



Advanced Photonics

Joint doctoral course / SMYLE / FEMTO-ST / EPFL

Optical Computing II



Niyazi Ulas Dinc

June 18, 2026

Main theme: 3D Optical Neural Networks

1. Motivation

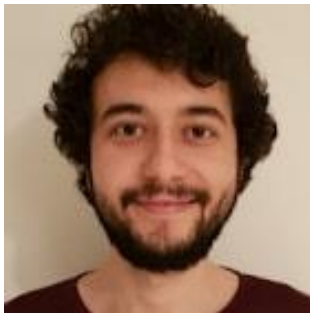
- Why optical computing? Why neural networks?
- Why 3D? 2D vs 3D

2. 3D optical storage (weight allocation)

3. OEO and nonlinear processing with linear optics

4. An example on optical generative models

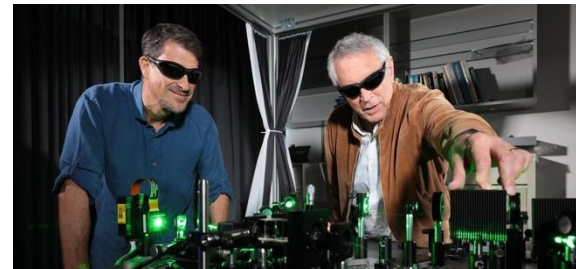
5. Examples on extreme learning machines



Dr. Niyazi Ulas Dinc
niyazi.dinc@epfl.ch

Sciencepreneur
Zigzag Interconnect: Reliable Circuit-Switched Optical
Networking

PhD, EPFL, January 2024
MSc, EPFL, August 2019
BSc, METU (Ankara, Turkey), June 2017



Prof. Moser & Prof. Psaltis

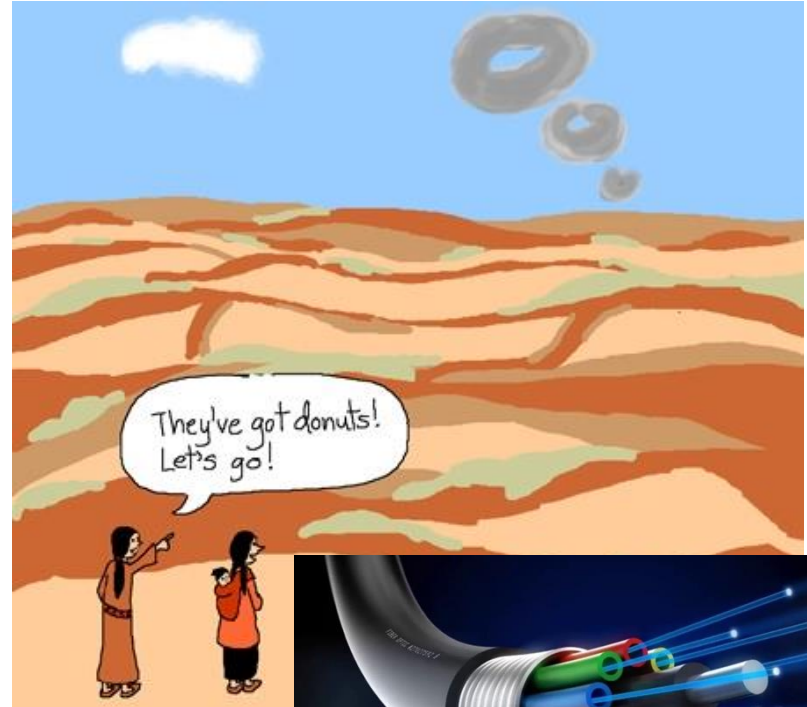
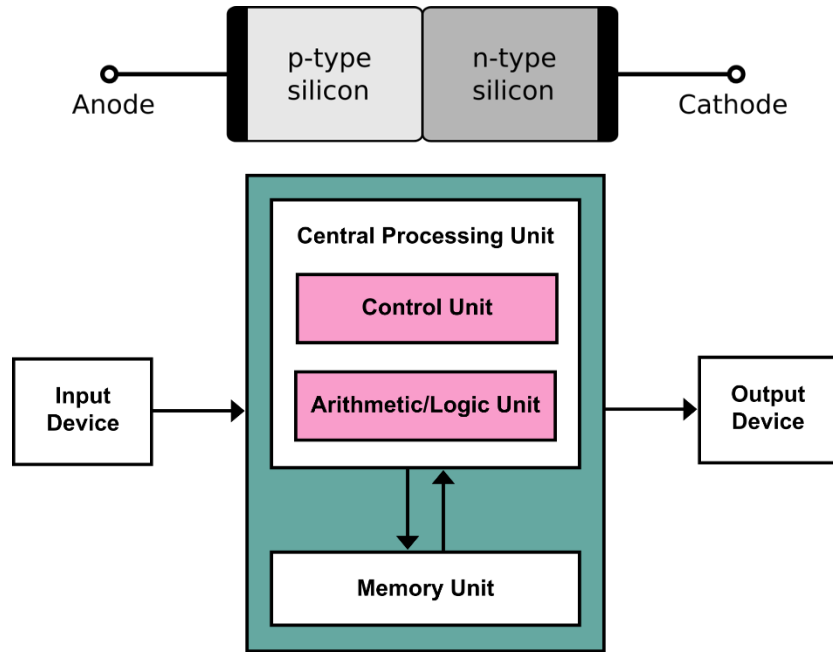


Dr. M. Yildirim



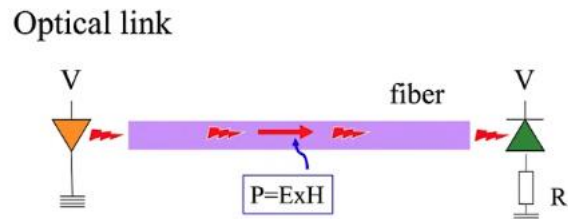
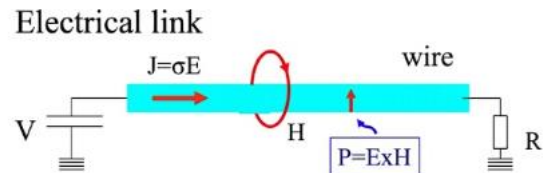
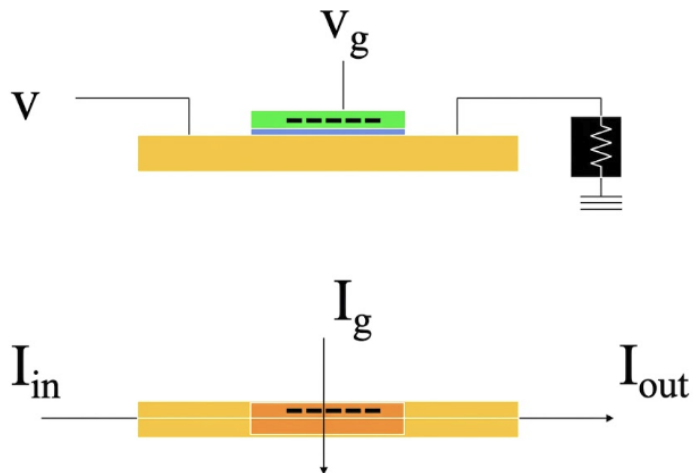
Dr. I. Oguz

“Conventional wisdom says that electrons compute and photons communicate”

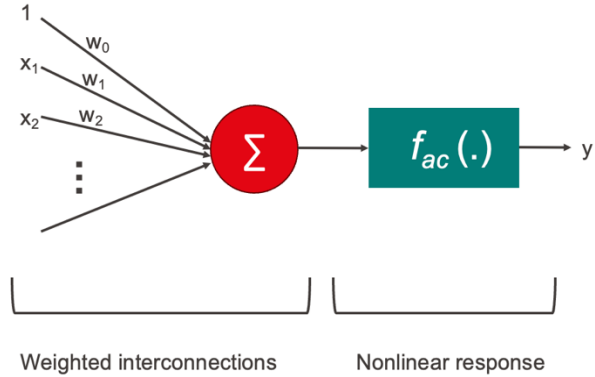
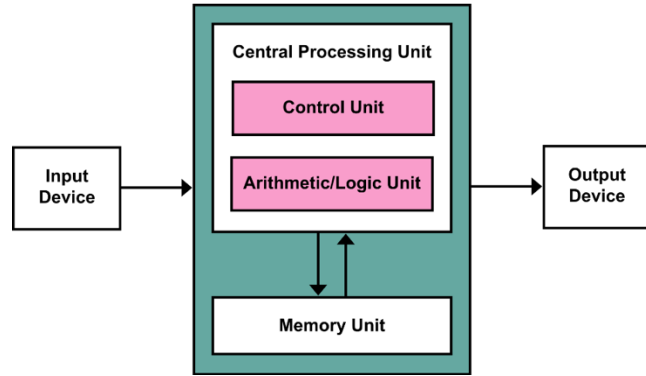


- Athale & Psaltis (2016) Optical computing: past and future. *Optics and Photonics News* (27.6) 32-39. Images: Various sources from web

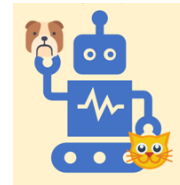
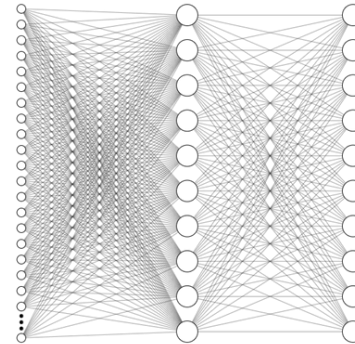
- Quasi-static Coulomb interactions
- Strong, localized forces => Boolean logic
- Moore's law

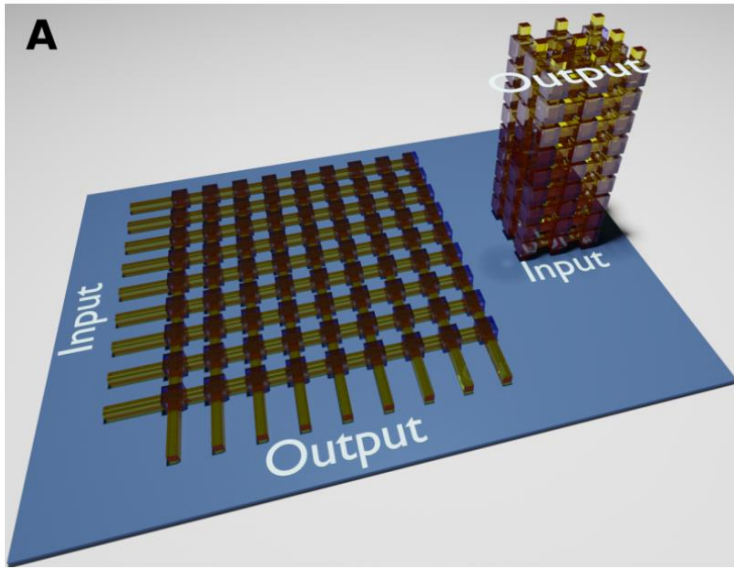


- No loss, no RC constant
- High carrier frequency
- No photon-photon interaction in linear regime

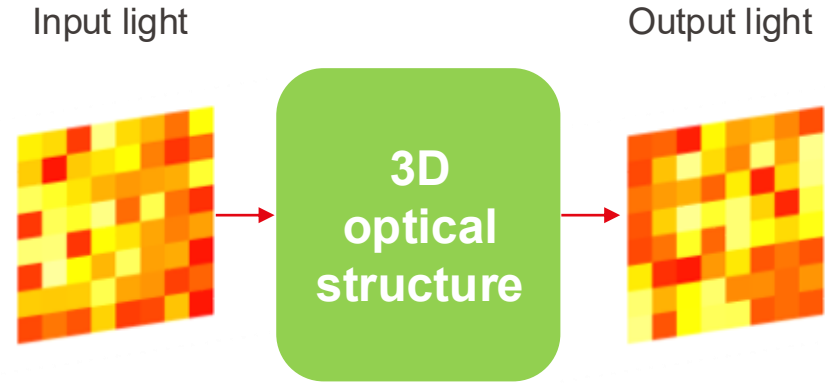


$+$ $-$
 \div \times





3D optics can increase the number of stored/used weights without being limited to a planar interconnect layout

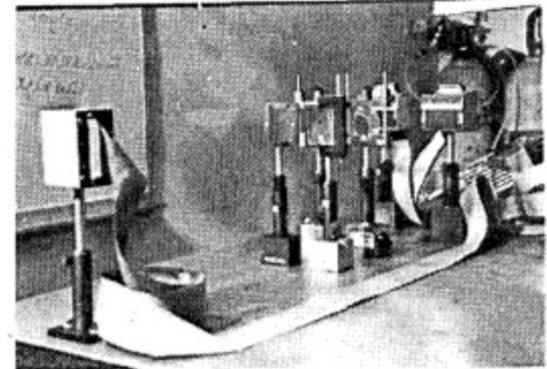
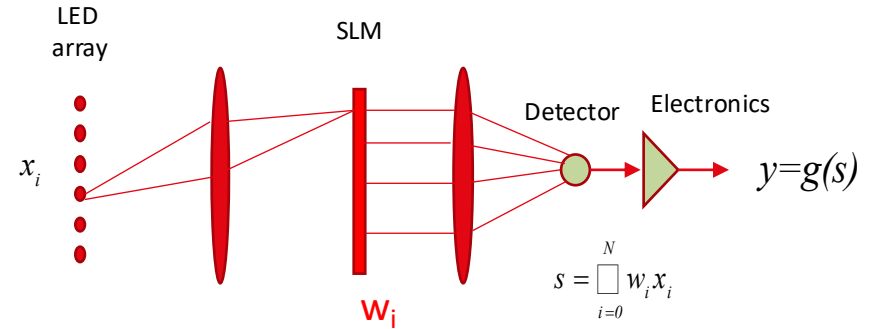
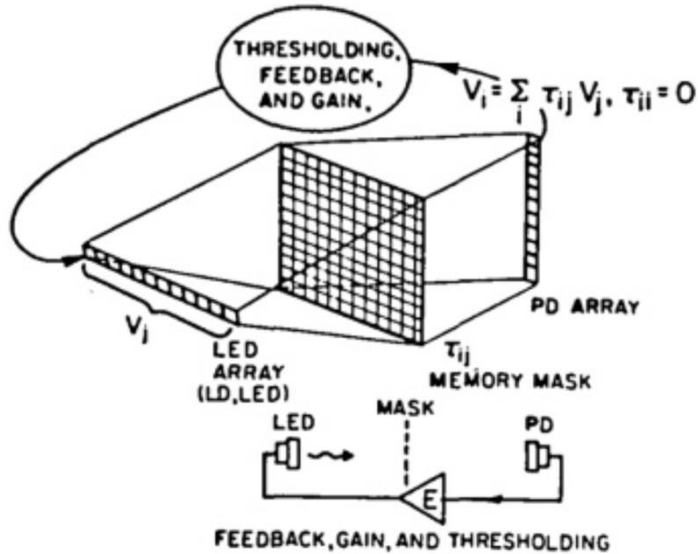


1 cm² MZI mesh
10² MZIs => 10² weights, 10 input dimensions
10² MZIs x 100 GHz = 10 TOP/S

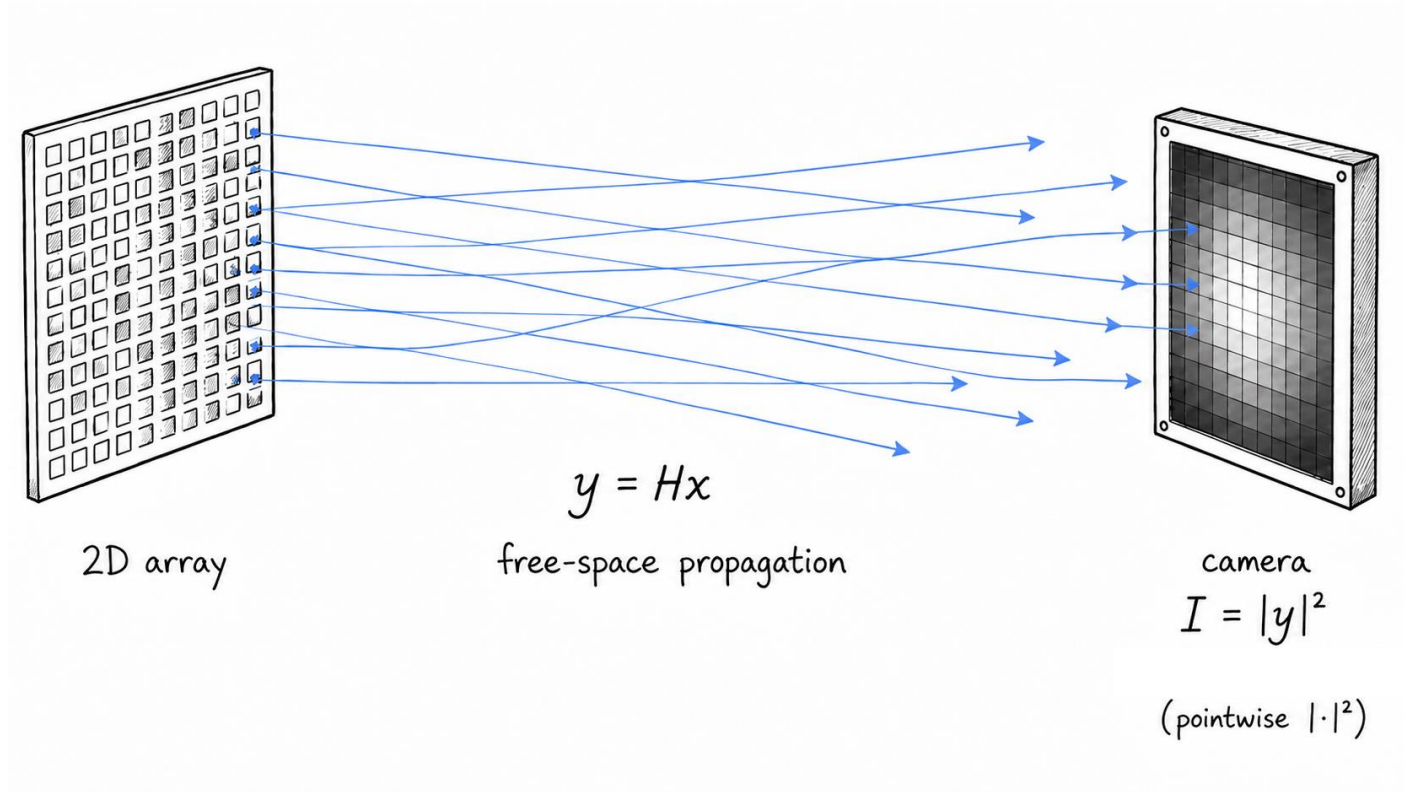
1 cm² LC or DMD display
10⁶ pixels, 10⁶ weights or 10⁶ input dimensions?
10⁶ pixels x 1 kHz = 1 GOP/S
 $\mathcal{O}(n^2)$ => 1 POP/s

NVIDIA A100 35 TOP/s/cm²
1 pJ/Operation

EPFL The first optical neural network



Psaltis and Farhat, Optics Letters Vol. 1985
 Farhat, Psaltis, Prata, and Paek, Applied Optics, 1985



Main theme: 3D Optical Neural Networks

1. Motivation

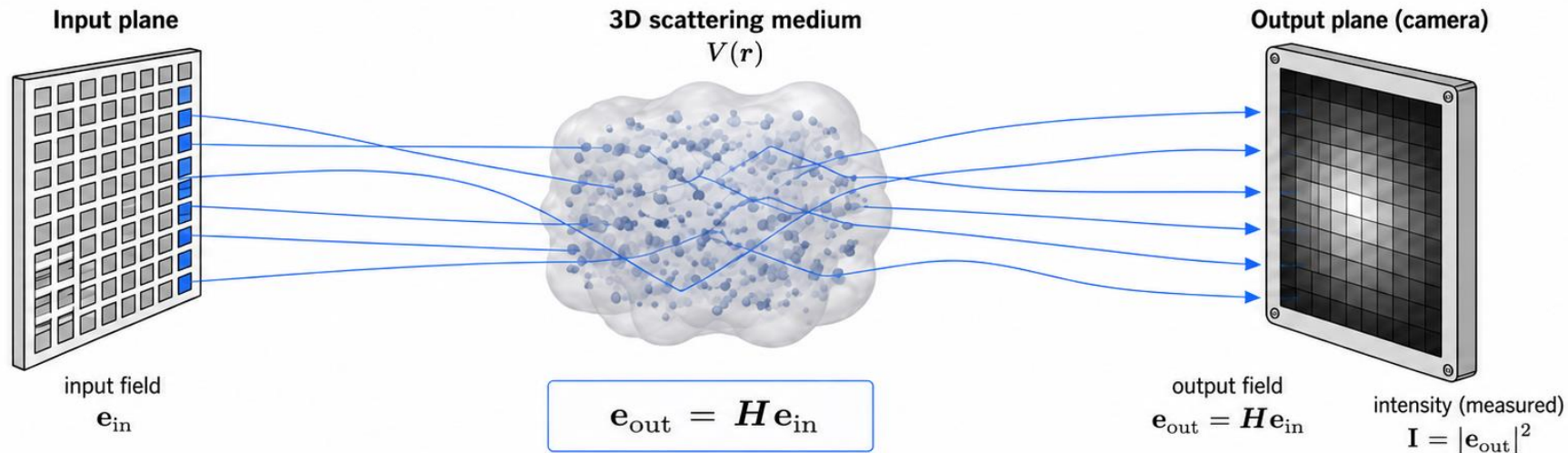
- Why optical computing? Why neural networks?
- Why 3D? 2D vs 3D

2. 3D optical storage (weight allocation)

3. OEO and nonlinear processing with linear optics

4. An example on optical generative models

5. Examples on extreme learning machines



Continuous-wave model

$$(\nabla^2 + k_0^2) \mathbf{E}_s(\mathbf{r}) = -V(\mathbf{r}) \mathbf{E}(\mathbf{r})$$

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_{in}(\mathbf{r}) + \mathbf{E}_s(\mathbf{r})$$

$$V(\mathbf{r}) = k_0^2 (n^2(\mathbf{r}) - n_b^2)$$

Discrete (matrix) model

$$\mathbf{e}_{out} = H\mathbf{e}_{in}$$

H = transmission matrix of the 3D medium

$$H = V + VG_0V + V[G_0V]^2 + V[G_0V]^3 + \dots$$

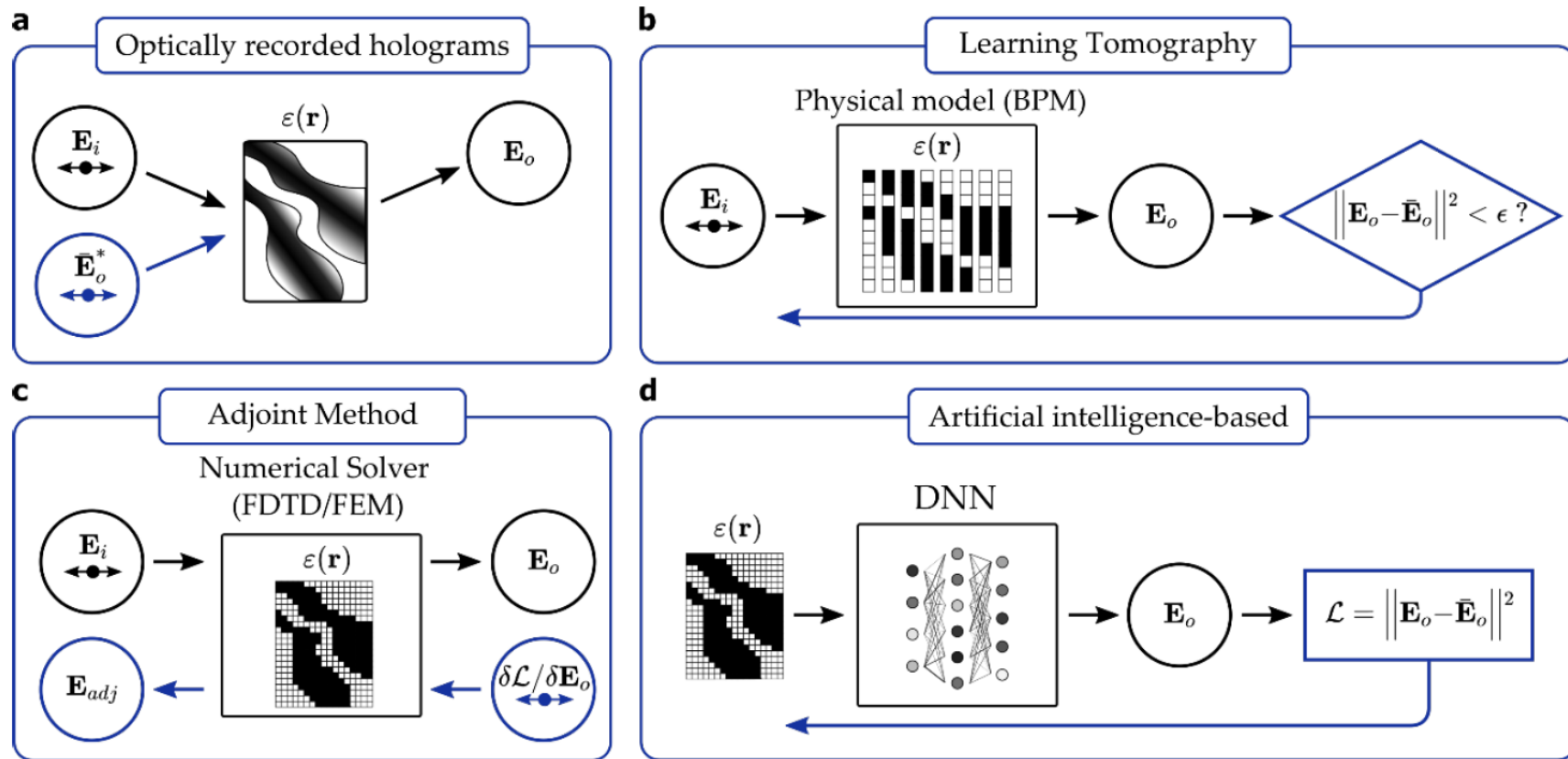
H : transmission matrix (input \rightarrow output)

V : scattering potential matrix
(diagonal, from $V(\mathbf{r})$)

G_0 : background Green's matrix
(free-space / background medium)

$\mathbf{e}_{in}, \mathbf{e}_{out}$: discretized complex field vectors
(at input and output planes)

\mathbf{I} : detected intensity (camera)
(pointwise magnitude square)



EPFL Volume hologram implementation of neural network

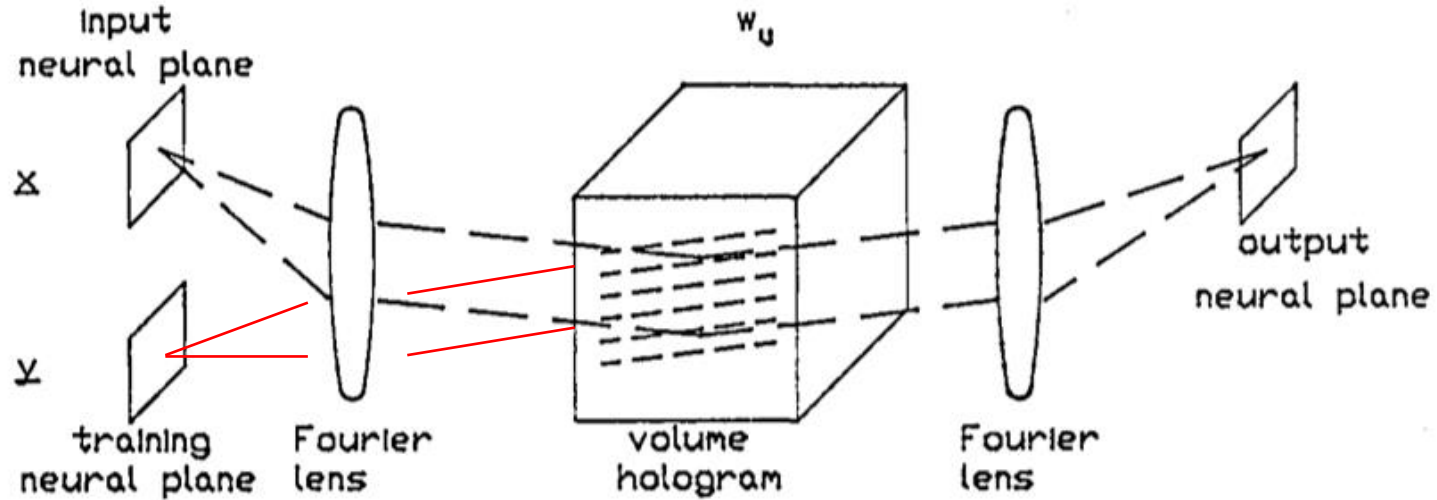
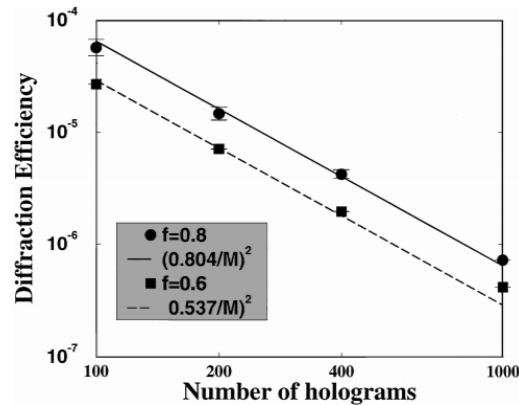
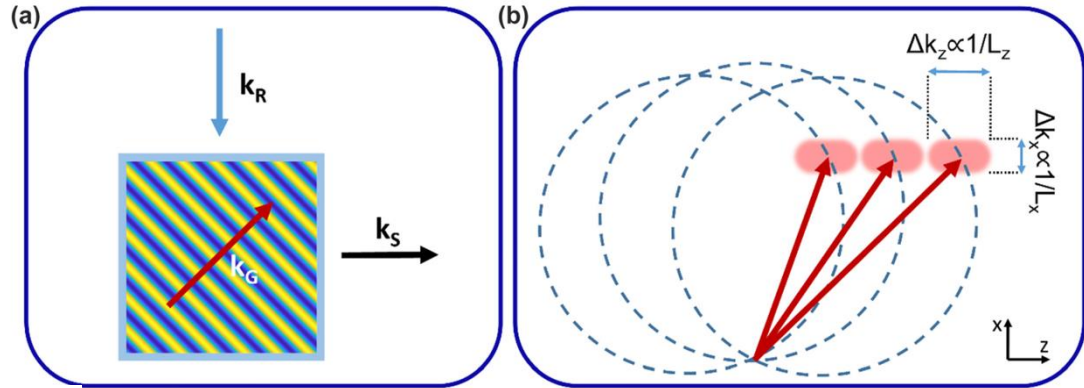
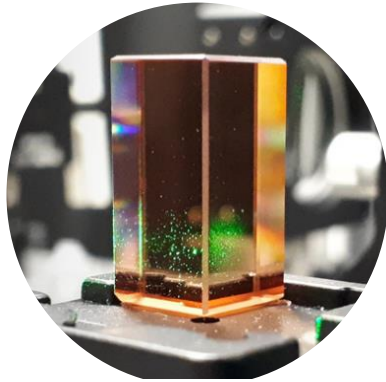


Fig. 3. Optical neural computer architecture.

Adaptive optical networks using photorefractive crystals,
Psaltis, Brady and Wagner, Applied Optics, 1988.



$$\propto M$$

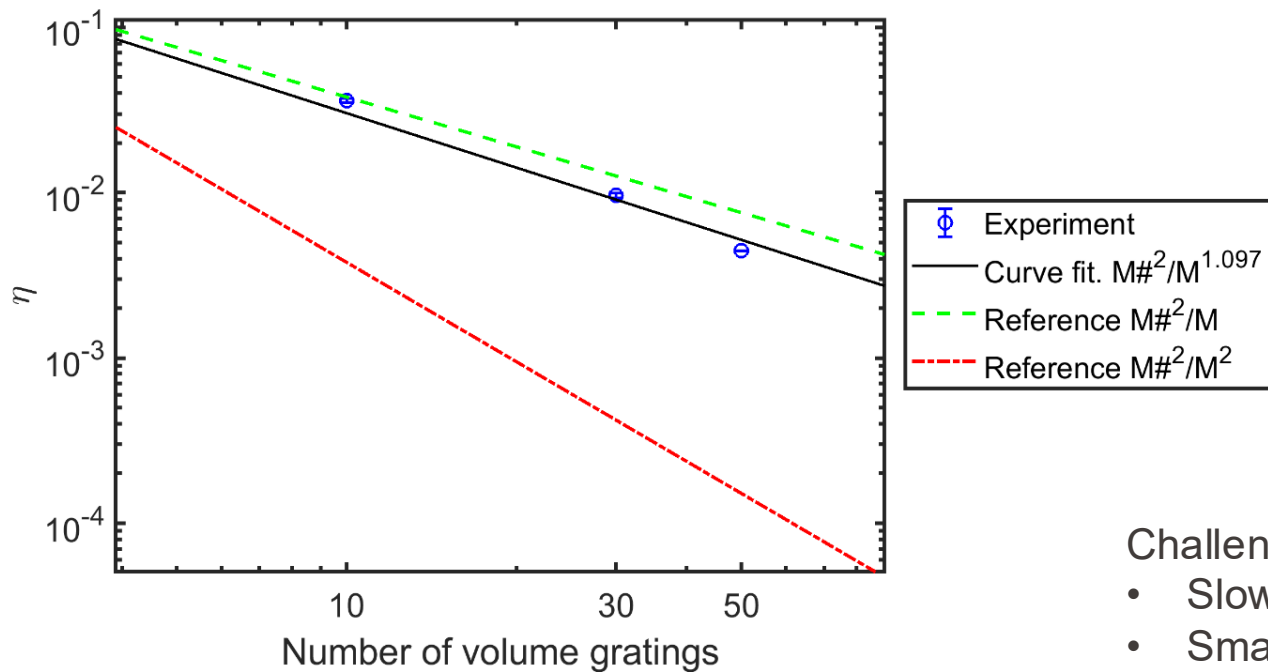
$$\propto \sqrt{M}$$

$$A_m = |E_R + E_S|^2 = \underbrace{E_R E_R^*}_{\text{DC}} + \underbrace{E_S E_S^*}_{\text{Noise}} + \underbrace{E_R^* E_S}_{\text{Reconstruction}} + \underbrace{E_R E_S^*}_{\text{Conjugate}}$$

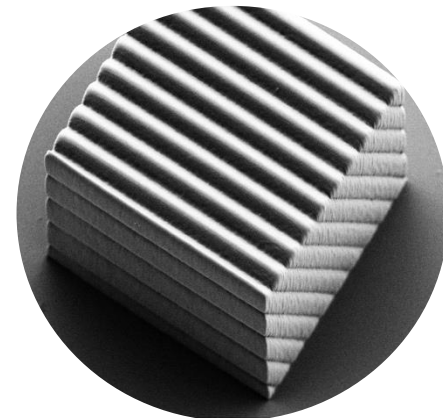
$$\eta = \left(\frac{M\#}{M} \right)^2$$

M: number of holograms multiplexed

Multiplexed volume holograms trade storage density against diffraction efficiency and crosstalk.

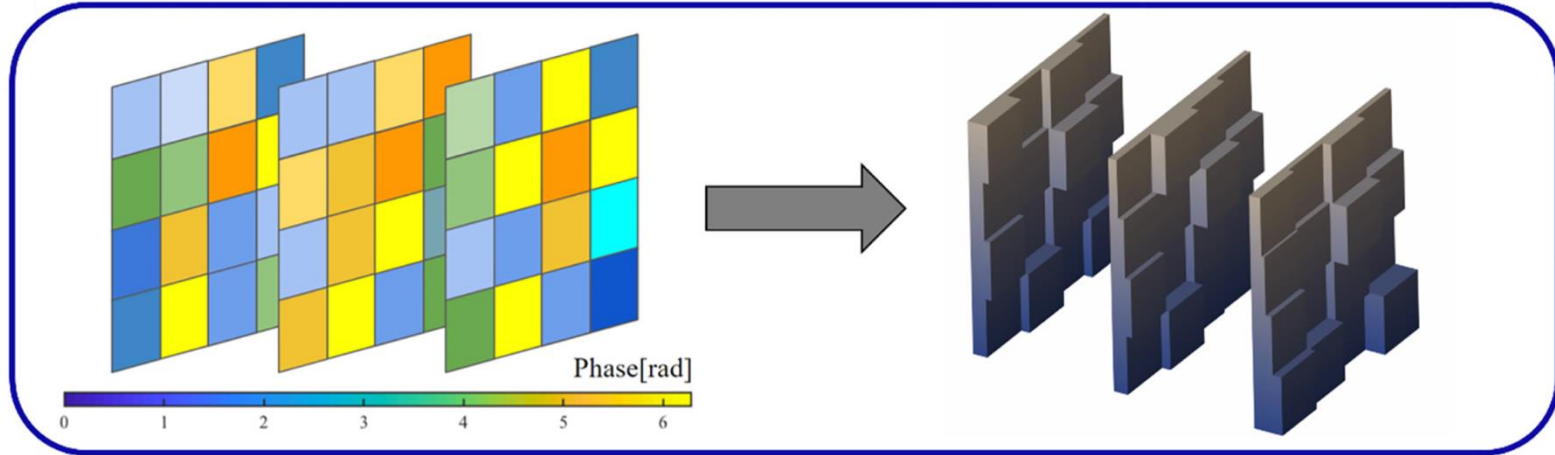


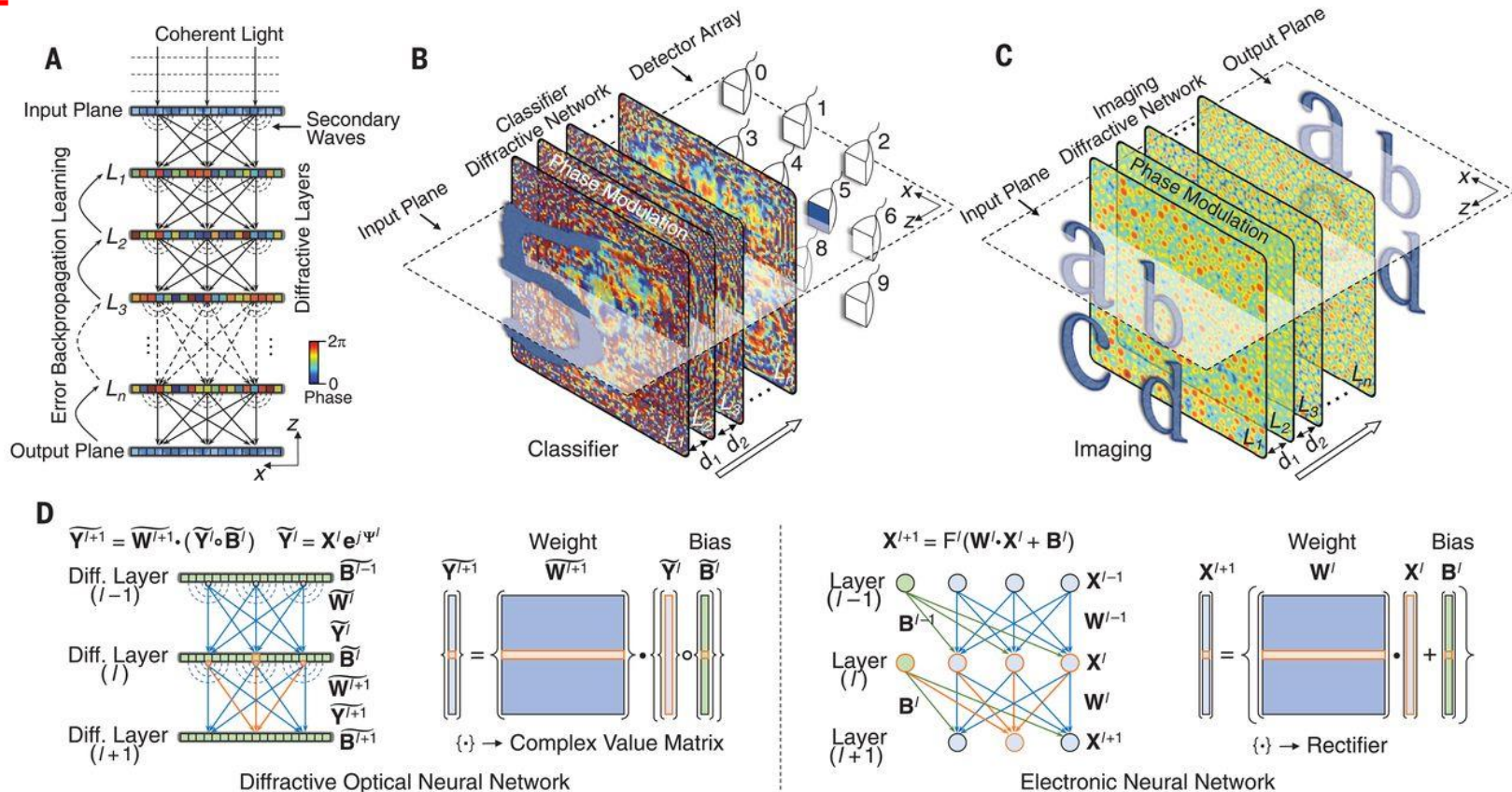
(3+1)D printed using
Nanoscribe



Challenges:

- Slow fabrication
- Small volume
- Δn ?





Xing Lin *et al.*

All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004-

1008(2018). DOI: [10.1126/science.aat8084](https://doi.org/10.1126/science.aat8084)

Main theme: 3D Optical Neural Networks

1. Motivation

- Why optical computing? Why neural networks?
- Why 3D? 2D vs 3D

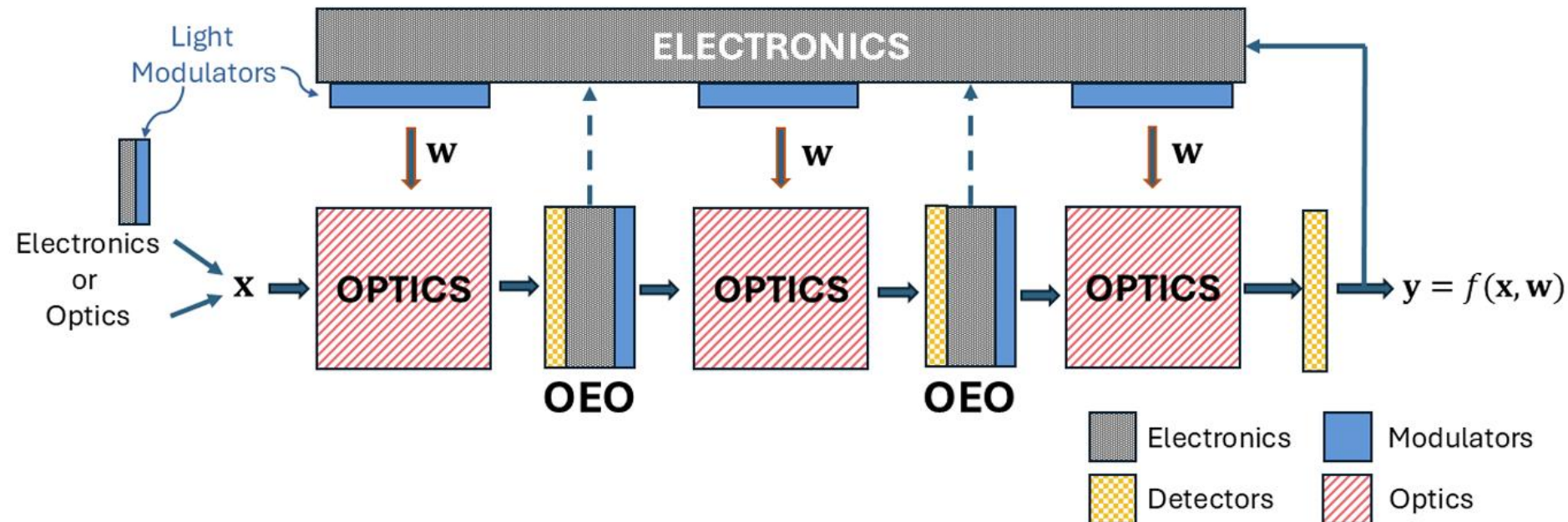
2. 3D optical storage (weight allocation)

3. OEO and nonlinear processing with linear optics

4. An example on optical generative models

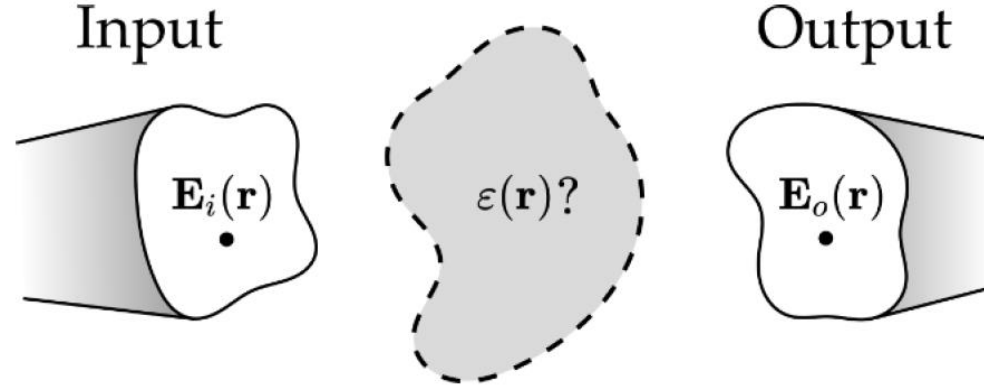
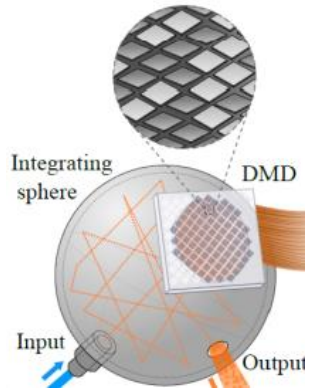
5. Examples on extreme learning machines

Opto-electronic compute: optical connections and electronic nonlinear activations



Brunner, Daniel, et al. "Roadmap on neuromorphic photonics." *arXiv preprint arXiv:2501.07917* (2025).

- Dinc, Niyazi Ulas, et al. "On the implementation of nonlinearities in optical neural networks: opinion." *Optical Materials Express* 14.10 (2024): 2413-2418.



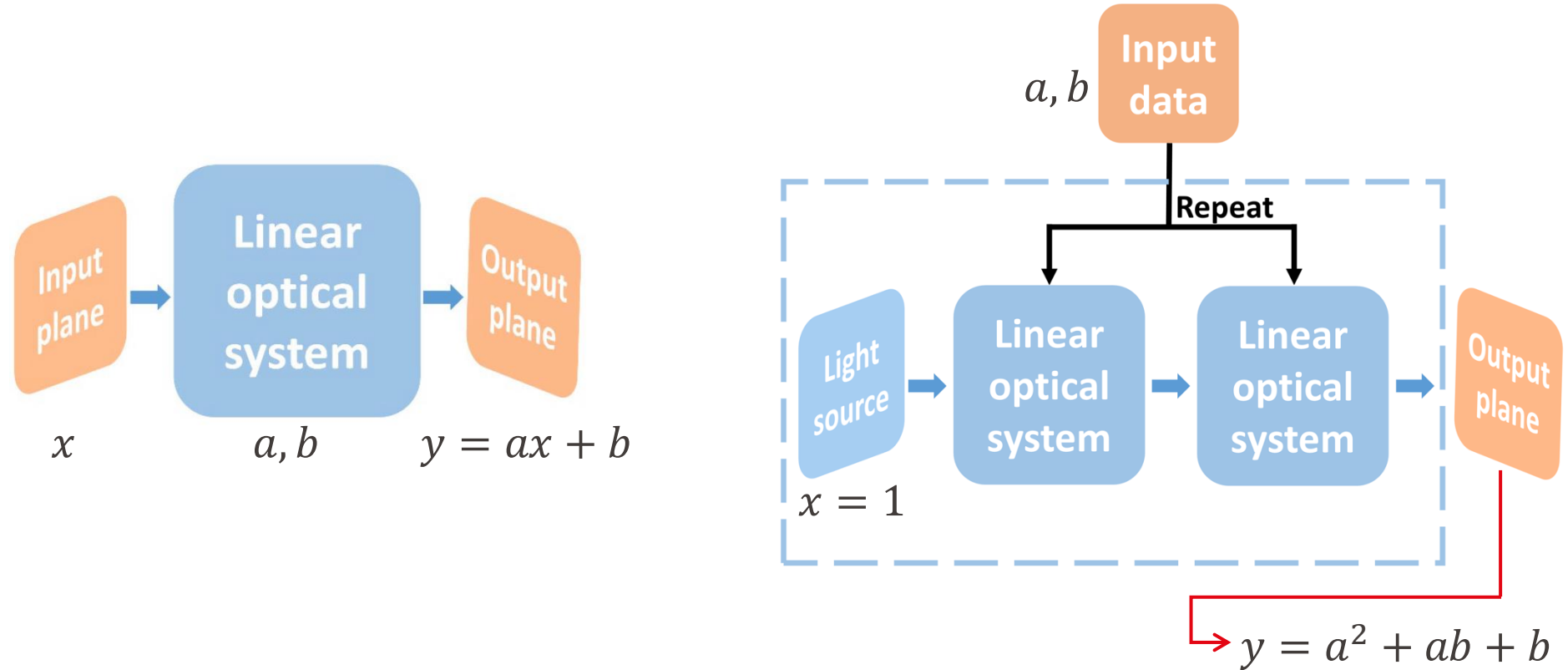
$$E_o = \mathbf{T}E_i = \left(\mathbf{V} + \mathbf{V}\mathbf{G}_0\mathbf{V} + \mathbf{V}[\mathbf{G}_0\mathbf{V}]^2 + \mathbf{V}[\mathbf{G}_0\mathbf{V}]^3 \dots \right) E_i$$

Eliezer, Ruhrmair, Wisiol, Bittner, and Cao.
Exploiting structural nonlinearity of a reconfigurable multiple-scattering system.
arXiv preprint arXiv:2208.08906 (2022).

\mathbf{T} : transmission matrix
 \mathbf{V} : scattering potential
 \mathbf{G}_0 : free space Green's matrix

EPFL **n**onlinear **P**rocessing with **O**nly **L**inear **O**ptics

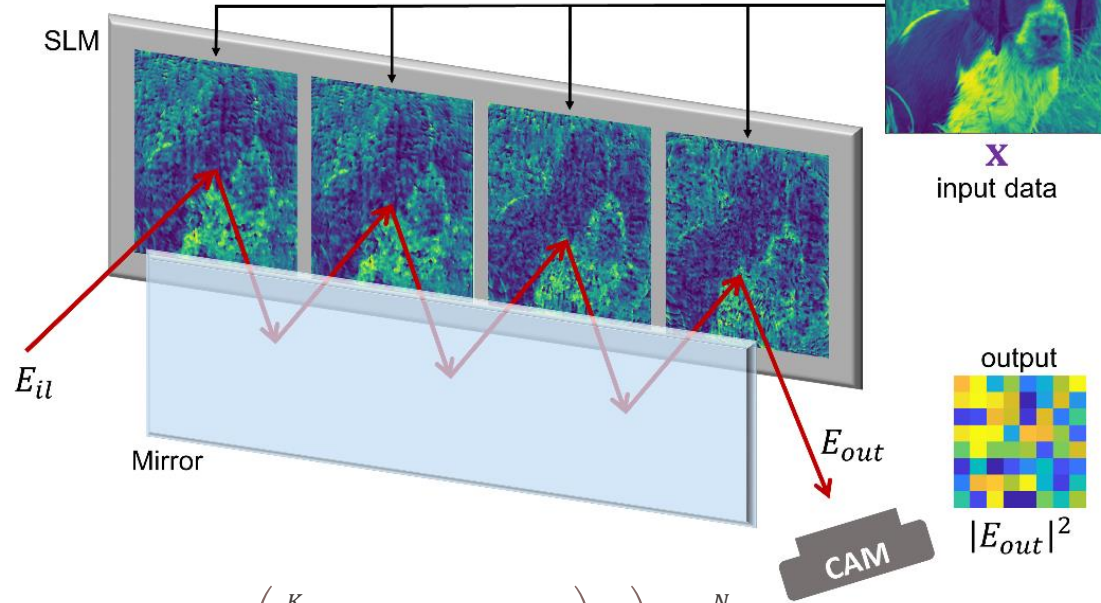
nPOLO



The nonlinearity comes from repeated data-dependent modulation and propagation, not from optical material nonlinearity.

$$E_{out} = \overbrace{HT_{LN}(\mathbf{x})}^{\text{N}^{\text{th}} \text{ layer}} \dots \overbrace{HT_{L2}(\mathbf{x})}^{\text{2}^{\text{nd}} \text{ layer}} \overbrace{HT_{L1}(\mathbf{x})}^{\text{1}^{\text{st}} \text{ layer}} E_{il}$$

$$T_{Ln} = e^{j(\mathbf{x} \odot \mathbf{S}_{Ln} + \mathbf{B}_{Ln})}$$

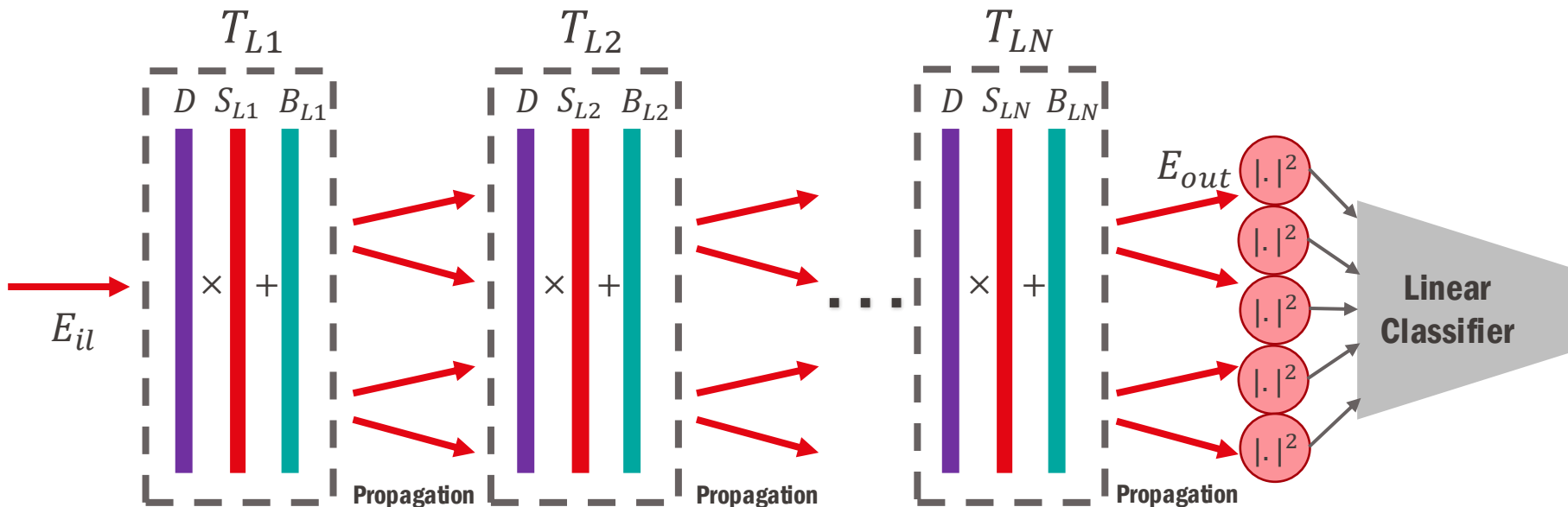


This implementation:

$$o_i = \sum_{j=1}^K h_{ij} (s_j^{(N)} x_j + b_j^{(N)}) \left(\dots \sum_{j=1}^K h_{ij} (s_j^{(2)} x_j + b_j^{(2)}) \left(\sum_{j=1}^K h_{ij} (s_j^{(1)} x_j + b_j^{(1)}) \right) \dots \right) = \sum_{m=0}^N \alpha_{i,m}(\mathbf{s}, \mathbf{b}) \mathbf{x}^m$$

Multi-layer perceptron:

$$o_i = g \left(\sum_{j=1}^K w_{ij} x_j + b_i \right) = \sum_m \beta_{i,m}(\mathbf{w}, \mathbf{b}) \mathbf{x}^m$$

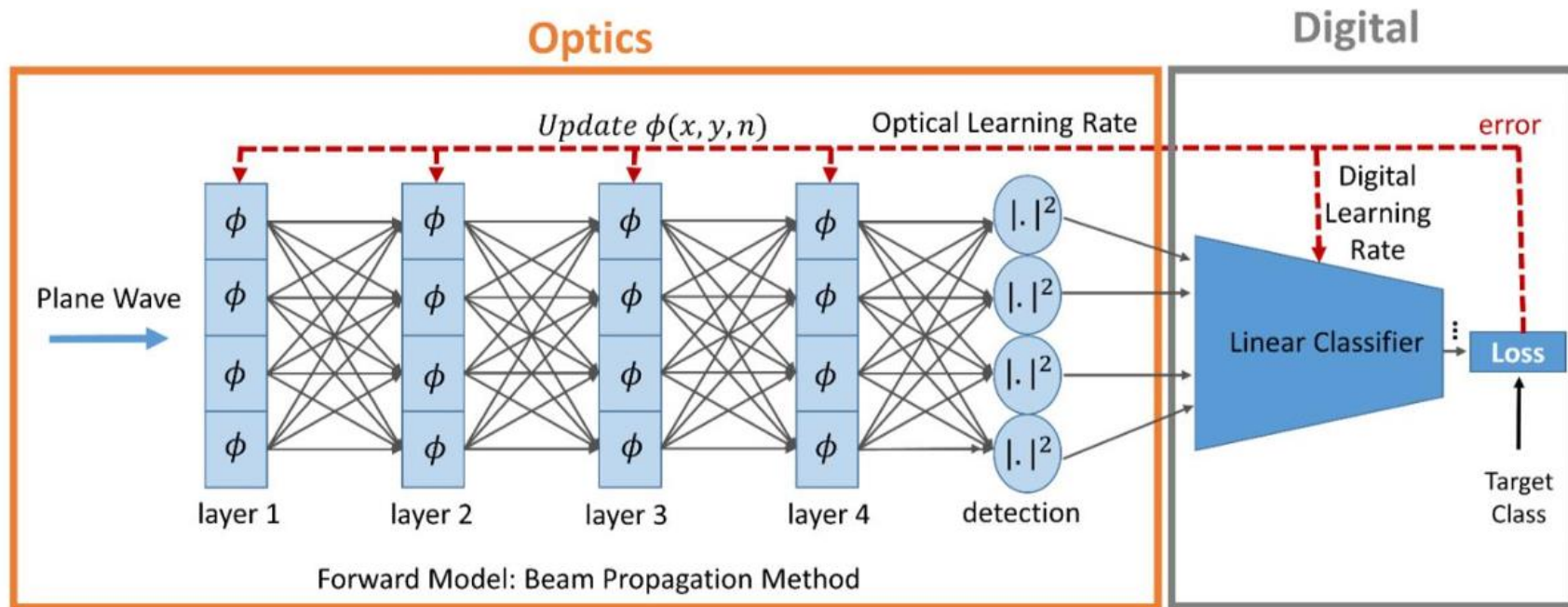


\mathbf{x}
Input Data



$$E_{out} = \overbrace{HT_{LN}(\mathbf{x})}^{\text{N}^{\text{th}} \text{ layer}} \dots \overbrace{HT_{L2}(\mathbf{x})}^{\text{2}^{\text{nd}} \text{ layer}} \overbrace{HT_{L1}(\mathbf{x})}^{\text{1}^{\text{st}} \text{ layer}} E_{il}$$

$$T_{Ln} = e^{j(\mathbf{x} \odot \mathbf{S}_{Ln} + \mathbf{B}_{Ln})}$$

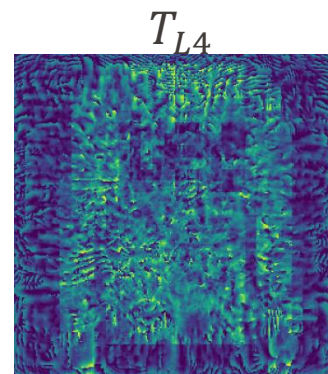
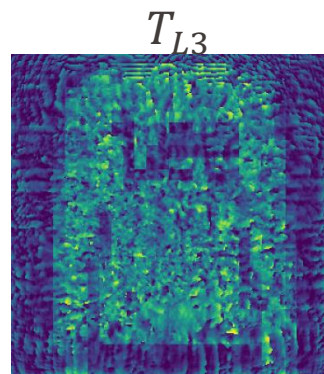
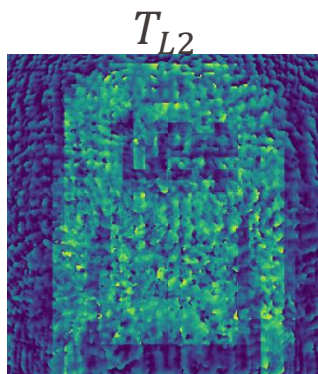
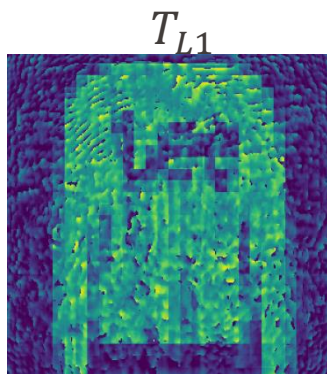


$$T_{Ln} = e^{j\phi}$$

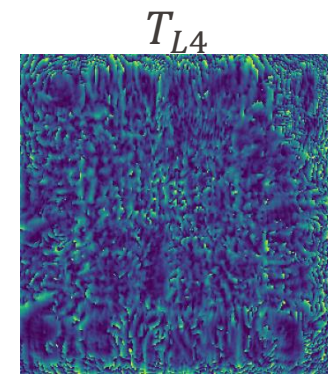
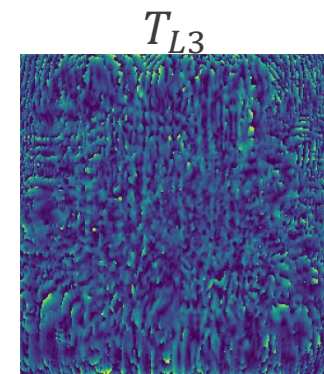
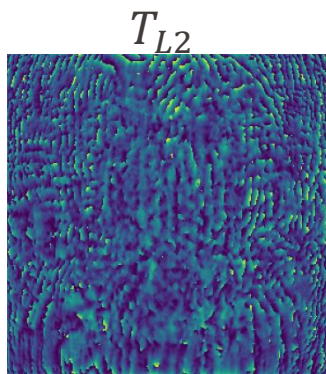
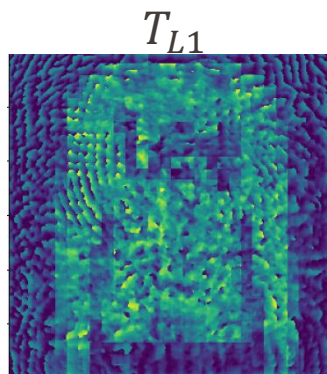
$$\phi = (D \times S_{Ln} + B_{Ln})$$

Trainable parameters: S_{Ln}, B_{Ln}

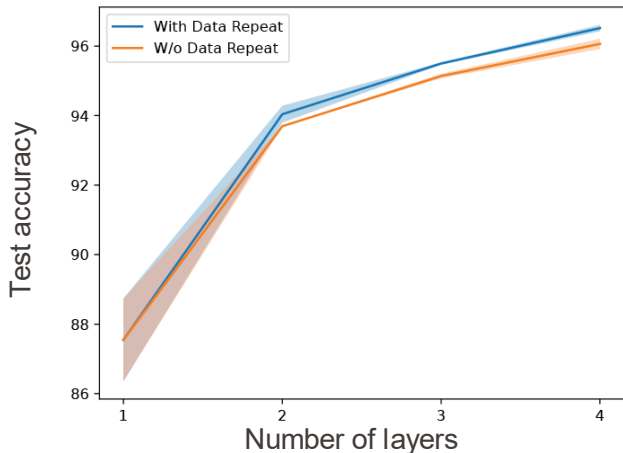
With
Data
Repeat



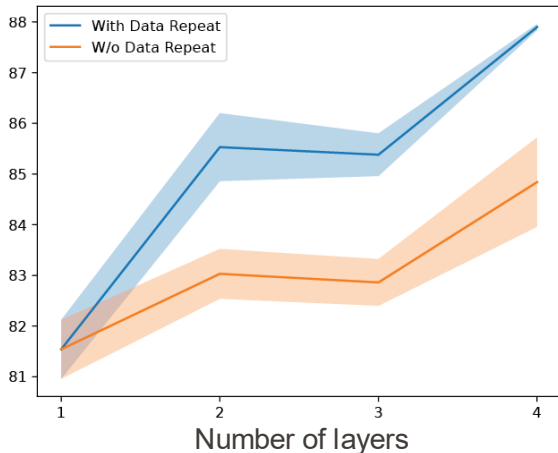
Without
Data
Repeat



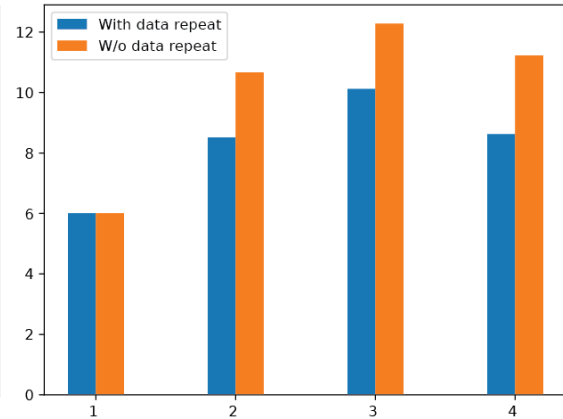
Simulation



Experiment



Sim. vs Exp. difference



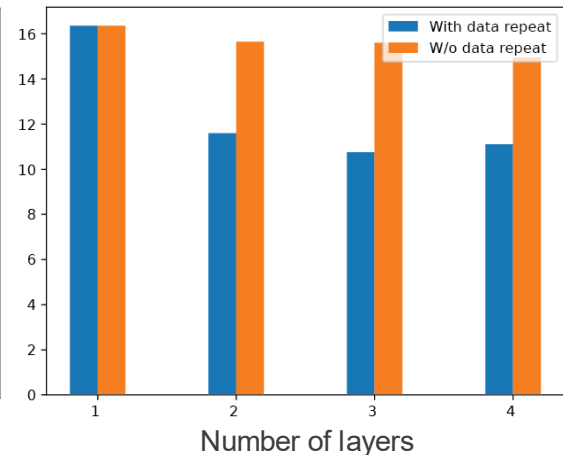
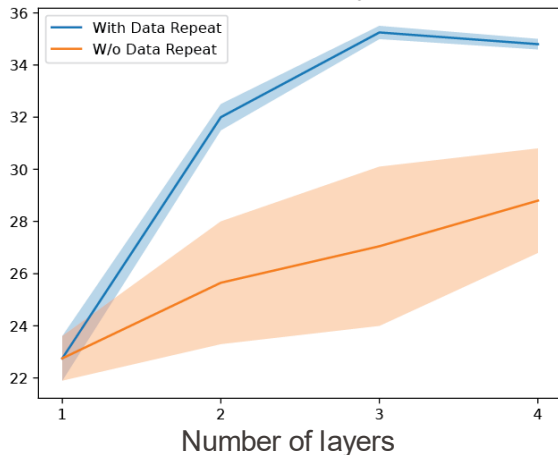
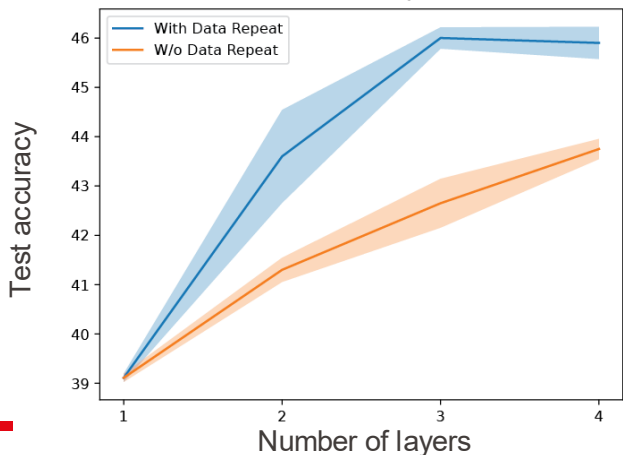
Number of layers

Number of layers

Number of layers

Digit MNIST

Imagenette



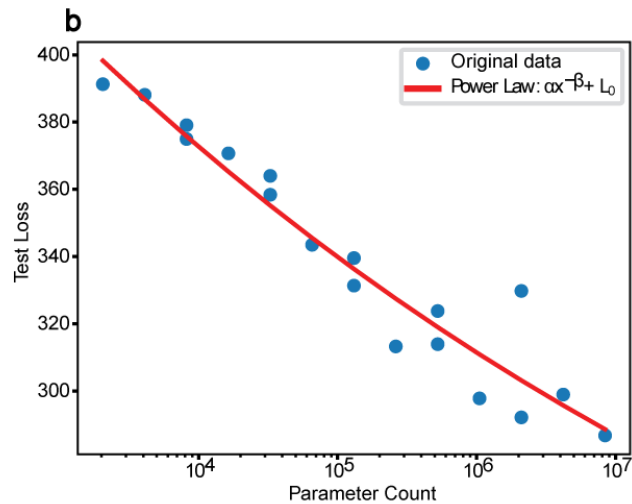
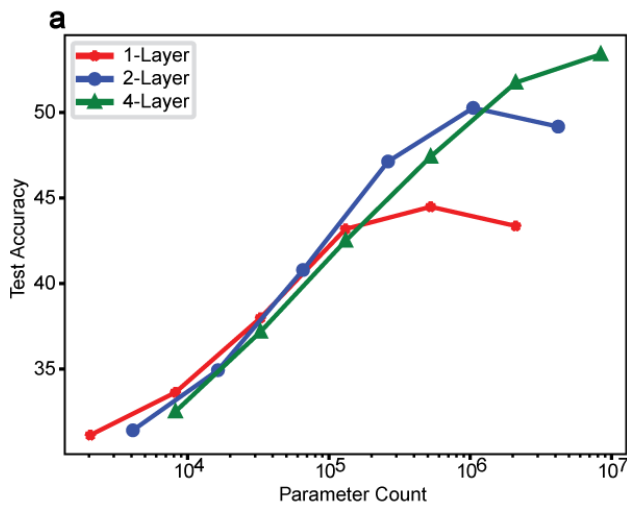
Number of layers

Number of layers

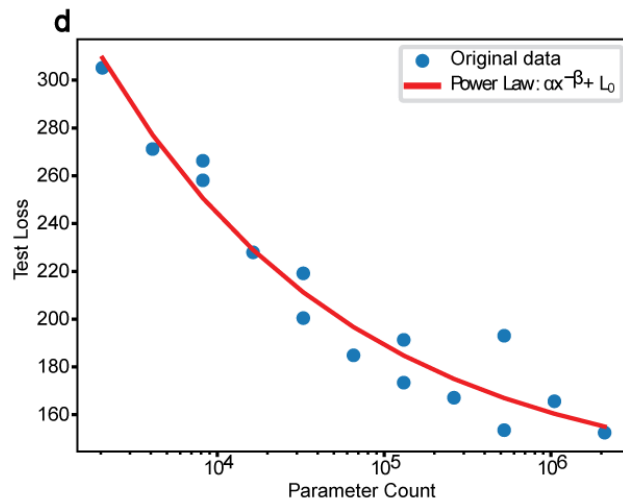
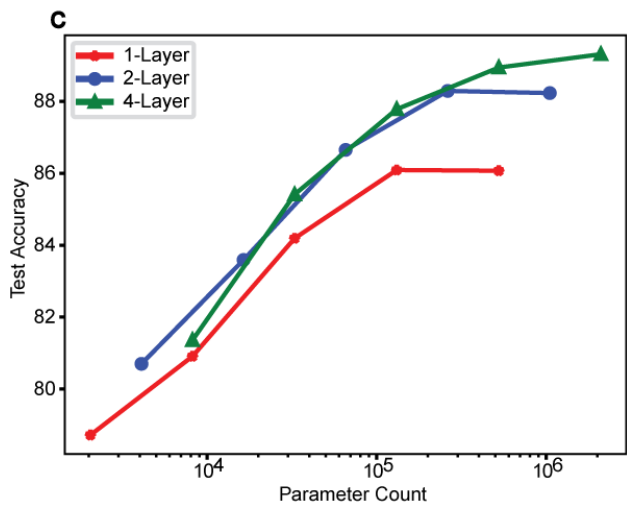
Number of layers



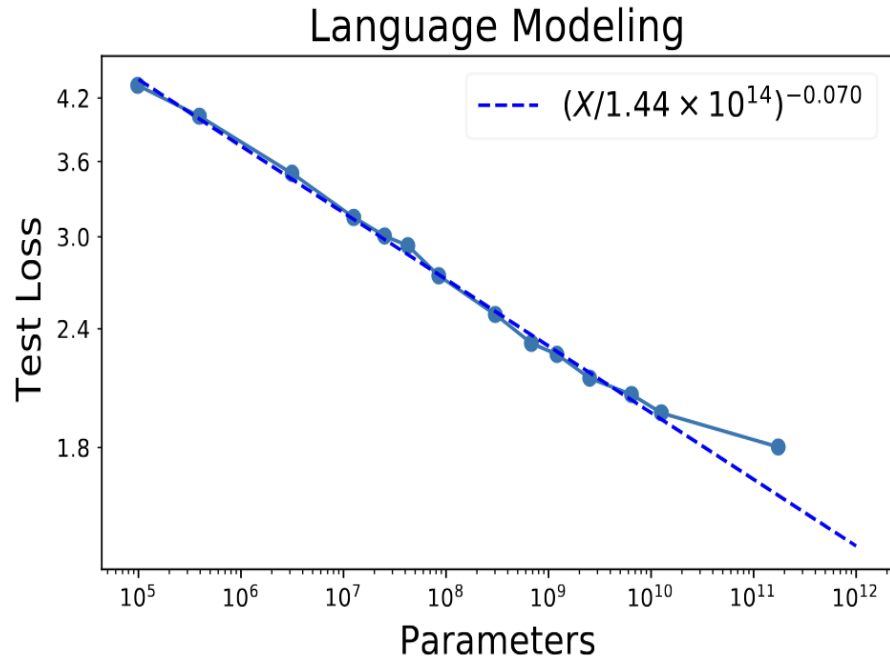
Imagenette dataset



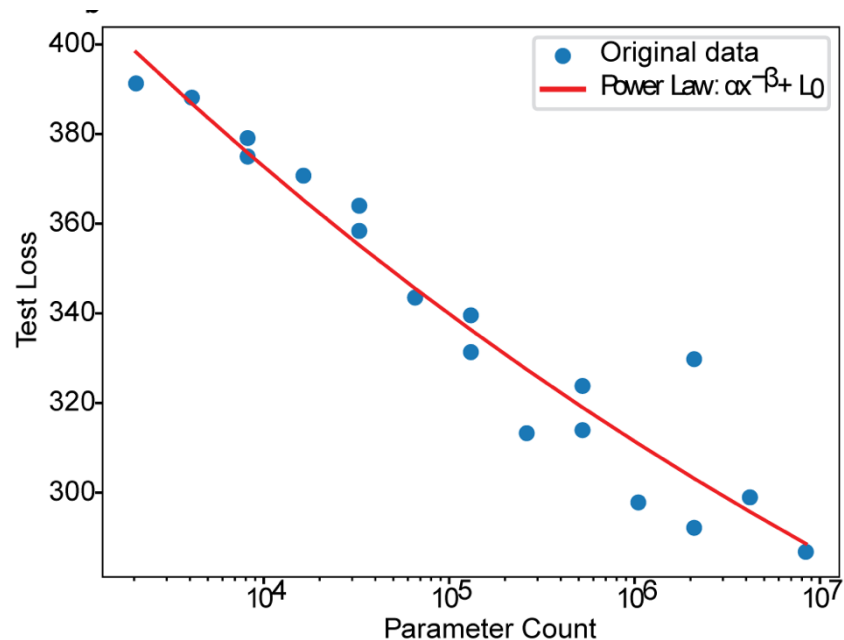
Fashion MNIST dataset

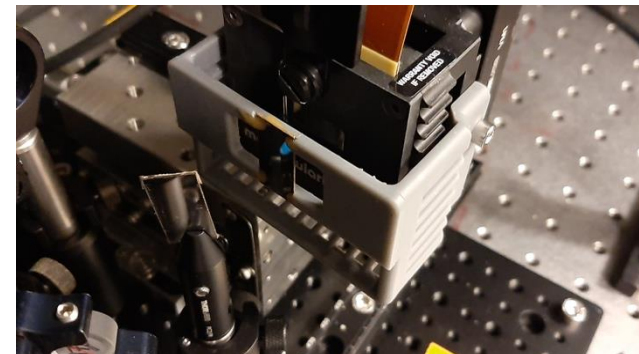
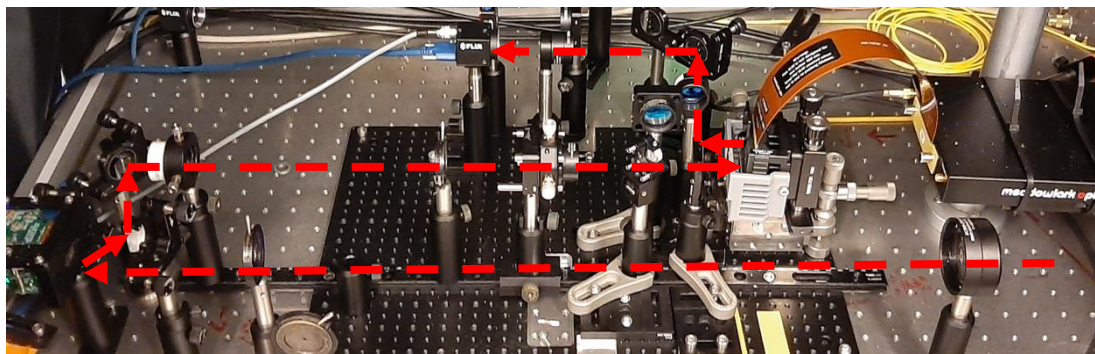
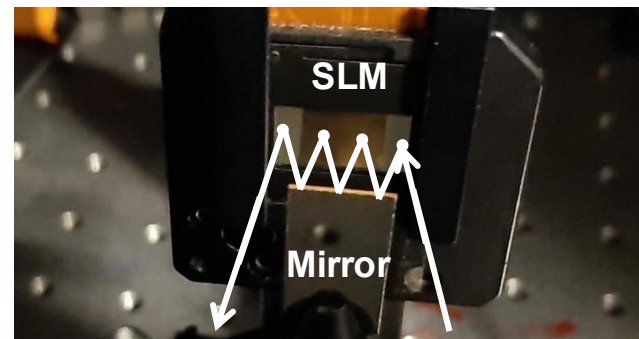
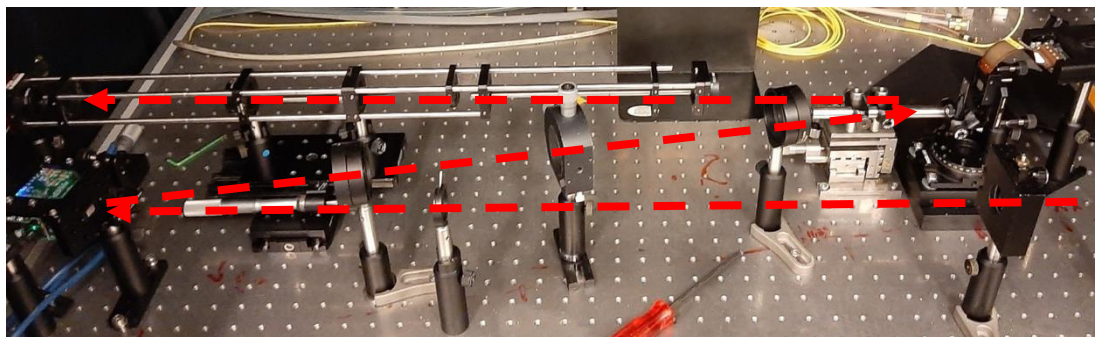


Scaling



$Parameter\ Count \propto Layer\ Width^2 * Layer\ Count$





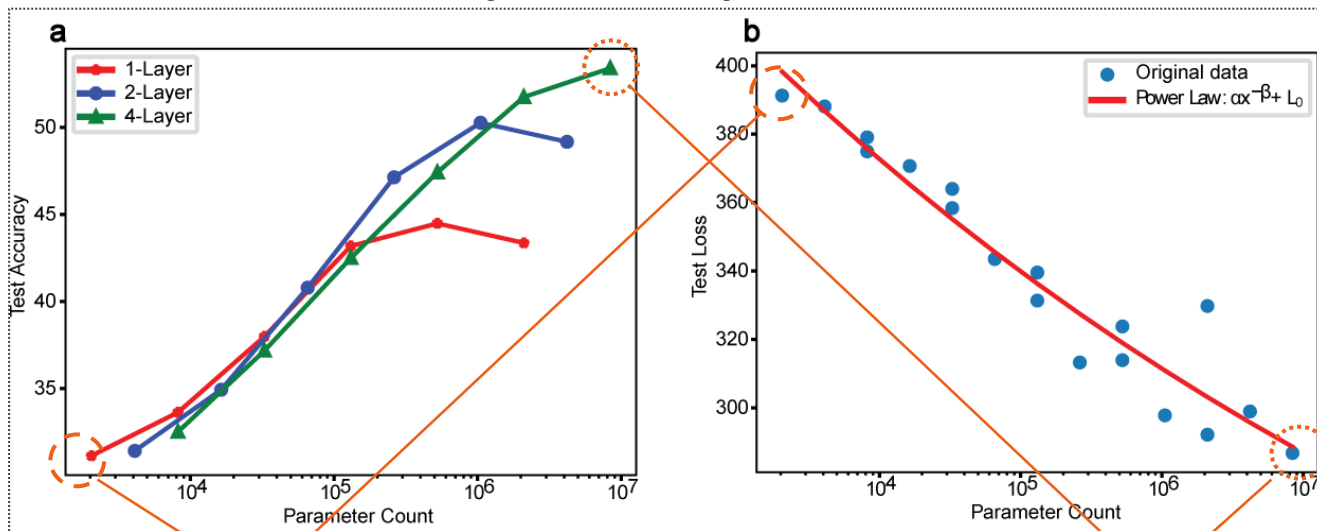
- Better SLM
 - Improved mirror mount
 - Better engineered setup
- ❖ Experimental accuracy using 4 layers for Imagenette dataset: $\approx 36\%$ to $\approx 42\%$

■

EPFL Power consumption

nPOLO uses 64 discrete sampling regions at the output plane. Let's assume we have 64 detectors (8-bit) with 50% efficiency at $\lambda=850$ nm and we consider only the shot noise. We also assume 90% reflectivity for the SLM.

Scaling of nPOLO for Imagenette dataset



Minimum number of input photons required: 9.2×10^6
Total energy of the photons in Joules: 2.2×10^{-12}
Parameter number: 2048
Photons per parameter: 4500

Minimum number of input photons required: 12.7×10^6
Total energy of the photons in Joules: 3×10^{-12}
Parameter number: 8.4×10^6
Photons per parameter: 1.5

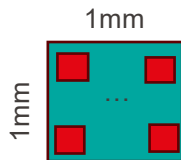
Compute Performance

■ NVIDIA A100

- Compute efficiency= 1pJ/Operation
- Compute density= 0.35teraOPmm⁻²s⁻¹

Assumption: **One parameter is equal to two operations (multiply and sum)**

LC- SLM



= 15625 pixels (parameters) meaning **31250 OP_{mm²}** (compute per area)

Pixel pitch = 8μm
Refresh rate = **1kHz**

Compute density = **compute per area x refresh rate** ≈ 30 megaOP_{mm²}s⁻¹

10x smaller pixel pitch → 100x

PhC spatial modulator (>100MHz) → 10⁵x

In principle, it can reach to 100petaOP_{mm²}s⁻¹

Compute Performance

▪ NVIDIA A100

- Compute efficiency= 1pJ/Operation
- Compute density= 0.35teraOPmm⁻²s⁻¹

Assumption: **One parameter is equal to two operations (multiply and sum)**

LC- SLM

1 million pixels/params
20W power consumption
1kHz refresh rate

Compute: 2 gigaOPs⁻¹

Energy cons. per second: 20J

Compute efficiency = 10nJ/OP
(Energy/Compute)

Energy \propto *Modulation physics* * *Refresh rate*⁻¹

Compute \propto *Pixel count* * *Refresh rate (bandwidth)*

\rightarrow *Compute eff.* \propto *Refresh rate*⁻²

Main theme: 3D Optical Neural Networks

1. Motivation

- Why optical computing? Why neural networks?
- Why 3D? 2D vs 3D

2. 3D optical storage (weight allocation)

3. OEO and nonlinear processing with linear optics

4. An example on optical generative models

5. Examples on extreme learning machines



17.4 secs, a few kJ/image

Prompt: A scene in an international conference. The talk will be about the optical implementation of denoising diffusion models

- Given a training dataset with a distribution $p(x)$
- Generative machine learning models **try to learn** this distribution as best as they can
- New samples are generated by **sampling** the learned distribution $\hat{p}(x)$

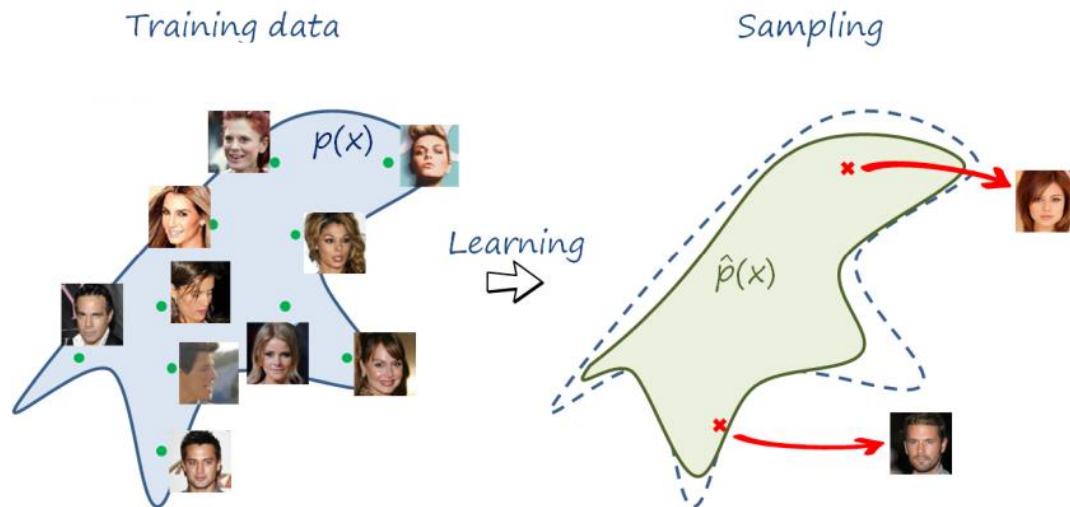
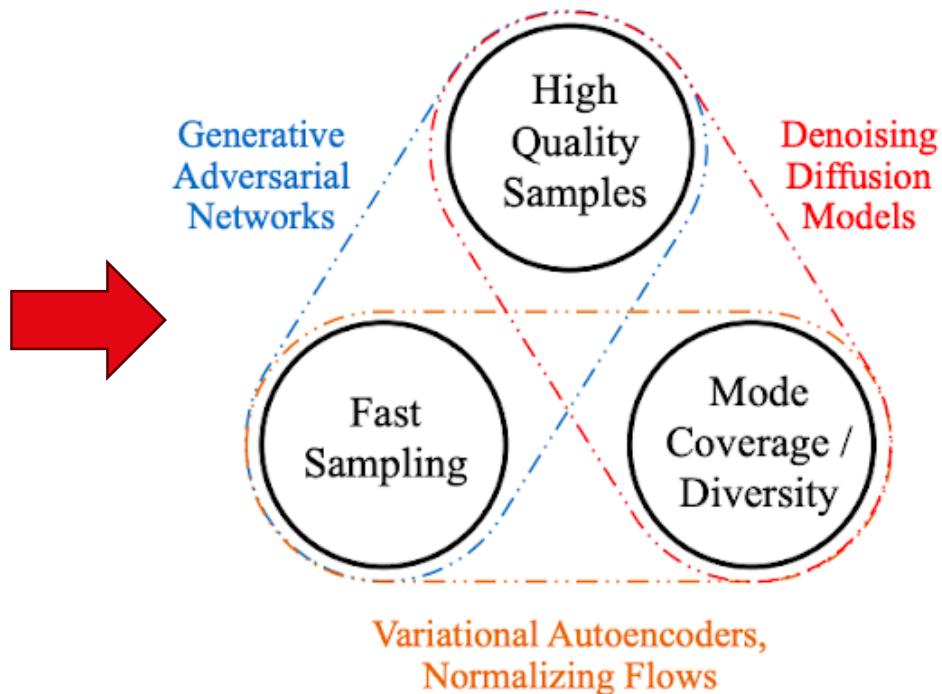


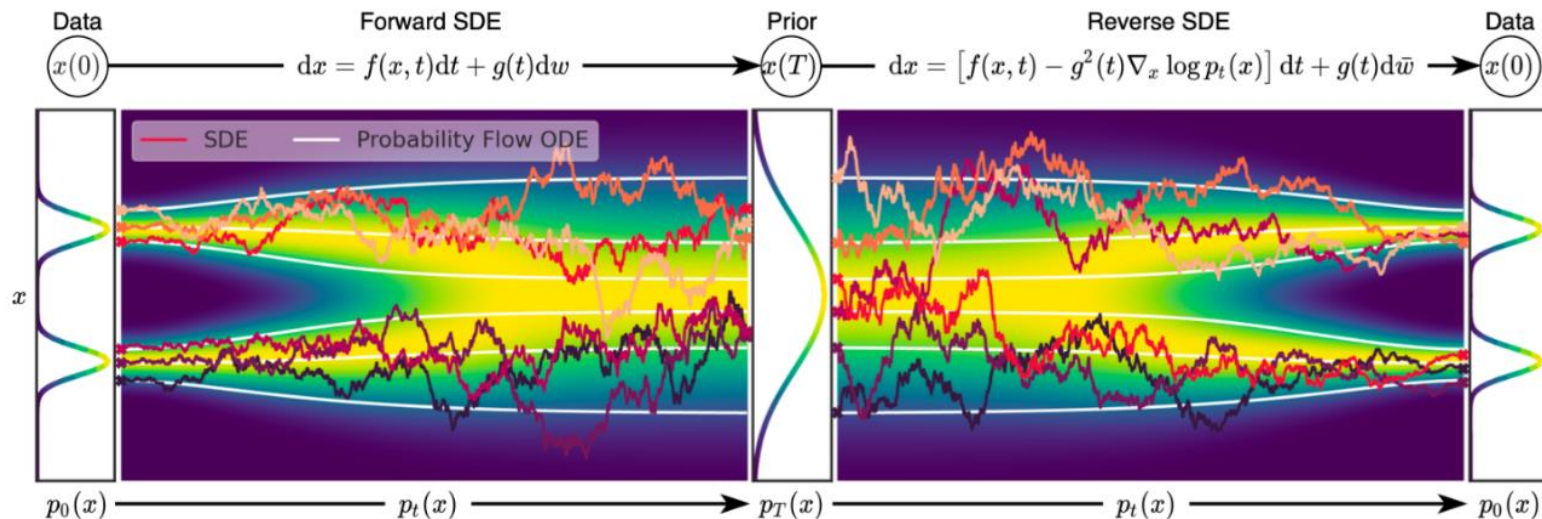
Image Generation Models

- Diffusion models outperforms previous SOTAs on broad datasets
- Due to iterative sampling, generation is time & energy expensive.



Generative Diffusion Models

- Forward SDE gradually adds noise to the data.
- A reverse SDE that can recover the original probability densities $p_0(x)$ exists.

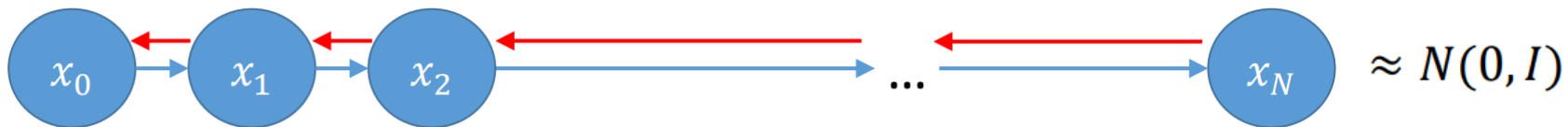


Denoising Diffusion Probabilistic Models

- The discretized version of the algorithm for image generation:

$$r_{t-1} = \frac{1}{\sqrt{1 - \beta_t}} (r_t - \beta_t \epsilon_{\theta}(r_t, t))$$

Backward Process, (learned by NN) noise prediction/removal:



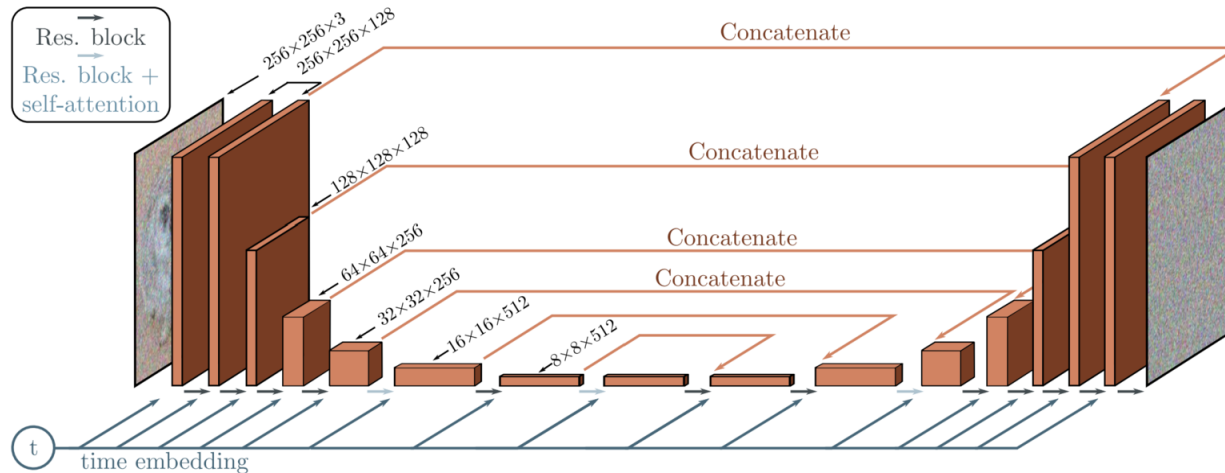
Forward Process, Noise addition: $r_t = \sqrt{1 - \beta_t} r_{t-1} + \sqrt{\beta_t} \epsilon_t$, $\epsilon_t \sim \mathcal{N}(0, I)$



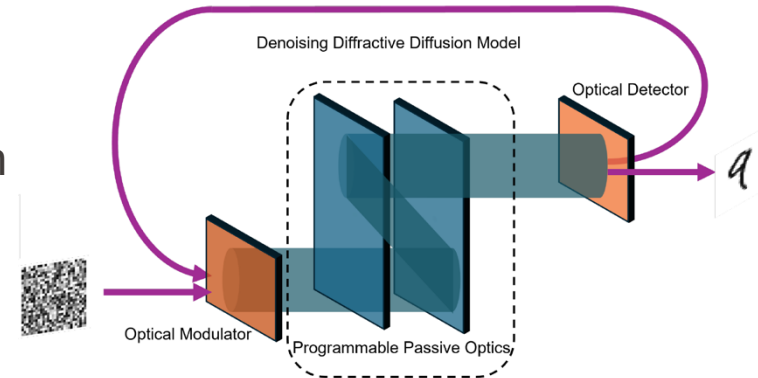
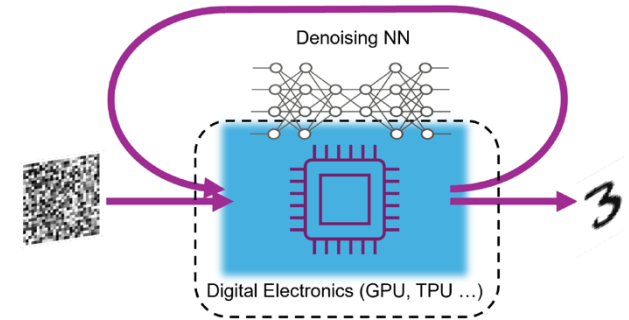
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33, 6840-6851.

Denoising Diffusion Probabilistic Models

- Well-known U-Net architecture is suitable for the image denoising task.
- Time embedding** informs each layer of the expected noise level in the given sample.

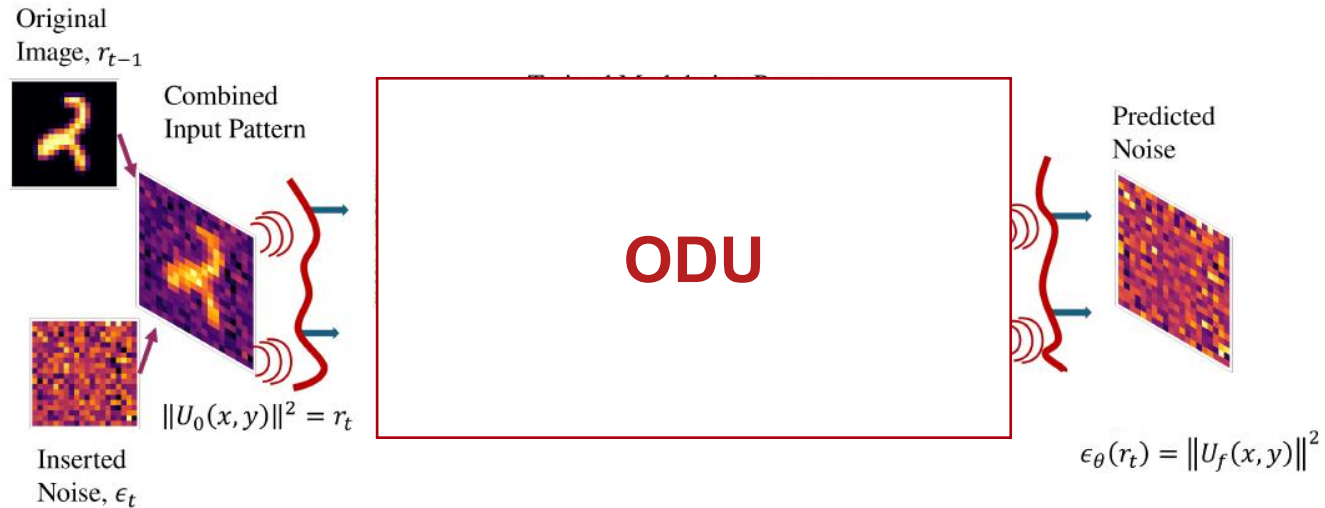


- General purpose electronic hardware may run the denoising NN >1000 times for sampling an image.
- Application-specific and passive optical denoising units(ODU) to realize high performance and efficiency sampling with diffusion models.



Optical Diffusion Unit (ODU)

- Optical denoising framework combines the trainable modulation with patterns $L_n(x, y)$, and the connectivity of free space propagation $h(x, y)$.



$$U_n = [U_{n-1}(x, y) * h(x, y)]L_n(x, y)$$

- Different layer sets for different noise time periods → time-aware operation with minimal layer updates

- Loss function:

$$\mathcal{L}_t = \mathbb{E}_{x_0, \epsilon} \left[\|\epsilon - \epsilon_{\theta_m}(x_t)\|^2 \right]$$

- Trainable parameters:

$$\{L_n^m\}_{n=1}^N(x, y)$$

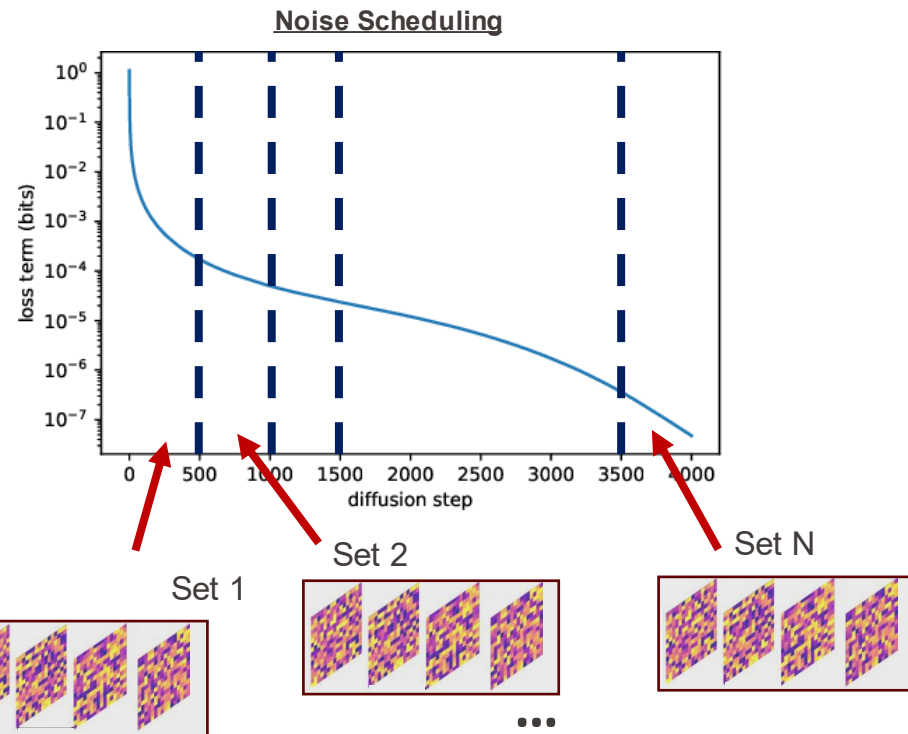
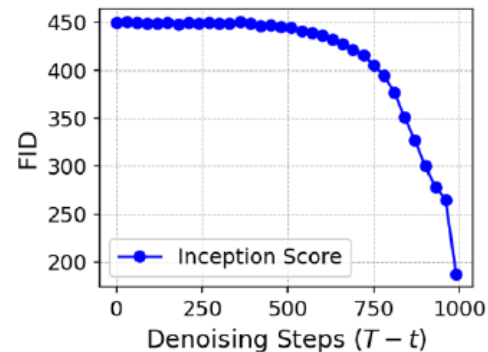
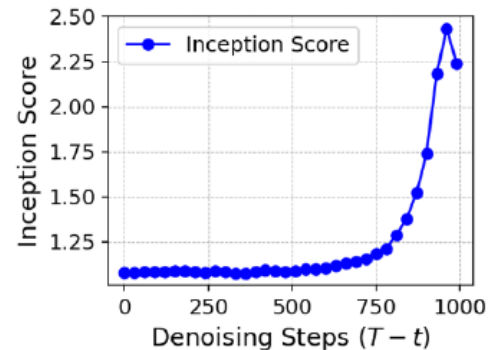
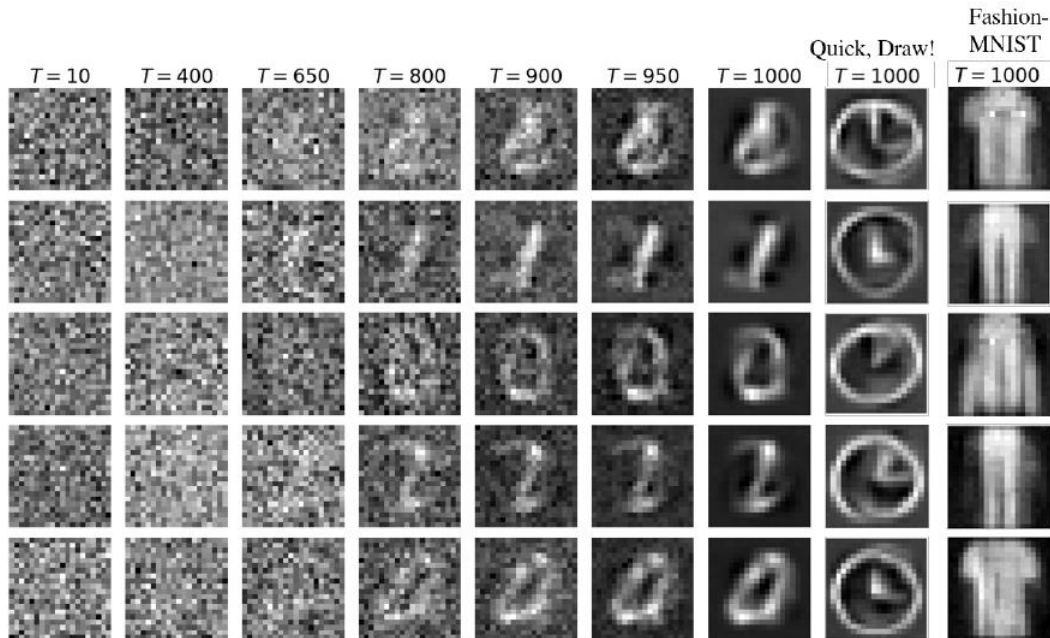


Image Generation with ODU

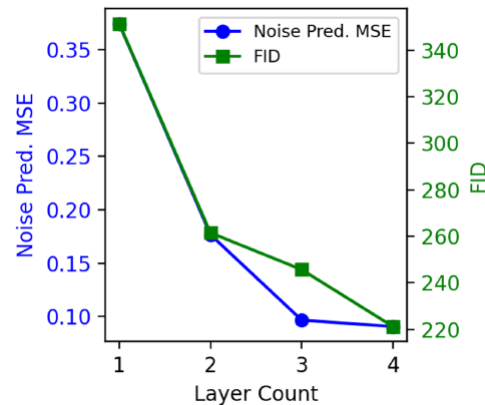
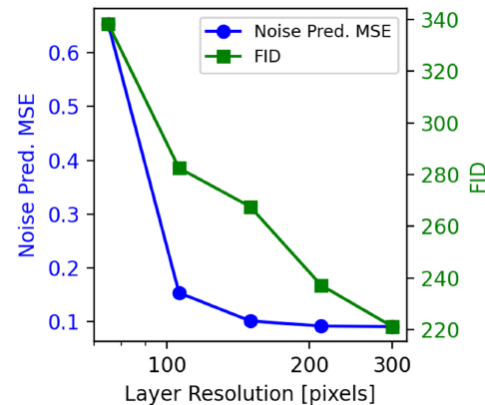
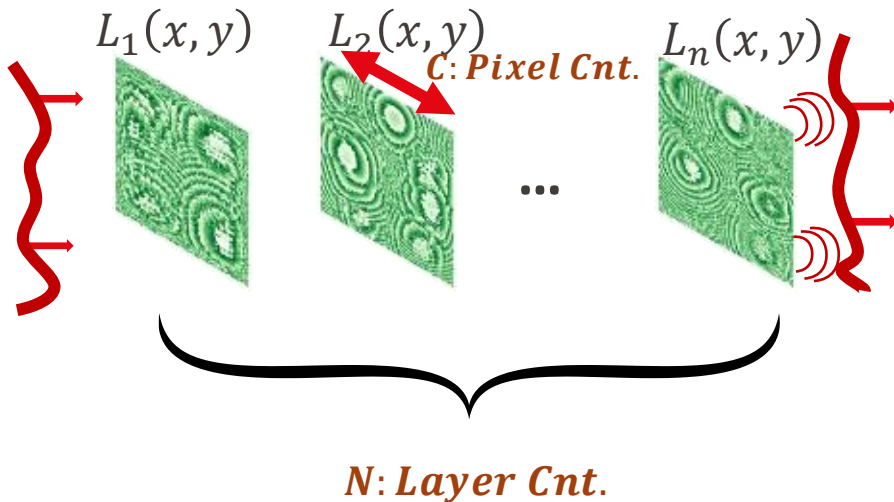
- Same sampling with DDPM, $T = 1000$, linear noise schedule from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$.



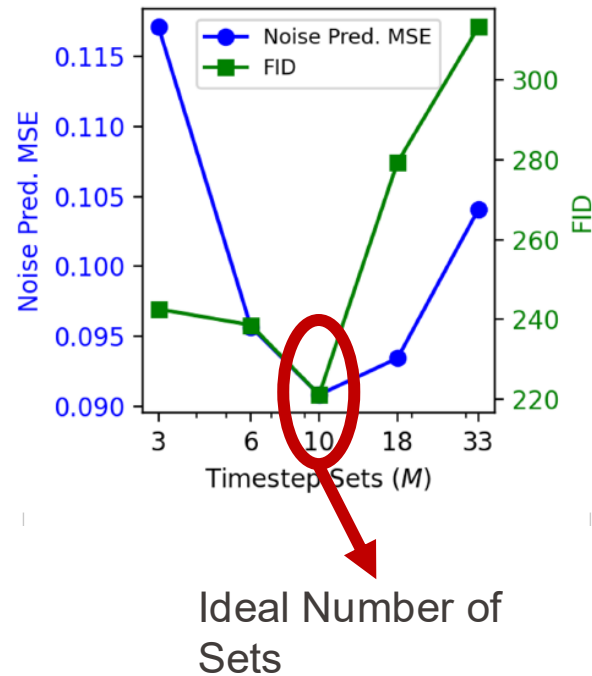
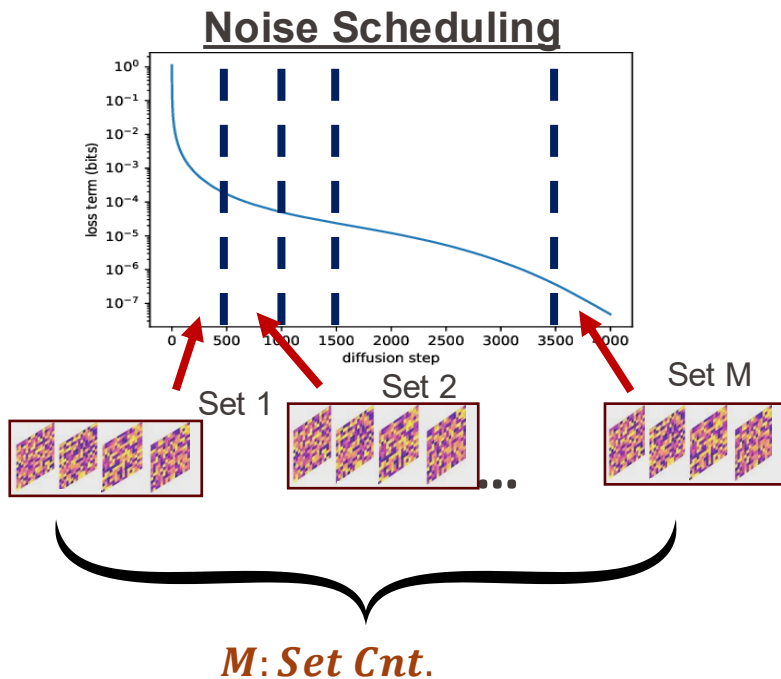
Scaling ODU: Layer Resolution and Count

Larger number of

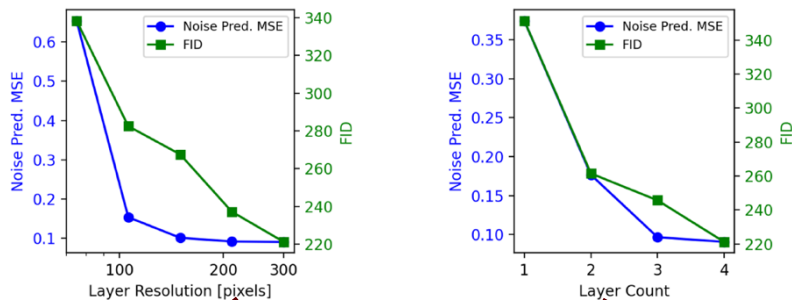
- modulation layers, N ,
 - pixels on modulation layers, C ,
- tends to yield better denoising and generation.



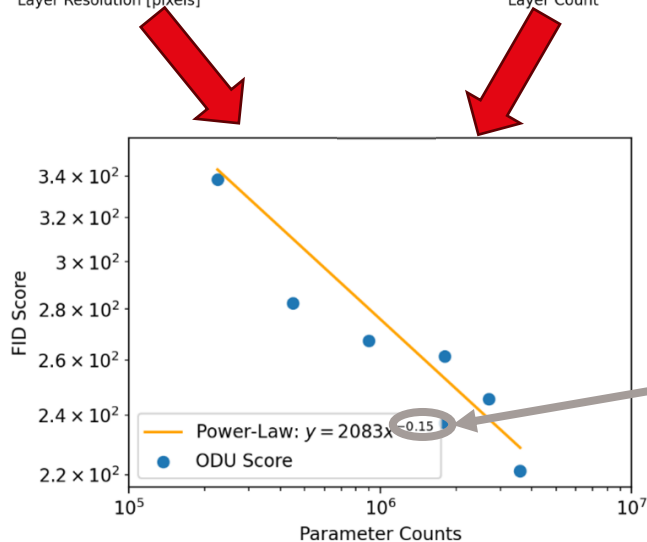
Scaling ODU: Number of Layer Sets



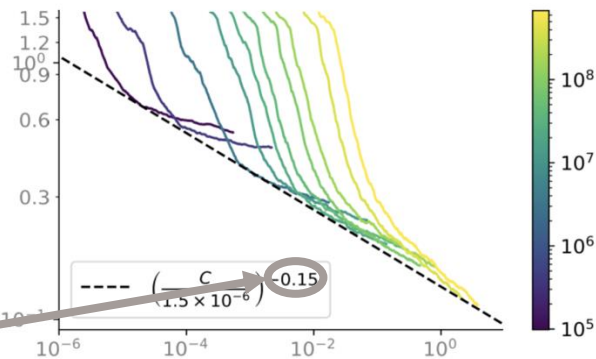
Scaling ODU: Total Parameter Count



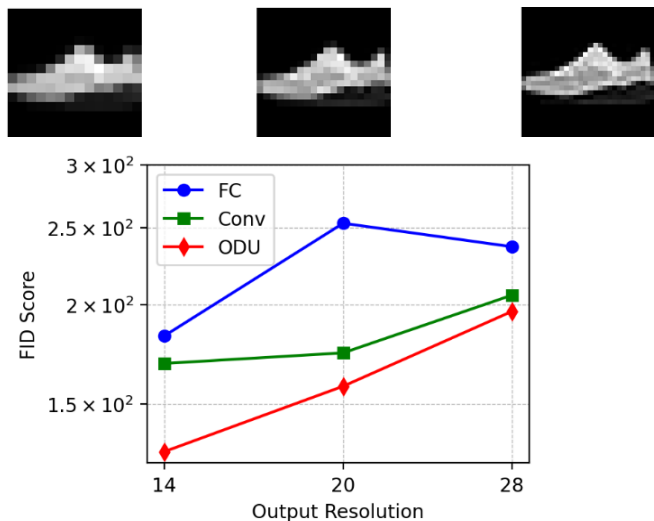
ODU scales at the same speed with transformers in terms of parameter count vs generation quality.



Text \rightarrow Image



Henighan, Tom, et al. "Scaling laws for autoregressive generative modeling." *arXiv preprint arXiv:2010.14701* (2020).



	# PARAM	# FLOPS	ENERGY [J]	IMAGES/S
MLP	19.6 M	39.0 M	1.74	41.3
U-NET	220 K	3.11 M	5.37	13.9
ODU	3.6 M	N/A	0.23	23.0

ODU scales to larger dimensional outputs at a similar speed with other neural network types.



Junjie Ke



Qifei Wang



Ilker
Oguz



Niyazi
Ulas Dinc



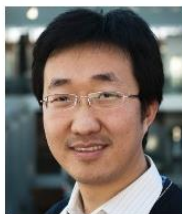
Mustafa
Yildirim



Published in NeurIPS '24



Infnarn Yoo



Feng Yang



Demetri
Psaltis



Christophe
Moser

Main theme: 3D Optical Neural Networks

1. Motivation

- Why optical computing? Why neural networks?
- Why 3D? 2D vs 3D

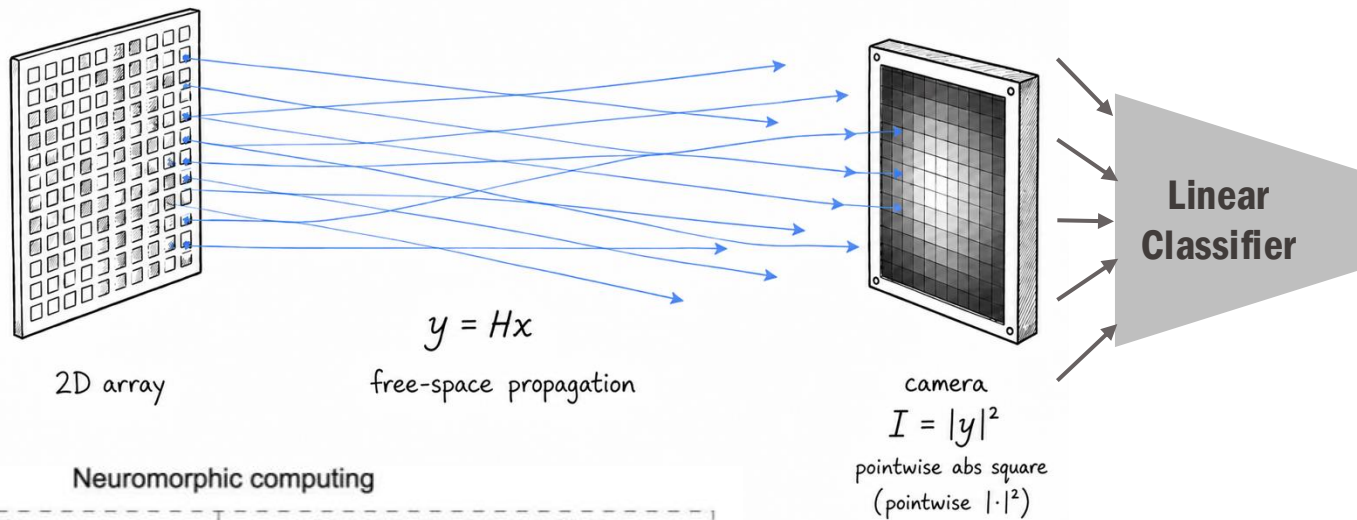
2. 3D optical storage (weight allocation)

3. OEO and nonlinear processing with linear optics

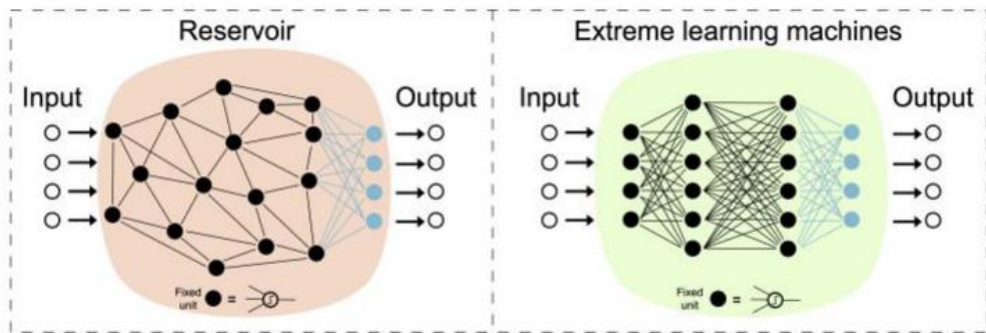
4. An example on optical generative models

5. Examples on extreme learning machines

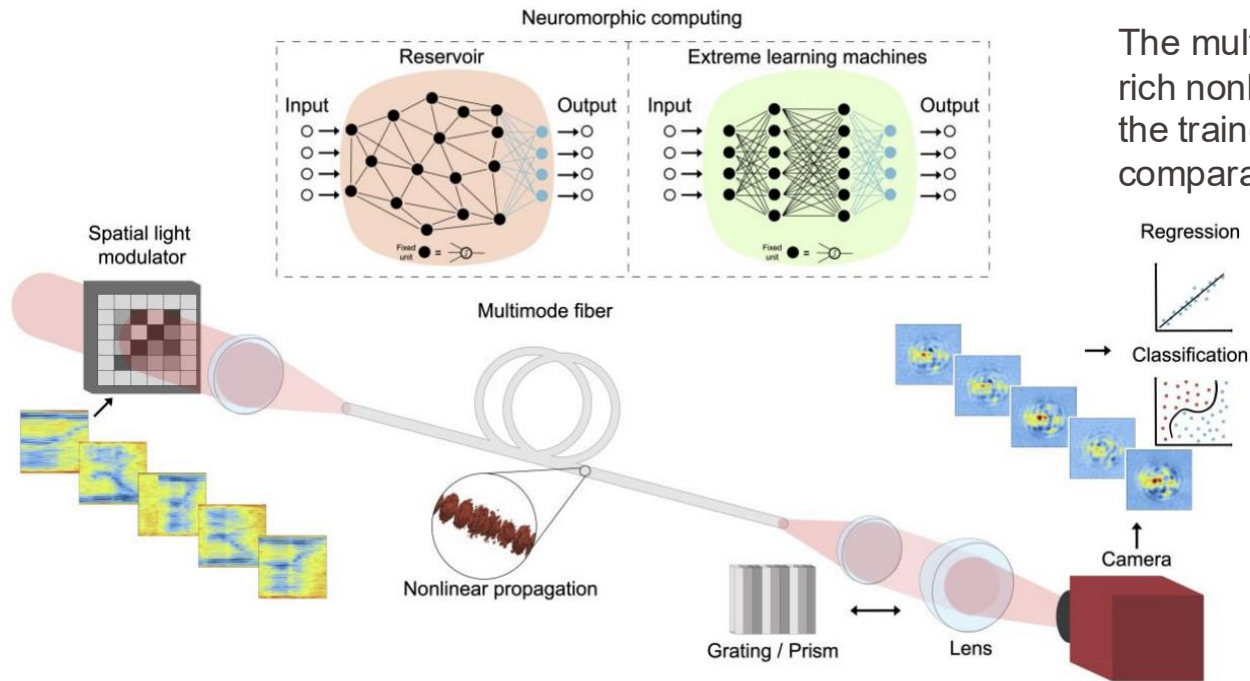
EPFL A basic 3D optical computer



Neuromorphic computing



Multi-Mode Fiber Optical Neural Network



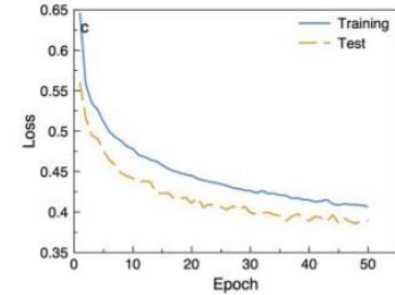
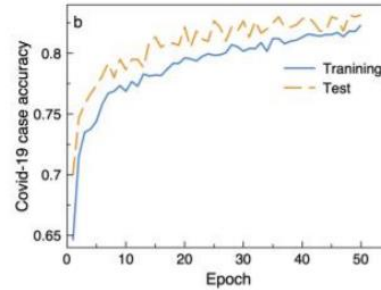
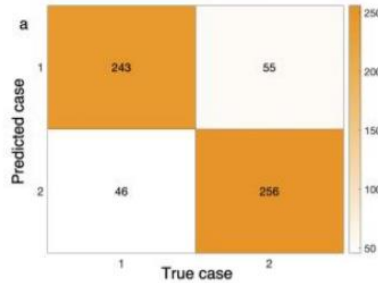
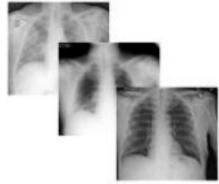
The multimode fiber acts as a rich nonlinear feature map; the trained part is comparatively lightweight.

$$\frac{\partial A_p}{\partial z} = \underbrace{i\delta\beta_0^p A_p - \delta\beta_1^p \frac{\partial A_p}{\partial t} - i\frac{\beta_2}{2} \frac{\partial^2 A_p}{\partial t^2}}_{\text{Dispersion}} + \underbrace{i \sum_n C_{p,n} A_n}_{\text{Linear Mode Coupling}} + \underbrace{i\gamma \sum_{l,m,n} \eta_{p,l,m,n} A_l A_m A_n^*}_{\text{Nonlinear Mode Coupling}}$$

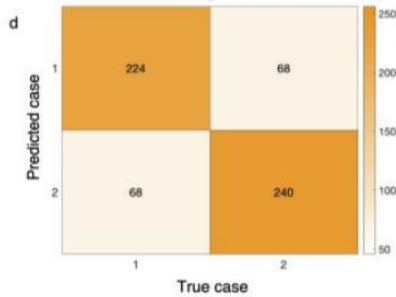
- Scalable Optical Learning Operator (SOLO), Nature Computational Science, August 2021

Computing with the Spatiotemporal Nonlinearities of MMFs

SARS-CoV-2 X-ray dataset

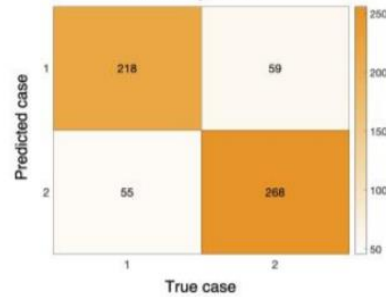


Accuracy: 77.3%

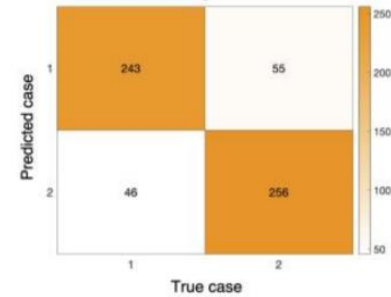


1.8

Accuracy: 81%

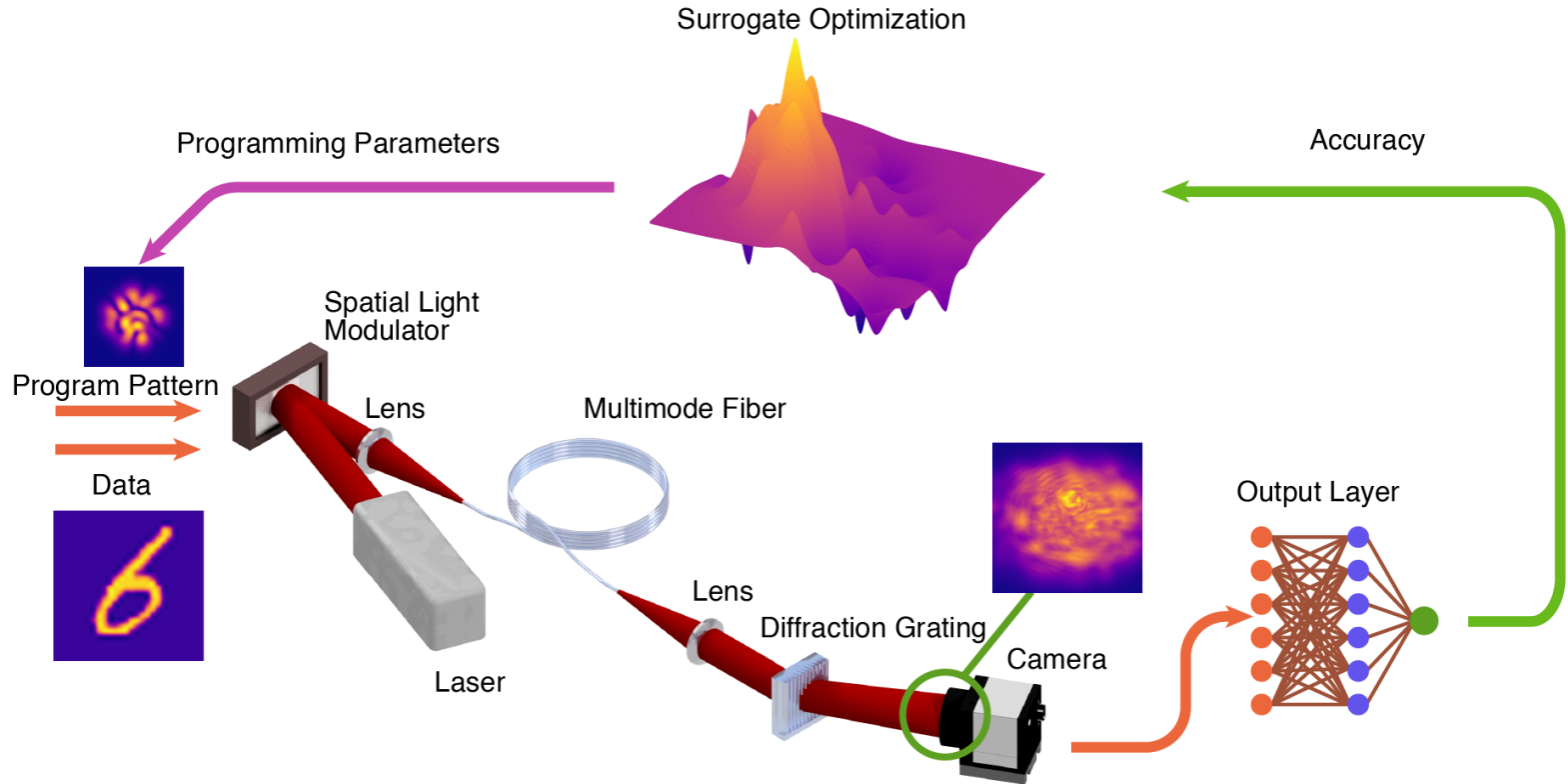
3.0
Pulse peak power (kW)

Accuracy: 83.2%

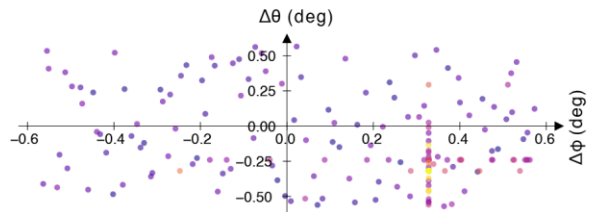
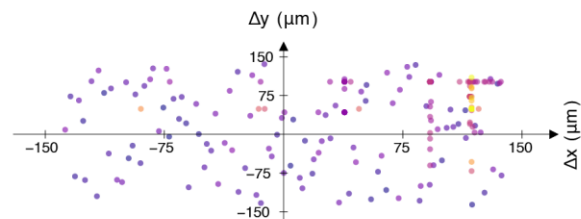
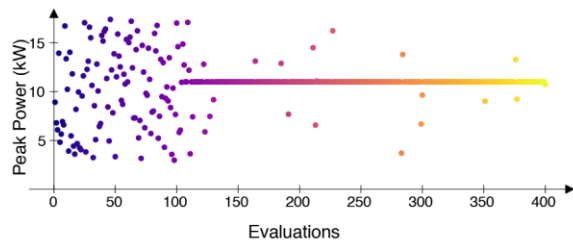
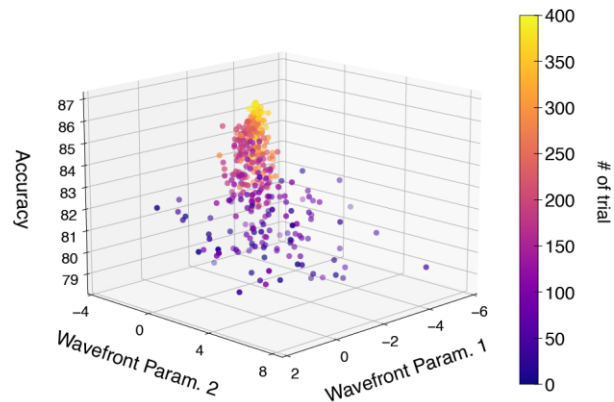
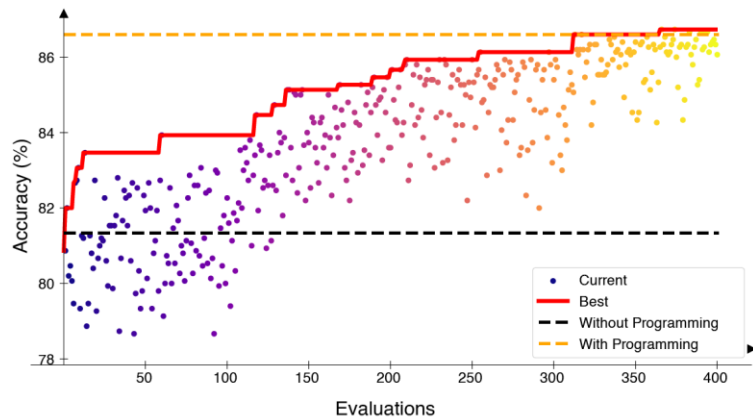


3.45

Programming SOLO



Programming SOLO

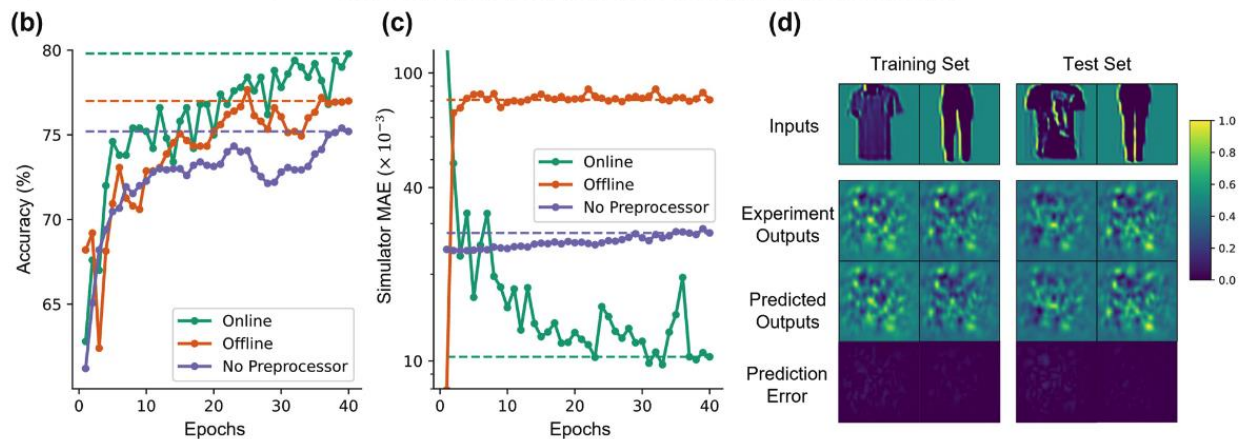
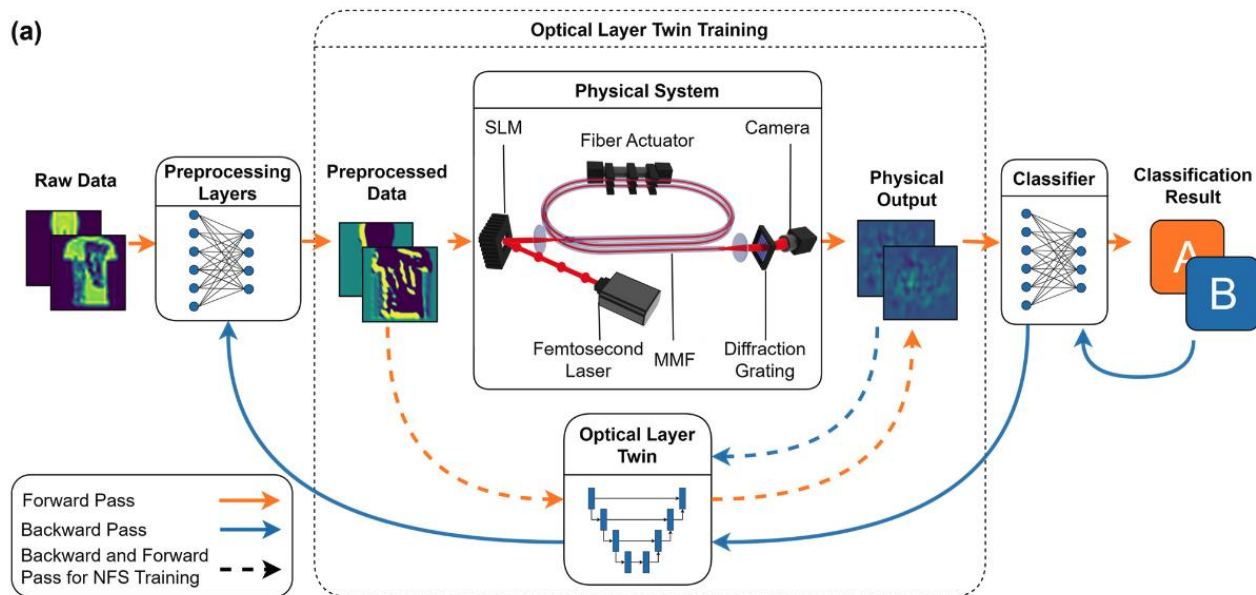


- COVID-19 classification on radiography images

Comparison with digital DNNs

Network Structure	Total Number of Weights	Operations per Sample(FLOPS)	Accuracy on COVID(1500 samples,40x40pixels) (%)
16 filters(3x3) conv + 32 filters(3x3) conv + fully connected	~10k	~870k	85.7 ± 0.7 (≈ Programmed SOLO)
8 filters(3x3) conv + fully connected	~1.7k	~18k	80.6 ± 0.9 (≈ Simple SOLO)
1 fully connected	~3.2k	~6.4k	75.5 ± 0.5
3 Conv layers + 2 fully connected layers with batch norm., max poolings and dropouts	~390k	~17.5 M	92.3 ± 0.8

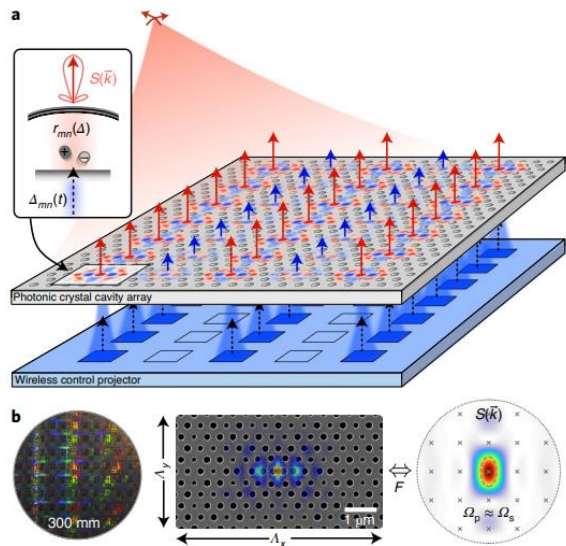
Training with a digital twin



Oguz, Ilker, Suter, Louis J.E., Hsieh, Jih-Liang, Yildirim, Mustafa, Dinc, Niyazi Ulas, Moser, Christophe and Psaltis, Demetri. "Training hybrid neural networks with multimode optical nonlinearities using digital twins" *Nanophotonics*, vol. 14, no. 16, 2025, pp. 2787-2797.

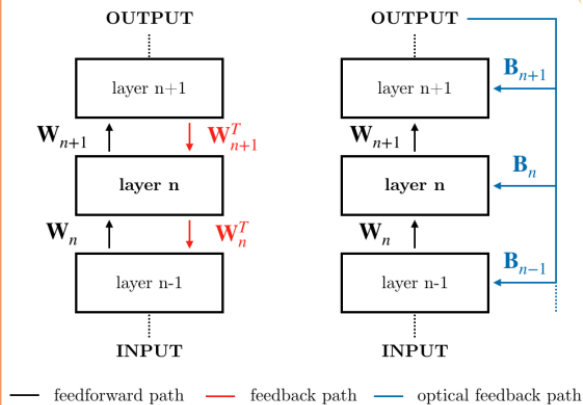
■ <https://doi.org/10.1515/nanoph-2025-0002>

Spatial modulators



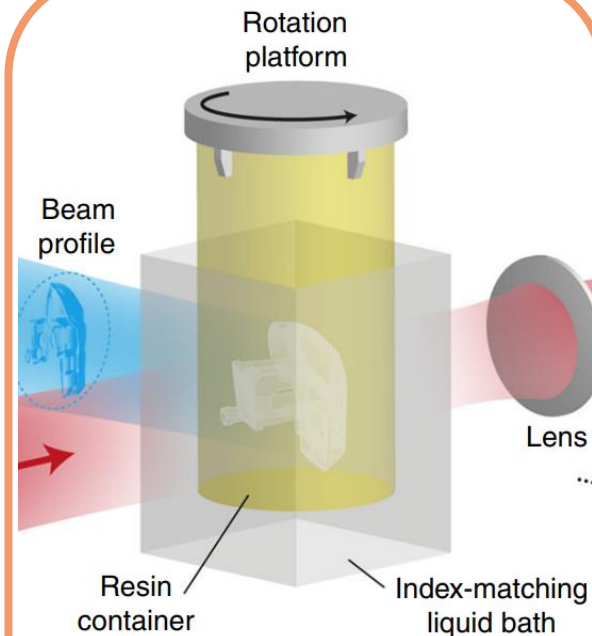
Panuski et al. Nature
2022

Algorithms



Launay et al. NeurIPS
2020

Fabrication



Loterie et al. Nat Comm
2020