ABSTRACT

In this work, we show future research directions about optical computing systems based on Arrayed Waveguide Grating Routers. We analyze the optical integration at different layers of large-scale architectures. At board-level, we aim to demonstrate significant execution time improvements and energy efficiency of optically interconnected systems running real benchmark applications on a cycle-accurate simulator. We simulate tiled optical Multi-Socket Blades (MSBs), and we implement multiple optimization techniques to achieve different tradeoffs between performance and energy consumption and to prove the importance of chip-level optical integration. In the large-scale we show, through a network simulator, significant latency and throughput improvement running uniform distributed traffic for both intra- and inter-cluster optical-based architecture. A software framework, based on trace-driven approach, will be developed for benchmarking large-scale solutions. We show an experimental hardware setup for next future testbeds to demonstrate the efficiency, in terms of throughput and energy consumption, of the proposed optical architectures.

1. INTRODUCTION

Current data center and High-Performance Computing (HPC) systems utilize many electronic switches in cascaded stages supporting only a single data stream on each link using a store-and-forward method. Due to limitations in port-count and bandwidth of these switches, the inefficiency of the cascaded switch stages compounds, especially in terms of latency, throughput, scalability, and power consumptions. Furthermore, most of this energy consumption is independent of the load, and thus savings are impossible to come from any load balancing/scheduling techniques. Compared to electrical interconnects, optical provide (1) higher transmission bandwidth and lower energy consumption independently of distance, (2) inherent parallelism, (3) low interference and crosstalk, and (4) low parasitic. Also, optics offers wavelength routing capability not available in electronics with comparable size and, therefore, can achieve all-to-all interconnection between computing nodes without contention. As a result, optically interconnected computing systems can potentially achieve (1) higher scalability and energy efficiency, (2) high-density parallel links and buses overcoming input/output pin density limits, and (3) low-latency avoiding the need for including repeaters or switches with store-and-forward architectures.

2. RESEARCH OBJECTIVE

The proposed research pursues energy-efficient silicon photonic interconnects towards enhancement in energy efficiency compared to the current state-of-the-art, while supporting extremely high-throughput with low-latency, and power-aware operation adapting the resources to the workload, to greatly reduce energy consumption. These goals can be achieved exploiting arbitration-free, and contention-less wavelength routing capabilities offered by Arrayed Waveguide Grating Routers (AWGRs) [1]. Our research will aim to study through simulations and testbeds innovative large-scale optical interconnect architectures that exploit the unique properties of AWGR at different levels, i.e., inside a blade (computing node) and rack, and between racks and clusters composing the full architecture. The architecture is based on passive AWGR interconnection for intra-rack and intra-cluster communication and active AWGR-based topology for inter-cluster transmission. We will conduct architecture and performance studies via benchmark simulations to evaluate the performance and energy efficiency of the proposed technologies under realistic traffic scenarios. Furthermore, we will perform experimental testbed studies to confirm the achieved simulation results.

2.1 Multi-Socket Blades

Large-scale infrastructures typically adopt MSBs to increase both the computation density as well as the amount of resources that applications can access with limited overhead [2]. We'll analyze the integration of optical and electronic components enabled by advanced techniques such as flip-chip bonding and Through-Silicon-Vias (TSVs), where 2D/3D integration will reduce the distance between cache controllers on separate chips. Figure 1 shows a scheme of the proposed MSB architecture, with N sockets. We’ll consider optics at the processor packages and utilize a Network on Chip switch (named Hub in Figure 1), acting as the interface between the electronic (intra-chip) and the optical (inter-chip) AWGR-based all-to-all network.

![Figure 1. Optical MSB architecture with AWGR.](image-url)
various electronic and optical interconnect options (i.e., MSBs).

the core instruction set architecture, memory hierarchy and
(i.e., PARSEC-2.1, SPLASH-2 [5]) in order to allow detailed
The cycle accurate simulator can run multi-threaded applications
system computer architecture simulator [4], capable to model
grain approaches and simulators [6, 7] will be identified for
configurations made of thousands of cores. Therefore coarser
traces also focus on architectural issues trying to collect and replicate traces capable of maintaining lower-level traffic pattern details even in the large-scale simulations. Figure 3 shows a representation of the discussed software framework.

3.2 Hardware Demos

The potentials of the proposed silicon photonic technologies, system architectures, optimal operation, and control planes will be integrated into an experimental testbed. Towards this goal, we will develop silicon photonic LIONS with prototype compute nodes instantiated in Field Programmable Gate Arrays (FPGAs). The testbed will involve eight such FPGA boards, each with at least four compute nodes instantiated to interface with silicon photonic transceivers and AWGRs, utilizing multiple wavelengths from the silicon-photonic optical frequency comb generator. The proposed testbed demonstration will culminate in a successful experimental demonstration of the optical interconnected ≥ 32 compute node system in three hierarchies of LIONS with anticipated throughput > 95% and a significant reduction in the energy consumption.

Figure 2 shows some of the simulated achieved results for the MSB architecture in terms of execution time (a) and Energy Delay Product (EDP) (b) under real benchmark traffic and with different optimizations (i.e., Clock Data Recovery (CDR) and Dynamic Voltage and Frequency Scaling (DVFS)) compared to a multi-socket, state-of-the-art electronic baseline.

2.2 LIONS Architecture

We will develop on a simulation and testbed platform the proposed MSB architecture basing on silicon photonic Low-latency Interconnect Optical Network Switch (LIONS) [3]. This solution consists of silicon photonic AWGRs, modulators and detectors, silicon photonic lasers and amplifiers, and silicon photonic multi-wavelength optical frequency comb generators to pursue low-power interconnects. A passive $N \times N$ AWGR provides all-to-all communication among $N$ computing nodes in a flat topology using $N$ wavelengths. Thus extremely high-throughput, low latency, and minimal energy can be expected in such interconnected clusters. Further, we propose to pursue hierarchical optical interconnection to scale the network beyond $N$ nodes while minimizing the number of additional hops.

3. APPROACH

3.1 Simulations

The fine-grain investigation will be supported by specific tools for intra-blade architectural modeling and coarse-grain methodologies will be adopted for the global-scale analysis. A framework will be developed based, at the small-scale, on a full system computer architecture simulator [4], capable to model the core instruction set architecture, memory hierarchy and various electronic and optical interconnect options (i.e., MSBs).

The cycle accurate simulator can run multi-threaded applications (i.e., PARSEC-2.1, SPLASH-2 [5]) in order to allow detailed understanding of the interconnection network and its effect on application performance and energy consumption at the blade-level. However, these features come at the price of prohibitive simulation time that limits the scope of possible explorations of configurations made of thousands of cores. Therefore coarser grain approaches and simulators [6, 7] will be identified for supporting the design-space exploration of the extreme-scale MPI-based architecture. In our simulation framework, we will also focus on architectural issues trying to collect and replicate traces capable of maintaining lower-level traffic pattern details even in the large-scale simulations. Figure 3 shows a representation of the discussed software framework.

3.2 Hardware Demos

The potentials of the proposed silicon photonic technologies, system architectures, optimal operation, and control planes will be integrated into an experimental testbed. Towards this goal, we will develop silicon photonic LIONS with prototype compute nodes instantiated in Field Programmable Gate Arrays (FPGAs). The testbed will involve eight such FPGA boards, each with at least four compute nodes instantiated to interface with silicon photonic transceivers and AWGRs, utilizing multiple wavelengths from the silicon-photonic optical frequency comb generator. The proposed testbed demonstration will culminate in a successful experimental demonstration of the optical interconnected ≥ 32 compute node system in three hierarchies of LIONS with anticipated throughput > 95% and a significant reduction in the energy consumption.

Figure 2. MSB time (a) and EDP (b) results.

This work was supported in part under DoD Agreement Number: W911NF-13-1-0090.

5. REFERENCES