Experimental Demonstration of 10 Gigabit Ethernet-Based Optical Interconnection Network Interface for Large-Scale Computing Systems

Wenjia Zhang*†, Ajay S. Garg*, Howard Wang*, Caroline P. Lai*, Jian Wu†, Jintong Lin†, and Keren Bergman*

*Department of Electrical Engineering, Columbia University, 1300 S. W. Mudd, 500 West, 120th St., New York, New York 10027, USA
† Key Lab of Information photonics and Optical Commun. (MoE), Beijing University of Posts and Telecom, Beijing, 100876, China
wz2207@columbia.edu

Abstract: A novel 10GE-based optical network interface card is proposed to connect high performance processors and an optical interconnection network. We demonstrate end-to-end link setup, and TCP and HD video streaming utilizing 4×3.125Gb/s wavelength-striped WDM payloads.

1. Introduction

The accelerated growth in the performance of microprocessors and the emergence of chip multiprocessors, which are widely leveraged in current data centers (DCs) and high-performance computing (HPC) systems, have motivated the need for developing novel technologies for interconnection networks. Characterized by high bandwidths, low power dissipations, time of flight latencies, and distance immunity, optical interconnection networks (OINs) have been proposed as a promising solution [1]. In recent years, there have been considerable research efforts in designing OINs including OSMOSIS [2], Data Vortex [3] and DOS [4]. These networks leverage the inherent high bandwidths provided by wavelength-division multiplexing (WDM) and exploit the high level of photonic capabilities enabled by innovative optical technologies. However, in order to truly capitalize on the benefits of an optical network, an efficient interface must be developed to support a transparent interconnection between high-performance electronic processors and the OIN. This interface must not only support high-bandwidth, low-latency communication; it must also represent itself as a switch compatible with traditional standardized networking protocols. To some extent, this concept has been largely overlooked by researchers both in developing the HPC systems and OINs. Indeed, this has become one of the main challenges facing the migration from current electronic interconnected large scale computing architectures to a next-generation optical interconnect paradigm [5].

We propose an Optical Network Interface Card (O-NIC) that can handle protocols widely used in today's interconnection networks, such as 10 Gigabit Ethernet (10GE), Infiniband and PCI-E [6, 7]. The envisioned O-NIC will have the ability to translate the electronic packets from the aforementioned protocols to the specific formats required by various prospective optical networks, including hybrid (electronic/optical and packet/circuit switched) networks, which are capable of supporting the diverse requirements imposed by large scale computing systems. In particular, by implementing virtual switching, flow control, and traffic statistics monitoring, the O-NICs create a virtual networking platform that features flexibility and scalability in dealing with topology discovery and failure notification by centralized or distributed control [5].

Ethernet in its current form has evolved from its implementation in Local Area Networks (LANs) and continues to be deployed today in large scale computing systems. Given the dominance of Ethernet and its compatibility with TCP/IP, we proposed the following design for a 10GE-based O-NIC (Fig.1). The 10GE-based O-NIC as implemented here mainly contains a 10GE network interface card (NIC) in end host and a development board, connected by Quad Small Form-factor Pluggable (QSFP) cables. In this work, we experimentally demonstrate: (1) the setup of a transparent physical layer Ethernet link, consisting of 4×3.125 Gb/s lanes, between commercially available commodity computers across optical links, and (2) the end-to-end transmission of Transport Control Protocol (TCP) and (3) high definition (HD) video over optics.

2. 10GE based Optical Network Interface Card

The NIC is a specialized hardware component, originally intended to connect CPUs to a network. In our design (as shown in Fig.1), a 10GE-based O-NIC consists of a commercialized 10GE NIC extended by a high speed Field Programmable Gate Array (FPGA) on the development board connected via a 10-Gigabit Attachment Unit Interface (XAUI). The XAUI supports four lanes of 8b/10b encoded 3.125-GBaud signals with an aggregate date rate of 10 Gb/s.

The logic design in the FPGA is illustrated in Fig. 1. On the electronic side, an Ethernet packet is transmitted from the CPU to the O-NIC, where data is de-serialized, aligned, 8b/10b decoded in the transceiver, and passed to self-defined modules. These modules are responsible for parsing the Ethernet header information, transferring clock domains and buffering effective data packets. The parsed information is then delivered to a virtual network function module for further analysis and the eventual control of an optical switch performed through optical network control interface and/or by generating optical headers for
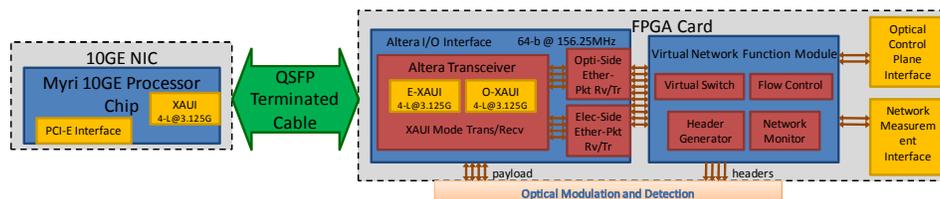

Fig.1 Optical Network Interface Card Architecture

distributed network routing [8], while concurrently submitting statistics to a management server through network measurement interface. On the optical side, the XAUI-based Ethernet payload capitalizes on the wavelength parallelism provided by the optical data channels (C36-C39 in the ITU grid). The 4×3.125 Gb/s optical data are transmitted over optics in this experiment, with link training via Align/Sync/Idle sequences (K28.0, K28.3, and K28.5) transmitted during channel idle time.

3. Experimental Demonstration and Discussion

In order to verify the viability of 10GE packets over an optical network, we construct a complete end-to-end testbed (Fig. 2 (a)), which incorporates two 10 Gb/s Myri-NIC equipped 64-bit PCs and two O-NICs implemented PLDA development boards connected by optical links. As shown in Fig. 2 (a), the Ethernet link originates from the 10GE NIC in the host computer. The NIC is then connected to an Altera Stratix II GX FPGA embedded development board through transceivers configured to the XAUI protocol. The transceivers then drive LiNbO3 modulators used to map the data onto four optical channels, namely 1548.51nm (C36), 1547.72nm (C37), 1546.92nm (C38), and 1546.12nm (C39). Finally, the data is received by transceivers on a second FPGA via p-i-n receivers with transimpedance amplifier (TIA) and limiting amplifier (LA) pairs. A DC block is inserted inline to achieve the required AC coupling to the development board. A standard reference clock (156.25MHz), which is used by the FPGAs, is provided by a signal generator through an RF splitter. In this experiment, the upstream traffic is looped back using electronic links.
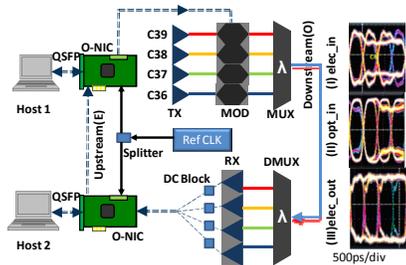

Fig.2 (a) Schematic diagram of experiment setup (b) Eye Diagram of C36

First, we confirm error-free transmission across four lanes over optics. Eye diagrams for the payload channel corresponding to C36 at the electronic transmitter, optical transmitter and electronic receiver are shown in Fig.2 (b). Given the pre-emphasis generated by the FPGA's analog transmitter, the eye diagram for the data mapped onto the optics is distorted as shown in Fig.2 (b.I). The eye diagram measured at the output of the receiver (Fig.2 (b.III)) is truncated due to the binary nature of the limiting amplifier. The open eye diagram shown in Fig.2 (b) confirms the signal integrity of the interface hardware over optics.

Second, we demonstrate an end-to-end TCP stack transaction over our implemented O-NIC. NTTCP [9] is used as an evaluation tool to measure TCP throughput. A standard 1500-byte maximum transfer unit (MTU) is used. Fig. 3 shows the TCP throughput increasing with sending data size. This is due to larger data lengths imposing fewer

loads on the CPU. It also shows that the throughput performance is improved by increasing the socket buffer size, giving more room to packets ready for transmission. This work does not aim to achieve optimized TCP throughput, which is related to performance bottlenecks in the end node. However, the stabilized TCP throughput (about 1500 Mb/s) as demonstrated here is the evidence of the potential for a 10GE-based optical interface system.
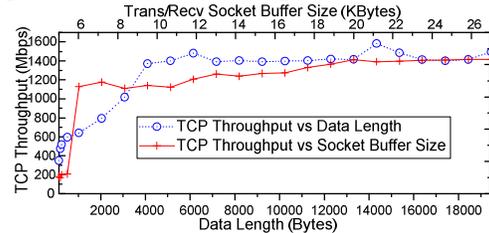

Fig.3 Throughput Performance with Different Sending Data Size and Different Transmission and Receiving Socket Buffer Size
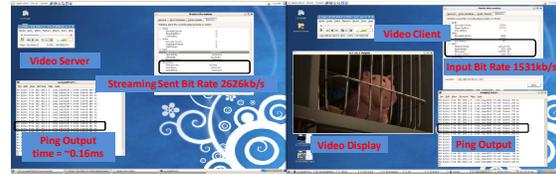

Fig.4 HD Video Streaming Demonstration (a) VLC server (b) VLC client

Lastly, we successfully demonstrate HD video streaming between a VLC [10] server and client via the O-NICs and optical testbed. Fig. 4 presents screen shots captured at hosts. The video stream is smooth without distortion or frame loss, verifying that the interface is performing as designed. We also disconnect the optical link manually and then reconnect after several seconds, resulting in the pausing of video and subsequent resumption as soon as the link is restored, demonstrating the robustness of this O-NIC design in dealing in the presence of link failures.

4. Conclusions

In this work, we successfully demonstrate the setup of an end-to-end link, and TCP and HD video transmission via the proposed 10GE based O-NIC assisted optical testbed. The experimental results show signal integrity in the physical links, with a peak throughput of TCP transmission of approximately 1500 Mb/s, and uninterrupted streaming of HD video. The O-NICs, which are able to handle traditional protocols and represent transparent, high-bandwidth and low-latency virtual network platform, are indispensible in future for large-scale optically-interconnected computing systems.

5. References

1. D. A. B Miller, Proc. IEEE, vol. 88, no. 6, pp 728-749, Jun 2000.
2. R. Luijten, et al, OFC 2009, paper OTuF3.
3. O. Liboiron-Ladouceur, et al, JLT vol. 26, no. 13, pp.1777–1789, 2008.
4. X.Ye, et al, 2010 ACM/IEEE Symposium on ANCS.
5. K. Bergman, ECOC 2010, paper Mo.2.D.1, Sep. 2010.
6. H.Wang, et al, OFC2009, paper OTuA4, Mar. 2009.
7. O. Liboiron-Ladouceur, et al. Opt. Express 17 (8), 6550-6561 (2009).
8. C. P. Lai, et al, PHO, paper ThQ2, Nov.2010.
9. http://freeware.sgi.com/Installable/nttcp-1.47.html
10. http://www.videolan.org/vlc/