Invited Paper

# 4F Optical Neural Network Acceleration: An Architectural Perspective

Shurui Li[a] and Puneet Gupta[a]

[a]Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, USA

## ABSTRACT

Low latency, high throughput inference on Convolution Neural Networks (CNNs) remains a challenge, especially for applications requiring large input or large kernel sizes. 4F optics provides a solution to potentially accelerate CNN inferences with Fourier optics and the well-known convolution theorem. However, existing 4F CNN accelerators suffer from various limitations that make the implementation of a multi-channel, multi-layer CNN not scalable or even impractical. In this paper, we discuss the limitations of 4F CNN accelerators including the positive sensor readout, intensity-only modulation and slow modulation frequency and methods to address them. We also propose the channel tiling method that can address an important throughput and precision bottleneck of high-speed, massively-parallel optical 4F computing systems, not requiring any additional optical hardware.

**Keywords:** Deep learning, 4F Optics, CNN

## 1. INTRODUCTION

Neural networks are the core of modern AI technologies and have shown significance in many fields. However, efficiently deploying neural networks remains a challenge due to the massive amount of computation required. 4F optics demonstrates the possibility for low latency inference of Convolution Neural Networks (CNNs) by converting expensive convolution operations into Fourier transforms and point-wise multiplications that are computationally 'free' in the optical domain. Free-space 4F systems offer massive parallelism due to the high resolution of SLMs and phase masks, as well as efficient convolution computation using the convolution theory. The 4F system can be implemented using two Fourier lenses, an input source, a device for multiplication and an output sensor. Fourier transform, point-wise multiplication and inverse Fourier transform in the 4F system is essentially constant time (the time of flight of light). However, there are still many bottlenecks and limitations that need to be addressed to allow 4F optics to outperform state-of-art electronic accelerators. In this paper, we discuss the various limitations of 4F systems and how they could be mitigated to boost system efficiency. We also proposed the channel tiling method that can improve the accuracy and throughput of 4F CNN accelerators, under the condition that full light modulation (amplitude + phase) is supported.

## 2. DMD-BASED 4F SYSTEMS

The main advantage of 4F systems is the 'free' convolution operation carried out by the Fourier lens, which makes them suitable for applications that require low latency. A fixed phase mask can be placed at the Fourier plane to compute the point-wise multiplication. In this case, the convolution can be computed in one pass, at the speed of light. However, the weights encoded on the phase masks are fixed, which severely restricts the learning capability of the system. Only a single filter or at most a single convolution layer can be implemented with phase masks. Clearly, weight programmability is necessary for most applications.

Digital Micromirror Device (DMD) can be used to replace phase masks to allow weight programmability. DMD can modulate the intensity of incoming lights by flipping its mirrors and it can operate faster than liquid crystal Spatial Light Modulators (SLM). Two main advantages of DMDs are their high resolution and relatively

Further author information: (Send correspondence to Shurui Li)
Shurui Li: E-mail: shuruili@ucla.edu
Puneet Gupta: E-mail: puneetg@ucla.edu

AI and Optical Data Sciences III, edited by Bahram Jalali, Ken-ichi Kitayama, Proc. of SPIE Vol. 12019, 120190B
 · 0277-786X · doi: 10.1117/12.2614731

high frequency. With weight modulation, multi-layer CNN can be implemented by feeding the outputs back to the system. A reccent work[1] demonstrates a prototype of 4F CNN accelerator using DMDs. However, there are still some issues with DMD based 4F system. DMD only supports intensity modulation, causing normal spatial filters used in CNNs cannot be mapped on DMDs correctly as filters will become complex after Fourier transform. This can be addressed by training with real-valued Fourier filters so that the weights can be directly mapped on DMDs. Another issue is low DMD utilization. DMDs have up to 4K resolutions, which is normally much higher than the resolution of inputs and filters, especially for later layers of a typical CNN where inputs are heavily downsampled. If only one input and kernel (one channel of a filter) are loaded on DMDs at the same time, the system will be heavily underutilized and void any potential benefits.

Tiling can be used to improve system utilization. Ideally, the system can compute multiple convolutions in parallel by tiling multiple inputs or kernels on the DMD. However, as DMDs only support real-valued weights, kernel (weight) tiling is not possible (Tiling kernels in space domain will lead to complex kernel in Fourier domain). Input tiling will also suffer from crosstalk between tiled inputs, making individual convolution results cannot be recovered accurately. The crosstalk is caused by training the kernels in the Fourier domain with real-valued weights, as there is no control over the size of the corresponding space-domain kernels (can be treated as convolving tiled inputs with a single large kernel). In practice, the crosstalk between inputs can be minimized by separating tiled inputs with enough zero padding and applying a high-pass filter during training.[2] The high-pass filter removes the low-frequency components that tend to be the crosstalk between inputs. With input separation and high-pass filter, input tiling can be approximated hence improving the DMD utilization. A drawback of this approach is the accuracy will be impacted, as the high-pass filter also removes some useful information. In the simulation, we observe around 5% accuracy drop on a single-layer CNN using the CIFAR-10 dataset.

## 3. SYSTEM SCALING AND CHANNEL TILING

### 3.1 System scaling

The performance of the current generation of 4F CNN accelerators is bounded by the DMD switching frequency. With DMD's around 20KHz operating frequency,[3] 4F accelerators are not able to compete against state-of-art GPUs in terms of throughput. However, high-speed light modulation is promising as many research works have demonstrated concepts or prototypes of various types of high-speed SLM. Analog Micromirror Arrays (MMA) could be used as a replacement for DMD, they naturally support multi-bit mode without sacrificing operating frequency by controlling the tilt angle of mirrors through voltage. Analog MMAs are under active research and can achieve much higher switching speeds than commercial DMDs. For instance, two works (11M pixels, 2.3MHz and 2.2M pixels, 1MHz)[4,5] have demonstrated high-speed, high-resolution MMAs. The light modulation frequency could be further scaled with advanced materials. Another work[6] proposed a phase-only SLM architecture that uses microcavities with barium titanate to achieve GHz operation frequency with high-pixel resolution. If the concept can be materialized, 4F systems could potentially operate in the GHz regime.

### 3.2 Channel Tiling

The real-valued weight constraint is another bottleneck of 4F systems as it can lead to worse throughput and/or accuracy. By modulating both amplitude and phase, complex-valued Fourier kernels can be represented and tiling can be implemented in a lossless manner. Even though most SLMs can either modulate intensity[3–5] or phase,[6] full modulation can be achieved through Mach-Zehnder Interferometer (MZI) concept[7] (for intensity modulation) or methods that can support full modulation using phase-modulation SLMs[8,9] (for phase modulation). If complex-valued kernels are supported, besides input and kernel tiling, channel tiling can also be implemented to improve system utilization. Channel tilling essentially tiles the channels (z-dimension) of inputs and weights on input and weight DMD. The illustration of channel tiling is shown in Figure 1.

#### 3.2.1 Channel Tiling Procedure

Consider the case where $N_c$ input channels with size $M \times M$ convolve with $N_c$ corresponding filters with size $N \times N$. The normal convolution process (same mode) can be formulated by

$$Y(i,j) = \sum_{a=0}^{N} \sum_{b=0}^{N} \sum_{c=0}^{N_c} X(a+i, b+j, c) \times F(a,b,c) \tag{1}$$

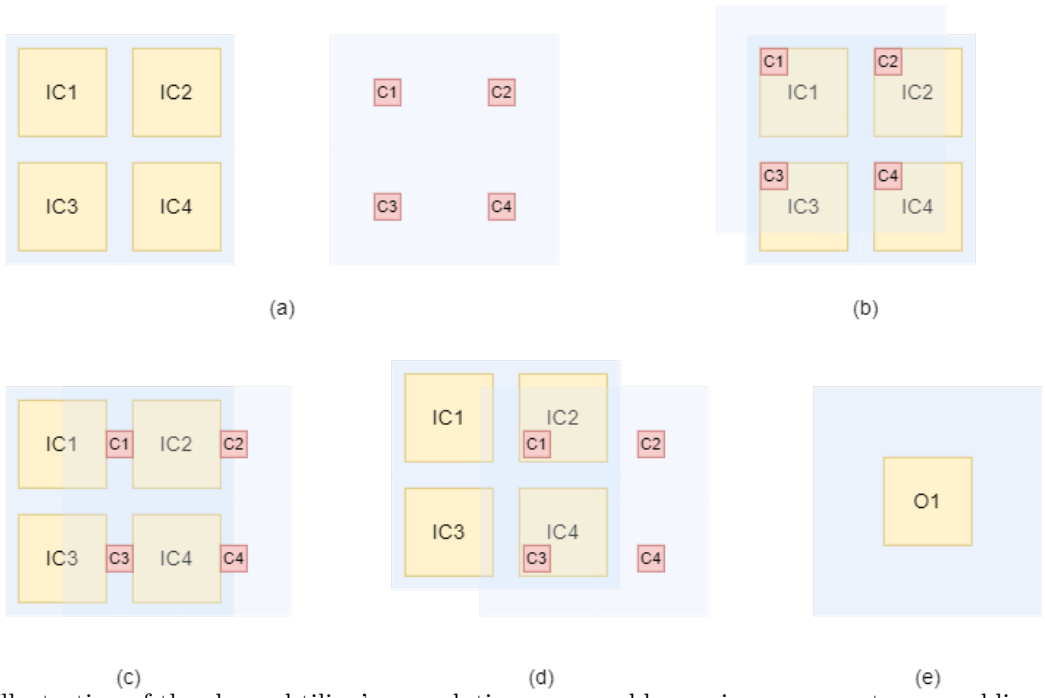



Figure 1: Illustration of the channel tiling's convolution process, blue regions represent zero padding. (a): Tiled input and kernel blocks. (b): Start point of the valid region. (c): Effect of padding, kernels will not overlap with multiple input channels. (d): Example of the invalid region, kernels are not convolved with their corresponding input channels. (e): Output format of channel tiling.

Table 1: Table of common notations

| Description | Notation |
|---|---|
| Input size | M × M |
| Filter size | N × N |
| SLM size | D × D |
| Total number of input channels | $N_c$ |
| Number of blocks can be tiled on SLM | T |

where $X$ is the zero padded input and $F$ is the convolution filter. Channel tiling method tiles input and filter channels on 2D planes thus the summation over channels in the above formula can be removed. The channels of both input and filter need to be zero-padded into blocks with size $(M+N-1)\times(M+N-1)$, to avoid overlap between single filter channel with multiple input channels. Then both the input channel blocks and filter channel blocks are tiled in the same order to form two large blocks $X_T$ and $F_T$ with size $\lceil\sqrt{N_c}\rceil(M+N-1)\times\lceil\sqrt{N_c}\rceil(M+N-1)$. For simplicity, denote the size of tiled input and filter blocks as $M_t \times M_t$, where $M_t = \lceil\sqrt{N_c}\rceil(M+N-1)$. The formula of computing convolution output can be written as

$$Y_T(i,j) = \sum_{a=0}^{M_t-1}\sum_{b=0}^{M_t-1} X_{TP}(a+i, b+j)\times F_T(a,b) \tag{2}$$

where $X_{TP}$ is the tiled input block $X_T$ and circular padded to size $(2M_t-1)\times(2M_t-1)$, which is inherently generated by Fourier transform. $F_T$ is the tiled filter block. The size of $Y_T$ and $F_T$ is $M_t \times M_t$.

When $i, j$ are in range of $\frac{M_t-M}{2}$, $\frac{M_t-M}{2}$ to $\frac{M_t-M}{2}+M$, $\frac{M_t-M}{2}+M$, the convolution result in this region can be expressed as

$$Y_{Tvalid}(i,j) = \sum_{a=0}^{M-1}\sum_{b=0}^{M-1} X_T(a+i, b+j)\times F_T(a,b) \tag{3}$$

where $X_T$ is the original tiled input block before circular padding. Since input and filter channels are tiled in the same order, each individual input channel on the tiled input block aligns with its corresponding filter on the tiled filter block in this region, as shown in Fig. 1 (b). The summation over channel dimension in equation 1 is effectively unrolled into the other two dimensions thus equation 3 has same output as equation 1. Therefore, within this region the convolution result $Y_T$ is the standard multi-channel convolution result.

When $i, j$ are outside of this region, as shown in Fig. 1 (c), the result is not valid, e.g., input channels are convolved with wrong filter channels. Those invalid results make the $(M_t - 1)/2$ extra zero padding of tiled input and filter due to inherent circular padding of Fourier transform unnecessary since circular padding only corrupts the results of invalid region. Therefore in this scheme only the center $M \times M$ region of the whole $M_t \times M_t$ convolution result is equivalent to the multi-channel convolution result of all input channels and should be extracted as the final result.

The output of this tiling scheme $Y_T$ is a single block with size $\lceil\sqrt{N_c}\rceil(M+N-1)\lceil\sqrt{N_c}\rceil(M+N-1)$ and only the center $M \times M$ region is valid. The valid region can be extracted by directly selecting $y(i, j)$ in the range of $\frac{M_t-M}{2}, \frac{M_t-M}{2}$ to $\frac{M_t-M}{2}+M, \frac{M_t-M}{2}+M$. Fig 1(e) visualizes the output format. Since only the center $M \times M$ part is used as the output, the camera's resolution requirement is massively reduced to the size of a single input. Such reduction can significantly reduce the output bandwidth and improve camera readout time.

For some CNN layers, applying one tiling method alone cannot guarantee full utilization. In this case, channel tiling can be applied together with input tiling or kernel tiling to further boost utilization, by applying channel tiling first and followed by input or kernel tiling. In this way, the system throughput can be maximized while still able to accumulate the channels in the optical domain at full precision.

### 3.2.2 Channel Tiling Benefits

In channel tiling, the kernels (channels of filters) are tiled in the space domain and the Fourier transform of tiled kernels is loaded on the kernel DMD. The accumulation of channel outputs is inherently carried out in the optical domain, unlike the other two tiling schemes where the channel outputs need to be accumulated in electronics, after photodetection and ADC. By computing the channel accumulation in the optical domain, channel tiling has two advantages compared to other tiling methods. One main advantage is the support of negative weights, which is a major limitation of optical neural network accelerators. For input and kernel tiling (or no tiling), the photodetection will apply the square function to individual channel outputs, making the convolution results invalid for negative weights. A workaround is to use the pseudo-negative approach [7] with positive-only weights, where each kernel is split into two parts and the convolution results of two parts are subtracted in electronics after readout to generate negative outputs. Doing so can address the positive readout issue but halves the throughput. By implementing channel tiling, the channel accumulation is computed at full precision in the optical domain hence will generate correct convolution output regardless of the sign of weights. The square function applied by the photodetection can be modeled as the non-linear activation function of the system and embedded into the training process.

Another advantage of channel tiling is by accumulating the channels in the optical domain at full precision, it is much more robust to camera/photodetector's quantization error and sensing noise, which further allows channel tiling to support much faster and lower bit-depth cameras/ADCs. For the free-space 4F system, a high-speed camera is usually used as the output detection device and it adds two kinds of errors to the system, namely the quantization error and random sensing noise. Getting high bit-depth (i.e., precision), high resolution, high speed and low noise is a tough challenge for cameras and photodetectors. Most high-speed cameras with reasonable resolution are limited to 8 or 12 bit precision (e.g., see[10, 11]).

Due to the limited precision or bit-depth of cameras, all outputs need to be quantized to 8-bit (or 12-bit) fixed-point format. While conventional CNNs usually do not require high precision for inference, the case is different for optical systems since the square function is applied to the activation during sensing which increases the dynamic range and leading to larger quantization error. The input and filter tiling methods quantize *each* channel as channel summation happens electronically after optical sensing (i.e., the partial sums themselves are quantized, not just the final activation value), thus the quantization errors are propagated during channel accumulation. In contrast, channel accumulation is carried out in the optical domain at full precision and only the accumulated results are quantized, leading to a smaller overall quantization error.

Table 2: Overall network inference time in seconds (per input) for different tiling schemes and network architectures. The results are for convolution layers only. Note for DeconvNet there's a convolution layer of 1x1 filters which is not suitable for 4F system acceleration and is not taken into account for the estimation.

| Network-dataset | GPU | Channel tiling | No Tiling | Speedup |
|---|---|---|---|---|
| VGG16-CIFAR10 | 5.07e-5 | 6e-6 | 8.17e-1 | **8.45** |
| VGG16-ImageNet | 1.41e-3 | 6e-5 | 8.17e-1 | **23.50** |
| AlexNet-ImageNet | 1.31e-4 | 7e-6 | 1.84e-1 | **18.71** |
| VGG16-SpaceNet7 | 1.79e-2 | 1.06e-3 | 8.17e-1 | **16.89** |
| DeconvNet | 3.76e-4 | 8e-6 | 1.02e-2 | **47.00** |
| SRCNN | 1.48e-3 | 2.4e-5 | 2.88e-4 | **61.67** |

Similarly, channel tiling is less susceptible to sensing noise in the camera. Sensing noise can be especially limiting for fast, high-resolution cameras needed for optical computing. Random sensing noise increases error in every channel in input/filter tiling unlike channel tiling. The error scales with the number of channels thus it impacts more for larger networks. Furthermore, camera SNR (Signal to Noise ratio) scales with the photon flux (or number of photons captured by a pixel).[12,13] Intuitively, if the physical size of camera's sensor is fixed, then the higher resolution it supports (more pixels), the fewer photons each individual pixel will receive. As discussed, channel tiling requires significantly less camera resolution compared to other tiling methods, which means it can have higher camera SNR than other methods.

# 4. EVALUATION

## 4.1 Performance Evaluation

We evaluate the 4F accelerator with channel tiling and compare it against GPU using inference time on real networks. Inference time is the average time for a single input to pass through the whole network. For 4f accelerators we assume a 4K, 2MHz SLM based system and for GPU we use Nvidia GTX-2080 Ti with fp-16 precision. We pick VGG-16[14] on CIFAR-10, ImageNet and SpaceNet,[15] and AlexNet on ImageNet. Besides the two image classification networks, we also include two other networks for image super-resolution and deconvolution. For super-resolution, the SRCNN[16] is used, which consists of two convolution layers with filter size $9 \times 9$ and $5 \times 5$ and the target image resolution is set to $512 \times 512$. The Deconv Net[17] contains five convolution layers with relatively large filter sizes ($121 \times 1, 1 \times 121, 16 \times 16, 1 \times 1$, and $8 \times 8$ respectively). Table 2 shows the evaluation results. The 2MHz SLM system is faster than GPU in all cases and is as much as 61.7X faster for the SRCNN case, which is better suited to the 4F system given its larger kernel sizes. For most conventional neural network architectures with relatively small filter sizes (e.g., AlexNet and VGG), the speedup is around 20X. It is interesting to compare VGG16 performance on CIFAR-10 vs. ImageNet. The smaller input size ($32 \times 32$ vs. $227 \times 227$) of the CIFAR-10 dataset leads to severe underutilization of the 4K SLM, especially in later layers of the network. This indicates that smaller (and therefore potentially cheaper, faster SLMs) may be a better design point for small input networks. As discussed in section 3, the frequency of 4F systems can be significantly improved with advanced SLMs, which makes the performance scaling of 4F systems promising.

Since the main goal of this paper is to focus on bringing 4F optical computing closer to reality, the complete analysis of energy and power is out of scope of this paper. The analysis of power and energy of various photonic and optical neural network implementations can be found in a survey paper.[18]

## 4.2 Network Accuracy Evaluation

### 4.2.1 Impact of Tiling Approaches on Network Accuracy

As discussed previously, different tiling schemes impose different restrictions on the network: input/kernel tiling places a square function on each channel's convolution result before summation, while channel tiling restricts the activation function to be the absolute value function (taking a square root after camera readout). The effect of these restrictions on network-level accuracy is reported in table 3, evaluating three datasets trained with a VGG-16 like model. The camera is assumed to have unlimited precision. All-positive filters are not used in

Table 3: Comparison of the accuracy of different datasets trained using VGG16 with different methods. For Input and kernel tiling the absolute value function is applied to each individual channel's convolution result while for channel tiling the absolute value function is applied after channel summation and acts as the activation function. For pseudo-negative cases, positive weights are used and subtraction is implemented after camera detection, then ReLU is applied on subtracted results as the activation function.

| Method | Fashion MNIST | SVHN | CIFAR10 |
|---|---|---|---|
| Input/Kernel Tiling | 75.4% | 78.5% | 55.6% |
| **Channel Tiling** | 93.2% | 95.1% | 89.3% |
| Pseudo Negative | 93.6% | 95.1% | 91.6% |

input/kernel tiling methods as they effectively nullify the purpose of activation and make deep CNNs like VGG-16 extremely hard to train properly. For the proposed channel tiling, the network is trained with the absolute value function applied to convolution results. All the results reported in table 3 are trained from scratch with floating-point precision.

The results clearly show that input or kernel tiling is unacceptably inaccurate for anything but the simplest of classification tasks. The proposed channel tiling approaches lose <3 percent accuracy compared to unconstrained electronic implementations or pseudo-negative approaches (which uses twice as many filters). This small gap in accuracy can be bridged in the future by better optical non-linearities (an active area of research[19,20]), improved training methods (e.g. different regularization terms during training to incentivize positive filter weights), or combining with the pseudo-negative approach,[21] all are part of our ongoing work but not explored further in this paper.

#### 4.2.2 Impact of Camera Limitations on Inference Accuracy

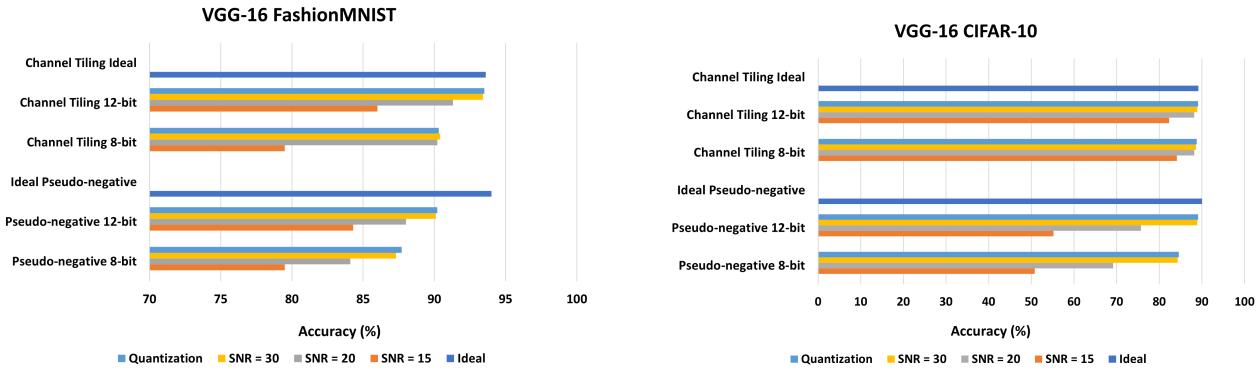
The previous accuracy results are ideal cases and do not consider camera bit-precision and sensing noise. To simulate quantization error due to the camera's limited bit-precision, square function is applied to the convolution results for simulating the intensity measurement and then the results are then quantized to scaled 8-bit and 12-bit format (256/4096 uniform intervals). The square root is taken on the quantized results to get the absolute value. We simulate the sensing noise using white Gaussian noise with average SNR at 15dB, 20dB and 30dB.[22] For both cases only partial sum and activations are quantized, while weights as assumed to be floating point precision. *

Figure 2 shows the Fashion MNIST and CIFAR-10 classification accuracy using VGG-16 for pseudo-negative with filter tiling and the proposed channel tiling, taking into account camera quantization error and different level of random noise. The results clearly show that channel tiling is far more robust to both quantization and sensing noise due to its error-free channel summation. The pseudo-negative (filter tiling) approach requires at least 12-bit camera precision as opposed to 8-bit for channel tiling to remain within *5 percent* accuracy drop from full precision. Interestingly, once the camera bit-depth is taken into account, channel tiling always has higher accuracy than a pseudo-negative approach despite being somewhat more restrictive.

For cases with random noise, the accuracy of the proposed channel tiling method is higher than the pseudo-negative method for almost all cases. 30dB or 20dB SNR is good enough for channel tiling accuracy to be within *5 percent* of the noiseless case while other tiling approaches need at least 30dB SNR. Moreover, the achievable SNR for channel tiling can be higher than filter tiling since it requires lower output resolution, making channel tiling more robust to sensing noise.

Altogether, our results indicate that the proposed channel tiling approach can reduce required camera precision by *33 percent* (8-bit vs. 12-bit) and improve noise tolerance by *10dB*. Though we do not explore it here, such relaxation can substantially improve the speed, energy, and cost of sensing in 4F computing systems.

*We do not use high SNR values as high resolution, high-speed cameras are likely to have higher photon noise.[22]




(a) FashionMNIST (b) CIFAR-10

Figure 2: Accuracy of VGG-16 on FashionMNIST and CIFAR-10 datasets using different setups.

## 5. CONCLUSION

In this paper, we introduce 4F CNN accelerators and discuss their issues and optimizations. We also provide scalability analysis and performance estimation of 4F CNN accelerators. We then propose the channel tiling methods for 4F systems to boost the performance and accuracy, without extra hardware or computation. By utilizing the properties of convolution and 4F systems' high resolution, channel tiling makes 4F systems able to accumulate all channel's convolution results in the optical domain, thus bypassing various constraints applied by output sensing. Compared to the recent pseudo-negative approach with kernel tiling,[21] the proposed channel tiling method gives similar accuracy (<3 percent difference on three datasets), 0.5X total filters required, 10-50X+ throughput improvement and reduction in required output camera resolution/bandwidth. The proposed channel tiling provides a simple and practical way to fully utilize the massive parallelism inherent in 4F optical computing systems to accelerate CNNs and bring it closer to reality.

## ACKNOWLEDGMENTS

We thank Prof. Volker J. Sorger, Dr. Mario Miscuglio, Zibo Hu, Jonathan George and Dr. Maria Solyanik-Gorgone from George Washington University for their collaboration and help. We also thank the Office of Naval Research for the funding support of this work.

## REFERENCES

[1] Miscuglio, M., Hu, Z., Li, S., George, J. K., Capanna, R., Dalir, H., Bardet, P. M., Gupta, P., and Sorger, V. J., "Massively parallel amplitude-only fourier neural network," *Optica* **7**, 1812–1819 (Dec 2020).

[2] Hu, Z., Li, S., Schwartz, R. L., Solyanik-Gorgone, M., Miscuglio, M., Gupta, P., and Sorger, V. J., "Batch processing and data streaming fourier-based convolutional neural network accelerator," *arXiv preprint arXiv:2112.12297* (2021).

[3] Texas Instruments, "Dlp9500 datasheet." https://www.ti.com/document-viewer/DLP9500/datasheet.

[4] Haspeslagh, L., De Coster, J., Pedreira, O. V., De Wolf, I., Du Bois, B., Verbist, A., Van Hoof, R., Willegems, M., Locorotondo, S., Bryce, G., Vaes, J., van Drieenhuizen, B., and Witvrouw, A., "Highly reliable cmos-integrated 11mpixel sige-based micro-mirror arrays for high-end industrial applications," in [*2008 IEEE International Electron Devices Meeting*], 1–4 (2008).

[5] Schmidt, J.-U., Dauderstaedt, U. A., Duerr, P., Friedrichs, M., Hughes, T., Ludewig, T., Rudloff, D., Schwaten, T., Trenkler, D., Wagner, M., Wullinger, I., Bergstrom, A., Bjoernangen, P., Jonsson, F., Karlin, T., Ronnholm, P., and Sandstrom, T., "High-speed one-dimensional spatial light modulator for Laser Direct Imaging and other patterning applications ," in [*MOEMS and Miniaturized Systems XIII*], Piyawattanametha, W. and Park, Y.-H., eds., **8977**, 167 – 176, International Society for Optics and Photonics, SPIE (2014).

[6] Peng, C., Hamerly, R., Soltani, M., and Englund, D. R., "Design of high-speed phase-only spatial light modulators with two-dimensional tunable microcavity arrays," *Opt. Express* **27**, 30669–30680 (Oct 2019).
[7] Mach, L., "Ueber einen interferenzrefraktor," *Zeitschrift für Instrumentenkunde* **12**(3), 89 (1892).
[8] Zhu, L. and Wang, J., "Arbitrary manipulation of spatial amplitude and phase using phase-only spatial light modulators," *Scientific reports* **4**, 7441 (2014).
[9] Wu, L., Cheng, S., and Tao, S., "Simultaneous shaping of amplitude and phase of light in the entire output plane with a phase-only hologram," *Scientific reports* **5**, 15426 (2015).
[10] Phantom, "S990 camera." https://www.phantomhighspeed.com/products/cameras/machinevision/s990.
[11] Phantom, "Flex4k camera." https://www.phantomhighspeed.com/products/cameras/4kmedia/flex4k.
[12] Healey, G. and Kondepudy, R., "Ccd camera calibration and noise estimation," in [*Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], 90,91,92,93,94,95, IEEE Computer Society, Los Alamitos, CA, USA (jun 1992).
[13] Healey, G. and Kondepudy, R. V., "Modeling and calibrating CCD cameras for illumination-insensitive machine vision," in [*Optics, Illumination, and Image Sensing for Machine Vision VI*], Svetkoff, D. J., ed., **1614**, 121 – 132, International Society for Optics and Photonics, SPIE (1992).
[14] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).
[15] Van Etten, A., Lindenbaum, D., and Bacastow, T. M., "Spacenet: A remote sensing dataset and challenge series," *arXiv preprint arXiv:1807.01232* (2018).
[16] Dong, C., Loy, C. C., He, K., and Tang, X., "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38**(2), 295–307 (2016).
[17] Xu, L., Ren, J. S., Liu, C., and Jia, J., "Deep convolutional neural network for image deconvolution," in [*Advances in Neural Information Processing Systems 27*], Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., eds., 1790–1798, Curran Associates, Inc. (2014).
[18] De Marinis, L., Cococcioni, M., Castoldi, P., and Andriolli, N., "Photonic neural networks: A survey," *IEEE Access* **7**, 175827–175841 (2019).
[19] Williamson, I. A. D., Hughes, T. W., Minkov, M., Bartlett, B., Pai, S., and Fan, S., "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," *IEEE Journal of Selected Topics in Quantum Electronics* **26**(1), 1–12 (2020).
[20] George, J., Mehrabian, A., Amin, R., El-Ghazawi, T., Prucnal, P. K., and Sorger, V. J., "Photonic neural network nonlinear activation functions by electrooptic absorption modulators," in [*Frontiers in Optics / Laser Science*], *Frontiers in Optics / Laser Science* , JW3A.123, Optical Society of America (2018).
[21] Chang, J., Sitzmann, V., Dun, X., Heidrich, W., and Wetzstein, G., "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Scientific reports* **8**(1), 1–10 (2018).
[22] Huang, W. and Xu, Z., "Characteristics and performance of image sensor communication," *IEEE Photonics Journal* **9**(2), 1–19 (2017).