

Photonics Chips for Machine Learning

Farhan Tanvir Khan

Boston University

Word Count: 1164

Photonics Chips for Machine Learning

Most of today's computers work with a centralized processing architecture. The central processing unit serves as the "brains" with a separate memory. For computational models that are distributed, parallel and adaptive, (such as those in neural networks) these architectures are inefficient. This is where photonic chips excel because of their internal connectivity and linear functions. In comparison to metal wire connections, a lot less energy is dissipated as heat, also. This literature review explores why and how photonic chips are used for machine learning in today's world and its future possibilities.

Among many other things, neuromorphic engineering tries to bring some components of machine learning and artificial intelligence to hardware. Synchronizing hardware to algorithms makes calculations much faster and at the same time making information processing more efficient. Unfortunately, although photons are heavily used as a means of communication in telecoms and data centres, it is not so popular in computations. This is due to the fact that the components used in said devices are incompatible with digital gates. Neural networks, however, are non-digital computing models and are more likely to be combined with photonics. Therefore, the objective of photonic chips should not be to replace conventional ones (at least the present ones available today), but work with them to enable faster computations where possible. Some examples where this could be integrated include:

- Solving partial differential equations
- Vector matrix multiplications
- Deep Learning
- Robotics

The main reasons as to why photonic circuits are so good at replicating neural networks are: interconnectivity and linear operations. We know that "neurons" in an artificial neural network are connected by a scalar weight. Therefore, the layout of interconnections can be thought of as a matrix-vector function where the input in every neuron comes from the output from its previous, connected neuron taking the weight factor into account. These approaches are nothing new, in fact they were pioneered decades ago by Psaltis (Optical information processing based on an associative-memory model of neural nets with thresholding and feedback) and Farhat (Optical implementation of Hopfield model). It is due to today's increase in computational demand and also because of the advancements in technology that the need for photonic neural networks has resurfaced.

"Lightmatter"- The photonic supercomputer company, as they call themselves, have created a photonic processor and interconnect that provides ultra fast processing of and doing it all the while keeping everything cool. They have focused on combining electronics, photonics and new algorithms to build a platform tailored for the purpose of artificial intelligence. They are claiming 5 times faster speeds and 7 times more energy efficiency than the NVIDIA A100 on BERT base in the same power footprint. Their products include the processor "ENVISE", an AI

inference accelerator. It brings together photonics & electronics under a single, compact packaging. They have an interconnect which they are calling “PASSAGE”, which allows different chips to communicate with each other at blinding speeds and being energy efficient at the same time. As for software stack, they named it “IDIOM”, and it interfaces with standard deep learning frameworks which provides all the necessary tools for those who deploy deep learning models. More information on the products and the codes can be found on their website. In an interview with John Koetsier, Nicholas Harris-CEO of Lightmatter speaks about the challenges faced in decreasing the size of transistors. While they are supposed to also decrease the energy used, that has not been the case for the past 15 years. So their mission statement is to use optical computing to reduce the energy used and to speed them up at the same time. Quantum tunneling is also another issue as transistors have been scaled down to the size of the electron and there are electron leakages happening very frequently. Currently, only one colour of light is being used for processing but as a future envision, Nicholas points out, multiple colors could be used to simultaneously run other processes and this would speed things up exponentially. This goes on to show that the future of photonics chips is indeed very bright. Lightmatter products are set to launch towards the end of 2021 and it will be the first photonic computer available for the general public.

There are also photonic circuit simulators available as open sources. “Symphony” is one of them. It is a fundamental package for designing and simulating PICs (Photonic integrated circuits) using Python programming language. Licenced under MIT and developed by CamachoLab at Brigham Young University, it is completely scriptable, runs on Windows, MacOS and Linux and includes model libraries from SiEPIC and SiPANN. Symphony comes with a built-in compact model library but is flexible to include custom ones as well. It is fast and simple to use. All that is needed is to define the components, add instances and lastly define the connection points before running the simulation. All the codes can be found on github.

Another simulator worth mentioning is “Noxim”- the Network on Chip simulator. Developed at the University of Catania in Italy. It is developed using SystemC (a description language based on C++) and is available for download under GPL license. They have a tutorial slide on the github which demonstrates how Noxim works. It illustrates how smaller and smaller technology and lower voltage leads to more noise. There are 2 main ways to use the simulator: 1) Performing simulations and 2) Modifying source code. Noxim deploys a command line interface for its many parameters of a NoC. The user can customize the size of the network as well as buffer size, size of the packet for distribution, the routing algorithm, selection strategy, etc. just to name a few. Different evaluation matrices can be collected including the number of packets, global average throughput, maximum or minimum global delay and the total energy consumed. The simulator comes with “Noxim Explorer”, a useful tool during the design space exploration stage. The explorer creates new configuration parameters for the user or even auto complete the exploration from a script.

Before we see an all-light chip we are in fact going to see a hybrid opto-electronic chip that uses the high speed of photons to transfer data quickly over large distances. And that future

might be sooner than later as transistors are reaching their lower size limit very fast and cannot, theoretically, be improved upon anymore.

References

Bhavin J. et al. *Photonics for artificial intelligence and neuromorphic computing*

Catania, V. et al. *Improving the energy efficiency of wireless Network on Chip architectures through online selective buffers and receivers shutdown*

Catania, V. et al. *Energy efficient transceiver in wireless Network on Chip architectures*

"Lightmatter". <https://lightmatter.co/>

"Beating Moore's Law: This photonic computer is 10X faster than NVIDIA GPUs using 90% less energy" *Youtube*. Uploaded by John Koetsier, 5 April 2021,
<https://www.youtube.com/watch?v=t1R7EIXEyag&t=183s>

"Symphony: A Simulator for Photonic circuits". <https://github.com/BYUCamachoLab/symphony>