Application Note

Photonic Integration for AI hardware

JePPIX Platform for Photonic Integration

Photonic AI Accelerators

Optical Computing for AI

The concept of optical computing is in no means a novel concept, but struggled to compete with rapidly improving electronic systems[1]. With higher data transmission rates and larger data storage volumes on the horizon, however, new solutions for high-speed information processing, which at the same time, meet the demand for reduced energy usage and decreasing latency, are required. Classical Von Neuman computing architectures separate the computing functions of memory and processing and thus physically limit the achievable speed and low-energy consumption.

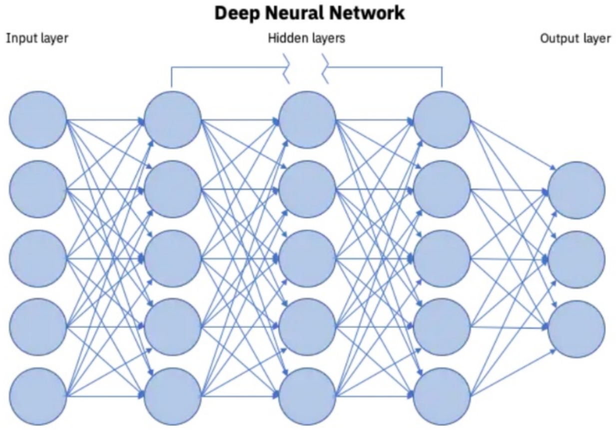

Figure 1 This schematic shows artifical neurons connected by artifical synapses. DNNs are comprised of multiple node layers including input, output and multiple hidden layers, from [2].

InP Photonic Integrated Circuits (PICs)

Optical chips or PICs can contain tens to hundreds of optical components. While electronic integrated circuits (EICs) consist of transistors, capacitors, and resistors, a PIC consists of, for example, lasers, modulators, photodetectors, and filters, all integrated on a single substrate. Several application fields, such as data- and telecom, sensing, and lidar are already using or are considering the use of PICs for their products. This PIC technology is accessible to users without a cleanroom, through so-called multi-project wafer runs and open access foundries. InP based technology is commercially available through SMART Photonics and Fraunhofer Heinrich-Hertz-Institut. Access is individually coordinated by JePPIX.

Deep Neural Networks

Deep neural networks (DNNs) have emerged as an attractive alternative hardware architecture, mimicking the brain's information processing, by adapting artificial neurons and artificial synapses connecting these neurons. The synaptic weights are defined by the strength of a particular connection. During the learning process, the DNN varies these weights while comparing its actual output to the desired output.

[1] https://www.nature.com/articles/d41586-019-01406-0

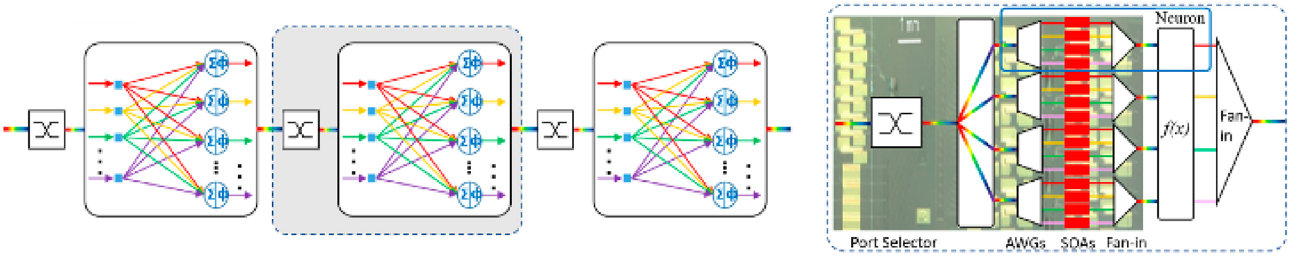

Figure 2 Scheme of a three-layer photonic deep neural network. The included port selector may be used to be used to select the desired input source. One full layer of neurons by exploiting onewavelength divisionmultiplexing (WDM) input,with a shaded photonic integrated circuit micrography at the back, to underline that part of the circuitry that is realized on chip. From [4].

Each neuron receives signals from a large number of other neurons and combines these inputs by means of a weighted sum. Only if this sum exceeds a certain threshold value, the neuron is activated and sends a signal to the next layer[2].

DNNs and other hardware accelerators are enablers for powerful artificial intelligence algorithms and deep learning applications, which are already used in everyday applications like face and speech recognition on mobilephones[3]. Photonic technologies are gaining more and more attention for the implementation of these hardware solutions. Optical fibers and optical waveguides offer vast bandwidths, allowing for wavelength division multiplexing (WDM), meaning that multiple signals can propagate together, yet be individually modulated. This allows for large-scale parallelization of processes and presents a scalable circuit architecture. Furthermore, integration of all-optical solutions allow direct processing of optical data, eliminating the electronic-to-optical and optical-to-electronic conversions.

Photonic Neural Networks

Photonic integrated circuits (PICs) can offer all of the required functions for neural networks and different approaches have been demonstrated. Micro-ring resonators have been implemented as arrays of synaptic weights[3], which have high potential for low-power multiply-accumulate operations but struggle with weight calibration[4]. Integrated Mach–Zehnder interferometers offer easier adjustment of weights and some early commercially available optical chips for artificial intelligence computing are currently emerging[5]. While these previous examples predominantly use the silicon photonic integration platform, the indium phosphide (InP) PIC platform can offer active photonic building blocks, introducing gain and leveraging the potential large-scale integration that includes laser sources and WDM systems on chip. Shi et al. [4] just recently demonstrated a DNN, realized by a InP-based, cross-connect circuit, using semiconductor optical amplifiers (SOAs) for weight calibration and array waveguide gratings (AWGs) to demultiplex the WDM input into individual channels. The 3-layer network is schematically shown in Figure 2. While still in early stage research, the combination of weighted addition functions, combined with on-chip non-linearities on the InP PIC platform shows a promising path for the future.

Discuss your application with us

If you are interested in knowing more about the capabilities and use of InP PIC technology for AI applications, contact JePPIX. The JePPIX Pilot Line provides low entrance-threshold to mature-manufacturing, enabling high-TRL development in a scalable design kit driven process, taking open access InP PICs from proof of concept to industrial prototyping levels.

[2] https://www.ibm.com/cloud/learn/neural-networks
[3] https://doi.org/10.1038/s41586-019-1157-8
[4] https://doi.org/10.3390/app10020474
[5] https://medium.com/lightmatter