SUMMARY OF Ph.D. DISSERTATION

School Science for Open and Environmental Systems	Student Identification Number	SURNAME, First name TSUJI, Akira

Title

A Study of Low Latency Communication Mechanisms on a Network Interface Attached into Memory Slot

Abstract

Recently, a PC (Personal Computer) cluster, which is consisting of many PCs connected each other with networks, has been a mainstream of high performance computing in most of enterprises and laboratories. For networks in such PC clusters, Myrinet, QsNET and InfiniBand as dedicated networks are used as well as Gigabit Ethernet. The network interfaces for such networks are attached into general I/O buses on PCs. However, the de-fact standard I/O bus; PCI bus running at 33 MHz with 32-bits data width supported only 132 MByte/s throughput, and it was far less than that of the networks. On servers and high-end workstations, high performance I/O buses; PCI bus with 64-bits data width and PCI-X bus were equipped. However, their expensive cost often spoiled the high degree of performance per cost, which is the fundamental benefit of PC clusters.

In contrast, the performance of memory system is higher than that of general I/O buses in order to adapt to the improvement trend of host processors. MEMOnet, a network interface attached into memory slot, was proposed in 1999 to make the best use of this performance. One of the forms of MEMOnet attached into DIMM slot is called DIMMnet.

This research is about designing a network controller logic for DIMMnet-2 which is the second generation of DIMMnet attached into the DDR-SDRAM memory slot, and supports the low latency and high throughput communication. The DIMMnet-2 prototype board used for implementation of the network controller has an FPGA (Virtex-II Pro) with a high speed serial transceiver, two DDR SO-DIMMs and an IEEE 10GBASE-CX4 connector. DIMMnet-2 is connected to InfiniBand network (4X: 10Gbps) using the transceiver and the connector.

All primitive operations on DIMMnet-2 have been implemented with hardwired logic in the FPGA to improve the communication performance. The performance at micro-benchmark level has been evaluated on the controller. The results indicate that the lowest unidirectional latency of BOTF (Block On-The-Fly) communication mechanism is 0.632us. Although the BOTF is for short messages using PIO (Programmed I/O), the throughput is reached at 631.11 MByte/s with unidirectional communication and 1163.70 MByte/s with bidirectional by issuing multiple BOTF requests continuously. They are even equal to those of other recent high performance networks. Thus, it is shown that the general PCs are able to get high communication performance by utilizing the memory slot.

Moreover, the message-receiving mechanisms, IPUSH (Indirect PUSH) and LHS (Limited-length Head Separation), are proposed and have been implemented. These mechanisms support the processing of message passing like MPI (Message Passing Interface). In the result of the evaluation, the efficiency of memory usage and the improvement of the overhead of comparing message tags are showed. These mechanisms are able to be applied to other networks attached into general I/O buses, and the performance improvement is expected on the parallel distributed computing systems using message passing.