StarT-jr: A Parallel System from Commodity Technology

James C. Hoe, Mike Ehrlich
Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139
jhoe@lcs.mit.edu, mikee@lcs.mit.edu

1 Abstract

StarT-jr is an experimental *parallel* system composed of a *network of personal computers (PCs)*. The system leverages the momentum of the microprocessor and PC industries to achieve excellent single node performance at a low cost. For parallel processing, StarT-jr uses the Flexible User-level Network Interface (FUNi) to provide low-overhead, *user-level interprocessor communication* over two *IEEE 1394 High Speed Serial Busses*. This efficient *message-passing* mechanism enables StarT-jr to exploit fine-grained parallelism for good parallel performance.

FUNi is based on an embedded coprocessor on a standard PCI card. Custom network hardware assembled from off-the-shelf 1394 silicon provides the FUNi coprocessor with direct access to the 1394 network. In order to avoid the long latency of accessing PCI devices from the host processor, FUNi's message-passing interface utilizes software-managed message buffers in the host memory. User programs directly manipulate the message buffers using cached memory reads and writes, instead of costly physical I/O operations to the FUNi device itself. Currently, FUNi can efficiently support both fine-grain message passing and direct memory-to-memory transfer of large blocked data. FUNi can also support globally coherent shared memory by capturing and responding to memory accesses within a designated global address range. The FUNi coprocessor locally maintains a globally coherent shared memory cache to minimize global memory access latency.

We have demonstrated a two-node prototype of StarT-jr and are awaiting fabrication of additional interface cards in order to assemble an eight-node system. StarT-jr currently supports a light-weight communication library, base on active-messages, for the C programming language. Preliminary measurements of the communication library demonstrated overheads of 1$\sim$3 μsec for sending and receiving of small ($\leq$ 40 bytes) messages, and an user-to-user latency of less than 30 μsec. Direct memory-to-memory transfers can sustain 6.5 MBytes/sec on an unloaded network. With regard to shared memory operation, a single-word read of a locally-cached shared-memory location takes around 2 μsec.

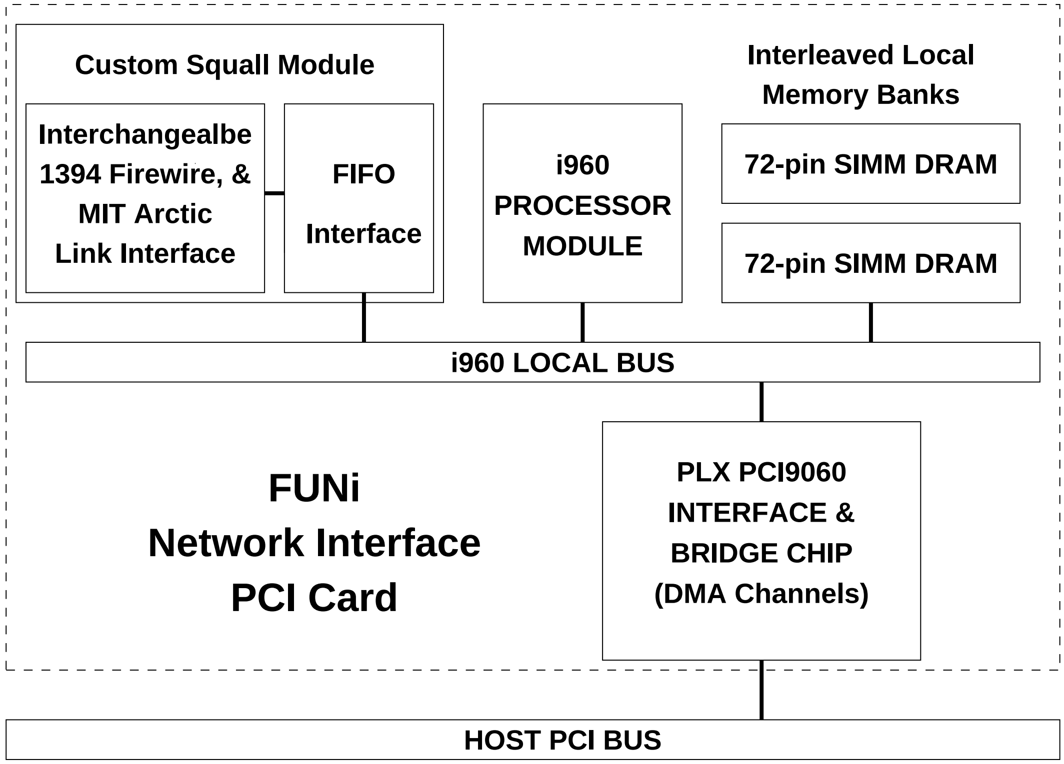

Figure 1: FUNi Network Interface PCI Card: Datapath

2 Acknowledgements

This research is supervised by Professor Arvind and is supported in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-92-J-1310 and by an Office of Naval Research Graduate Fellowship. Special thanks goes to Intel for donating the machines used in this project and to Texas Instruments for their assistances in manufacturing the 1394 interface cards.

References

[1] D. Chiou, et. al. StarT-NG: Delivering seamless parallel computing. Technical Report CSG Memo 371, MIT Laboratory for Computer Science, February 1995.

[2] J. C. Hoe. Network interface for message-passing parallel computation on a workstation cluster. In *Proceedings of Hot Interconnects II*, August 1994.