调研：
硅光芯片深度学习

2019.09.25

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

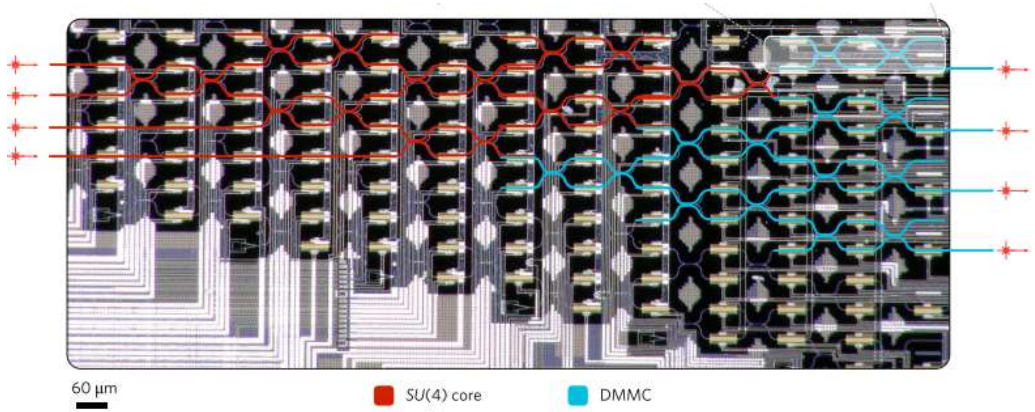
- Research Laboratory of Electronics, MIT

Marin Soljacic

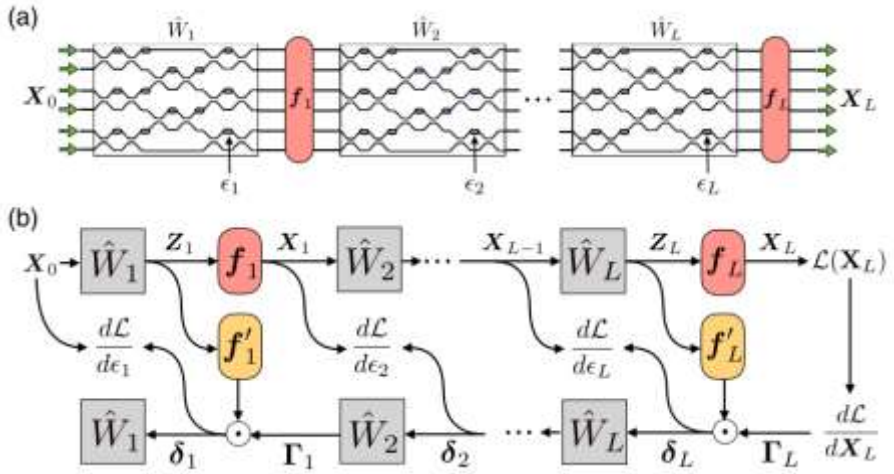
- Large-Scale Optical Neural Networks Based on Photoelecric Multiplication
- *Gated orthogonal recurrent units: On learning to forget
- Deep learning with coherent nanophotonic circuits

- 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)

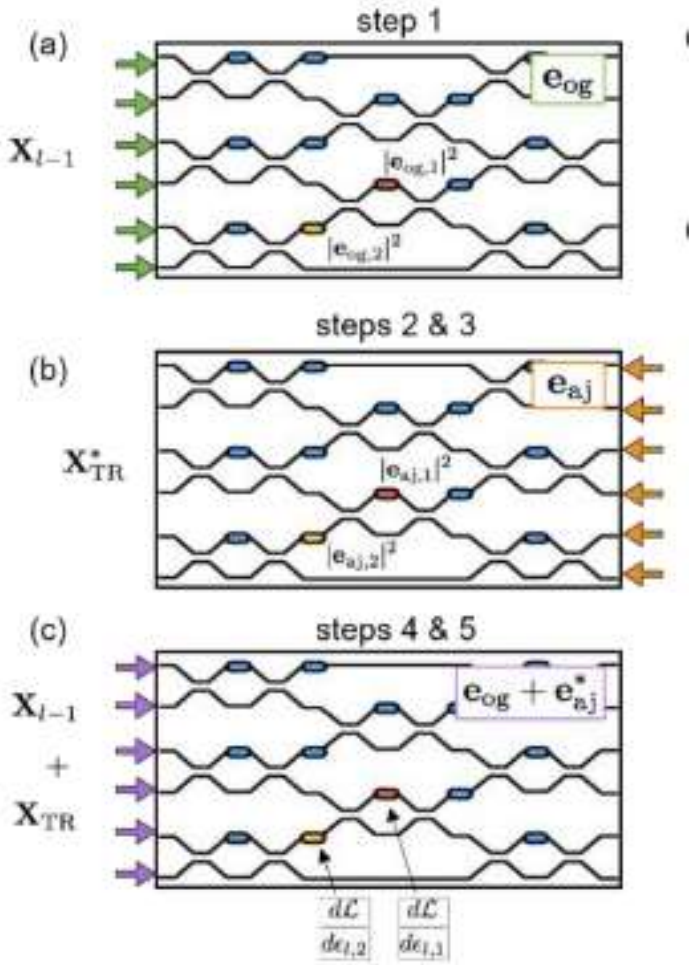
- in situ intensity measurements
- Additional components:
- physically implementing adjoint variable method

- 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.

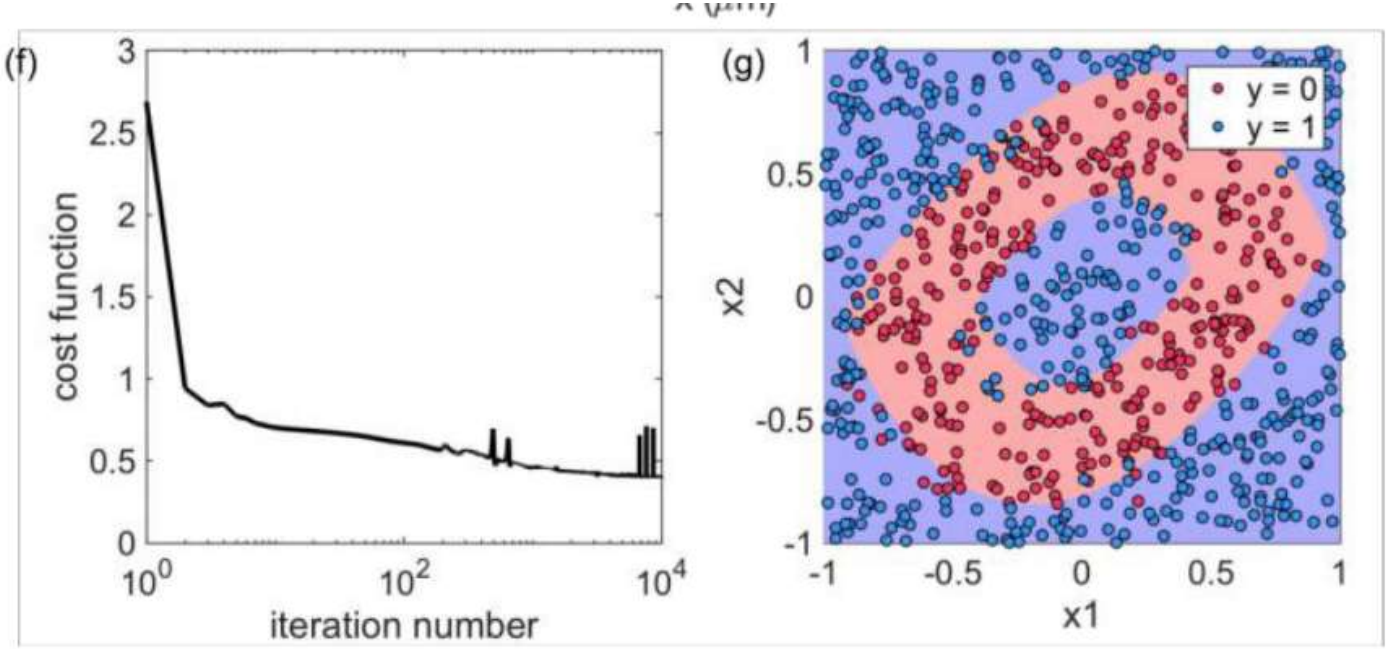
Inserting this into Eq. (21), we thus find that the gradient is given by the overlap of the two fields over the phase-shifter positions:

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

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)

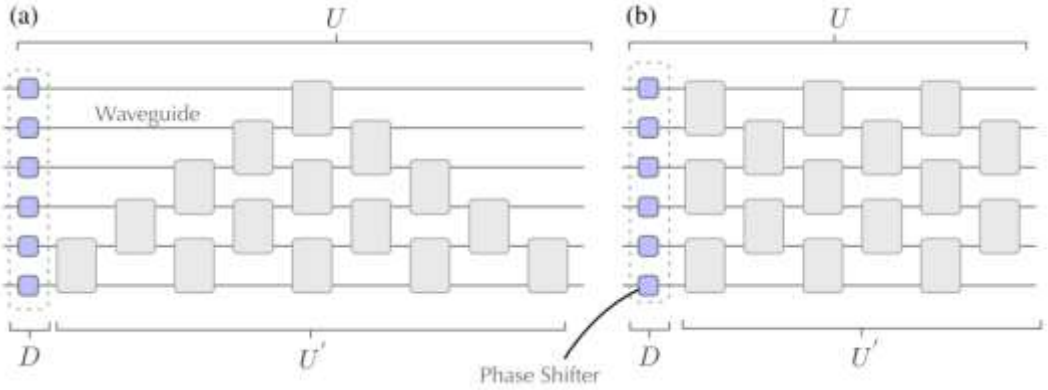


Shanhui Fan et al., "In-situ Backpropagation in Photonic Neural Networks", Frontiers in Optics/ Laser Science, 2018.

Linear programmable nanophotonic processors

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

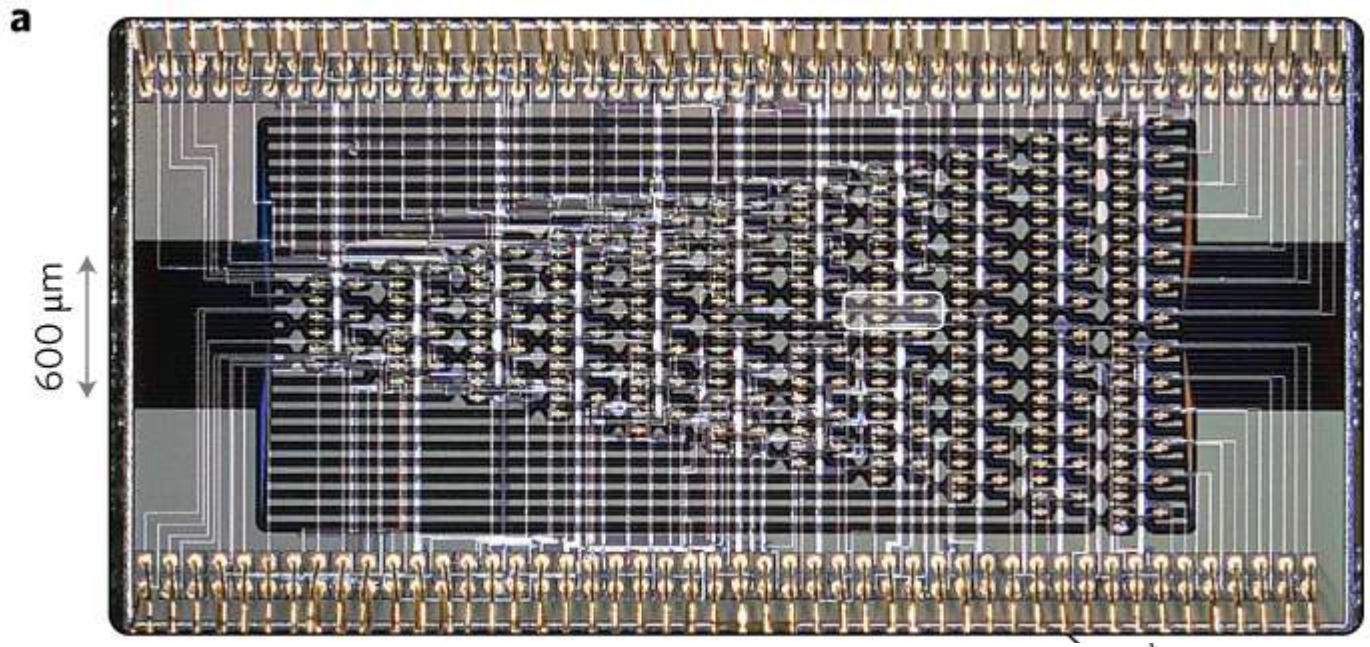
- N input to N output problem up into 2 x 2 mode transformers – Mach-Zehnder interferometers(MZI).
- $\Sigma_n = N(N-1)/2$ MZI needed. For instance, n=6, $\Sigma_n = 15$.



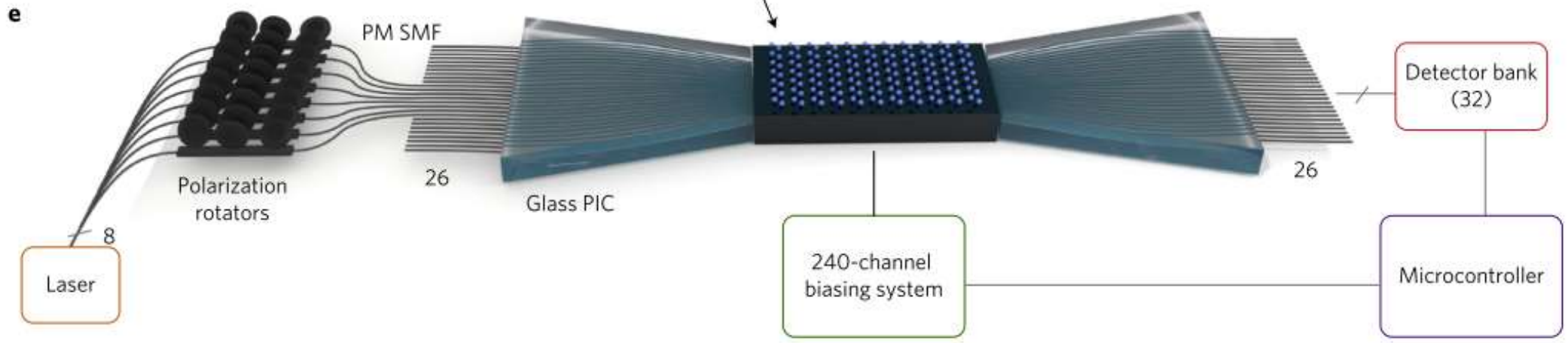
(a) reck(1994) (b)clement(2016)

- 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].



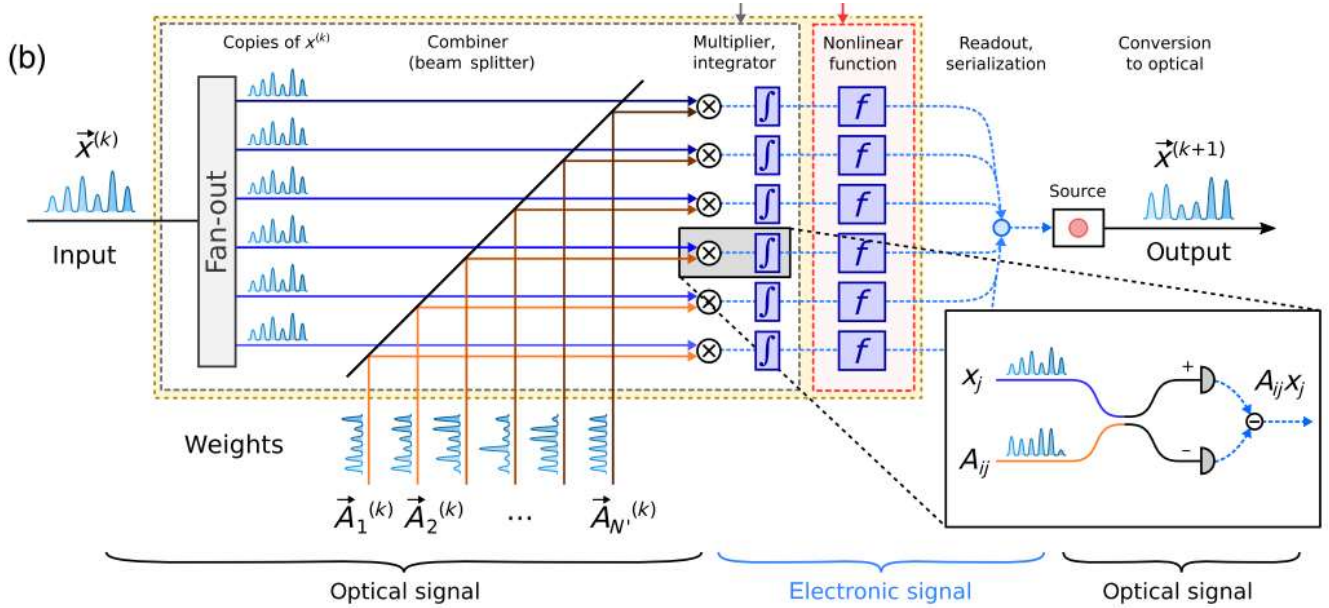
Processor composed of 88 MZIs, 26 input modes, 26 output modes and 176 phase shifters

(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

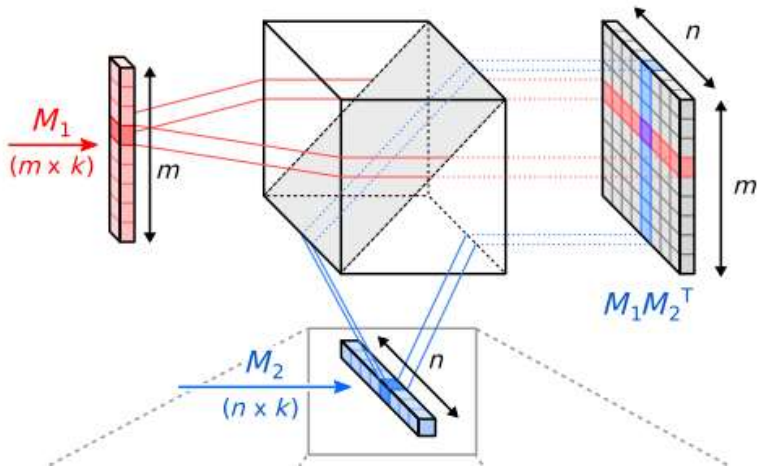
Ryan Hamerly et al., Large-Scale Optical Neural Networks Based on Photoelectric Multiplication, Physical Review X, 2019



- 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

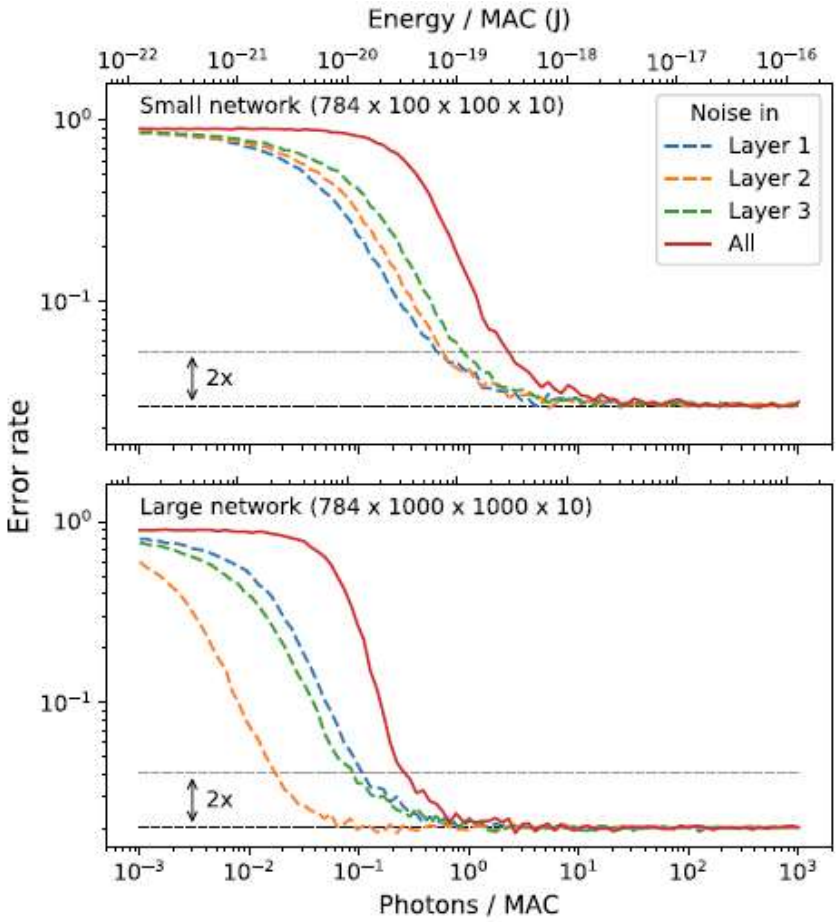


FIG. 3. MNIST digit classification. Error rate for neural-network inference as a function of photons per MAC n_{MAC} (equivalently energy $E_{\mathrm{MAC}} = (hc/\lambda)n_{\mathrm{MAC}}$; here, $\lambda = 1.55\ \mu$m).

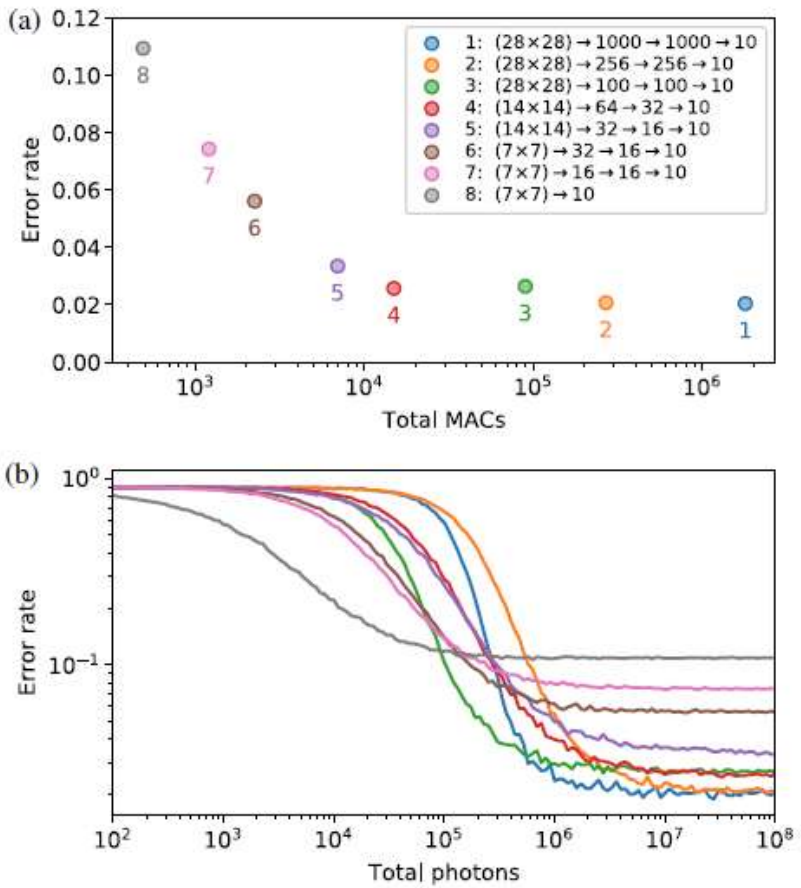
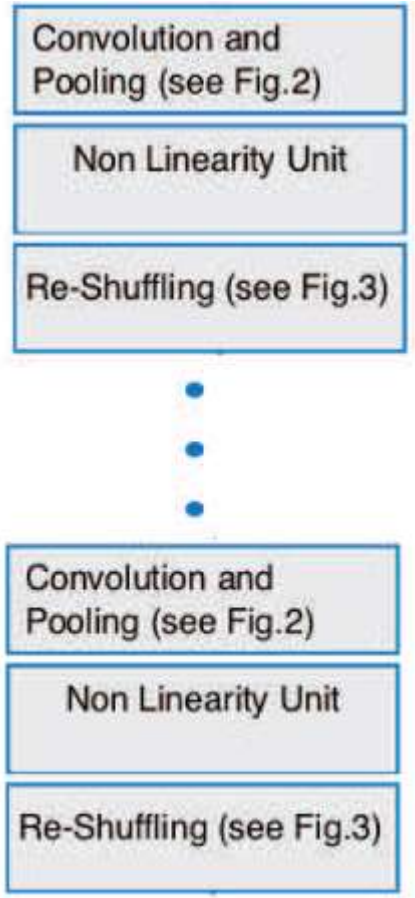


FIG. 4. (a) Conventional picture. Error rate as a function of number of MACs for different fully connected MNIST neural networks. (b) SQL picture. Error rate as a function of total number of photons, for the same networks.

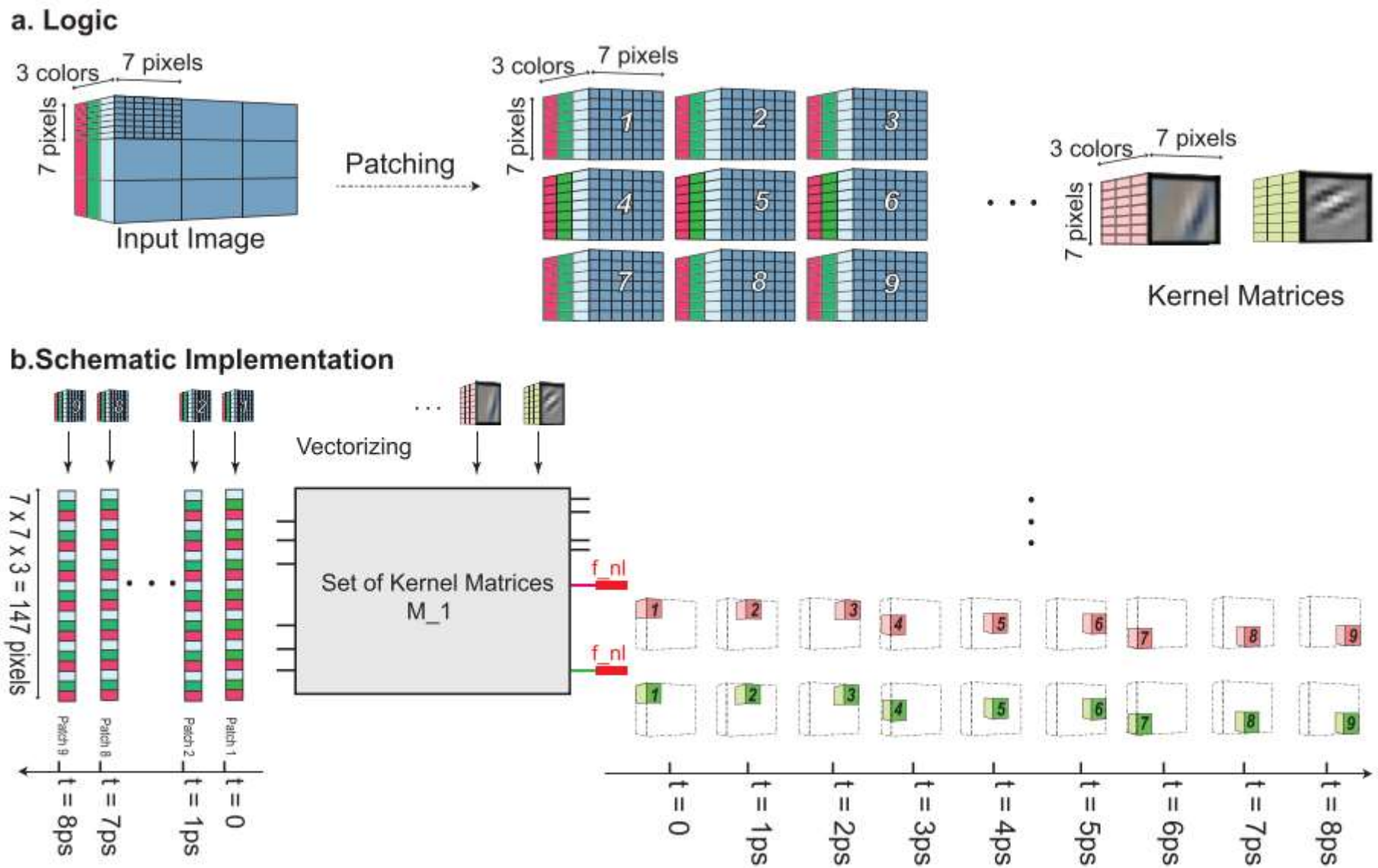
Ryan Hamerly et al., Large-Scale Optical Neural Networks Based on Photoelectric Multiplication, Physical Review X, 2019

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

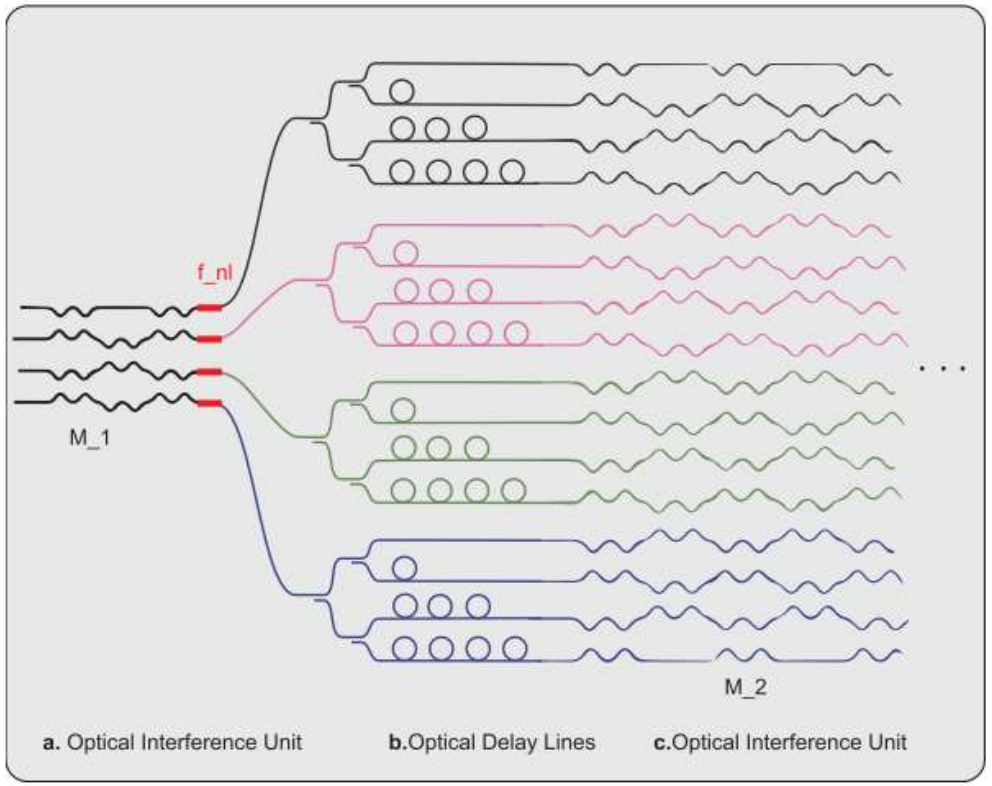
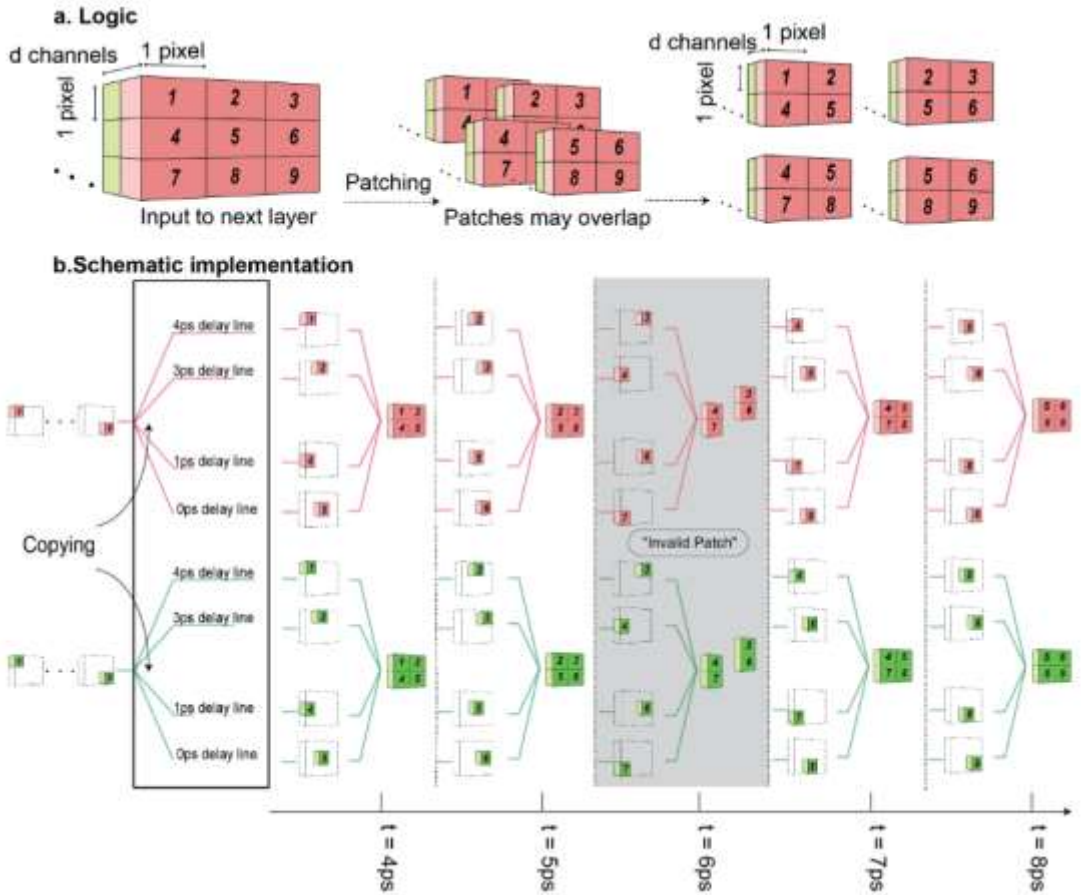
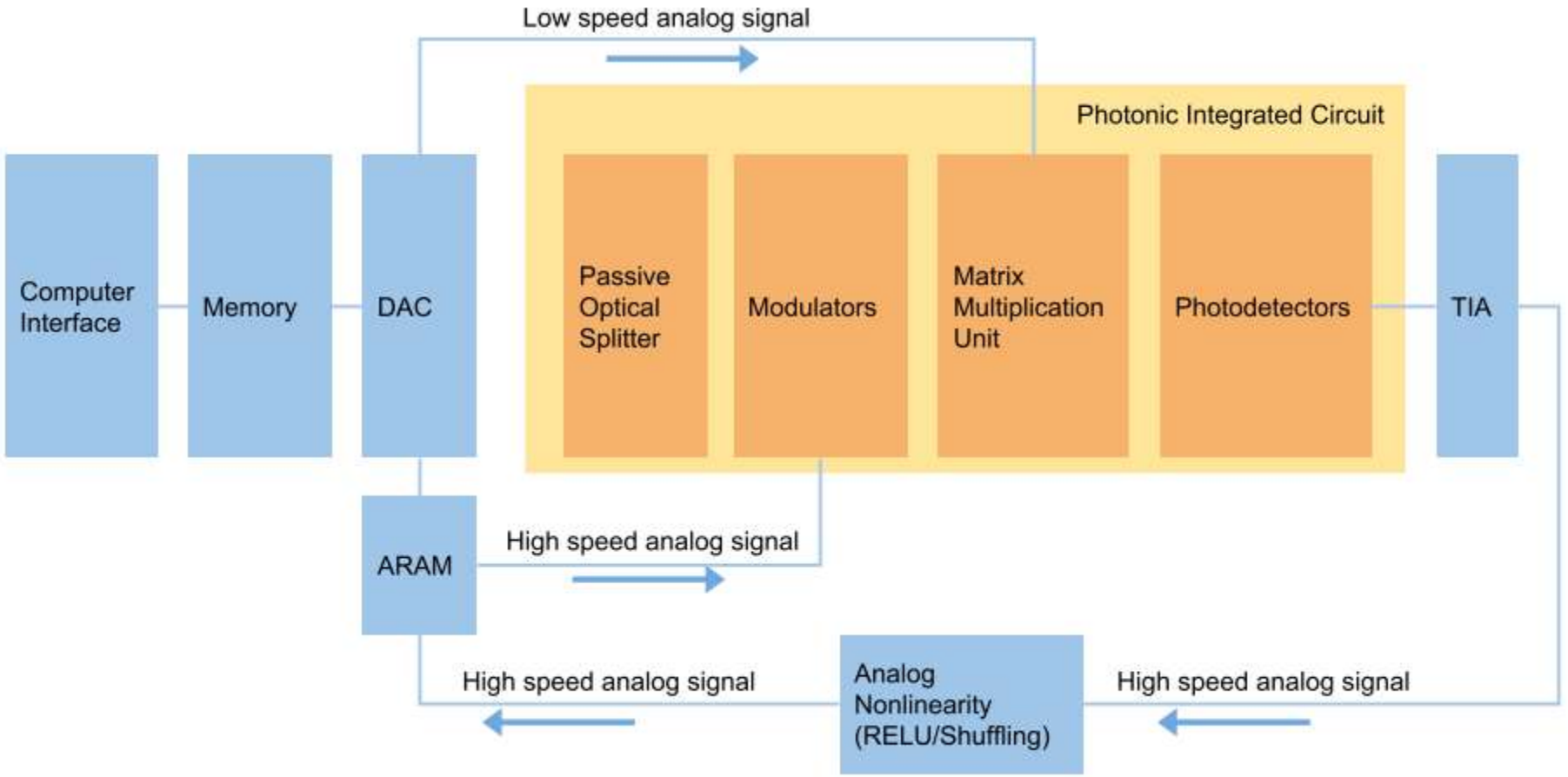


Fig. 4. Illustration of optical interference unit with delay lines a. Optical interference unit: In the first stage an optical interference unit is used to implement a kernel matrix M_1 which processes the patches from the original image. The red segments on the output of M_1 are optical nonlinearity. b. Optical Delay Lines: In the second stage, optical delay lines properly reform the sequence of kernel dot products into new patches for input into the second kernel matrix M_2. c.Optical interference unit: In the third stage the next optical interference unit is used to implement M_2 (partially depicted here). For clarity the actual number of inputs and outputs have been reduced and the attenuator stage and subsequent additional optical interference units have been omitted from M_1 and M_2. Additionally we have omitted optical amplifiers required in each layer which boost the power sufficiently to trigger the optical nonlinearity.



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

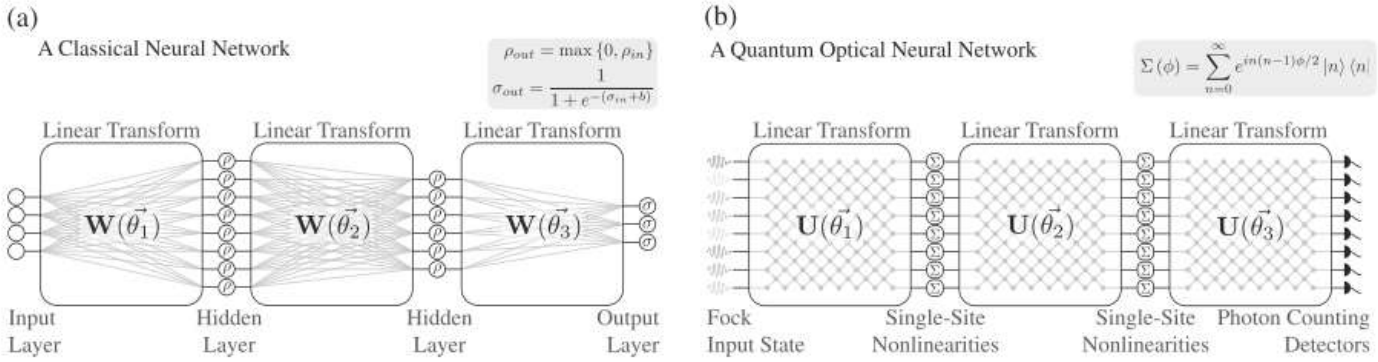


Fig. 1 Quantum optical neural network (QONN). **a** An example of a classical neural network architecture. Hidden layers are rectified linear units (ReLUs) and the output neuron uses a sigmoid activation function to map the output into the range (0, 1). **b** An example of our quantum optical neural 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

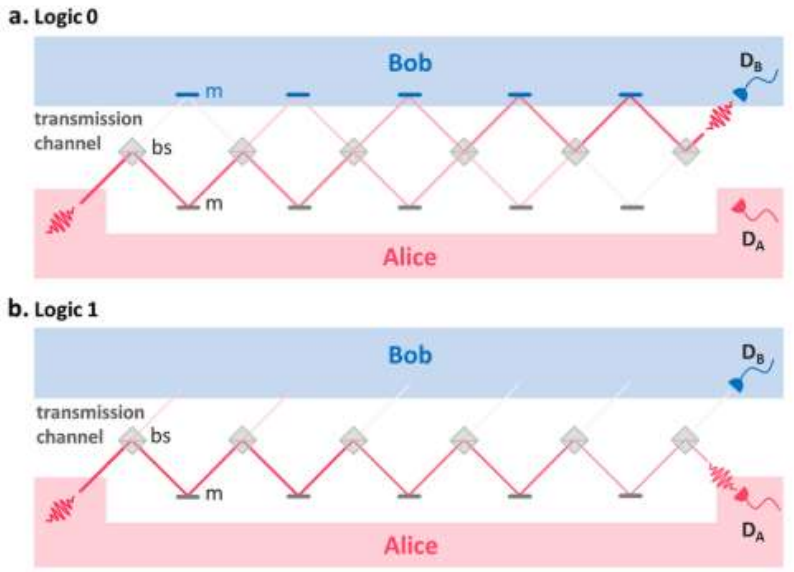
Trace-free counterfactual communication with a nanophotonic processor

- 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 $\hat{V}(\vec{\emptyset})$ to find the time reversed condition$\hat{V}(\vec{\emptyset}) = U^{\dagger}$ for a known input state

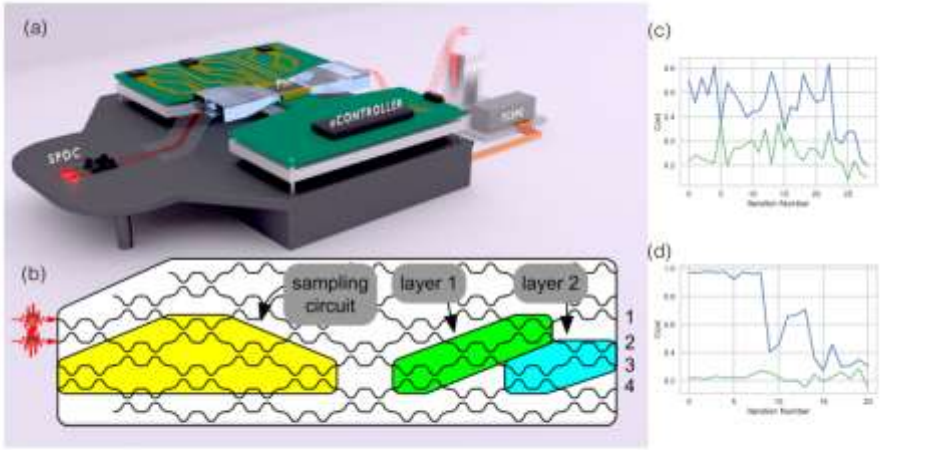


Fig. 1. (a) Schematic of the full experimental setup. (b) The structure of the full VQU protocol, with the optical sampling circuit (yellow) and the subsequent unsampling layers (green, blue). Experimental results for (c) the first layer of unsampling, and (d) the final layer of the protocol.

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

问题：

- 电路中的精度是什么？
- 光路中的精度与哪些因素有关？
- 这些因素在电路中是否也要考虑？为什么？
- 为什么和其他因素无关？

电路中精度

- 计算的精细程度
- 电平高低表示“0”和“1”
- 精度不受限于硬件
- 理论上可以实现无线高精度的计算（耗时）

光路中精度

- 模仿电路中使用高低电平来表示“0”和“1”
- 所以理论上只要能分辨出“0”和“1”，该光路就可以用于计算，但是存在计算速度的问题。
- 达到精度要求：可以分辨出“0”和“1”
- 相关因素：取决于具体结构
  - 探测器灵敏程度
  - 芯片设计的复杂程度

(以linear programmable processor为例)

Split ratio

- performance is limited by beam splitters that deviate from the ideal 50：50 split.

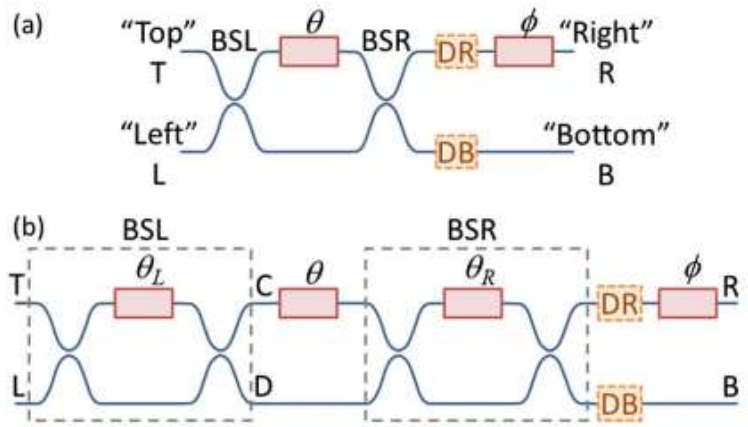
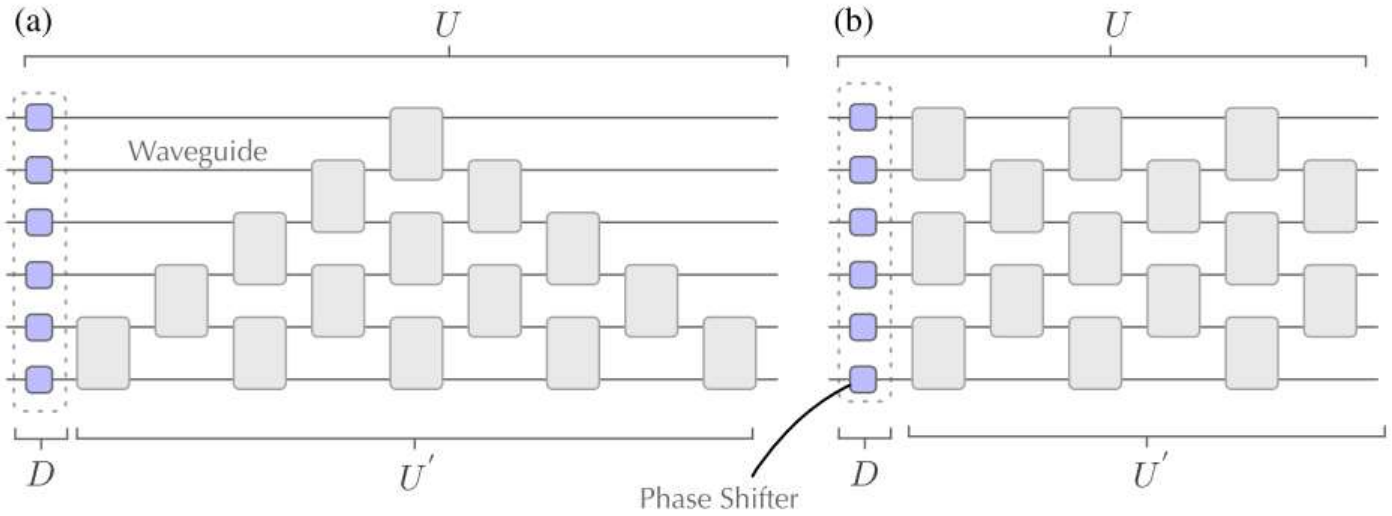


Fig. 2. MZI block configurations. (a) "Top," "Bottom," "Left," and "Right" label waveguides in correspondence with the faces of a conventional cube beam splitter. BSL and BSR are the nominally 50:50 beam splitters. The fabricated split ratios of all beam splitters may differ from this ideal split ratio. DR and DB are optional detectors, which will be mostly transparent, sampling a small amount of the power in their respective waveguides to give the signal for the feedback loops used to set up the block's function. (b) shows those beam splitters themselves implemented with additional MZIs that may also include beam splitters with fabricated split ratios unintentionally different from 50:50.

1. setting them up using a "beam splitter 50: 50 setup algorithn" (BFSA) to be 50: 50 in a MZI based only on maximizing or minimizing power on a detector
2. using a double Mach-Zehnder interferometer (DMZI) configuration as if it were a perfect MZI with 50: 50 beam-splitter ratios
3. "mesh 50: 50 setup algorithm" (MFSA)

DAVID A. B. MILLER, Perfect optics with imperfect components, Optical Society of America, 2015

MZI的排列方式

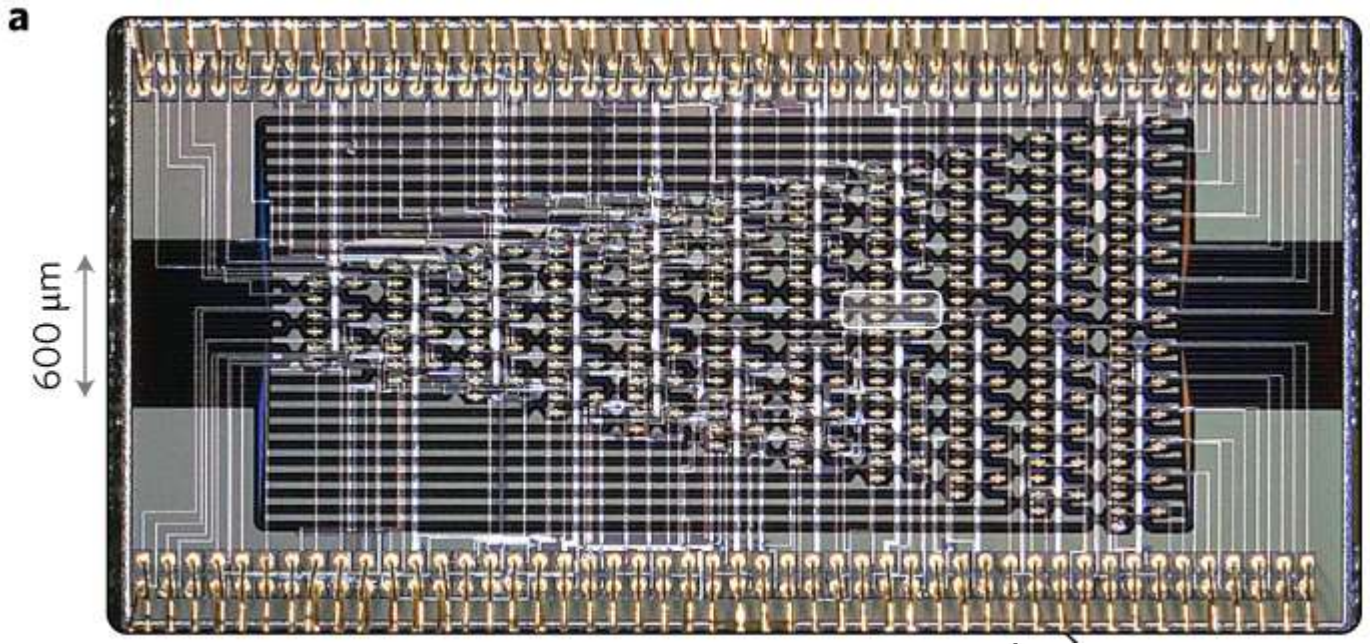


(a) reck(1994) (b)clement(2016)

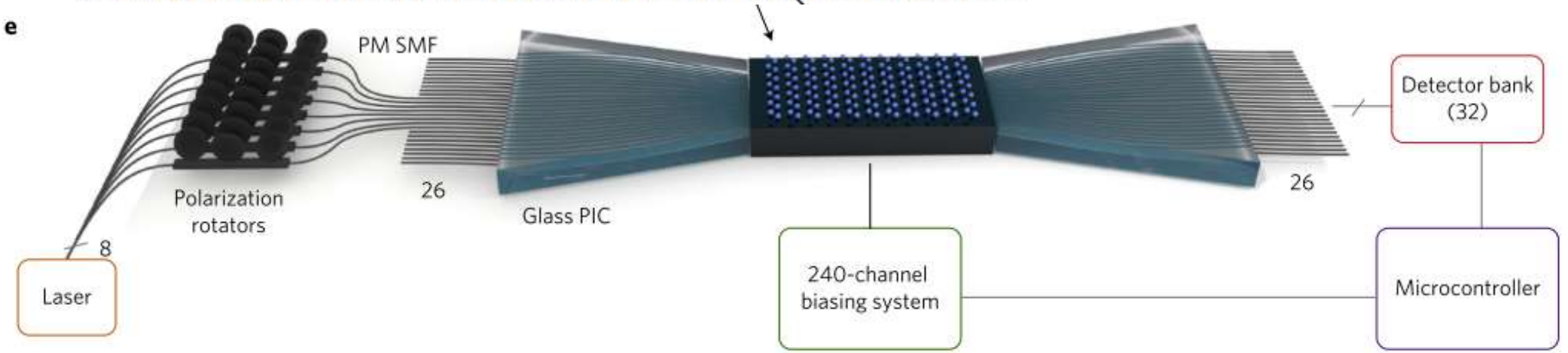
- new design based on an alternative arrangement of beam splitter and phase shifters
- half the optical depth of the Reck design and is significantly more robust to optical losses

Willian R. Clements, Optimal design for universal multiport interferometers, Optica, 2016

Losses



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].



问题:

- fiber tip的自定义加工（polish等）

- 长飞 和 烽火未找到

Axicon lenses

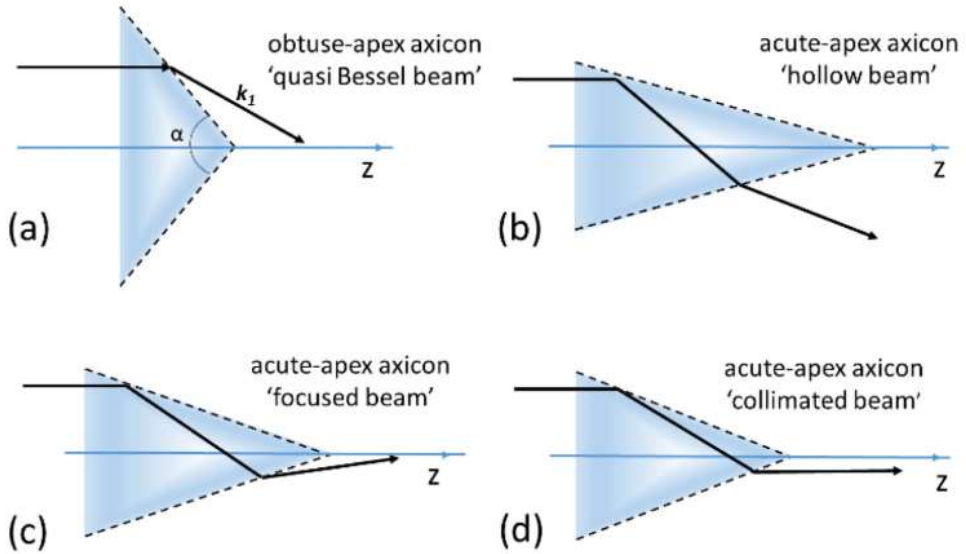


Fig. 1. Schematic illustration of ray paths for axicon lenses, including (a) obtuse apex-angle and (b)–(d) acute apex axicon with single internal reflection.

- 产生Bessel光束

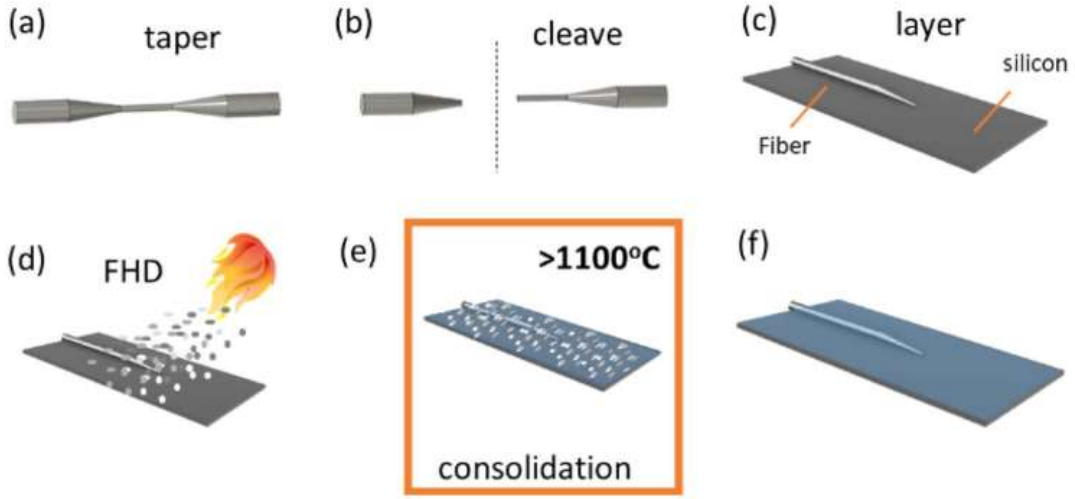


Fig. 2. Schematic illustration showing consecutive steps (a)–(f) of integrated optical fiber fabrication and self-assembly of the axicon lens.

加工过程

1. tapering using micro-heaters
2. flame hydrolysis deposition (FHD) processes

Christopher Holmes et al., Self-assembled axicon lens in integrated optical fiber, Optics Letters, 2019

Deep seated negative axicon (DSNA) probe

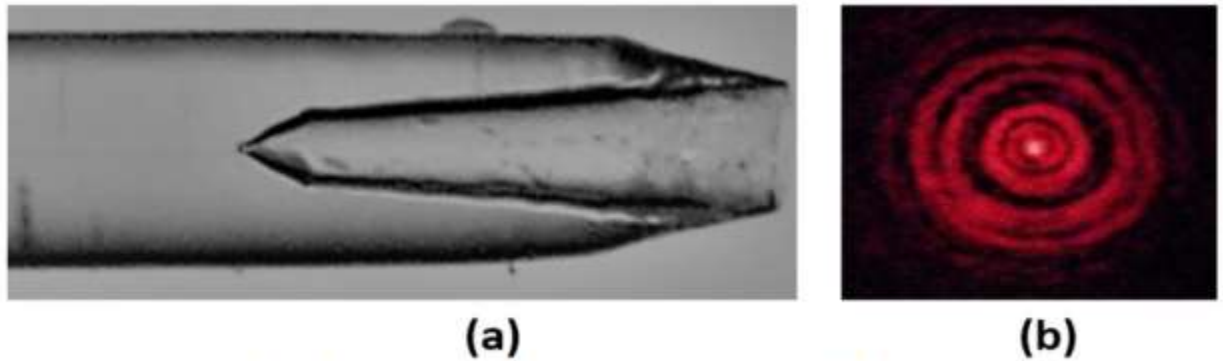


Figure 1 (a) Deep seated negative axicon tip, (b) Bessel beam

- fabricated at the distal end of a single mode optical fiber
- highly Ge doped inside both core and cladding by a chemical etching procedure

In this process, the fiber end is cleaved after removing the jacket of the fiber such that only core and cladding part of fiber undergo the etching and placed on the surface of the convex meniscus formed by 48% HF and toluene in a tube of ~2mm diameter under the controlled conditions. The etching occurs both side-wise and inside the fiber due to the capillary action, but the later action is foremost in DSNA formation.

P. Gupta et al., Application of Bessel Beam from Deep Seated Negative Axicon in Optical Coherence Tomography of Tissue Structure, SPIE, 2019