# Electro-Optical Hybrid Fourier Neural Network with Amplitude-Only Modulation

**Zibo Hu[1], Mario Miscuglio[1], Shurui Li[2], Jonathan George[1], Puneet Gupta[2], and Volker J. Sorger[1]**
[1]*Department of Electrical and Computer Engineering, George Washington University, Washington DC, DC, USA*
[2]*Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, USA*
*sorger@gwu.edu*

**Abstract:** We demonstrate an electro-optical hybrid massively-parallel Fourier Neural Network exploiting Digital Micromirror Devices performing amplitude-only filtering, achieving ~10,000 2-Megapixel convolutions per second, preserving the inference accuracy level similar to phase-based approaches. 
**OCIS codes:** 200.0200.200.4260, 070.0070.070.5010, 200.0200.200.4960

## 1. Instruction

### *1.1. Convolutional Neural Network*

In the recent years, deep learning has thrived due to its ability to learn patterns within data and perform thus pave the way for intelligent decision making, superior in some cases to human possibilities. [1-3] At the heart of many emerging machine learning applications, such as image and video recognition, image classification, natural language processing, are deep Convolutional Neural Networks (CNNs), which are inspired by the visual neural structures of animals; the small regions of cells are sensitive to specific visual features (i.e. foveated). In other words, the type of sensitivities can extract the features from images in order to allow an accurate pattern recognition. In a CNN, convolutional layers are responsible for the feature extraction function, and thus, can perform accurately classification. Since convolutional processes are computationally overhead hungry (i.e. power, latency), specialized hardware such as Graphic Process Units (GPUs) and Tensor Process Units (TPUs) are used for CNNs providing compute acceleration with parallelization paradigms. However, GPUs and TPUs are rather power-hungry and require longer computation time (>10's of ms). Moreover, smaller matrix multiplication for less complex inference tasks (e.g. MNIST[4]) are still challenged by a non-negligible latency, predominantly due to the access overhead of the various memory hierarchies and the latency in executing each instruction in the GPU[5]. These shortcomings motivate our work of seeking compute-alternatives for CNNs.

### *1.2. 4F system*

Processing convolutions scales with $\boldsymbol{O}(n^2k^2)$, where $n$ and $k$ are the matrix side-lengths of a squared data input matrix and the kernel, respectively, and is hence overhead heavy requiring many MAC operations (multiplications accumulate) in the spatial domain, and are therefore often conducted as point-wise dot-product multiplication in the Frequency domain instead. However, this requires Fourier-domain transformations (FT) which is costly in electronics, however is passively performed in optics (by a lens). Such Fourier Neural Networks (FNNs) can map onto a CNNs using. Here two convex lenses perform the FT and inverse FT domain crossings forming a 4$f$ system. This design paradigm allows for massive papalism using digital display technology, while being power efficiency. [6]

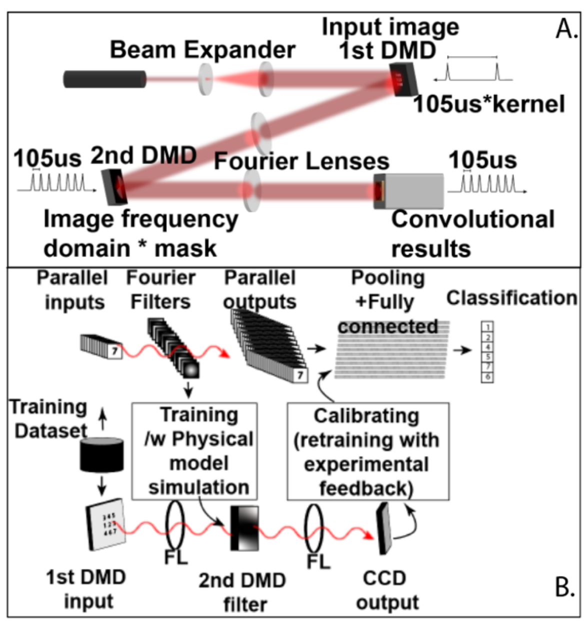


**Figure 1** (A) Experimental setup of massively parallel Fourier-optic CNN: a first DMD loads the inputs from MNIST or CIFAR-10 onto an coherent optical beam. A second DMD loads the trained frequency mask (kernel) with an update rate of 9.5kHz and filters the image. A high-speed camera completes one convolutional filter layer. (B) CNN Training Process: the entire optical system is modeled and used to optimized the systems-loss function on MNIST and CIFAR-10 data sets. The trained kernels are loaded into the optical system, processes and read. Completing one CNN layer, polling and a fully-connected layers are performed electronically, past the camera.

### *1.3. DMD based 4f system*

Typical 4*f* systems are based on the Spatial Light Modulator (SLM) or passive masks to control both amplitude and phase. Such approaches can be relatively energy efficiency compared to serial electronics, due to the $10^6$ channel optical parallelism. However the liquid crystal technology of SLMs are modulation-speed limited to <100Hz.

In contrast, the 4*f* CNN processor system introduced here is realized with Digital Micromirror Device (DMD, DLP6500, Taxes Instruments) offering rapid ~10kHz operation frequency. Other parts are: a sufficiently-long coherence laser (He-Ne 633nm), a 2-lens beam expander, 2 Fourier-lenses and a high-speed camera (IDT Y-7) (**Fig. 1A).** The processed images are sent to pooling and fully connected layer to complete classification performed electrically. Instead of using interferometric schemes or super-pixel approaches for considering also the phase, the DMD act as light valve performing amplitude-only convolutions.

## 2. Result

Our FNN has 3 layers, optical Fourier layer, electrical pooling layer and fully connected layer. We train our network in 2 steps. At first, we train in the computer by simulating the physical model. And then, we load the trained filters in the $2^{nd}$ DMD, read out the filtered images from camera and calibrate the fully connected layer to compensate the experimental errors such as misalignment, optical distortion and DMD angle distortion. (**Fig. 1B**).

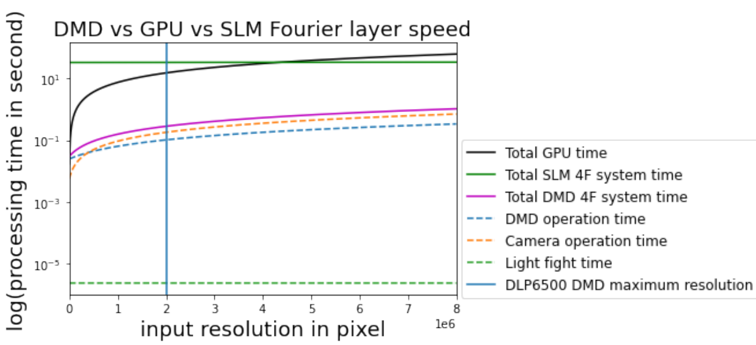


**Figure 2** Shows the operation time of our amplitude-only version, Nvidia P100 GPU and SLM 4F system. The vertical line is the maximum resolution for DLP6500 DMD. The DMD-based approach allows for an ~10x convolution processing speed up over the GPU.

| Model | MNIST | CIFAR |
|---|---|---|
| Space-domain convolution (full precision CNN) | 98% | 63% |
| Simulation model (Fourier convolution) | 98% | 62% |
| Hardware model (without calibration) | 92% | 25% |
| Hardware model (with calibration) | 98% | 40% |
| Spatial Light Modulator model[6] | NA | 44% |

**Table I**. Inference accuracies for a 1xlayer CNN: our simulated model, experimental realization without and with calibration, and SLM-based system.

By examining the processing speed, we found the amplitude-only 4*f* system enabling an about 10x speed improvement to process convolutions compared with a state-of-the-art GPU, when performing large matrix convolutions. The maximum resolution is 2K for DLP 6500 DMD, which is shown as vertical line in Figure 2. Despite the of the amplitude-only modulation system we find competitive inference accuracies compared to phase-based systems [6] (**Table I**).

## 3. Conclusion

Here we introduced an optical accelerator processing convolutions optically as dot-product multiplications in the Fourier domain. Using amplitude-only high-speed digital mirror displays we harness the best from massively parallel optics (million parallel channels), while offering an rapid ~10kHz kernel update rates, an speed-up of over 100x compared to SLM-based systems. Using this optical-electrical hybrid 1-layer CNN performing inference task on trained datasets (MNIST and CIFAR-10), we find a comparable accuracy performance to phase-based alternatives. Note, our actual accuracities testing was cut short due to COVID-related lab closures, and we expect this system to show higher performance using more data for system calibration, thus improving the physical model.

## References

[1] Assael, Yannis M., et al. "Lipnet: End-to-end sentence-level lipreading." *arXiv preprint arXiv:1611.01599* (2016).
[2] Google's AI Wizard Unveils a New Twist on Neural Networks. *Wired*
[3] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." *Nature* 518.7540 (2015): 529-533.
[4] Kochura, Yuriy, et al. "Batch size influence on performance of graphic and tensor processing units during training and inference phases." *International Conference on Computer Science, Engineering and Education Applications*. Springer, Cham, 2019.
[5] Volkov, Vasily. *Understanding latency hiding on gpus*. Diss. UC Berkeley, 2016.
[6] Chang, Julie, et al. "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification." *Scientific reports* 8.1 (2018): 1-10.