Zhaohan Song

Professor Alshaykh

EC601-Project 1

18 September 2022

Research of Photonics chips for machine learning

Before 2010, the growth in computing power required for training was in line with Moore's Law, doubling approximately every 20 months（Fig. 1）. Since the introduction of deep learning in the early 2010s, the amount of computing power required for training has grown rapidly, doubling approximately every 6 months. Traditional machine learning relies on the graphical processing unit (GPU) because of its efficient parallel computing, which has good computational advantages for matrix multiplication and convolution. Then a new trend emerged in late 2015 as the need for training computing power increased by 10 to 100 with the advent of large-scale Machine Learning models. With the increasing demand for computation, the traditional approach to machine learning often takes longer to simulate and has more power consumption. To meet the requirement of the large-scale Machine Learning models, dedicated hardware to do machine learning is needed. Tensor Processing Units (TPUs) are Google's custom-developed application-specific integrated circuits (ASICs) used to accelerate machine learning workloads[1]. However, it is still difficult to balance low power consumption and high computing speed. That is why Silicon Photonics was brought up to solve problems.

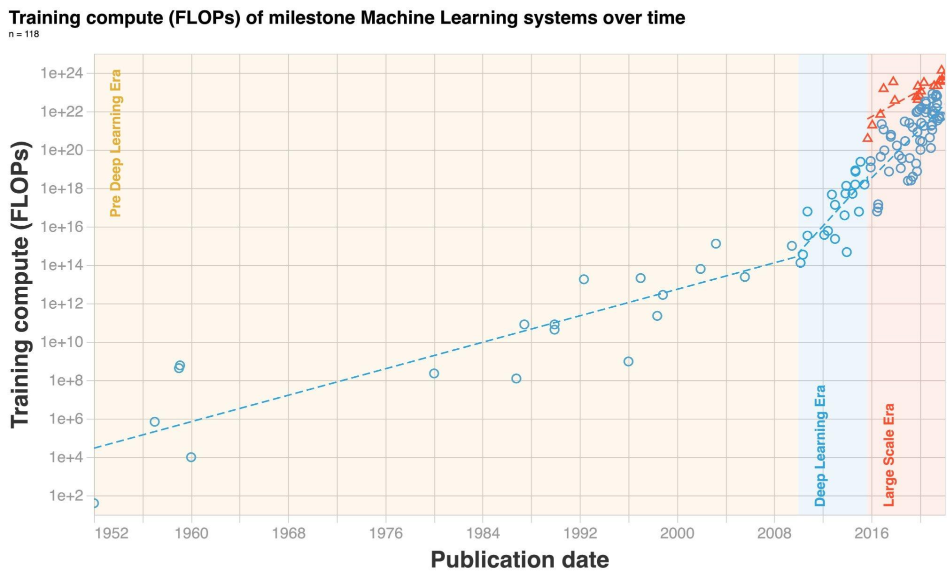
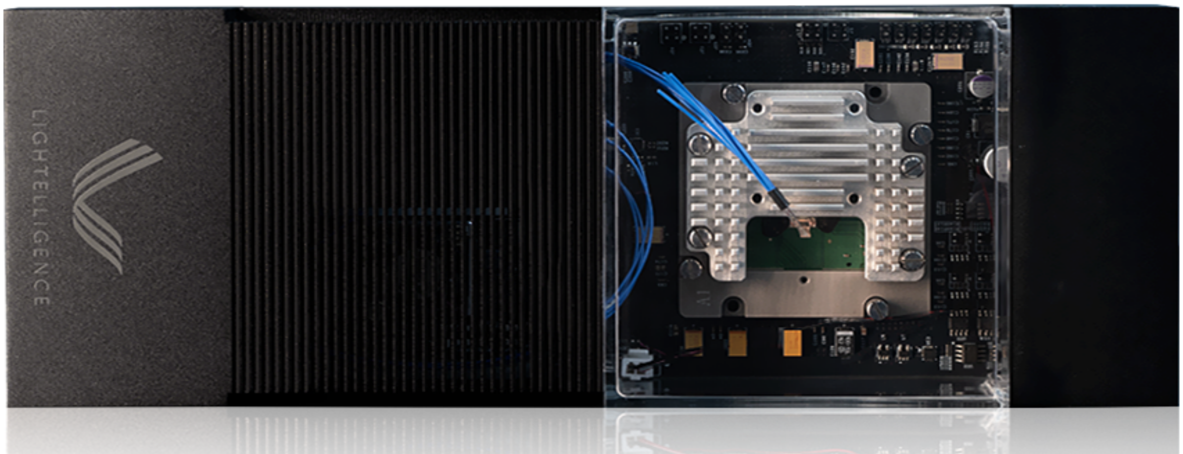


Fig. 1 Training compute (FLOPs) of milestone Machine Learning systems over time[2]

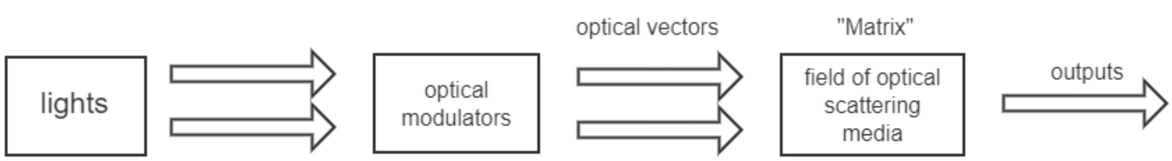
In traditional electronic computing, electrons are the carriers of information, but photons have long been considered a superior alternative. Because the spectrum covers a wide range of wavelengths, photons of different wavelengths can be simultaneously multiplexed (transmitted in parallel) and modulated (by changing the information they can carry) without the optical signals interfering with each other.

Since the convolution process involves passive transport, the computation of the photonic processing core can theoretically be performed at the speed of light and with low power consumption. This capability is valuable for energy-intensive applications, such as those used in cloud computing.

I learned how silicon photons could be used for machine learning through the photonic Arithmetic Computing Engine-PACE by lightelligence[3].



When light travels and scatters in nonhomogeneous media, it resembles the form of a linear mathematical operation. So we can use this feature to make photonics do matrix multiplication, which is widely used in machine learning. In a photonic matrix system, the light will be encoded by arrays of optical modulators to form input light vectors; so they can convey information by different light intensities and directions, which is quite similar to using voltages to represent digital 1 and 0. The optical vectors then enter a field of programmable light scattering media representing the matrix, and the output is the result of multiplication. Finally, using the photodetector to convert the light intensity into current and back into the digital system.



This system is basically a combination of an electronic device and a photonic device. The electronics contain digital and analog modules, the former used as control, I/O, and data storage, and the analog device is the bridge between the photonic device and the electronics, converting the signal stored in the digital module into an analog signal, and then transmitting the information into the light through an optical modulator. The photonic device is a dedicated computing module, which dramatically increases the overall system processing speed and significantly reduces energy consumption. But this new technology also poses several new problems. For instance, Silicon is an indirect band gap semiconductor. So when we want the light to connect to electronic devices by using lasers and photodetectors, All light sources must be fabricated with other materials such as III-V or Ge, which means that the full range of optics cannot be fabricated on the same standard CMOS SOI process. This introduces a number of cost issues. In addition, as the CMOS transistor size gradually decreases, the optics cannot continue to shrink due to wavelength and diffraction limitations, which is also worthwhile to continue to delve into the problem.

Works Cited

[1]https://cloud.google.com/tpu/docs/tpus?hl=en

[2] COMPUTE TRENDS ACROSS THREE ERAS OF MACHINE LEARNING

2202.05924.pdf (arxiv.org)

[3]Lightelligence Lightelligence

[4]https://spectrum.ieee.org/the-future-of-deep-learning-is-photonic

[5]https://www.nature.com/articles/s41586-020-03070-1

[6]https://www.nature.com/articles/d41586-020-03572-y

[7]https://arxiv.org/pdf/2203.02285.pdf

[8]https://journals.aps.org/prx/pdf/10.1103/PhysRevX.9.021032