

# Preliminary Investigation of Active Memory Operations

*Lixin Zhang, Zhen Fang, John B. Carter,  
Mike Parker*

UUCS-04-009

School of Computing  
University of Utah  
Salt Lake City, UT 84112 USA

June 1, 2004

## ***Abstract***

We are rapidly approaching a time when large-scale shared memory supercomputers will have remote memory latencies measured in the thousands of cycles and cross-section bandwidth will be a limiting performance factor. For these machines to scale, mechanisms that minimize inter-processor communication will be essential. We propose one such mechanism, *active memory*, which allows operations to be sent to and executed on the home memory controller of particular data items. Performing the operations near where the data resides, rather than moving it across the network, operating on it, and moving it back, eliminates significant network traffic, introduces opportunities for additional parallelism, and hides high remote memory latencies. Active memory provides many of the benefits of PIMs without the need for non-standard DRAMs, and enables significantly better application scaling than conventional shared memory synchronization and range operations.

In this paper we investigate an active memory design that supports three classes of memory-centric operations that benefit common parallel constructs: *atomic scalar*, *range*, and *reduction* operations. We present architectural and programming models for active memory and compare its performance against a baseline conventional shared memory system implementation and a variety of optimized memory architectures. We find that active memory easily outperforms the conventional shared memory and other architectures by factors of over 10x on a collection of parallel constructs.