-scale integration or printed circuit boards [2]. Since the area required for the processors is the same for all networks, we consider only the area necessary to achieve the interconnections. For each of the networks, we can analyze the area by expressing the side length with n processors as $S(n) = \sqrt{n} \cdot d \cdot W \cdot P$, where P is the wiring pitch and d can be thought of as an average wire density per row or column (with channel width 1) when the processors are laid out in a $\sqrt{n}$ by $\sqrt{n}$ grid.

Tables 1 through 3 give the bisection widths (B), pin-outs (PO), and average wire densities (d) for a range of values of n along with channel widths (W) to equalize cost.

n	*mesh*		*hypercube*		*BFT*		*fat-pyramid*	
	B	W	B	W	B	W	B	W
16	4	32	8	16	4	32	6	21
64	8	32	32	8	8	32	16	16
256	16	32	128	4	16	32	40	13
1024	32	32	512	2	32	32	96	11
4096	64	32	2048	1	64	32	224	9

Table 1: The bisection width with channel width 1 and the channel width to maintain constant bisection width across different networks.

n	*mesh*		*hypercube*		*BFT*		*fat-pyramid*	
	PO	W	PO	W	PO	W	PO	W
16	48	32	64	24	52	30	60	26
64	224	32	384	19	232	31	296	24
256	960	32	2048	15	976	31	1328	23
1024	3968	32	10240	12	4000	32	5664	22
4096	16128	32	49152	10	16192	32	23488	22

Table 2: The pin-out with channel width 1 and the channel width to maintain constant pin-out across different networks.

n	*mesh*		*hypercube*		*BFT*		*fat-pyramid*	
	d	W	d	W	d	W	d	W
16	1	32	2	16	2	16	3	10
64	1	32	5	6	3	10	4.5	7
256	1	32	10	3	4	8	6	5
1024	1	32	21	2	5	6	7.5	4
4096	1	32	42	1	6	5	9	4

Table 3: The wire density per row/column and the channel width under constant layout area constraints.

3 Experimental Results and Conclusions

It has been customary to use network latency as the primary performance measure because of its tendency to limit performance in practice in today's fine-grained parallel systems. The average latency is the average time to completely transmit a message from source to destination. It depends on load rate (the number of message bits generated per cycle per node), and simulations are run at a fixed load rate with random sources and destinations until average latency reaches a steady state. The average latency generally stays rather constant at low load rates and then increases rapidly as the network saturates. In practice, parallel networks should be designed to operate on the the flat portion of the latency curve. Maximum throughput is another important performance metric, and for certain applications such as sample sorting can be dominant [11]. Maximum throughput can be read from the latency graphs by looking for the load rate at which the network saturates.

Figures 2 through 4 show packet routing simulation results under constant bisection width, constant pin-out, and constant area constraints, respectively. Simulations are shown for several values of n up to $n = 4096$. Message lengths of 320 bits are used throughout.

The most striking aspect of the packet-routing simulation results in Figures 2 through 4 is that the mesh always performs very well in comparison to the other networks despite the use of the unit wire delay model. While the best low-load latency is obtained with the fat-tree under constant bisection and constant pin-out constraints (for large networks), it is surprising that the performance of the fat-tree is not generally better than what is shown by our simulations, particularly under the sort of area constraint that motivated study of the fat-tree. Performance of the fat-tree and fat-pyramid might be better with the more area-efficient variation in [8, Secs. II–III]. Also interesting is that for the most part, the packet routing graphs look qualitatively very similar to those obtained from wormhole routing with a reasonable range of worm lengths. (As would be expected, however, the packet routing results tend to show higher average latencies and higher maximum throughput.) Only in the case of constant pin-out did the choice of packet routing versus wormhole routing cause some change in the ranking of networks by low-load latency; our wormhole routing results for constant pin-out are shown in Figure 5.

References

[1] S. Abraham and K. Padmanabhan. Performance of multicomputer networks under pin-out constraints. *Journal of Parallel and Distributed Computing*, pages 237–248, Dec. 1991.

[2] H. B. Bakoglu. *Circuits, Interconnections, and Packaging for VLSI.* Addison-Wesley, 1990.

[3] P. Bay and G. Bilardi. An area-universal VLSI circuit. In *Proceedings of the 1993 Symposium on Integrated Systems*, pages 53–67, 1993.

[4] P. Bay and G. Bilardi. Deterministic on-line routing on area-universal networks. *Journal of the ACM*, 42(3):614–640, May 1995.

[5] J. Beecroft, M. Homewood, and M. McLaren. Meiko CS-2 interconnect Elan-Elite design. *Parallel Computing*, 20:1627–1638, Nov. 1994.

[6] G. Bilardi and F. P. Preparata. Horizons of parallel computation. *Journal of Parallel and Distributed Computing*, 27(2):172–182, June 1995.

[7] W. J. Dally. Performance analysis of k-ary n-cube interconnection networks. *IEEE Trans. Computers*, 39(6):775–785, June 1990.

[8] R. I. Greenberg. The fat-pyramid and universal parallel computation independent of wire delay. *IEEE Trans. Computers*, 43(12):1358–1364, Dec. 1994.

[9] R. I. Greenberg and C. E. Leiserson. Randomized routing on fat-trees. In S. Micali, editor, *Randomness and Computation.* Volume 5 of *Advances in Computing Research*, pages 345–374. JAI Press, 1989.

[10] R. I. Greenberg and H.-C. Oh. Universal wormhole routing. 1994. Submitted. Earlier versions of portions in University of Maryland technical reports UMIACS-TR-93-60 and UMIACS-TR-93-102 and *Proceedings of the Fifth IEEE Symposium on Parallel and Distributed Processing*, 1993.

[11] F. T. Hady. *A Performance Study of Wormhole Routed Networks Through Analytical Modeling and Experimentation.* PhD thesis, University of Maryland Electrical Engineering Department, 1993.

[12] M. W. Incorporated. Computing surface CS-2HA technical description, 1994.

[13] F. T. Leighton, B. M. Maggs, A. G. Ranade, and S. B. Rao. Randomized routing and sorting on fixed-connection networks. *Journal of Algorithms*, 17(1):157–205, July 1994.

[14] C. E. Leiserson. Fat-trees: Universal networks for hardware-efficient supercomputing. *IEEE Trans. Computers*, C-34(10):892–901, Oct. 1985.

[15] C. E. Leiserson, Z. S. Abuhamdeh, D. C. Douglas, C. R. Feynman, M. N. Ganmukhi, J. V. Hill, W. D. Hillis, B. C. Kuszmaul, M. A. S. Pierre, D. S. Wells, M. C. Wong, S.-W. Yang, and R. Zak. The network architecture of the connection machine CM-5. In *Proceedings of the 4th Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 272–285. Association for Computing Machinery, 1992.

[16] C. D. Thompson. Area-time complexity for VLSI. In *Proceedings of the 11th ACM Symposium on Theory of Computing*, pages 81–88. ACM Press, 1979.

[17] C. D. Thompson. *A Complexity Theory for VLSI.* PhD thesis, Department of Computer Science, Carnegie-Mellon University, 1980.

[18] L. G. Valiant. A scheme for fast parallel communication. *SIAM Journal on Computing*, 11(2):350–361, May 1982.

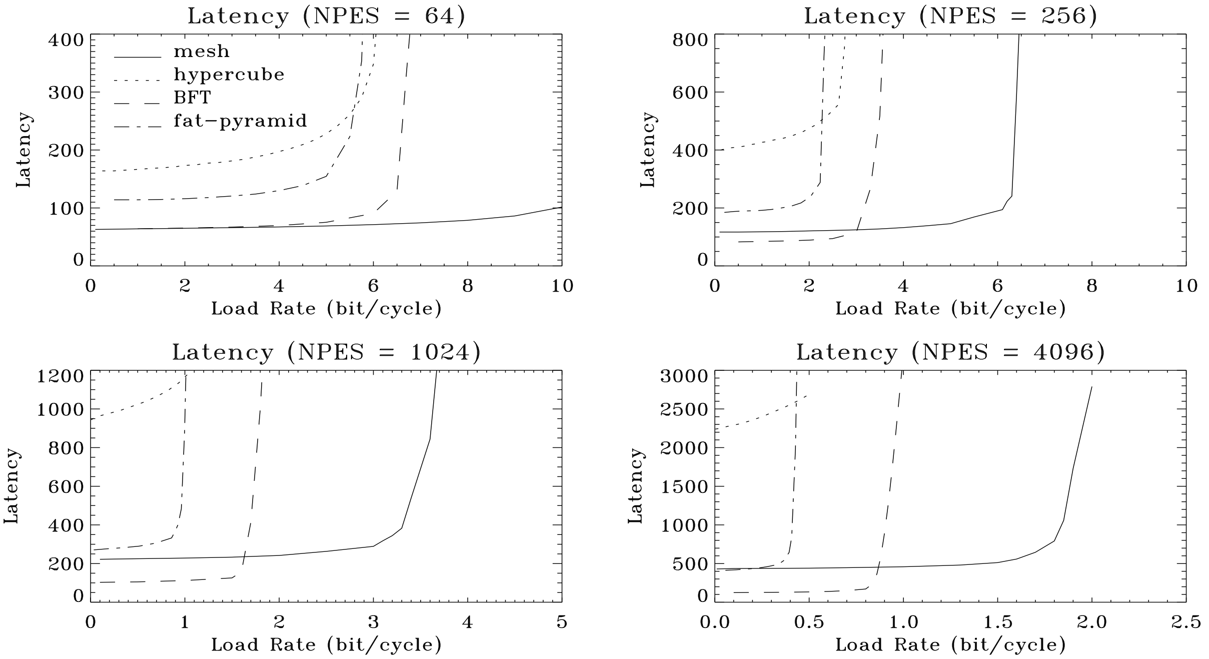

Figure 2: Comparison of packet routing latency under constraint of equal bisection width.

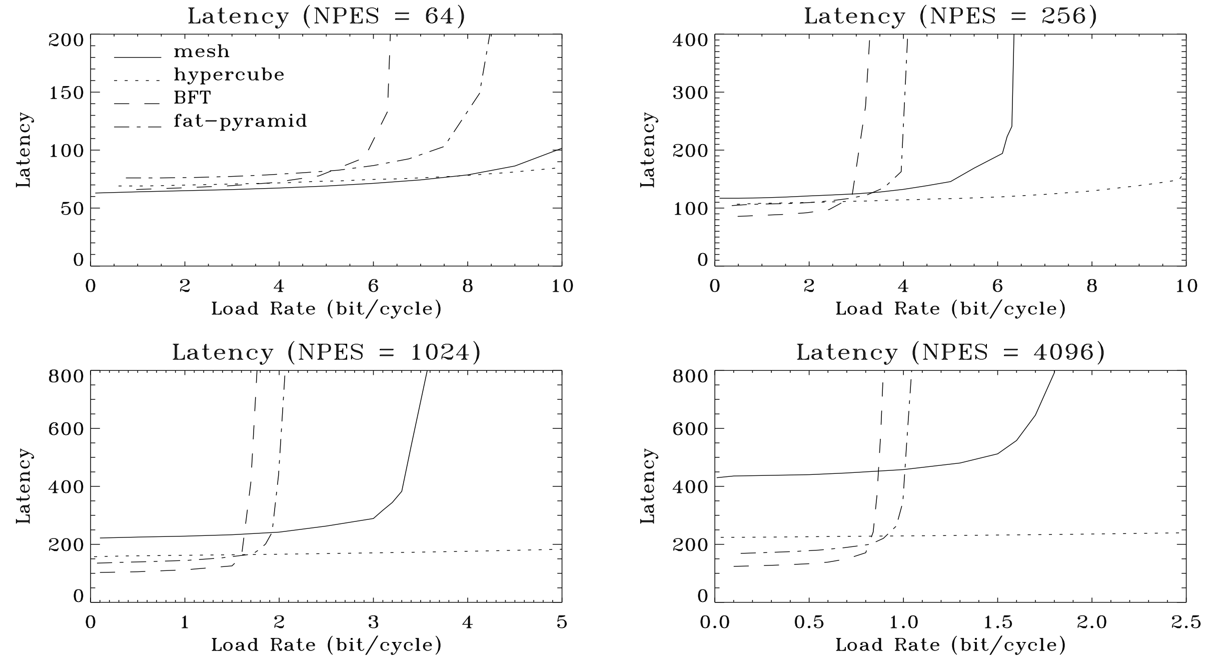

Figure 3: Comparison of packet routing latency under constraint of equal pin-out.

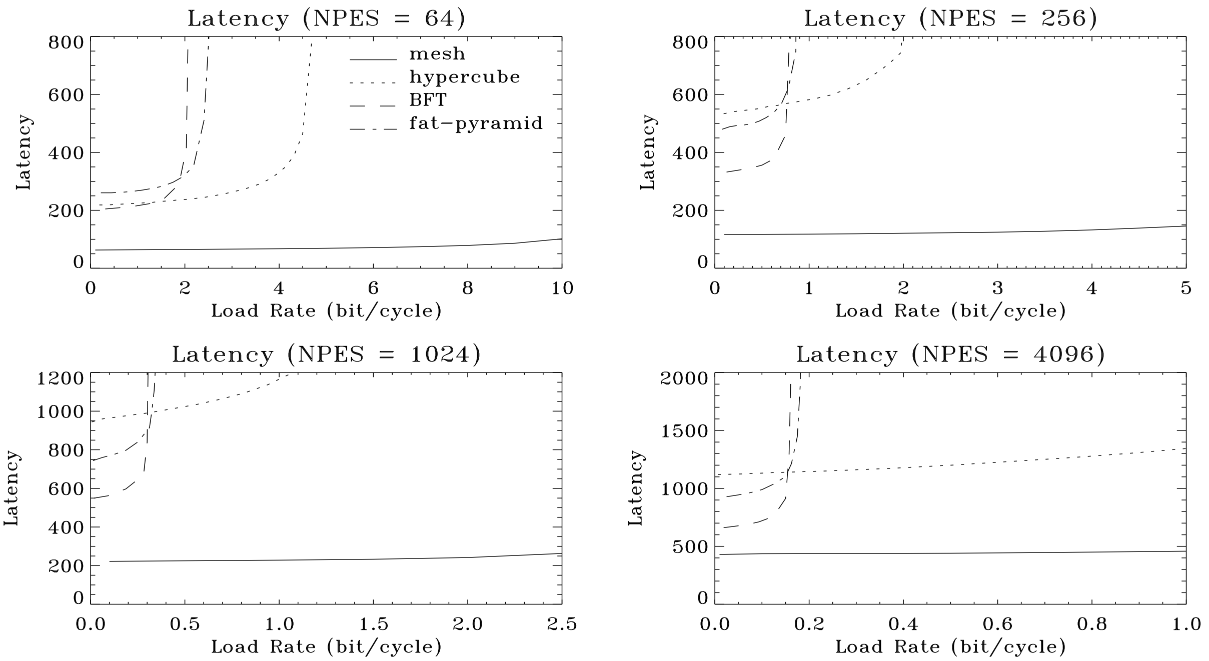

Figure 4: Comparison of packet routing latency under constraint of equal interconnect area.

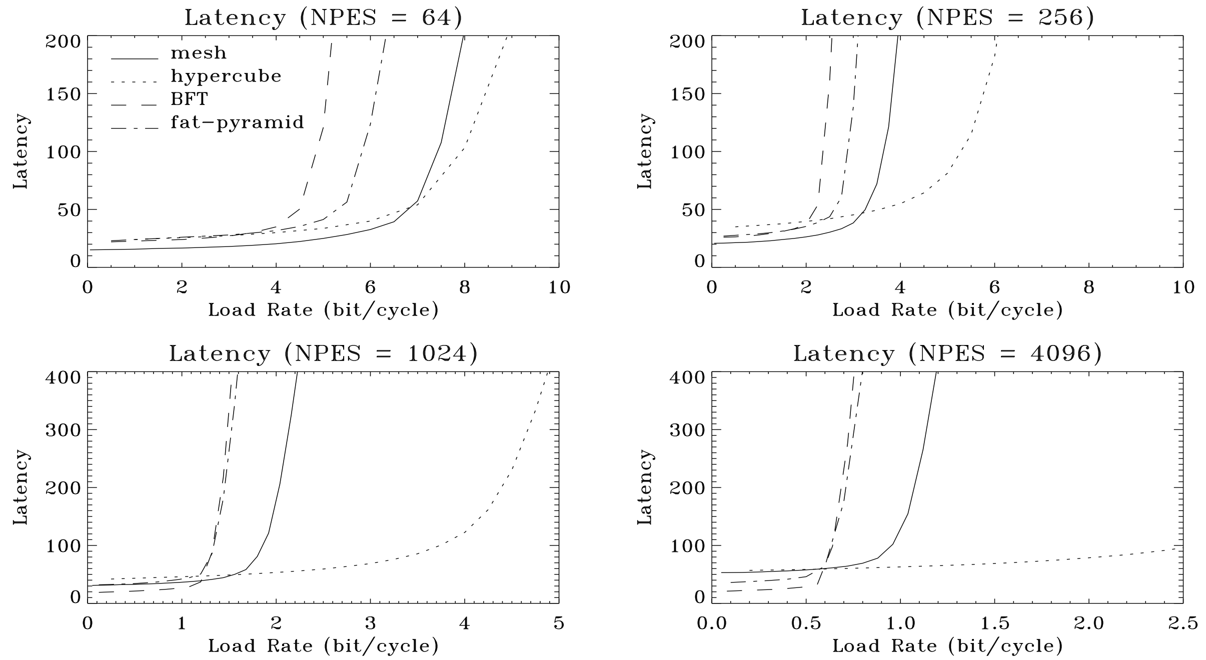

Figure 5: Comparison of wormhole routing latency under constraint of equal pin-out.