

Design of a Parallel Vector Access Unit for SDRAM Memory Systems

Binu K. Mathew, Sally A. McKee, John B. Carter, Al Davis

{mbinu | sam | retrac | ald}@cs.utah.edu
Draft. Do not distribute

1

UUCS-99-006

Department of Computer Science
3190 Merrill Engineering Building
University of Utah
Salt Lake City, UT 84112
February 4, 2000

Abstract

Parallel Vector Access is a technique that exploits the regularity of vector or stream accesses to perform them efficiently in parallel on a multi-bank memory system. The performance of applications that have vector accesses may be improved using a memory controller that performs scatter/gather operations so that only the vector or stream elements that are accessed by the application are transmitted across the system bus. These scatter/gather operations can be speeded up by broadcasting vector operations to all banks of memory in parallel, each of which implements an algorithm to determine which elements of the requested vector they contain. This thesis presents the mathematical foundations behind one such algorithm for efficient parallel access of base-stride vectors on both word interleaved and cache-line interleaved memory systems. The design of a memory controller subcomponent that uses the Parallel Vector Access (PVA) algorithm to improve the performance of applications with strided access patterns is described. The hardware implementation issues behind such a memory controller are investigated. The the performance of such a memory controller on vector kernels is studied by gate level simulation and the results analyzed. Because of the parallel approach, the PVA is able to load elements up to 32.8 times faster than a conventional memory system and 3.3 times faster than a pipelined vector unit, without hurting normal cache line fill performance.

Keywords: memory architecture, memory latency, memory bandwidth, bus utilization, cache efficiency

¹This effort was sponsored in part by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) under agreement number F30602-98-1-0101 and DARPA Order Numbers F393/00-01 and F376/00. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of DARPA, AFRL, or the US Government.