

# **Avalanche: A Communication and Memory Architecture for Scalable Parallel Computing**

John B. Carter, Al Davis, Ravindra Kuramkote,  
Chen-Chi Kuo, Leigh B. Stoller, Mark Swanson

UUCS-95-022

Computer Systems Laboratory  
University of Utah

## **Abstract**

As the gap between processor and memory speeds widens, system designers will inevitably incorporate increasingly deep memory hierarchies to maintain the balance between processor and memory system performance. At the same time, most communication subsystems are permitted access only to main memory and not a processor's top level cache. As memory latencies increase, this lack of integration between the memory and communication systems will seriously impede interprocessor communication performance and limit effective scalability. In the Avalanche project we are re-designing the memory architecture of a commercial RISC multiprocessor, the HP PA-RISC 7100, to include a new multi-level context sensitive cache that is tightly coupled to the communication fabric. The primary goal of Avalanche's integrated *cache and communication controller* is attacking end to end communication latency in all of its forms. This includes cache misses induced by excessive invalidations and reloading of shared data by write-invalidate coherence protocols and cache misses induced by depositing incoming message data in main memory and faulting it into the cache. An execution-driven simulation study of Avalanche's architecture indicates that it can reduce cache stalls by 5-60% and overall execution times by 10-28%.