
ELSEVIER

Parallel Computing 27 (2001) 1079–1115

PARALLEL COMPUTING

www.elsevier.com/locate/parco

Performance analysis of the simultaneous optical multi-processor exchange bus

Constantine Katsinis

Electrical and Computer Engineering, Drexel University, Philadelphia, PA 19104, USA

Received 19 March 1998; received in revised form 31 July 2000

Abstract

The performance of a multi-computer system based on the simultaneous optical multi-processor exchange bus (SOME-Bus) interconnection network is examined using queuing network models under the message-passing and distributed-shared-memory (DSM) paradigms. The SOME-Bus is a low latency, high bandwidth, fiber-optic interconnection network which directly links arbitrary pairs of processor nodes without contention. It contains a dedicated channel for the data output of each node, eliminating the need for global arbitration and providing bandwidth that scales directly with the number of nodes in the system. Each of N nodes has an array of receivers, with one receiver dedicated to each node output channel. No node is ever blocked from transmitting by another transmitter or due to contention for shared switching logic. The entire N-receiver array can be integrated on a single chip at a comparatively minor cost resulting in $O(N)$ complexity. By supporting multiple simultaneous broadcasts of messages, the SOME-Bus has much more functionality than a crossbar, allowing synchronization phases and cache consistency protocols to complete much faster. Simulation results are presented which validate the theoretical results and compare processor utilization in the SOME-Bus, the crossbar and the torus, with and without synchronization. Compared to these two networks, the SOME-Bus performance is least affected by large message communication times. Even in the presence of frequent synchronization, processor utilization remains practically unaffected while it drops in the other architectures. Although it has a larger number of channels compared to the crossbar and the mesh, the SOME-Bus is much simpler and inexpensive because it is free of complex routing, congestion and blocking.

Keywords: Computer architecture; Interconnection networks; Performance analysis

E-mail address: ckatsini@ece.drexel.edu (C. Katsinis).

0167-8191/01/$ - see front matter
PII: S0167-8191(01)00071-0

1. Introduction

High performance computing is required for many applications, including simulation of physical phenomena, simulation of integrated circuits and neural networks, weather modeling, aerodynamics, and image processing. It has been relying increasingly on microprocessor based computer nodes which are rapidly becoming the technology of preference. Groups of these nodes are interconnected to form a distributed-memory multi-computer system. Such systems are scalable and capable of high computing power. Difficulties in the design and use of such systems arise because the required computers need interconnection networks with high bisection bandwidth and low latency to connect hundreds of nodes. Current massively parallel computers use small degree networks with large diameters. Wormhole routing achieves lower latency when the network is not heavily loaded. Often performance is poor or moderate with many modern large-scale applications, mostly due to load imbalance, barrier synchronization, and communication patterns, such as all-to-all communication [26], which place an excessive load on the interconnection network. The major reason of the moderate success lies in the nature of currently available interconnection topologies (trees, hypercubes and mesh networks), regardless of their actual implementation medium, and in the mismatch between interconnection architecture and application structure. Most applications have irregular and dynamic communication patterns which change as the application progresses.

A review of many researchers' experiences with large applications on modern multi-computers reveals that, even after extensive efforts of software tuning, only moderate success is encountered. Experiments have been performed in various multiprocessor systems including IBM SP-1 and SP-2, the Cray T3D, Thinking Machines CM-5, Intel Paragon and Delta, and the SGI Power Challenge XL. Researchers have studied communication patterns and costs, and the effects of load imbalance on the performance of various applications, including atmospheric models (chemical tracer, general circulation), three-dimensional Navier–Stokes solvers, and N-body simulations. It has been observed that processing in this type of applications is based on two-dimensional and three-dimensional FFT and requires a series of complete exchange operations, as well as global reduction and gather operations. Individual steps are followed by global synchronizations.

Researchers find moderate to severe performance degradation [19], increasing effects of load imbalance and communication costs as the number of processors is increased [16], processor utilization between 30% and 40% even when attempting to minimize interprocessor communication through the use of a substantial amount of (infrequently performed) local bookkeeping [34], or even with code that has been optimized for years [5].

There is a large amount of research in multi-cast communications in popular architectures with path-based broadcasting [27], and trees and (multi-destination) wormhole routing [13,31,37]. Large efforts are focused on development of extensive algorithms to alleviate the fact that intense multi-cast communications cause wormhole routing to resemble store-and-forward routing. Experiments on the Paragon, SP-2 and CS-2 using the multi-phase complete exchange are described in [10],