调研：
硅光芯片深度学习

2019.09.17

Shanhui Fan

- Training of photonic neural networks through in situ backpropagation and gradient measurement, Optica, 2018.
- Stanford Univ, Ginzton Lab, Stanford; Stanford Univ, Dept Elect Engn, Stanford

Zongfu Yu

- Optimization of Nonlinear Nanophotonic Media for Artificial Neural Inference, Photonics Research, 2019
- Univ Wisconsin Madison, Dept Elect & Comp Engn

Yichen Shen

- Deep learning with coherent nanophotonic circuits, Nature Photonics
- MIT, Elect Res Lab
- Shen, YC; Harris, NC (通讯作者)

- Tunable Efficient Unitary Neural Networks (EUNN) and their application to RNN, arXiv, 2017
- MIT
- Jing, L (通讯作者)

Nicholas C. Harris

- Linear programmable nanophotonic processors, Optica, 2018
- MIT
- Englund, D (通讯作者)

- Deep learning with coherent nanophotonic circuits

Dirk Englund

- Large-Scale Optical Neural Networks Based on Photoelecric Multiplication, PHYSICAL REVIEW X 9, 2019
- Quantum optical neural networks, Quantum Information, 2019
- Trace-free counterfactual communication with a nanophotonic processor, Quantum Information, 2019
- Variational Quantum Unsampling on a Programmable Nanophotonic Processor, CLEO, 2019
- Deep learning with coherent nanophotonic circuits

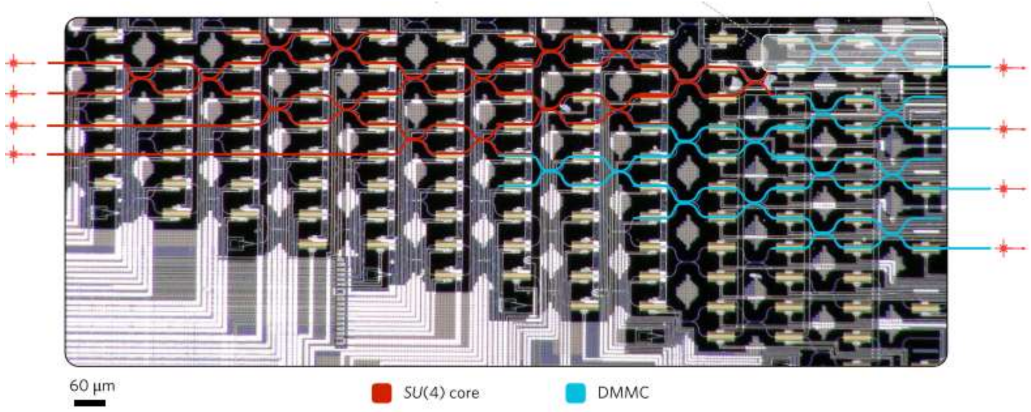
- Research Laboratory of Electronics, MIT

Marin Soljacic

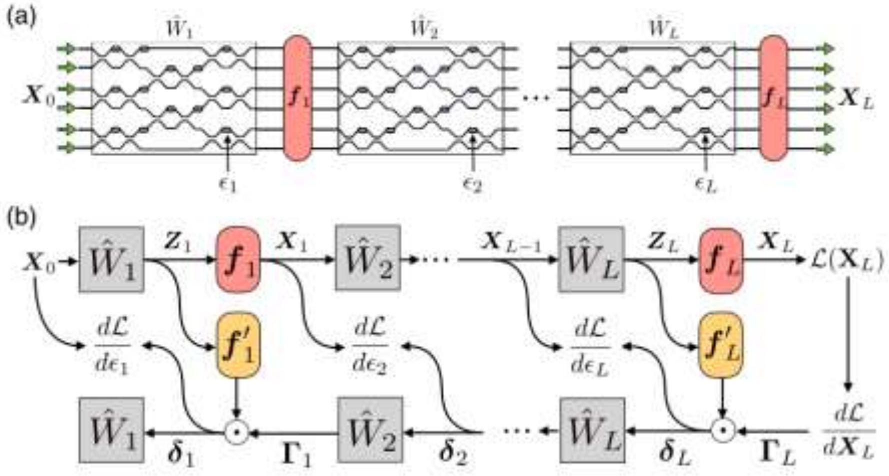
- Large-Scale Optical Neural Networks Based on Photoelecric Multiplication
- *Gated orthogonal recurrent units: On learning to forget
- Deep learning with coherent nanophotonic circuits

- Mitgrating Knowledge between Physical Scenarios Based on Artificial Neural Networks, ACS Photonics, 2019
- On-Chip Optical Convolutional Neural Networks
- WaveletNet, Logarithmic Scale Efficient Convolutional Neural Networks for Edge Devices, arXiv, 2018
- Department of Physics, MIT

In situ backpropagation and gradient measurement



Shen Yichen etal., Deep learning with coherent nanophotonic circuits, Nature Photonics, 2017



(a) Schematic of the ANN architecture.
(b) Illustration of operation and gradient computation in an ANN.

Shanhui Fan et al., “Training of photonic neural networks through in situ backpropagation and gradient measurement”, Frontiers in Optics/ Laser Science, 2018.

- training weights ex situ on a computer model of the system
- creating final weights in the physical device using an idealized model that relates the matrix elements to the phase shifters
- losing the potential advantages(对原文的评价)

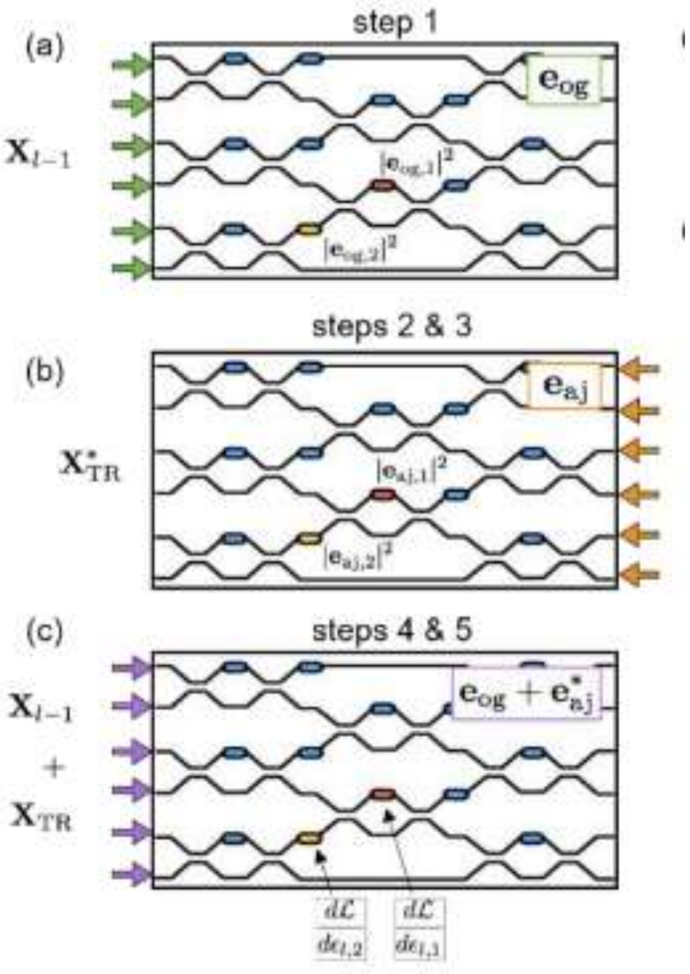
- in situ intensity measurements
- Additional components:
- physically implementing adjoint variable method

- ANN的实现
  - input vector ---> output vector via matrices
  - tuning matrix elements (weights) for minimized cost function
  - tuning is implemented by “backpropagation algorithm”
  - utilizing the chain rule from the output layer to the input layer

- 对原文献的评价：
  - the training of the phase-shifter settings for this system was performed using a model implemented on a standard computer, which does not take into account experimental errors, and furthermore loses all the potential advantages in time and energy of the photonic implementation.

- 改进：
  - The only additional component that is required is a means to measure the light intensity in the vicinity of each of the tunable phase shifters.

步骤



1. Send in the original field amplitude and measure and store the intensities at each phase shifter.
2. Send delta into the output ports and measure and store the intensities at each phase shifter.
3. Compute the time-reversed adjoint input field amplitudes.
4. Interfere the original and the time-reversed adjoint fields in the device, measuring again the resulting intensities at each shifter.
5. Subtract the constant intensity terms from steps 1 and 2 and multiply by k square to recover the gradient.

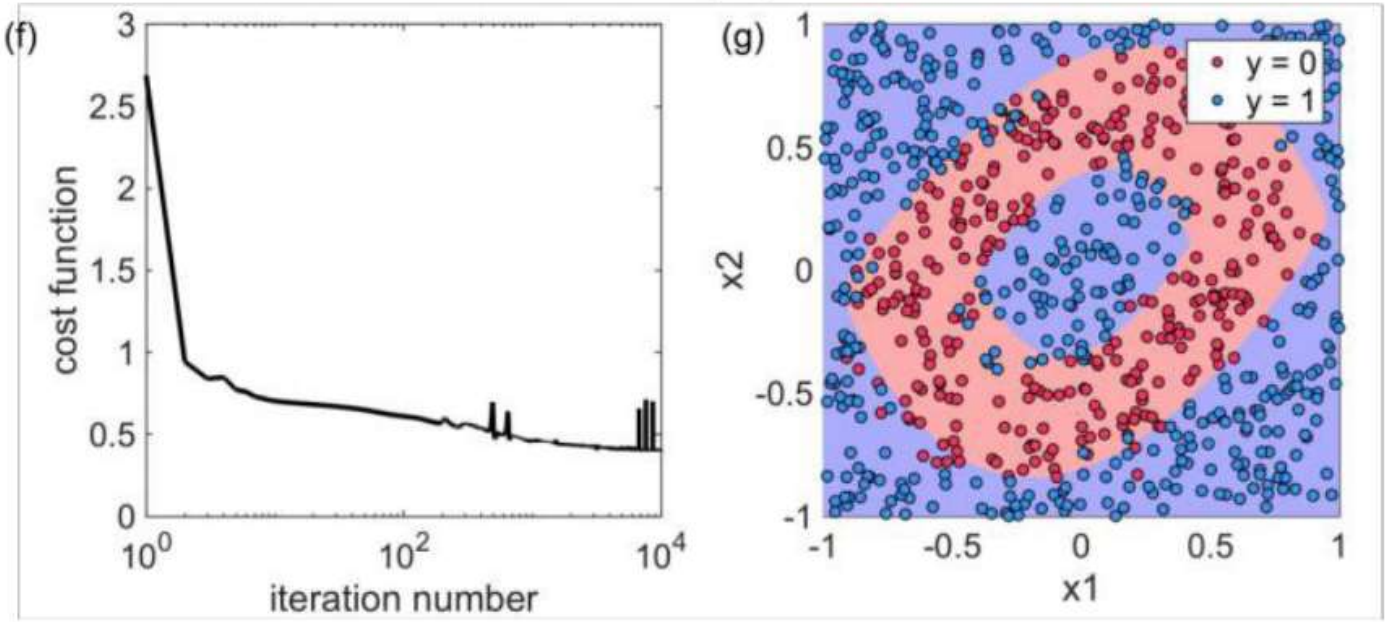
Inserting this into Eq. (21), we thus find that the gradient is given by the overlap of the two fields over the phase-shifter positions:

$$\frac{d\mathcal{L}}{d\epsilon_l} = k_0^2 \mathcal{R}\left\{\sum_{r\in r_\phi} \boldsymbol{e}_{\mathrm{aj}}(\boldsymbol{r})\boldsymbol{e}_{\mathrm{og}}(\boldsymbol{r})\right\}. \qquad (23)$$

Shanhui Fan et al., “Training of photonic neural networks through in situ backpropagation and gradient measurement”, Frontiers in Optics/ Laser Science, 2018.

Restriction and Result

1. Assuming linear, lossless, reciprocal, feed-forward propagation inside the OIU.
2. Mode-dependent loss limits the ability to accurately reconstruct the time-reversed adjoint field.
3. 40% of the light is lost due to back-scattering and radiation losses for 3 * 3 operation. (shown below)

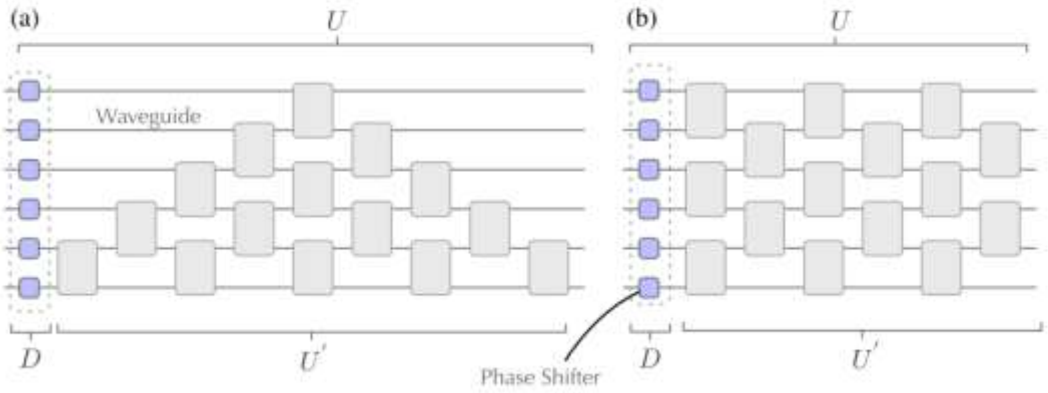


Shanhui Fan et al., "In-situ Backpropagation in Photonic Neural Networks", Frontiers in Optics/ Laser Science, 2018.

Linear programmable nanophotonic processors

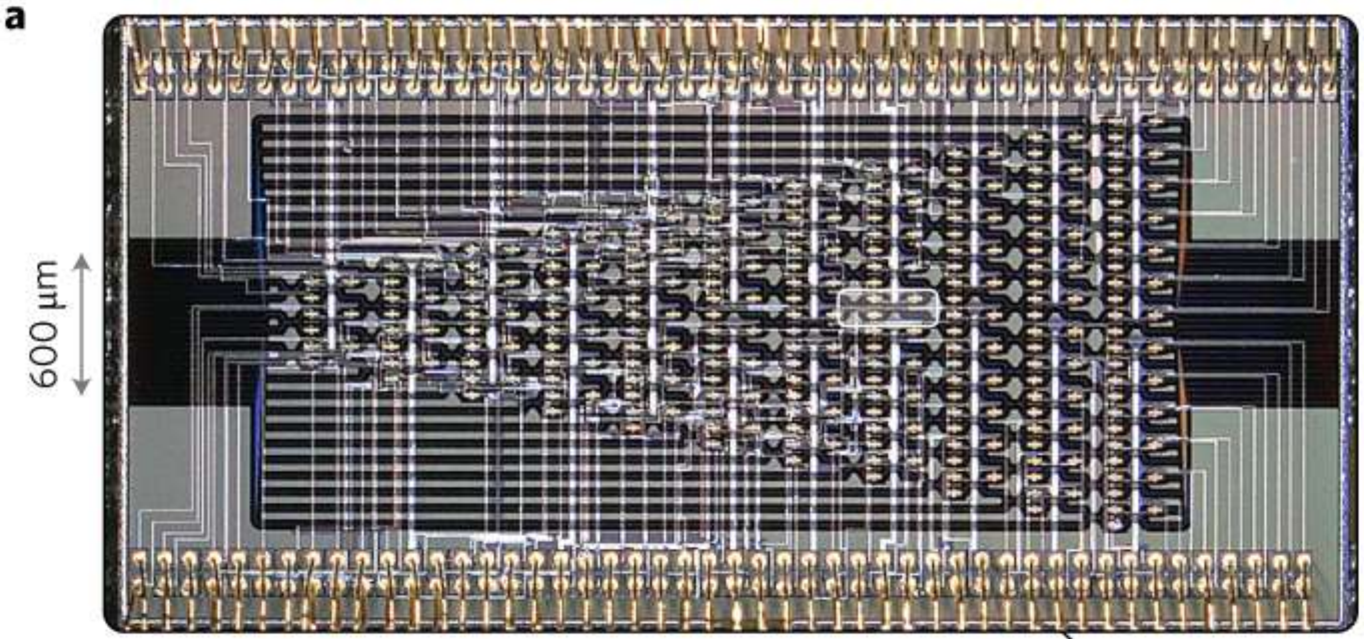
(Nicholas C. let al., "Linear programmable nanophotonic processors", Optica, 2018.)

- N input to N output problem up into 2 x 2 mode transformers – Mach-Zehnder interferometers(MZI).
- $\Sigma_n = N(N-1)/2$ MZI needed. For instance, n=6, $\Sigma_n = 15$.



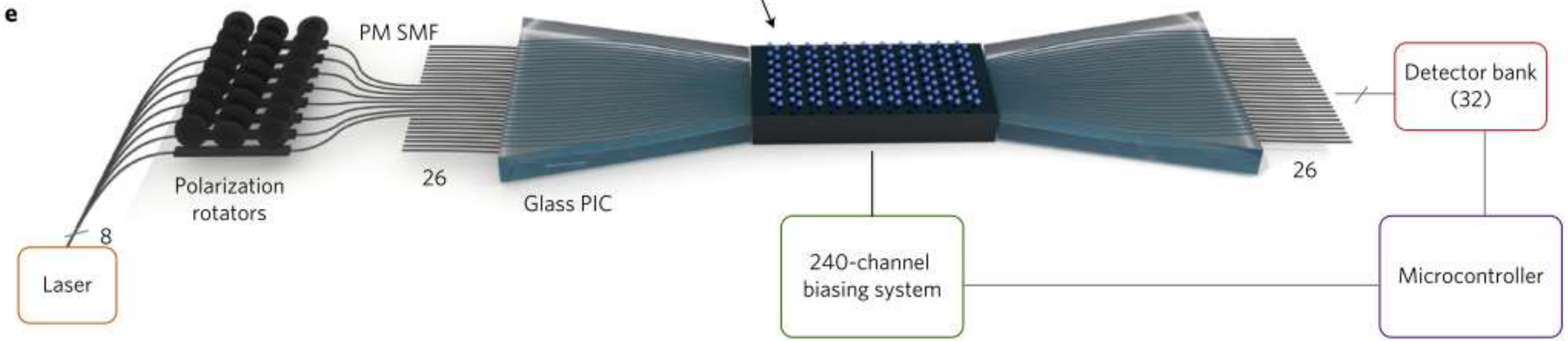
(a) reck(1994) (b)clement(2016)

- The SOI platform offers high index contrast of 3.4:1.5.
- Largest PNPs – 88 MZIs connecting 26 optical modes(4.9 mm x 2.4 mm)



损耗

Silicon waveguides in the PNP are inverse-tapered to a mode-field diameter of 2 μm with a mode spacing of 25.4 μm. The 52 modes of the PNP are coupled to optical fibres using two laser-written glass photonic integrated circuits as indicated in Fig. 2c (Supplementary Section 1). Mode-field diameter mismatch between the glass chips and the PNP, and fibre connectorization, result in a transmission loss of 3.5 dB per facet. The total loss through the PNP, including both input and output coupling, is 8 dB. Accounting for the coupling losses of 3.5 dB per facet, the PNP transmission is 80%. This matches the expected propagation loss for our silicon nanowire waveguides[44].



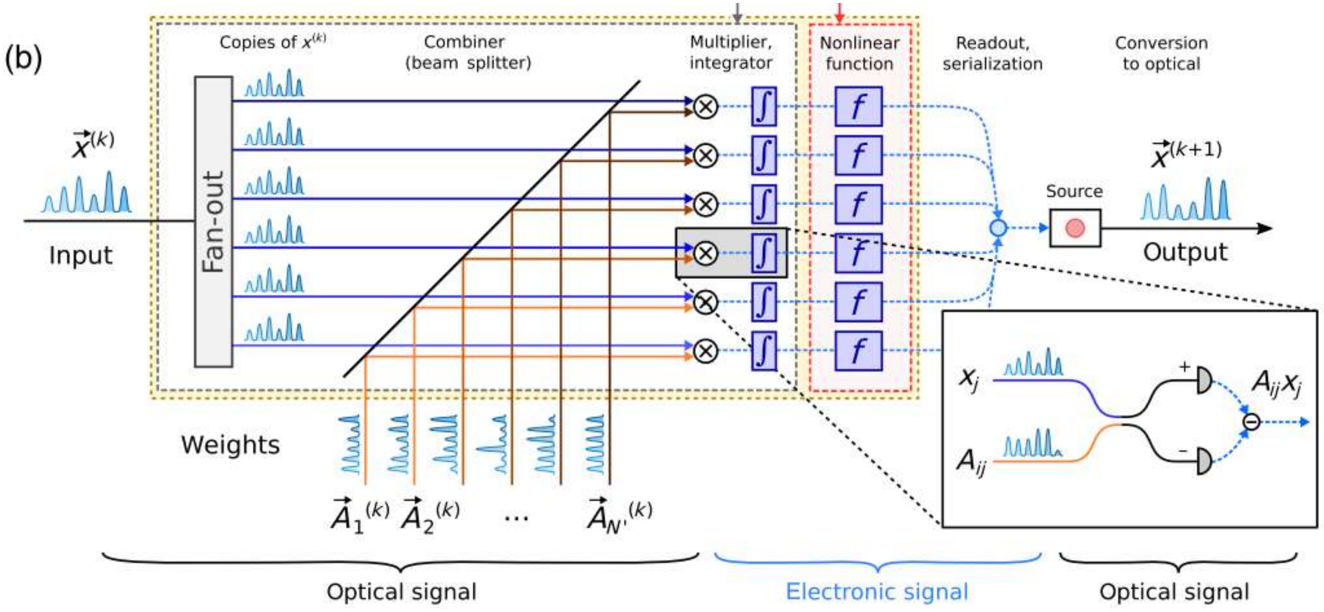
Processor composed of 88 MZIs, 26 input modes, 26 output modes and 176 phase shifters

(Nicholas C. let al., “Linear programmable nanophotonic processors”, Optica, 2018.)

Large-Scale Optical Neural Networks Based on Photoelectric Multiplication

- 特点
  - reduce E/MAC (the energy per multiply and accUmulate) from 20 pJ/MAC(ASICs, GPUs) to around 1 pJ/MAC.
  - 1 fJ/MAC for modulator, rise above 1 pJ/MAC once the driver electronics and memory access are included
  - naturally adapted to free space optics
- fundamental limits:
  - detector shot noise presents a standard quantum limit （光电探测器散粒噪声）
  - leading to classification error
    - C. M. Caves, Quantum-Mechanical Noise in an Interferometer, Phys. Rev. D 23. 1693 (1981)
- pretraining weights on a GPU
- computing Neural-network performance using Monte Carlo simulations

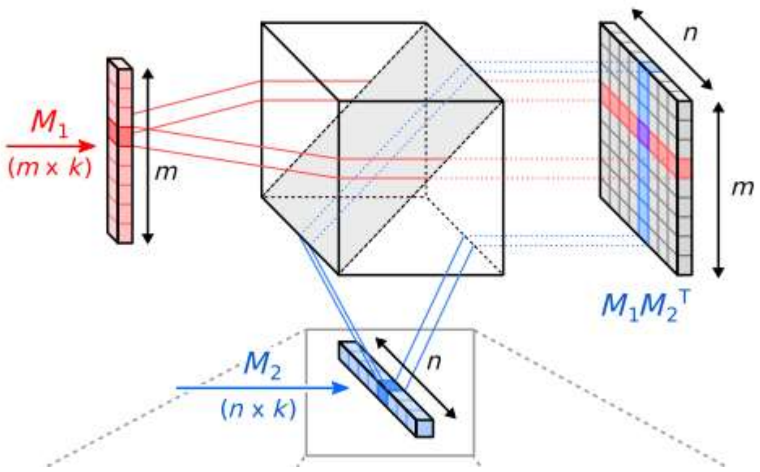
Ryan Hamerly et al., Large-Scale Optical Neural Networks Based on Photoelectric Multiplication, Physical Review X, 2019



- 乘积值正比于探测器收到的电荷值

$$Q_i = \frac{2\eta e}{\hbar\omega}\int \mathrm{Re}[E^{(\mathrm{in})}(t)^* E_i^{(\mathrm{wt})}(t)]dt \propto \sum_j A_{ij}x_j.$$

- 并行运算
  - running a batch of instances, X = [x1…xB], the output Y = [y1…yB] can be computed through the matrix-matrix product Y = AX



上图：计算过程

左图：并行计算

Ryan Hamerly et al., Large-Scale Optical Neural Networks Based on Photoelectric Multiplication, Physical Review X, 2019

Results

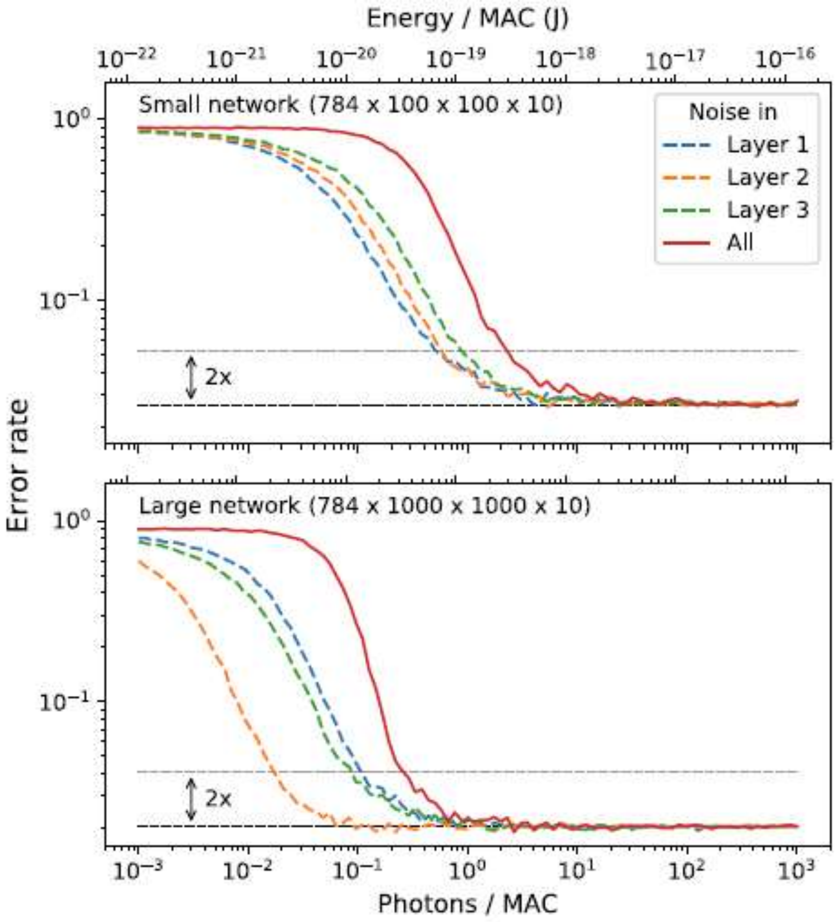


FIG. 3. MNIST digit classification. Error rate for neural-network inference as a function of photons per MAC n_{MAC} (equivalently energy $E_{\mathrm{MAC}} = (hc/\lambda)n_{\mathrm{MAC}}$; here, $\lambda = 1.55\ \mu$m).

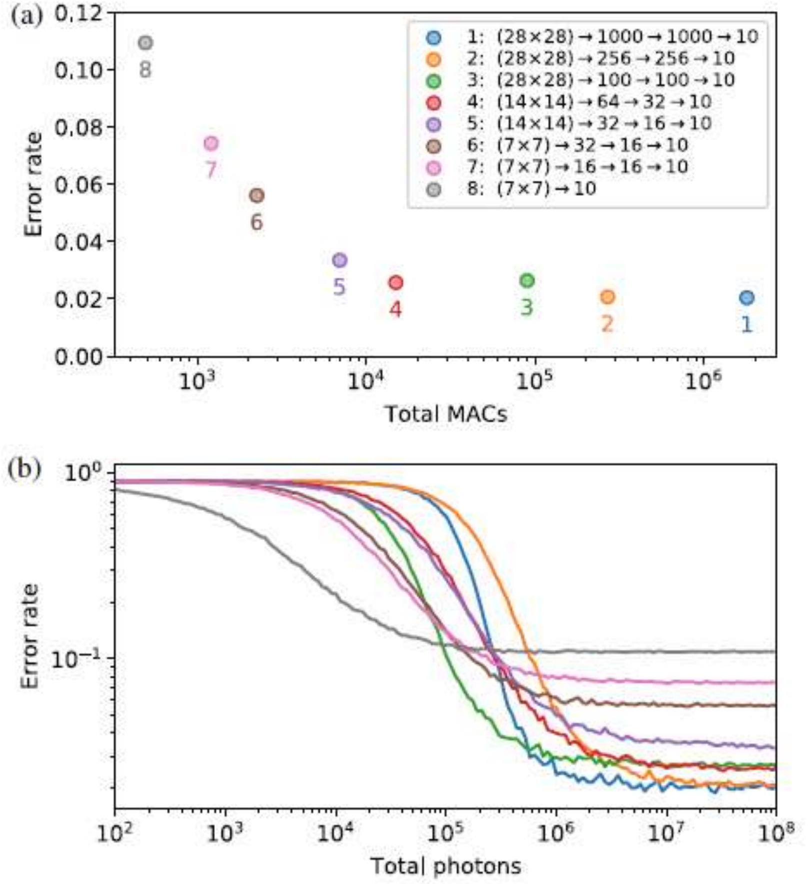
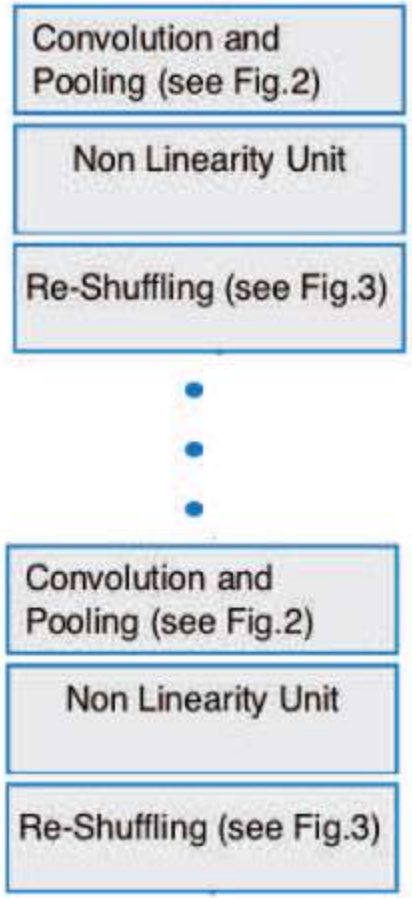


FIG. 4. (a) Conventional picture. Error rate as a function of number of MACs for different fully connected MNIST neural networks. (b) SQL picture. Error rate as a function of total number of photons, for the same networks.

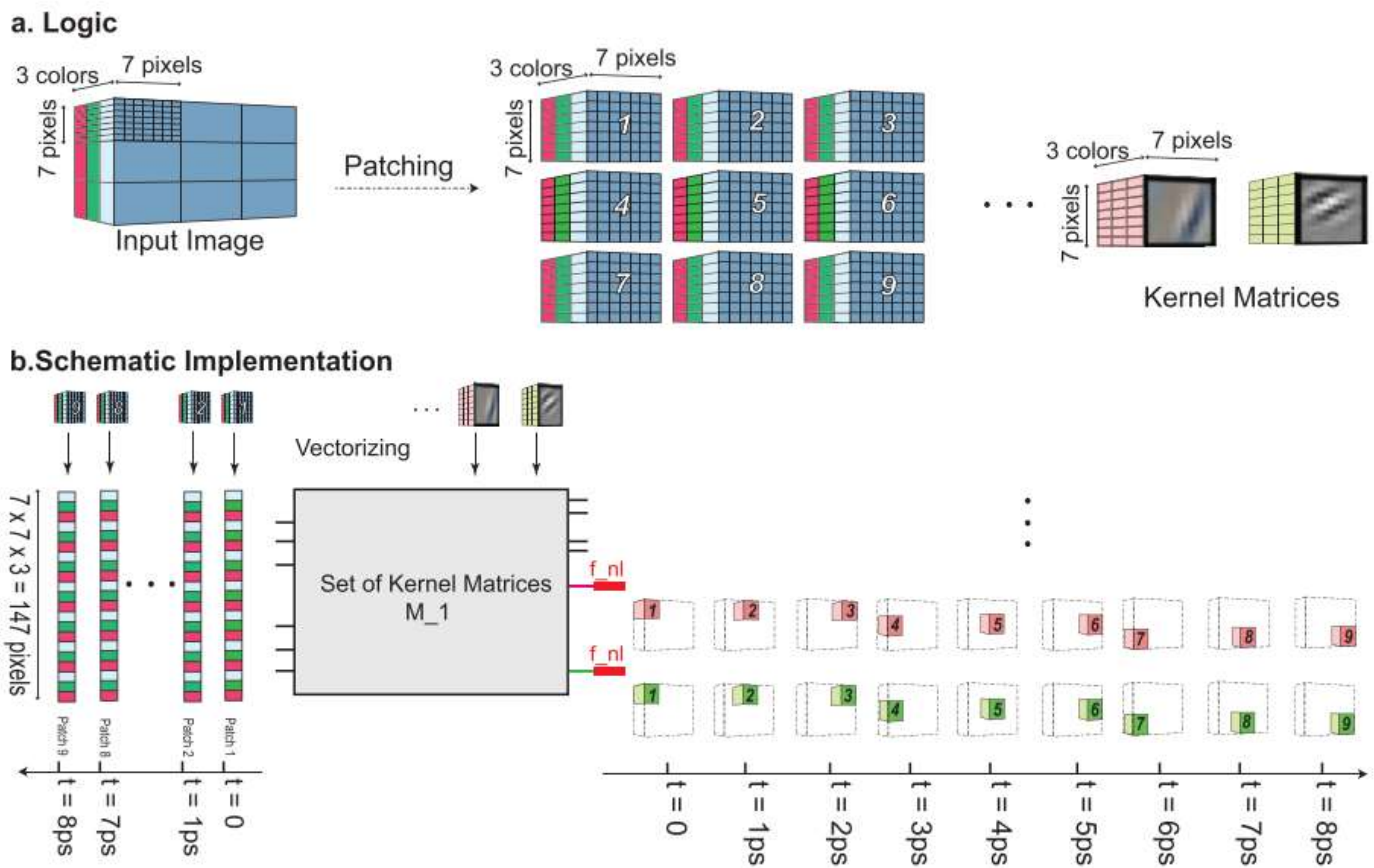
Ryan Hamerly et al., Large-Scale Optical Neural Networks Based on Photoelectric Multiplication, Physical Review X, 2019

On-Chip Optical Convolutional Neural Networks

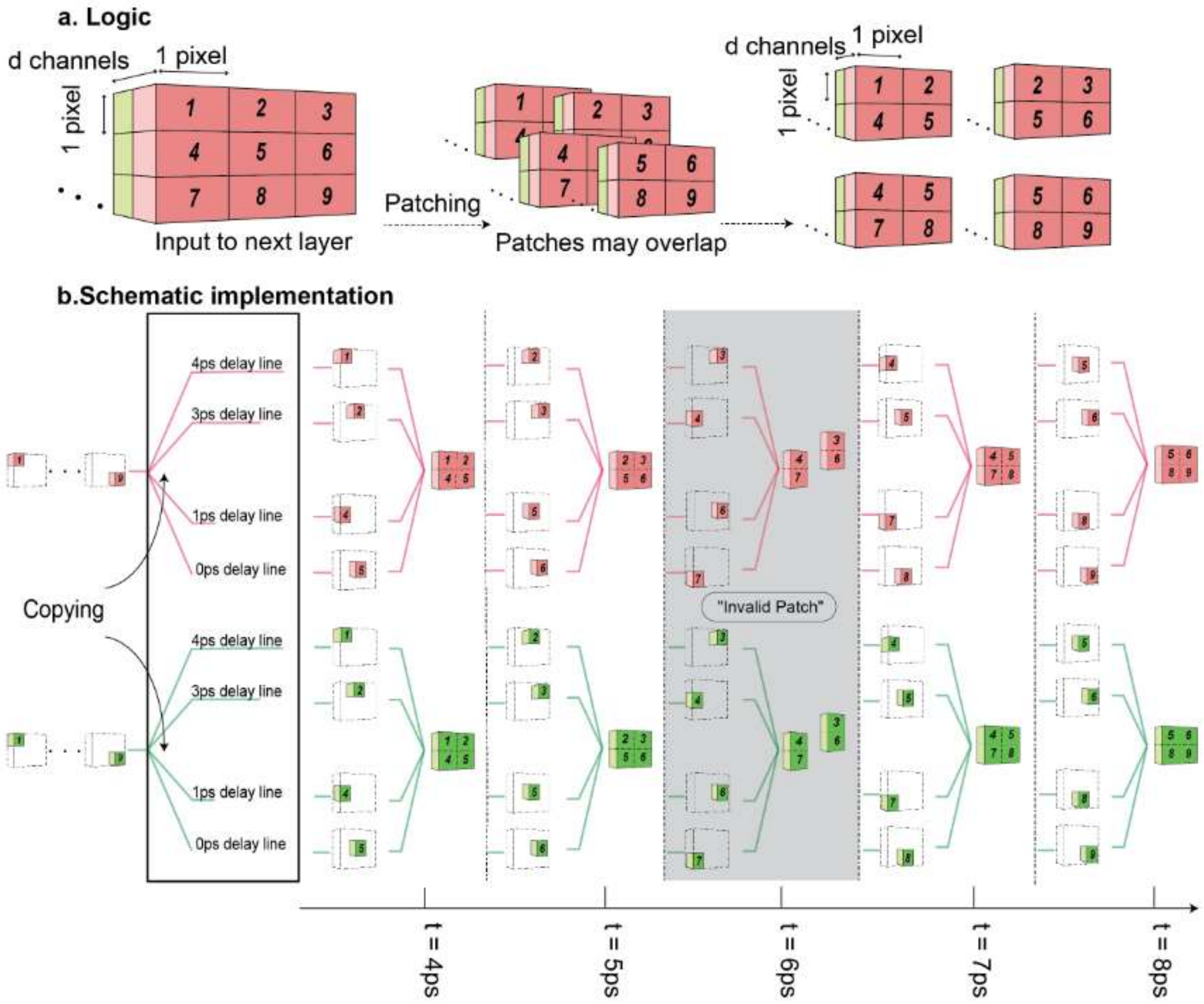
- Convolutional Neural Networks (CNN): convolving input images with filter-kernels for object recognition and classification purposes
- input image → convolution and pooling layers → mapping output to classification targets



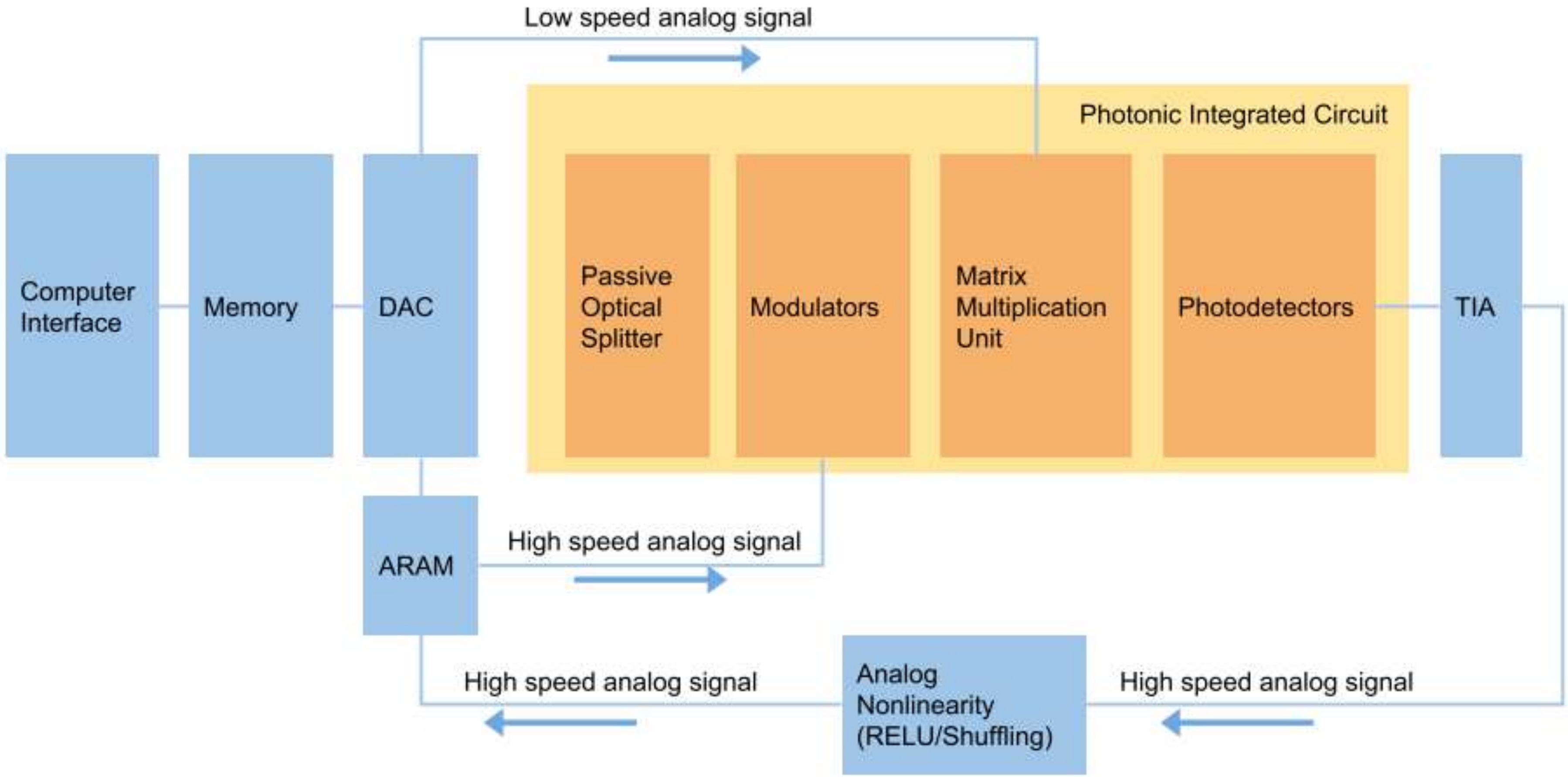
Hengameh Bagherian et al., On-Chip Optical Convolutional Neural Networks, arXiv, 2018



Hengameh Bagherian et al., On-Chip Optical Convolutional Neural Networks, arXiv, 2018



Hengameh Bagherian et al., On-Chip Optical Convolutional Neural Networks, arXiv, 2018



Hengameh Bagherian et al., On-Chip Optical Convolutional Neural Networks, arXiv, 2018

Quantum optical neural networks (Dirk Englund)

- mapping quantum optics features(mode mixing, optical nonlinearity) to neural networks
- similar architecture
- CMOS-compatible platform instead of photonics chips
- Application: quantum information processing tasks: quantum optical state compression for quantum networking and black-box quantum simulation

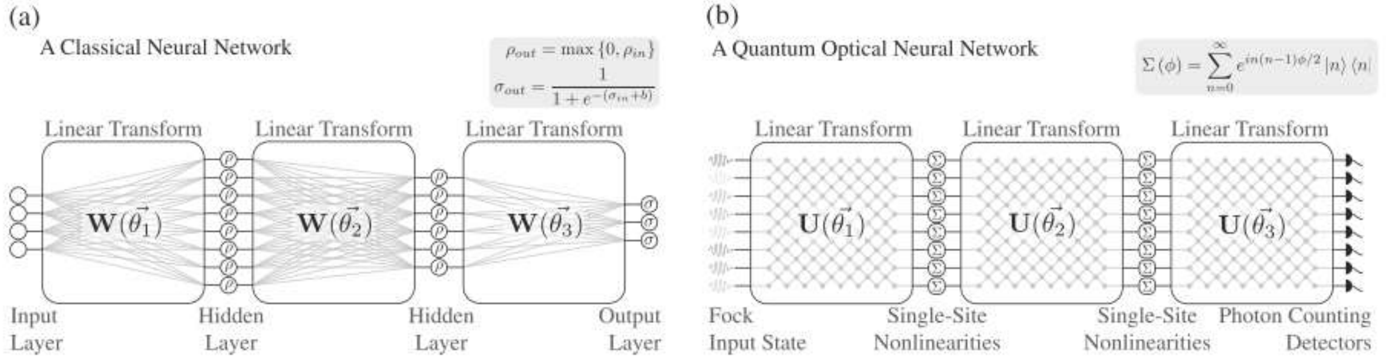
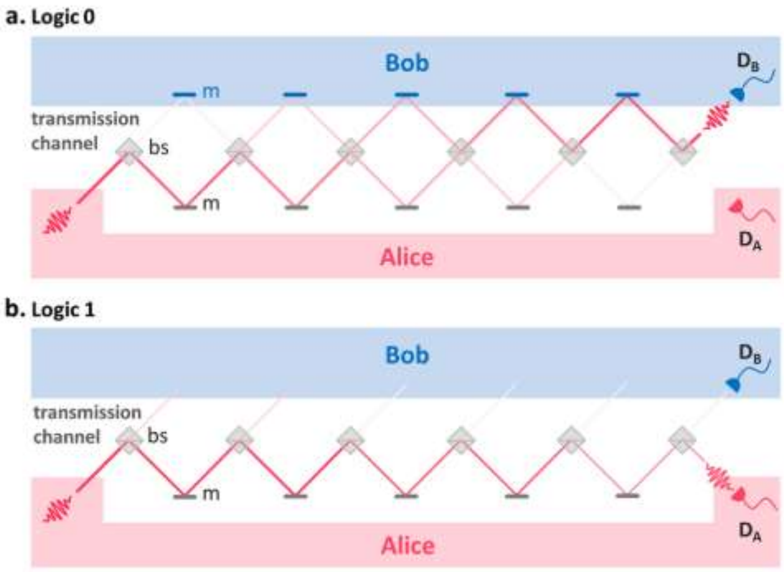


Fig. 1 Quantum optical neural network (QONN). **a** An example of a classical neural network architecture. Hidden layers are rectified linear units (ReLUs) and the output neuron uses a sigmoid activation function to map the output into the range (0, 1). **b** An example of our quantum optical neural network (QONN) architecture. Inputs are single photon Fock states. The single-site nonlinearities are given a Kerr-type interaction applying a phase quadratic in the number of photons. Readout is given by photon-number-resolving detectors, which measure the photon number at each output mode

Gregory R. Steinbrecher et al.,Quantum optical neural networks, Quantum Information, 2019

Trace-free counterfactual communication with a nanophotonic processor

- Performing Quantum communication
- Performing high-fidelity counterfactual communication protocol without post-selection enabled by a programmable nano-photonic programmable nanophotonic processor
- implement CFC protocol using two to six concatenated beam splitters on the same photonic chip
- high (99.94%) average visibility of the individual integrated interferometers allowed bit error probabilities as low as 1.5%



I. Alonso Calafell et al.,Trace-free counterfactual communication with a nanophotonic processor, Quantum Information, 2019

Variational Quantum Unsampling on a Programmable Nanophotonic Processor

- Implementing VQU protocol for verification and inference of near-term quantum circuits outputs
- performs optimization on $|\psi_{out}\rangle$using a sequence of auxiliary quantum circuits $\widehat{V}(\vec{\emptyset})$ to find the time reversed condition$\widehat{V}(\vec{\emptyset}) = U^{\dagger}$ for a known input state

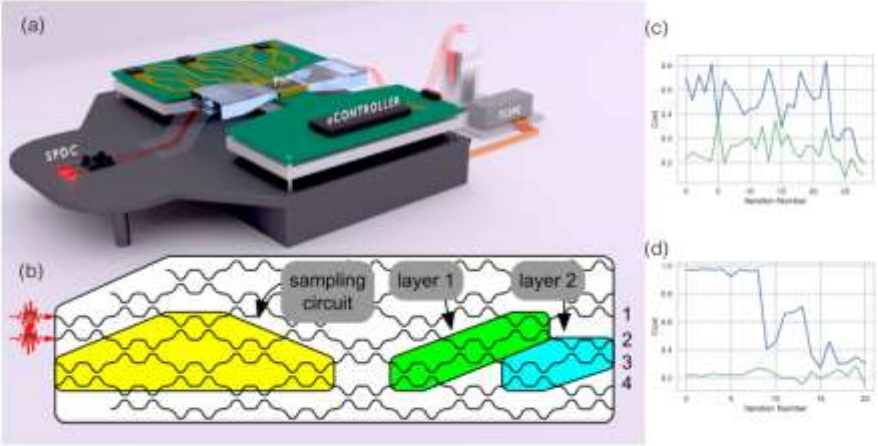


Fig. 1. (a) Schematic of the full experimental setup. (b) The structure of the full VQU protocol, with the optical sampling circuit (yellow) and the subsequent unsampling layers (green, blue). Experimental results for (c) the first layer of unsampling, and (d) the final layer of the protocol.

Jacques Carolan et al., Variational Quantum Unsampling on a Programmable Nanophotonic Processor, CLEO 2019