Abstract

Numerous challenges present themselves when scaling traditional on-chip electrical networks to large many-core processors. Some of these challenges include high latency, limitations on bandwidth, and power consumption. Researchers have, therefore, been looking for alternatives. As a result, on-chip nanophotonics has emerged as a strong substitute for traditional electrical NoCs. However, optical networks are not a panacea to all our problems. They still have significant issues with respect to high static power consumption and scalability.

In this work, we propose various techniques to decrease the static power consumption in on-chip optical interconnects. We start by proposing two novel techniques to reduce static power consumption in photonic on-chip networks – a neural network based method for laser modulation by predicting optical traffic, and a distributed and altruistic algorithm for channel sharing – that are significantly closer to a theoretically ideal scheme. In spite of this, a lot of laser power still gets wasted. We propose to reuse this energy to heat micro-ring resonators (achieve thermal tuning) by efficiently recirculating it. These three methods help us significantly reduce the energy requirements.

However, area and manufacturing yield restrict such designs to single chips with small number of cores (16-32). As a result, we propose a scalable photonic based multi-chip architecture by leveraging the non-uniform cache architecture (NUCA) scheme on top of a virtual chip in order to decrease inter chip communication and increase the hit rate in the last level cache.

After that we proceed to improve the performance of GPUs with the help of silicon nanophotonics. We propose an energy efficient and scalable optical interconnect for GPUs. We intelligently divide the components in a GPU into different types of clusters and enable these clusters to communicate optically with each other. In order to reduce the network delay, we use separate networks for coherence and non-coherence traffic. Moreover, to reduce the static power consumption in optical interconnects, we modulate the off-chip light source by proposing a novel GPU specific prediction scheme for predicting on-chip network traffic.

Finally, we highlight the fact that unless security concerns are addressed, it will not be possible for third-party vendors to sell optical communication solutions at a large scale. As a result, we proposed a novel, secure, and efficient optical network that is immune to eavesdropping, spoofing, replay, and message-removal attacks. We leveraged some properties of optical networks to bring the expensive cryptographic operations out of the critical path in order to limit the overheads associated with our scheme.