Bandwidth Enhanced Photopolymer Waveguide Hologram Based Optical Backplane

Hai Bi, Chuhua Tian and Ray T. Chen, Fellow, OSA, SPIE, IEEE
Microelectronics Research Center
Department of Electrical and Computer Engineering
The University of Texas at Austin
PRC/MER 1.606G, 10100 Burnet Road, Austin, TX, USA 78758
Tel: +1-512-471-7035, Fax: +1-512-471-8575
E-mail: raychen@uts.cc.utexas.edu

Xuliang Han, Member, SPIE
Brewer Science, Inc.
2401 Brewer Drive, Rolla, MO 65401

ABSTRACT

As multiprocessing comes into the mainstream, the board-to-board interconnects become even more critical. In a shared-memory multiprocessing system, the shared bus topology is the preferred interconnect scheme because its broadcast nature can be effectively utilized to reduce communication latency, lessen networking complexity, and support cache coherence. In the electrical domain, however, a major performance bottleneck is anticipated due to the restricted bus bandwidth. In this paper, an innovative architecture, optical centralized shared bus, is proposed for use in the multiprocessing systems. This architecture utilizes the terascale bandwidth capacity of substrate-guided optical interconnects, while at the same time, retaining the essential merits of the shared bus topology. Thus, a smooth migration with substantial multiprocessing performance improvement is expected. A conceptual emulation of the shared-memory multiprocessing scheme is demonstrated on a generic PCI subsystem with an optical centralized shared bus. The objective of this effort is to prove the technical feasibility from the architecture standpoint.

Keywords: Optical Interconnects, Multiprocessing, Shared Bus, Switched Medium, Electro-Optical Interface, Vertical-Cavity Surface-Emitting Laser (VCSEL)

1. INTRODUCTION

Backplanes are used in high performance computer systems to accommodate multiple processors and memories. There may be up to 100 processors on one backplane. For such a computing system, one dominant factor that influences the performance is the interconnect [1]. There are electrical switches or buses on a backplane to interconnect the processors and memories. Electrical interconnects face numerous challenges such as signal integrity, power consumption, electromagnetic interference (EMI), and skin effect at high speed. Currently a typical electrical backplane bus operates at a frequency of less than 400MHz, whereas the speed of the state-of-the-art microprocessors has already surpassed 3GHz. This trend of computing speed outpacing interconnect capacity is becoming more and more prominent. Meanwhile, the next generation networks are envisioned to deliver beyond 10Gbps throughput to terascale grid-based applications. Therefore, a major performance bottleneck is anticipated at the board-to-board hierarchical level. Optics is well known for its interconnect capability [2], [3]. The success of optical interconnects has emerged at the machine-to-machine hierarchical level. To prevent the projected bottleneck from throttling the board-to-board data transfers, a new opportunity exists for the further exploitation of optical interconnects to replace the conventional electrical interconnects inside a box [4].
Besides the interconnect data rate, the architecture of an interconnect also influences the performance of a system. One significant challenge in the design of a multiprocessing system is to efficiently fulfill the communications among several processes that are simultaneously running on multiple processors. The shared bus topology was the preferred interconnect scheme because its broadcast nature can be effectively utilized to reduce communication latency, lessen networking complexity, and support cache coherence in a multiprocessing system [5]. However, the physical length, the number of fan-outs, and the operation speed of the backplane bus impose strict limitations on electrical interconnects. Thus, the switched backplane with switch fabrics and simple point-to-point interconnections is currently being employed in the electrical domain. By changing the backplane topology from the shared bus to the switched medium, however, several crucial performance aspects are compromised. For example, multiple processors may share a common variable in their cache to speed up calculation. If the shared variable is changed, the new value should be broadcasted to all processors that have cached the original value. The switched medium cannot carry out broadcast as effectively as the shared bus; therefore, the switch fabric has to execute the cache value update one by one. Besides the additional latency of going through the intermediate switching nodes, signal routing introduces substantial delay and considerable complexity, which has become a throttling factor in the high-end multiprocessing systems [6]. Therefore, an innovative optical technology that can provide sufficient bandwidth capacity, while at the same time, retaining the essential merits of the shared bus topology is highly desirable for multiprocessing performance improvement.

As one of the most significant contributions to the efforts on optical backplane bus, an innovative architecture called optical centralized shared bus was developed [7]. To the best of our knowledge, this is the first architecture that is able to achieve equalized bus fan-outs in the optical domain. Since the critical optical/electrical interface becomes uniform across the entire backplane bus, this merit can considerably save the system power budget to maintain the required bit error rate (BER) and substantially ease the overall system integration. Based on this architecture, we propose in this paper to apply the optical centralized shared bus in the multiprocessing systems for performance improvement. After a brief review of the architectural features of the centralized shared bus in Section 2, a preliminary feasibility demonstration on a generic PCI subsystem will be presented in Section 3. Finally, a summary is given in Section 4.

2. OPTICAL CENTRALIZED SHARED BUS ARCHITECTURE

Figure 1 illustrates the architectural concept of the optical centralized shared bus [7]. For simplicity, only five slots (#A1, #A2, #B1, #B2, and #C) are drawn in this schematic. A distributor board is to be inserted into the central slot (#C), while the other slots (#A1, #A2, #B1, and #B2) on the backplane bus are for processor boards. The electrical backplane provides interconnects for the non-critical paths. The electro-optical transceivers, including VCSELs (vertical-cavity surface-emitting lasers) and photodiodes, are integrated at the bottom of the electrical backplane, and aligned with the underlying optical interconnect layer. Therefore, the insertion/removal of the boards during the normal operations does not affect the critical alignment. Different from other modules, the positions of the VCSEL and photodiode for distributor board are swapped as indicated in Figure 1. The configured optical interconnect layer consists of a waveguiding plate with the properly designed volume holographic gratings integrated on its top surface. The plate provides a turbulence-free medium for optical interconnects, and the waveguide holograms function as optical fan-in/fan-out devices. Underlying the central slot (#C) is an equal-efficiency double-grating hologram, while the others are single-grating holograms. By employing such a unique configuration, both broadcastability and bidirectionality of signal flows on the backplane bus are enabled [7], which are the essential obstacles to achieving equalized bus fan-outs in other optical shared bus architectures [8], [9].
The optical centralized shared bus fits well in the shared-memory multiprocessing scheme from the architecture standpoint. The memories can be integrated on the central board functions as the centralized shared memory or distributed in all processors boards. For a write operation upon the shared memory, as illustrated in Figure 1, the VCSEL of the source processor board emits the light that carries the data and projects it surface-normally onto its underlying waveguide hologram. This light is coupled into the optical waveguiding plate by the grating and propagates within the confinement of the plate under the total internal reflection (TIR) condition [10]. Then, it is surface-normally coupled out of the plate by the central double-grating hologram and detected by the central photodiode. Because there is only one photodiode inside the electro-optical transceiver module for memory board, the data or data packet deliveries from the processor boards on the backplane to the memory board are forced in a sequential order as in the centralized shared-memory multiprocessing scheme, i.e., write serialization. The central VCSEL generates the outbound optical signal that carries the updated data and projects it surface-normally onto its underlying double-grating hologram. This light is coupled into the plate and equally diffracted into two beams by the hologram, propagating along the two opposite directions within the confinement of the plate under the total internal reflection (TIR) condition [10]. During the propagation, a portion of the light is surface-normally coupled out of the plate by the single-grating hologram underlying each processor board on the backplane and detected by the photodiode. By snooping on the shared bus, all processor boards can immediately obtain the updated data from the centralized shared bus and then either invalidate or update the cached copies. In this manner, cache coherence is consistently maintained across the whole system. There is only one write operation from the distributor board in order to broadcast any cache update. Compared to switches used in electrical backplane, the centralized optical backplane bus can reduce the delay to keep cache coherence.

The volume holographic gratings integrated on the top surface of the waveguiding plate function as optical fan-in/fan-out devices. Their diffraction properties in the Bragg regime can be analyzed with Kogelnik’s Coupled Wave Theory [11]. By balancing the diffraction efficiency of the waveguide holograms in use, the bus fan-outs across the entire optical interconnect layer can be equalized as demonstrated in Figure 2 [12]. This merit is highly desirable from the system integration standpoint because of the reduced constraint on the dynamic ranges of the electro-optical transceiver modules in use.
Compared with electrical interconnects, the most significant benefit of optical interconnects is the tremendous gain in the bandwidth capacity. We chose commercial available VCSEL and photodiode in our system. A 2.5G signal source is used to generate eye diagram of the system, shown in Figure 3. An 8.79 Q factor indicates that the bit error rate is below $10^{-18}$. However, the bandwidth of signal source is not the limit of our optical bus. The main factors that influence the optical bandwidth include the line width of the VCSEL and the dependence of hologram diffractive efficiency on wavelength. Since the technology of VCSEL and external modulators is improving, we can expect that the speed limit of the polymer hologram based optical backplane bus is imposed by the polymer hologram itself. Using a 150fs laser pulse, we can measure the FFT of the original pulse and compare it with a FFT of output pulse. The bandwidth capacity per substrate-guided optical line was then experimentally characterized to be approximately 2.5THz [13] as shown in Figure 4. With such an enormous bus bandwidth, the interconnect bottleneck in the shared-memory multiprocessing scheme would be completely eliminated by employing the optical centralized shared bus.
3. DEMONSTRATION ON PCI SUBSYSTEM

The real implementation of the shared-memory multiprocessing scheme by employing the optical centralized shared bus certainly involves too many processor-specific issues. Meanwhile, the advanced microprocessors are upgrading at a rapid pace, probably with different micro-architectures from one generation to the next. With the focus on demonstrating the technical feasibility in a general scenario, a conceptual emulation of the optically interconnected shared-memory scheme was carried out on a generic PCI subsystem that incorporated an optical centralized shared bus, as showed in Figure 5.

![Figure 4. Signal Bandwidth of one Optical Channel](image)

![Figure 5. Conceptual Emulation Scheme on PCI Subsystem](image)
A PCI memory card is positioned between the SBC card and NIC card, functioning as the shared memory as in the multiprocessing scheme. The SBC card contains a 1.2GHz microprocessor and a North Bridge that controls the interface to the PCI backplane. The microprocessor can access to the PCI memory card through the North Bridge. The communications between the NIC card and the SBC card can proceed via the PCI memory card on the shared bus in a conceptually equivalent manner to the shared-memory communications among multiple processors. It is important to note that the word “centralized” used in this architecture doesn’t entail a centralized memory. It only means that there is a distributor in the central slot in order to balance the received optical power in each daughter board. Therefore, the optical centralized backplane bus can be used for systems with distributed shared memories also.

As shown in Figure 6, the optical centralized shared bus was integrated underneath the system chassis and aligned with the passive PCI backplane, where the equalized bus fan-outs were established across the entire optical interconnect layer. As a preliminary attempt, only PCI bus line AD02 was replaced by the optical interconnection link while the other electrical wires on the passive PCI backplane were remained. In order to incorporate the optical interconnect layer into the generic PCI subsystem, a special extension interface was developed, as shown in Figure 7, to be integrated with the AD02 pins of the PCI slots on the backplane. It contains an electro-optical transceiver module and the required logic controls in consistent with the PCI protocol. A single electrical PCI bus line carries bi-directional signal transmissions. Meanwhile, the commercial PCI core does not explicitly indicate the actual data transfer direction. The data transaction type, either read or write, is negotiated between the master and the target in an implicit manner involving several PCI bus signals [15]. Thus, a new PCI core or a logic-interpreting circuit should be implemented to generate the RACTIVE and TACTIVE control signal to appropriately coordinate the operations of the electro-optical transceiver modules during the PCI data transfers.

![Diagram showing shared-memory multiprocessing on PCI subsystem using optical interconnect](image)

Figure 6. Shared-Memory Multiprocessing on PCI Subsystem using Optical Interconnect
From the operating system standpoint, the PCI memory card was actually treated as a RAM disk after mounting a file system. To conceptually emulate the shared-memory communications among multiple processors in a shared-memory multiprocessing system, the same file was transferred from the NIC card to the PCI memory card, and then from the PCI memory card to the SBC card. On the shared bus, the signal waveforms during the PCI data transfers were captured in the real time by the bus analyzer card. This bus analyzer card was connected to a logic analyzer (HP1660ES) for the logic timing verification. In particular, the signal waveforms presented at the AD02 pins of the NIC card and the PCI memory card were displayed on an oscilloscope for the direct visualization of the implemented optical interconnection during the PCI data transfers. Figure 8 is one of the captured results during such tests, where Channel 1 displays the signal waveforms at the AD02 pin of the NIC card, which were the modulation inputs to the VCSEL driver inside its extension interface module, and Channel 2 displays the signal waveforms at the AD02 pin of the PCI memory card, which were the outputs from the edge detector inside its extension interface module. The obtained results of these tests verified the correct connectivity of the implemented optical interconnection link.
4. CONCLUSION

The optical centralized shared bus utilizes the enormous bandwidth capacity of substrate-guided optical interconnects, while at the same time, retaining the desirable architectural features of the shared bus. Its unique topological configuration enables the fulfillment of equalized optical bus fan-outs across the entire architecture, and thus a uniform electrical/optical interface can be obtained. This significant achievement is highly desirable from the system integration standpoint. Meanwhile, it is particularly pointed out in this paper that this innovative architecture fits well in the shared-memory multiprocessing scheme. As a preliminary attempt, a conceptual emulation of the shared-memory multiprocessing scheme was carried out on a generic PCI subsystem that incorporated an optical centralized shared bus. Since this research prototype originated from the existing system, the actual data transfers were still at the same standard PCI bus speed (33MHz) as without using optical interconnects. Apparently, the compromise is that it cannot exhibit any performance improvement since the terahertz bandwidth potential of optics is not utilized in the constructed prototype. Nonetheless, the objective of the demonstration presented herein is to prove the technical feasibility from the architecture standpoint. Because there is no doubt on the interconnect capability of optics, which has been confirmed both theoretically and experimentally, it can be projected for sure that the interconnect bottleneck of the shared bus in the shared-memory multiprocessing system would be completely eliminated by employing the optical centralized shared bus.

ACKNOWLEDGEMENT

The authors would like to thank BMDO, DARPA, ONR, AFOSR, and the ATP program of the State of Texas for supporting this work.

REFERENCES


