# Specialized Network Topologies for Efficient Communication in Computer Clusters[1]

Urban Borštnik
urban@cmm.ki.si

Milan Hodošček
milan@cmm.ki.si

Dušanka Janežič
dusa@cmm.ki.si

National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

**Keywords**: hierarchical hypercube topology, clusters of servers, parallel processing
**Subject Area**: Technologies for Advanced Networking

## Abstract

A topology of point-to-point interconnections is an efficient way to network a cluster of computers with well-defined communication patterns. Such a specialized topology allows parallelized programs to run with little penalty. Combining a specialized topology with a local area network (LAN) decreases traffic on the LAN and increases redundancy in the entire combined network.

We developed a hierarchical hypercube topology for a cluster of personal computers (PCs). A hierarchical hypercube is a hypercube topology with links of different speeds. Links representing the first dimension are the fastest, whereas those representing higher dimensions are successively slower. This sequence of speeds forms the hierarchy in this topology.

The cluster's topology was designed for the macromolecular mechanics program CHARMM [1], specifically the distributed global sum and broadcast operations [2] that account for most of the data exchange among processes in our cluster. CHARMM's implementation on a hypercube topology has the property that it transmits and receives different amounts of data through different dimensions [2], as depicted in fig. 1(a). The amount of data transferred in each successive dimension is halved. For this communication pattern not all of the links are used efficiently if they are all of the same speed. Only links representing the first dimension are saturated, therefore these limit the performance of the whole topology. However, if the underused links representing higher dimensions are replaced with slower (and cheaper) ones, then only some performance is lost, yet the cost of implementing the topology is greatly reduced. Care must be taken to ensure the replacement links remain fast enough so as not to be saturated and hence limiting. The resulting topology would have links with speeds proportional to the data transferred on them.

Our cluster is composed of 32 processors in 16 PCs with gigabit and fast Ethernet networking. We used the system bus between two processors for the first dimension of the hypercube, gigabit point-to-point links for the second, and a fast Ethernet switch for the remaining three dimensions. This configuration is depicted in fig. 1(b). The fast Ethernet switch is also used for general network traffic on the LAN.

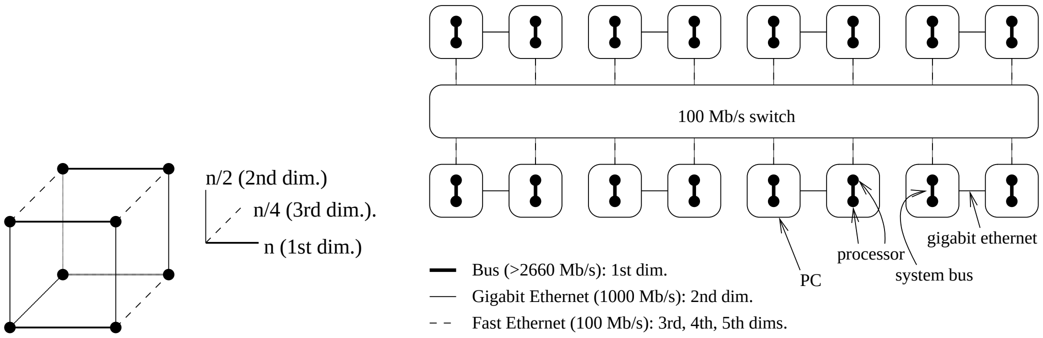
We performed comparisons of the described hierarchical hypercube topology to other topologies, the most interesting of which is solely a fast Ethernet switch, as it is a popular PC clustering solution [4]. The benchmark was a macromolecular simulation using the CHARMM program [1]. We tested configurations with hypercubes of up to 32 processors.

A similar approach to the hierarchical hypercube may be employed for a cluster of servers to allow running parallel programs, to reduce general network traffic, and to increase redundancy.

Using point-to-point links between computers reduces the need for LAN networking gear such as switches and hubs. Any two servers that exchange much data are connected with a direct link. In an existing setup, such connections serve mostly to increase the redundancy of a cluster, yet also to alleviate the traffic on a hub or switch. Complete redundancy of a cluster can be achieved if every server has at least one direct link to another server [3, 4].

Servers, such for dynamic web content or computational requests, in a cluster with dedicated point-to-point links may serve requests by running several serving processes on different servers while using

[1] Contract grant number: P01-0104-503 sponsored by the Ministry of Education, Science and Sports of Slovenia.


n/2 (2nd dim.)
n/4 (3rd dim.).
n (1st dim.)
100 Mb/s switch
Bus (>2660 Mb/s): 1st dim.
Gigabit Ethernet (1000 Mb/s): 2nd dim.
Fast Ethernet (100 Mb/s): 3rd, 4th, 5th dims.
PC
processor
system bus
gigabit ethernet

(a) Amount of data transferred in a three-dimensional hypercube.

(b) The physical connections in our hierarchical hypercube topology. The fastest links (system bus) form the first dimension of the hypercube, the gigabit point-to-point links form the second dimension, while the fast ethernet switch forms the third, fourth, and fifth dimensions.

the links to communicate among them. Instead of using traditional load balancing to distribute requests among multiple computers, parallelized serving software handles a request in parallel. The cluster's links should reflect the communication between the computers running the serving software. Requests can thus be processes faster, but the total throughput of the system remains approximately the same [3, 4].

We have determined that the hierarchical hypercube topology is more efficient than using only a switch to connect the component PCs. The computational speedup compared to other topologies becomes more significant as the number of processors is increased, which demonstrates the effectiveness of the topology for the cluster [5] and allows a greater number of processors to be used. However, it is not efficient to use too many processors in parallel as the communication overhead may become too great, leading to an actual decrease in performance [3]. In addition, the number of processors is limited by the number of links each processor may have. With typical PCs and network interface cards, a hypercube with point-to-point links is limited to around five dimensions, or six for multiprocessor machines. Such a limit can be overcome by using a hierarchical hypercube in which the higher dimensions use a switch.

A topology designed according to the characteristics of a sy stem can greatly enhance its performance. Specialized topologies of point-to-point links allow parallelized programs to efficiently run on a cluster of servers while also decreasing LAN traffic and making the network more redundant.

# References

[1] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[2] B. R. Brooks and M. Hodošček. Parallelization of CHARMm for MIMD machines. *Chemical Design Automation News*, 7:16–22, 1992.

[3] D. W. Heermann and A. N. Burkitt. *Parallel Algorithms in Computational Science.* Springer-Verlag, Berlin, 1991.

[4] D. H. M. Spector. *Building Linux Clusters: Scaling Linux for Scientific and Enterprise Applications.* O'Reilly & Associates, Sebastopol, CA, 2000.

[5] M. Hodošček, U. Borštnik, and D. Janežič. CROW for large scale macromolecular simulations. *Cellular & Molecular Biology Letters*, 7(1):118–119, 2002.

# Vitae

**Urban Borštnik** received his B.Sc. in Computer and Information Science at the University of Ljubljana. He is now a Young Researcher at the National Institute of Chemistry in Ljubljana and a graduate student at the Faculty of Computer and Information Science of the University of Ljubljana. His main research interest is developing computer clusters to perform high-speed macromolecular simulations.

**Milan Hodoščcek** is a research scientist at the National Institute of Chemistry, Ljubljana, Slovenia. From 1990 to 1993 and in 1998 he was a Visiting Fellow at the National Institutes of Health, Bethesda, MD, USA. He is currently working as a codeveloper of a CHARMM software system for macromolecular simulations. His main interest is in developing QM/MM methods and parallelizing the program CHARMM.

**Dušanka Janežič** is a senior research scientist and Head of the Center for Molecular Modeling at the National Institute of Chemistry, Ljubljana, Slovenia. In 1988 she was a Guest Researcher at the National Institutes of Standards and Technology, Gaithersburg, MD, USA. From 1989 to 1991 she was a Fogarthy Visiting Fellow and in 1994/95 a Senior Fulbright Scholar at the National Institutes of Health, Bethesda, MD, USA. As a holder of a Deutscher Akademischer Austauschdienst fellowship in 1999 she spent two months at the Technical University Munich, Garching, Germany. In 1999 she received the Ambasador of the Republic of Slovenia in Science award. In 2001 she became the Associate Editor of the Journal of Chemical Information and Computer Sciencies, an American Chemical Society publication. Her main research interest is developing symplectic integration algorithms for biomolecular simulations and their applications to harmonic analysis and molecular dynamics simulations of macromolecules.