AIST
Create the Future, Collaborate Together

MPSoC'25, June 16, 2025

Optical Streaming Processor Utilizing Integrated Photonic Circuits

Guangwei Cong

Photonics-Electronics Integration Research Center
National Institute of Advanced Industrial Science and Technology (AIST), Japan.
E-mail : gw-cong@aist.go.jp

Outline

- Why optical computing is reignited?
- How to do optical computing in photonics?
- Our works: optical streaming processors implementing novel models in photonics

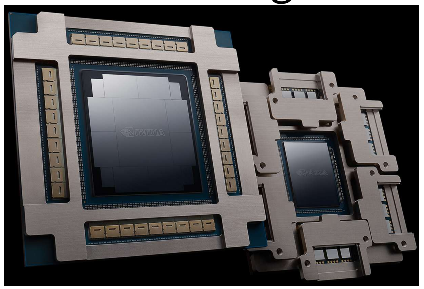
Why optical computing is reignited?

(1) Power issue and demands for computational power
- Computational power doubles per 10 months
- Energy consumption is already substantial, and is increasing rapidly
- End of Moore's law→ stagnation of digital processor performance
- End of Dennard law → energy efficiency gains slows

(2) Advances in optical interconnects and photonics-electronics co-integration

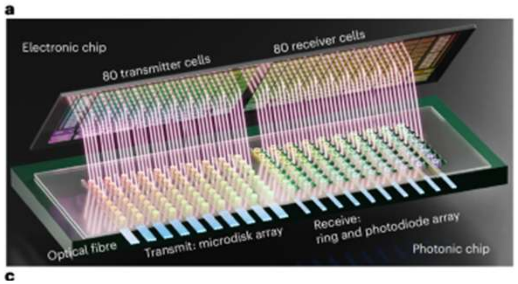
(3) Physical system computes better in energy efficiency

On-Package



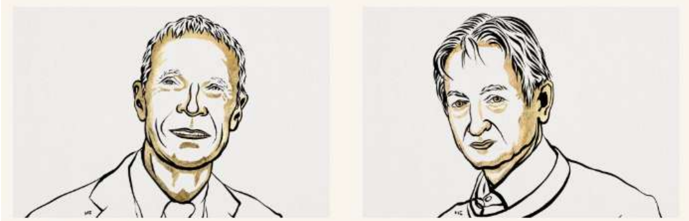
https://nvidianews.nvidia.com/news/nvidia-spectrum-x-co-packaged-optics-networking-switches-ai-factories

On-Chip



https://www.nature.com/articles/s41566-025-01633-0

Physical concepts + machine learning



John Hopfield. Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey Hinton. Ill. Niklas Elmehed © Nobel Prize Outreach

"They used physics to find patterns in information"

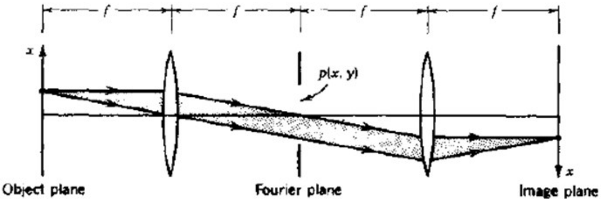
https://www.nobelprize.org/

Intrigue constructing optical physical analog AI system

Optical computing: Hype or hope?

Energy saved or not?

Example: Optical Fourier transformation



I'd to call it lens instead of computing since

Three elements required

- On-demand use input
- Reconfigurable
- Result interpretation

➔ Overhead consumes additional powers

Pros and cons

Pros:

Low energy
Low latency
High clock speed
High bandwidth
High parallelism
Network compatibility

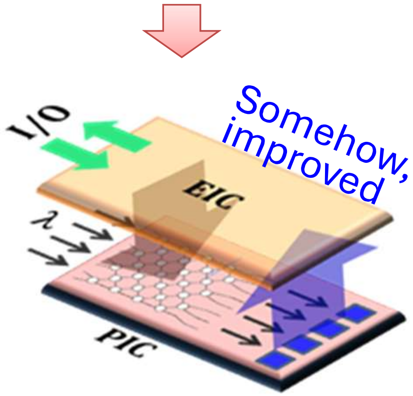
Cons:

Scalability
Programmability
Nonlinearity
Trainability
Stability
Low resolution
Area efficiency

Why it failed?

Lost to VLSI due to

- Scalability
- Low integration
- Programmability
- Not easy to use



What we are expecting

- One-shot
- All-optical
- High-throughput
- Applicable scale
- System-level energy efficiency
- Less digital intervene

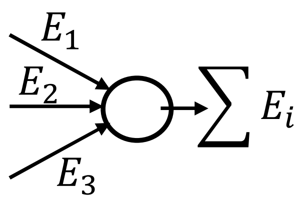
Comments:

- System evaluation is of necessity
- Concerns of heavy overhead related to digital processors and memory
- It is time to re-evaluate the possibilities of optical computing

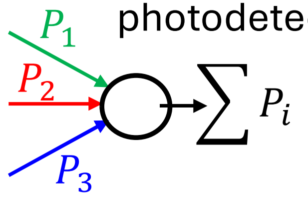
How to do computing in photonics?: scalar operations

- Addition
  - Electrical field addition (coherent)
  - Optical power addition (incoherent)

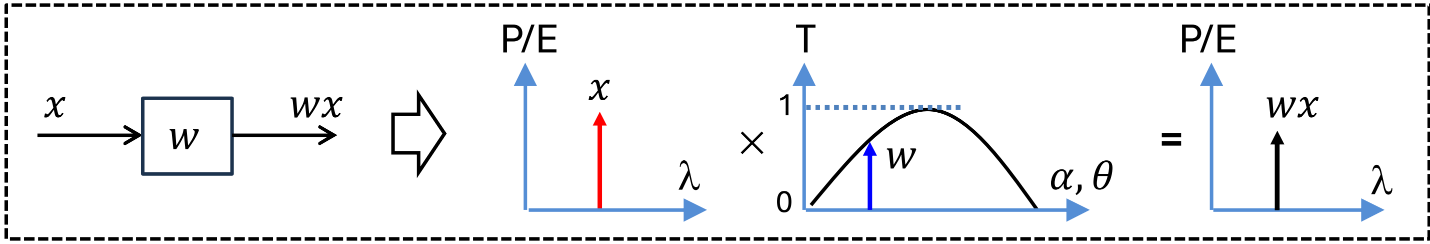
interferometer



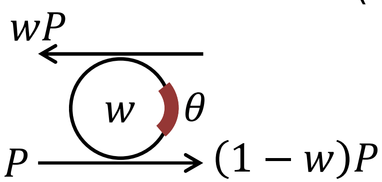
photodetector



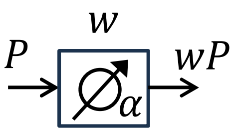
- Scalar multiplication



Incoherent (i.e., power)

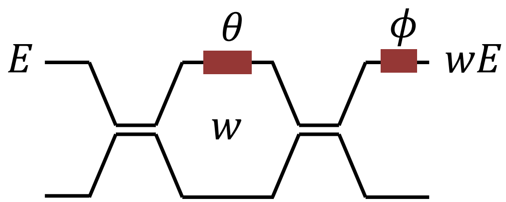


Ring

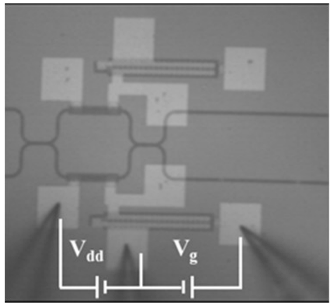


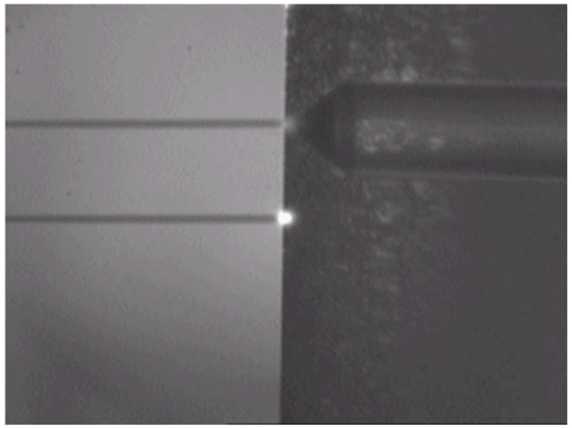
VOA

Coherent (i.e., electrical field)



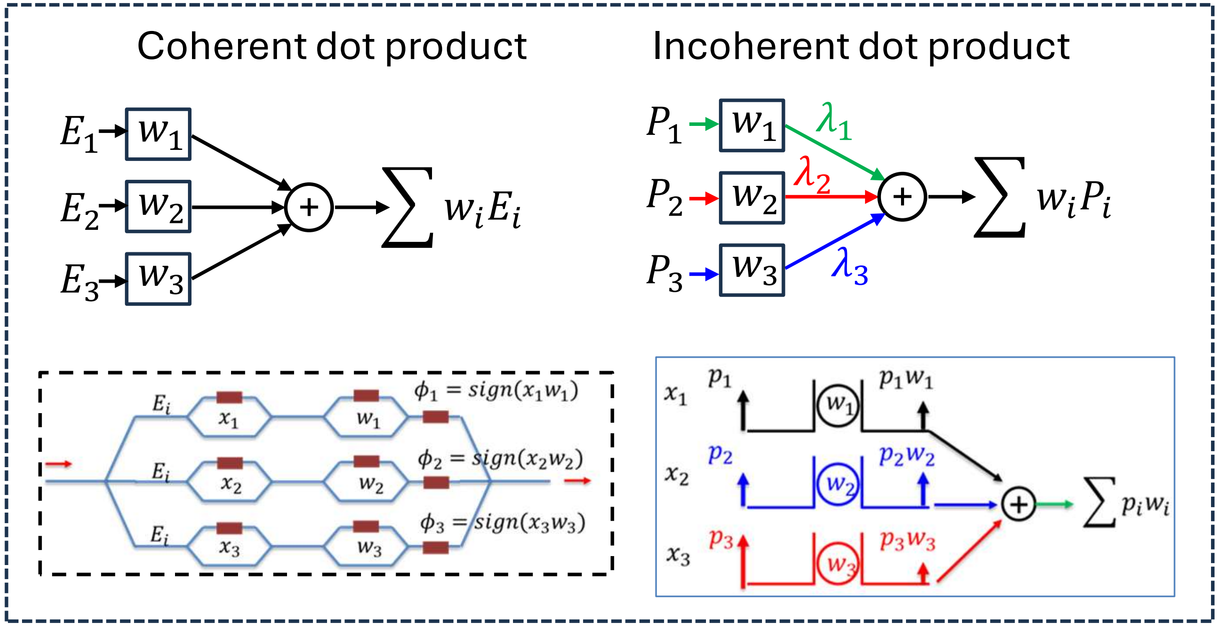
MZI

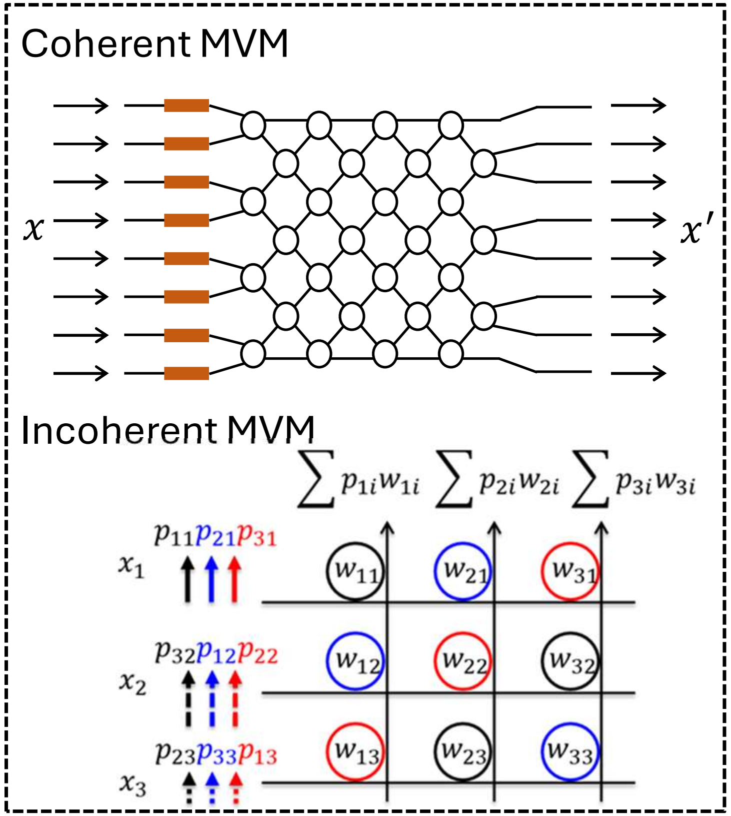


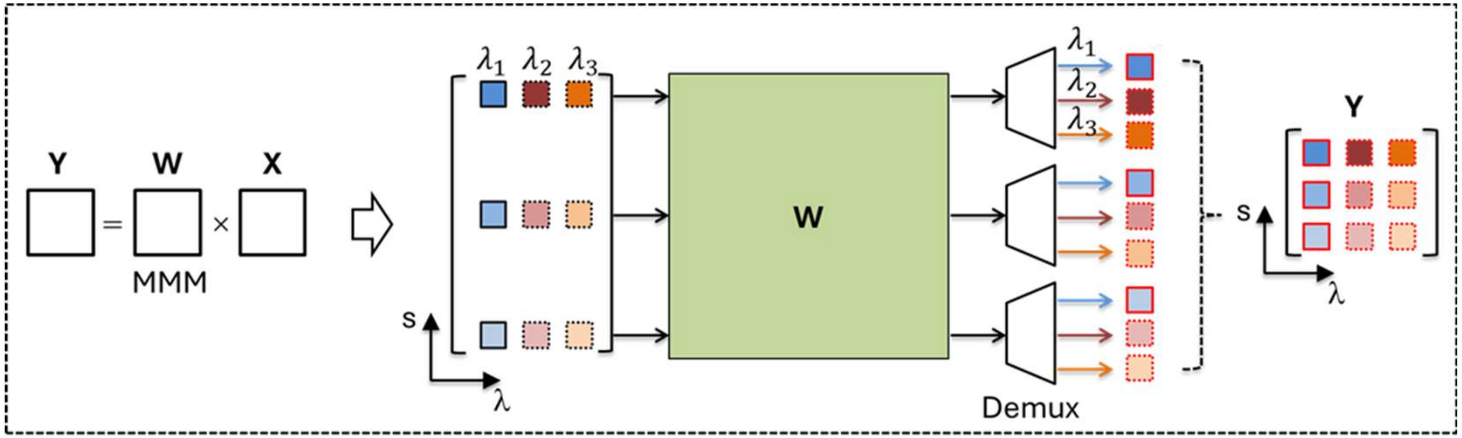


AIST

Multiplexing for linear operations





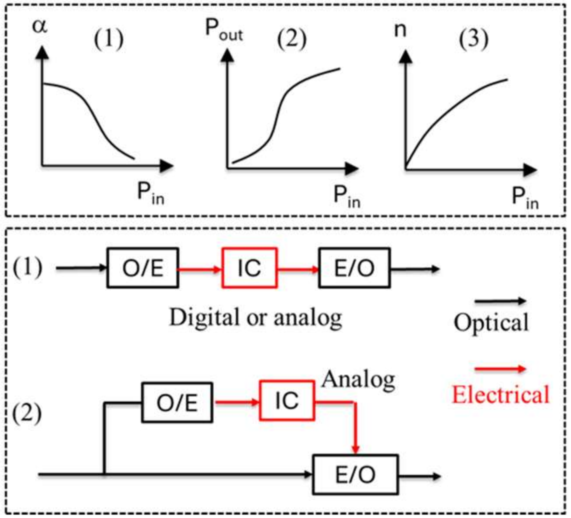


Massive parallelism by sharing hardware, offering scalable and energy efficient linear operations.

How nonlinear functions can be achieved in photonics?

All-optical

OEO



- Saturation absorber
- TPA
- SOA
- $\chi^{(3)}$...

O→PD → TIA/ADC → E

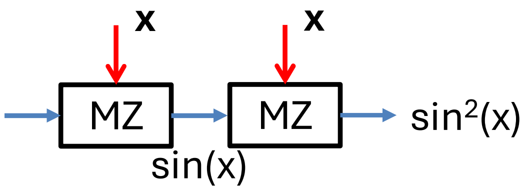
Digital circuits

Analog circuits

E→DAC/Driver →Laser/modulator → O

Towards traditional activation functions

Data-encoding enabled nonlinearity



- We are focusing on wide nonlinearities to construct new models instead of replacing traditional ones
- Despite of same technology, new possibility and effects are achievable
- Offering new merits in solving some remaining challenges

Support vector machine (SVM)-like nonlinear projection

Bilinear projection

Vertically hierarchical PNN

EO Hopfield network

Towards challenges of :

Nonlinearity

Scalability
Small-sample learning

Our works: optical streaming processors implementing novel models

Linear transformation

Linear + EO-hybrid nonlinearity

Linear + All-optical nonlinearity

2022 2023 2024 2025

Our approach

Data-encoding nonlinearity

Multiple operators

Bilinear

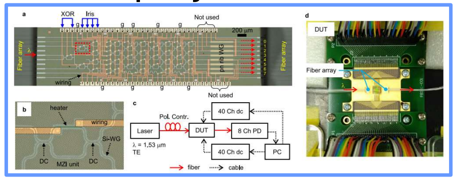
Nonlinearity

Spatiotemporal encoding
Physical intelligence

Loss-robust architecture

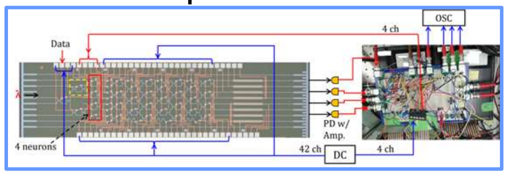
Scalability

SVM (support vector machine)-like nonlinear projection-based PNN

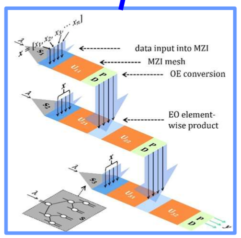


Nature Communications 13, 3261 (2022)

EO-Hopfield network

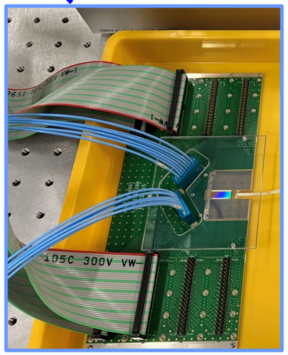


OFC 2023, W3G.2.(Invited talk)



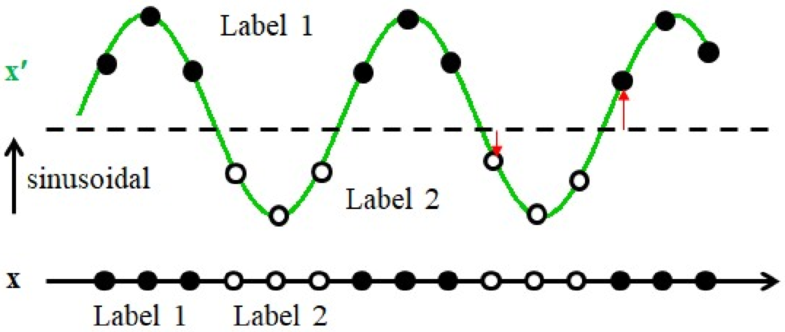
Vertically hierarchical PNN

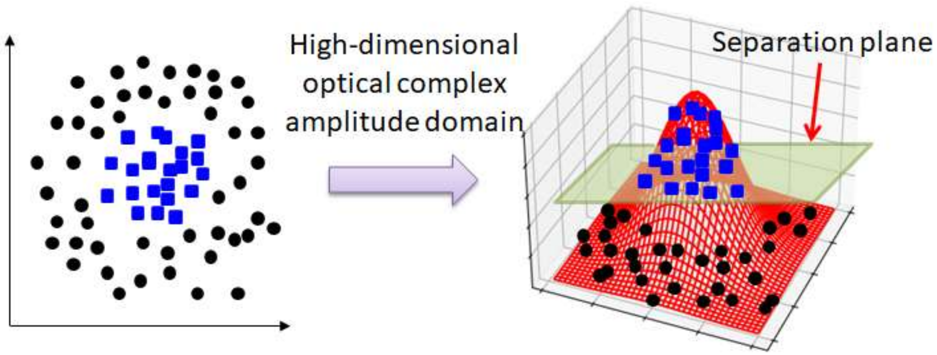
APL Photonics 9, 056110 (2024)



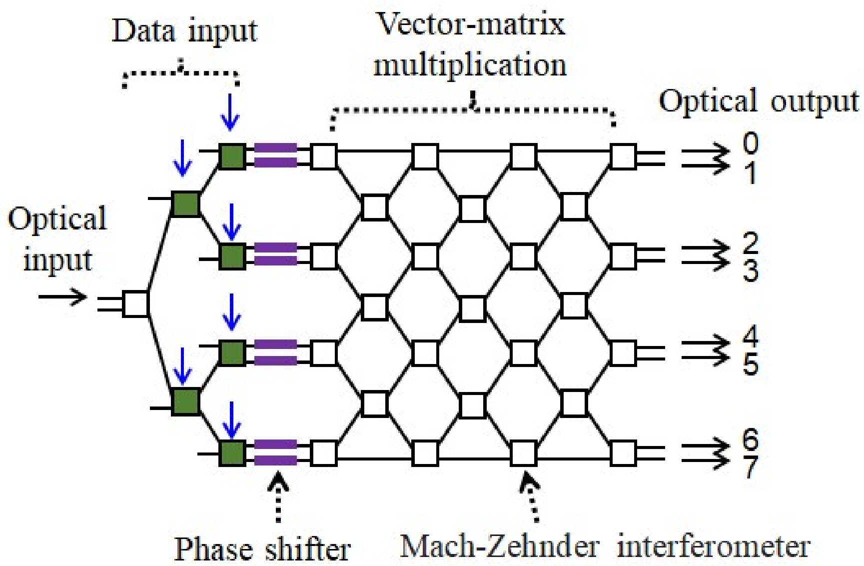
Gate modulated PNN

(1) Nonlinear projection-based PNN





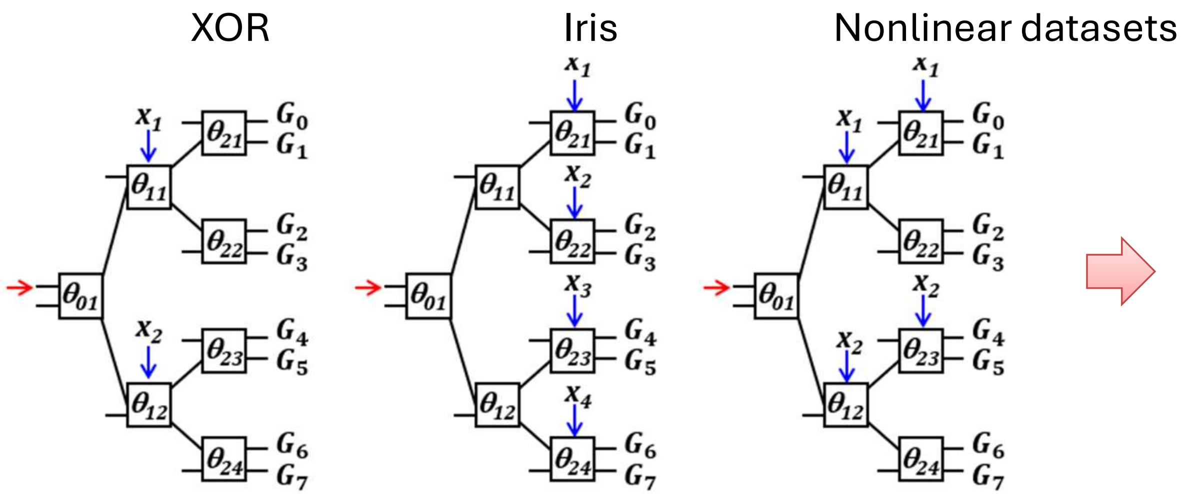
- For multilayer perceptron (MLP), it is using $y = wf(\ldots(wf(wx + b) + b))$ to approximate a nonlinear function.
- We are thinking reversely: constructing nonlinear projection function firstly, that may enable easy separation.

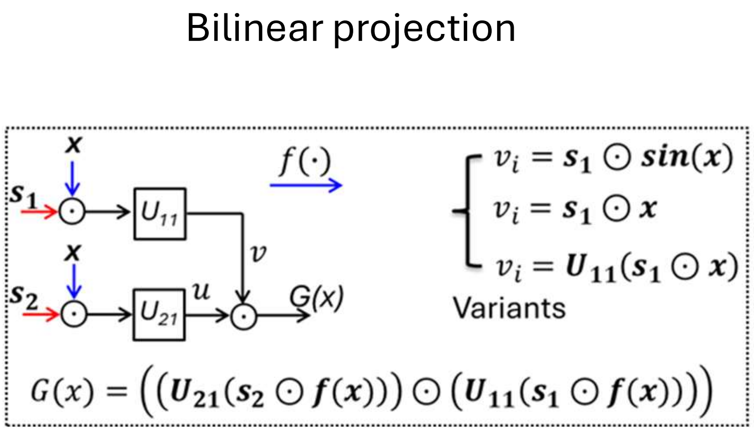


Nonlinear projection + linear separation

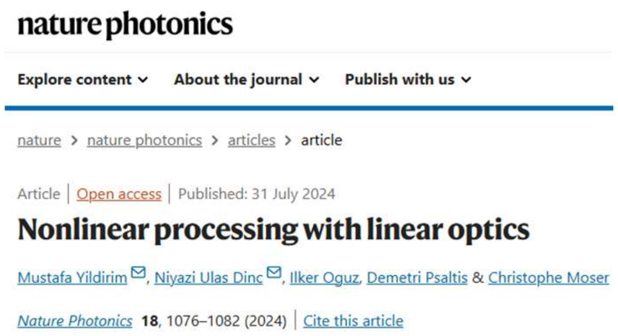
- Leveraging phase-amplitude (EO) nonlinearity of MZI
- Linear separation by VMM afterwards
- Maximum-power port position indicates the result
- Only passive circuits

Encoding the data into MZI network





- Encoding the data into MZI networks
- Leveraging EO nonlinearity associated with data encoding


nature photonics

Explore content About the journal Publish with us

nature > nature photonics > articles > article

Article | Open access | Published: 31 July 2024

Nonlinear processing with linear optics

Mustafa Yildirim, Niyazi Ulas Dinc, Ilker Oguz, Demetri Psaltis & Christophe Moser

Nature Photonics 18, 1076–1082 (2024) | Cite this article

$$E_{out} = HT_{Ln}(\mathbf{x}) \ldots HT_{L2}(\mathbf{x}) HT_{L1}(\mathbf{x}) E_{il}$$


nature photonics

Explore content About the journal Publish with us

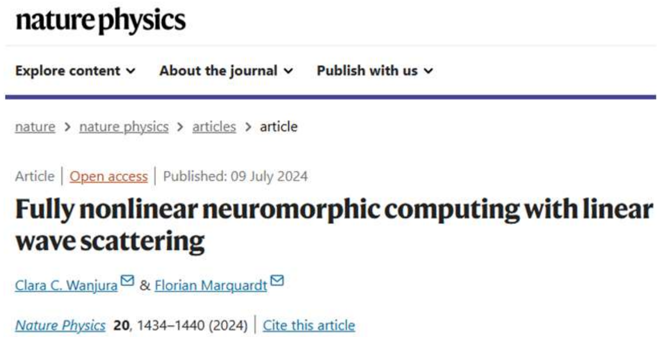
nature > nature photonics > articles > article

Article | Open access | Published: 31 July 2024

Nonlinear optical encoding enabled by recurrent linear scattering

Fei Xia, Kyungduk Kim, Yaniv Eliezer, SeungYun Han, Liam Shaughnessy, Sylvain Gigan & Hui Cao

Nature Photonics 18, 1067–1075 (2024) | Cite this article


nature physics

Explore content About the journal Publish with us

nature > nature physics > articles > article

Article | Open access | Published: 09 July 2024

Fully nonlinear neuromorphic computing with linear wave scattering

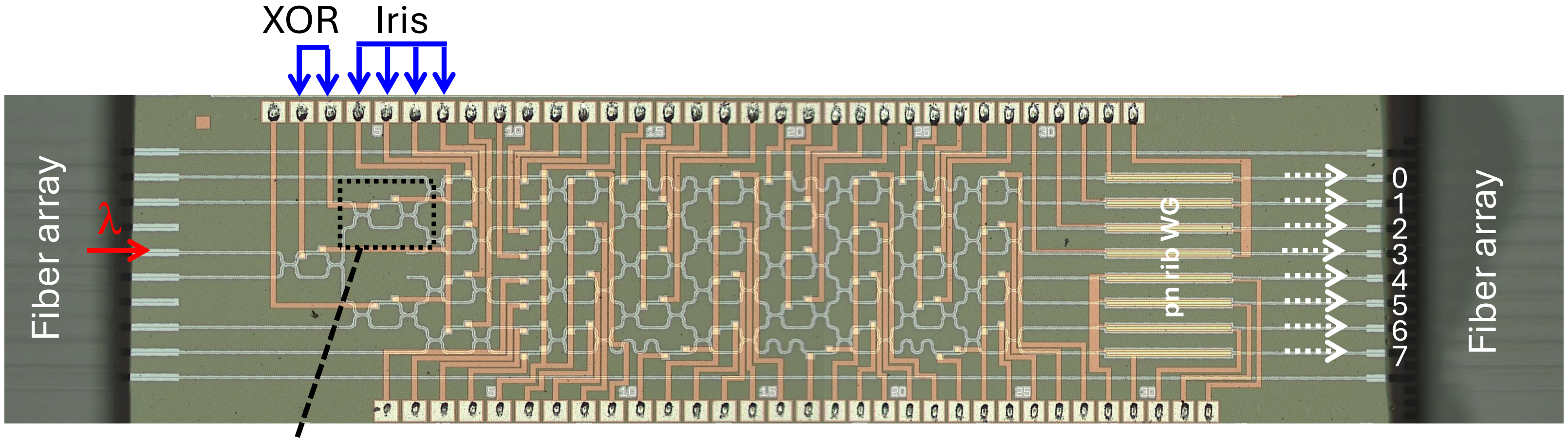
Clara C. Wanjura & Florian Marquardt

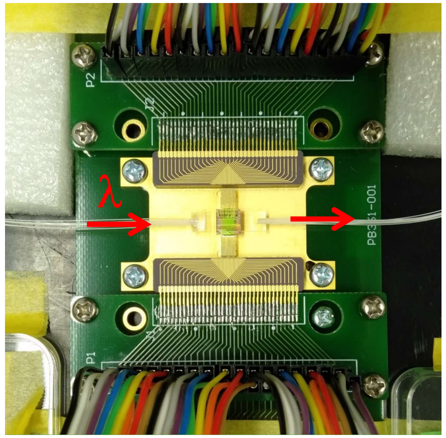
Nature Physics 20, 1434–1440 (2024) | Cite this article

Input replication improved nonlinear expressivity

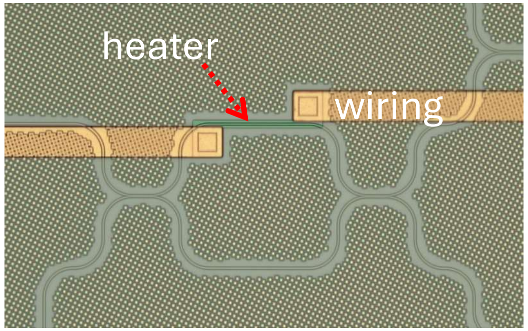
Device and module

Nature Communications 13, 3261 (2022)

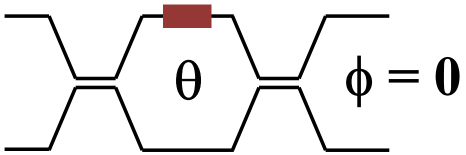


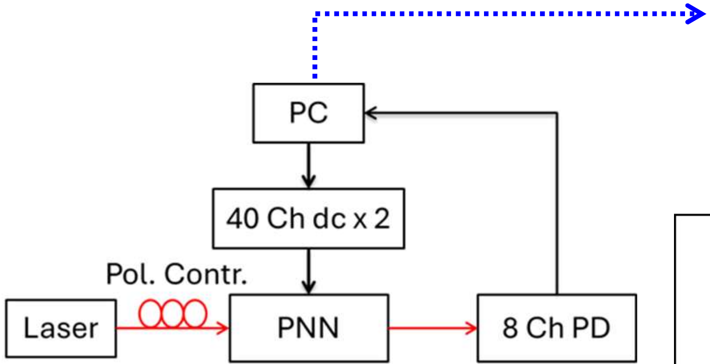


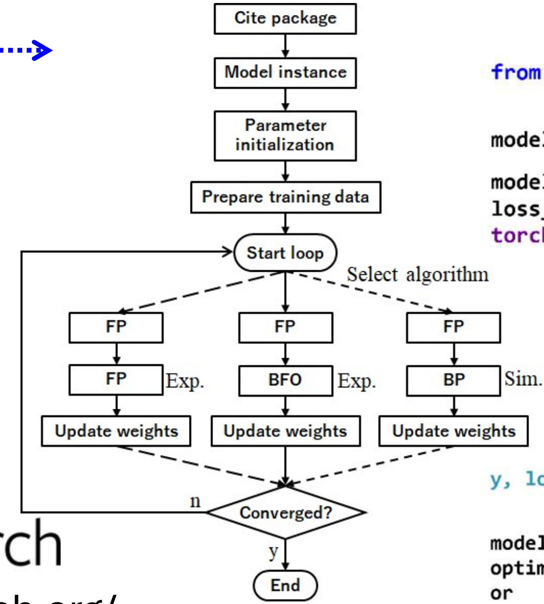
PNN module



MZI unit



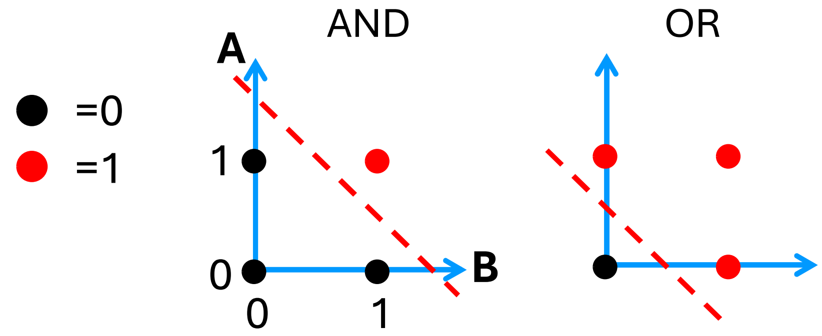


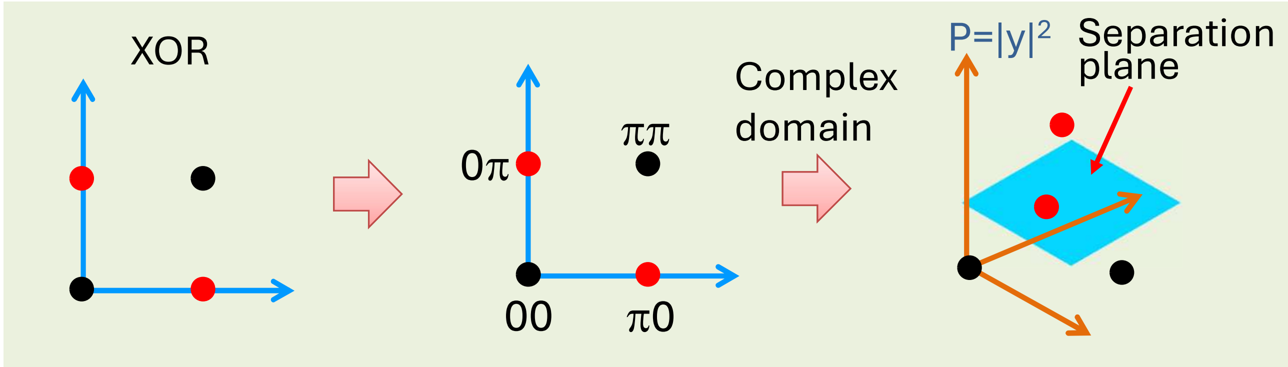


PyTorch
https://pytorch.org/

```
from pcnet import PhotonicClassifier

model = PhotonicClassifier()

model.weight_init()
loss_func =
torch.nn.MSELoss(reduction='sum')

y, loss = model(x)

model.FFGrad(x, yt, loss, 0.001)
optimizer.step()
or
model.BFOEng(xb, yb, loss, 0.001, …)
```

Classification experiment: Boolean logic

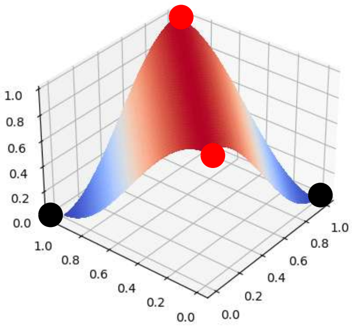




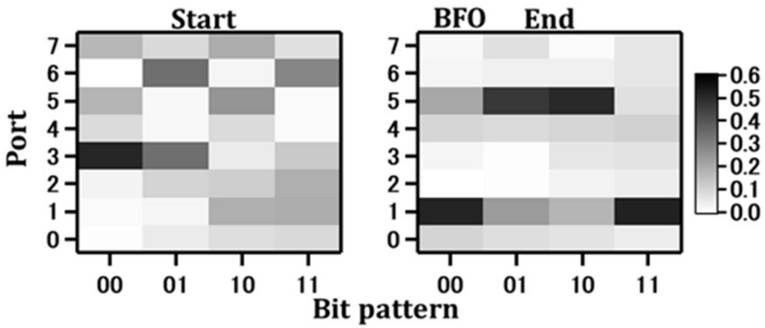
XOR

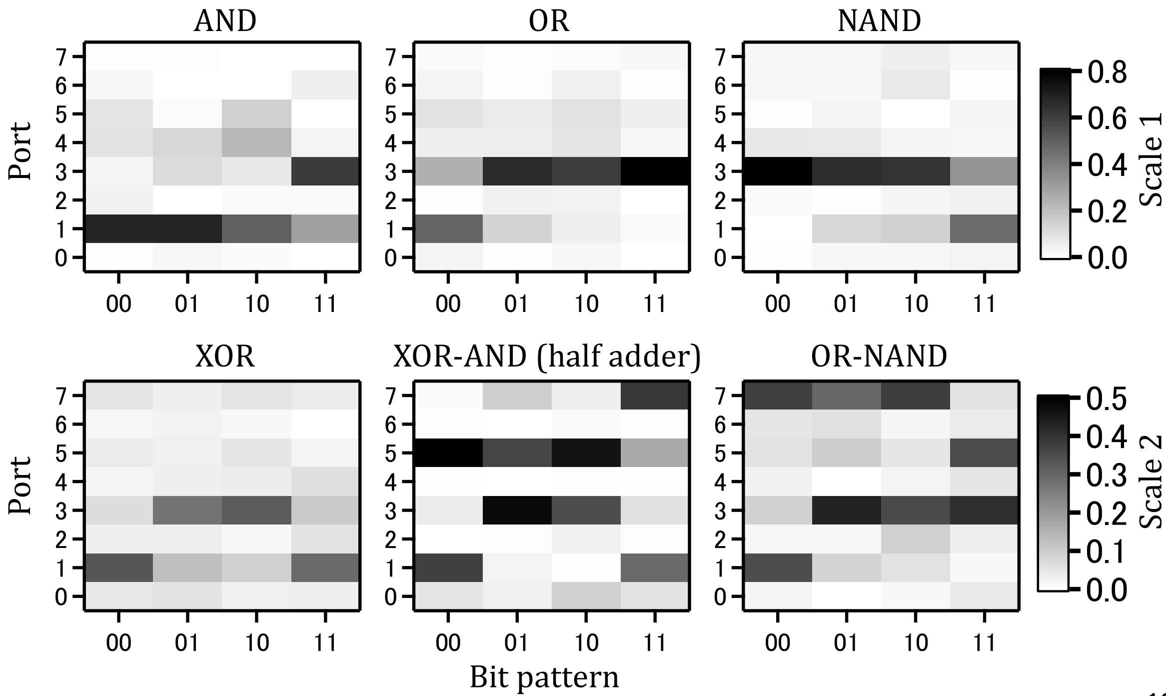
A	B	X
0	0	0
0	1	1
1	0	1
1	1	0

- Port 1 → 0
- Port 5 → 1



Optical power mapping



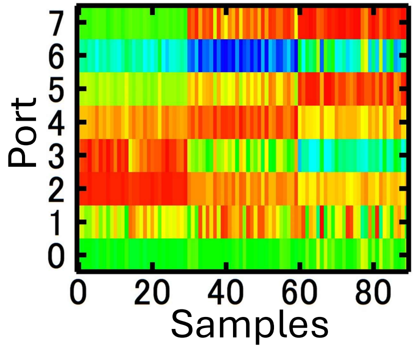


Classification experiment: Iris dataset

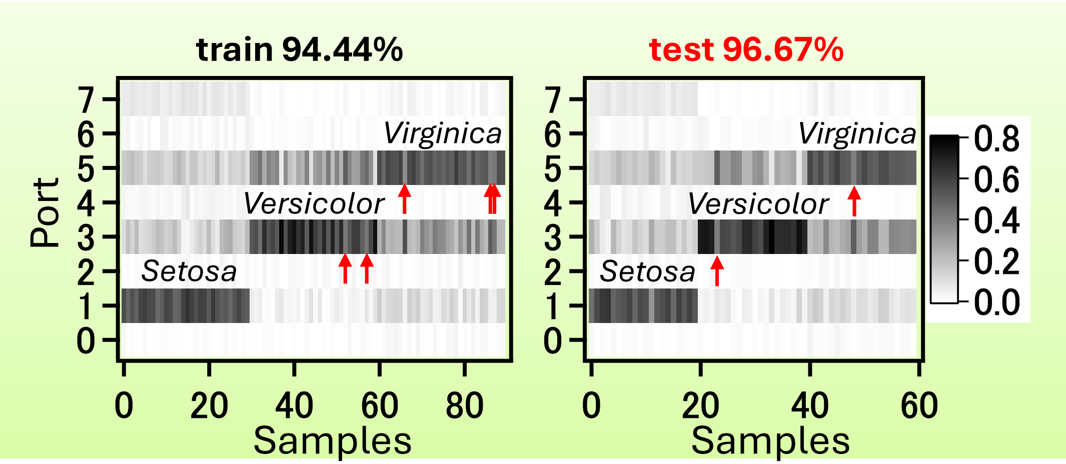
https://archive.ics.uci.edu/ml/datasets/iris

No.	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3	1.4	0.2	Setosa
3	4.7	3.2	1.3	0.2	Setosa
:	:	:	:	:	:
51	7	3.2	4.7	1.4	Versicolor
52	6.4	3.2	4.5	1.5	Versicolor
53	6.9	3.1	4.9	1.5	Versicolor
:	:	:	:	:	:
101	6.3	3.3	6	2.5	Virginica
102	5.8	2.7	5.1	1.9	Virginica
103	7.1	3	5.9	2.1	Virginica
:	:	:	:	:	:

Optical power before training



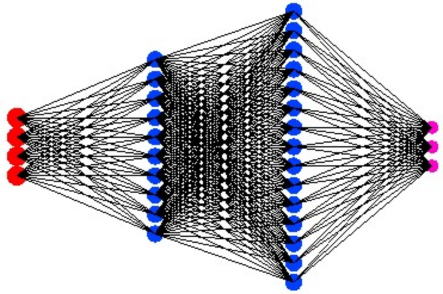
- Data normalization
- Port assignments
  - Port 1 → *Setosa*
  - Port 3 → *Versicolor*
  - Port 5 → *Virginica*



Optical power mapping after training

Two examples of digital NN

https://python.atelierkobato.com/variety/



- 4×10×15×3
- ReLU
- 235 weights
- ~97%

Conferences > 2018 International Conference...

A Model of Deep Neural Network for Iris Classification With Different Activation Functions

Publisher: IEEE Cite This PDF

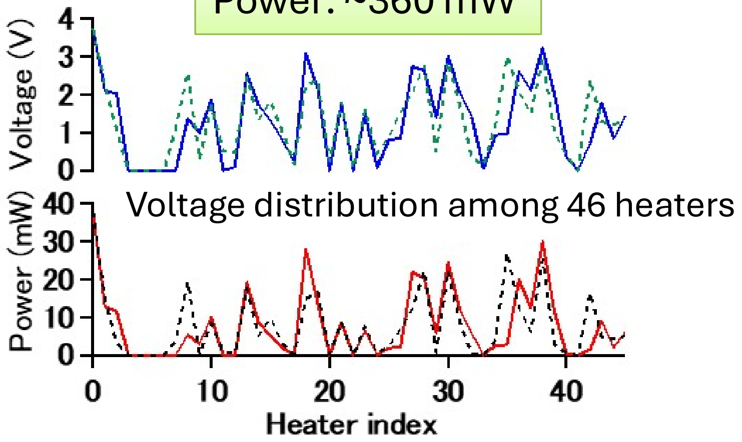
DOI: 10.1109/IDAP.2018.8620866

1st Hidden Layer	2nd Hidden Layer	Epoch Number	Accuracy Rate
Relu	Sigmoid	100	86%
		200	86%
		300	86%
		400	86%
Relu	Tanh	100	90%
		200	90%
		300	90%
		400	90%
Relu	Relu	100	90%
		200	90%
		300	90%
		400	90%

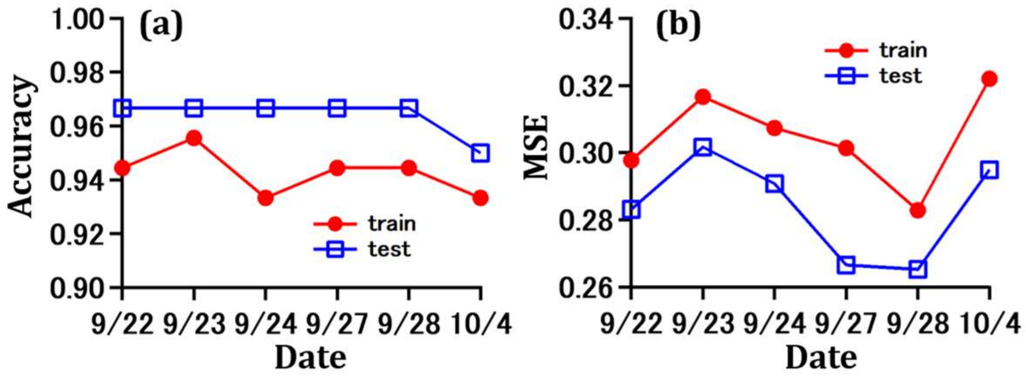
86-90%

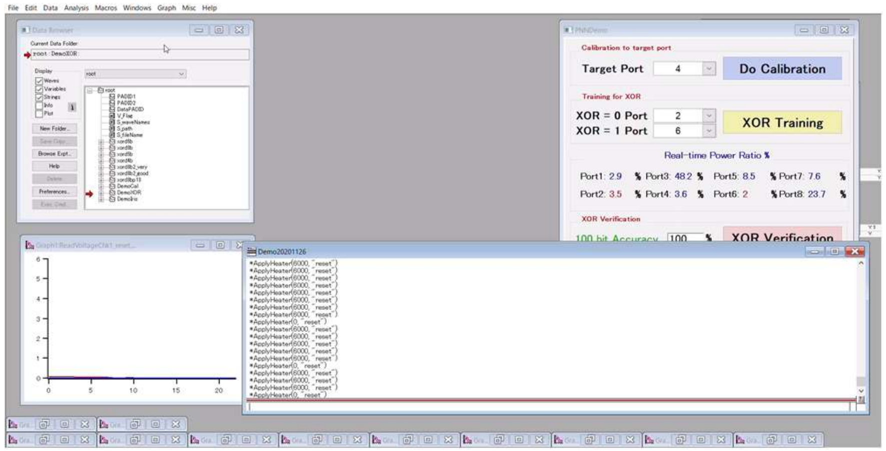
Details of experimental results

Power: ~360 mW

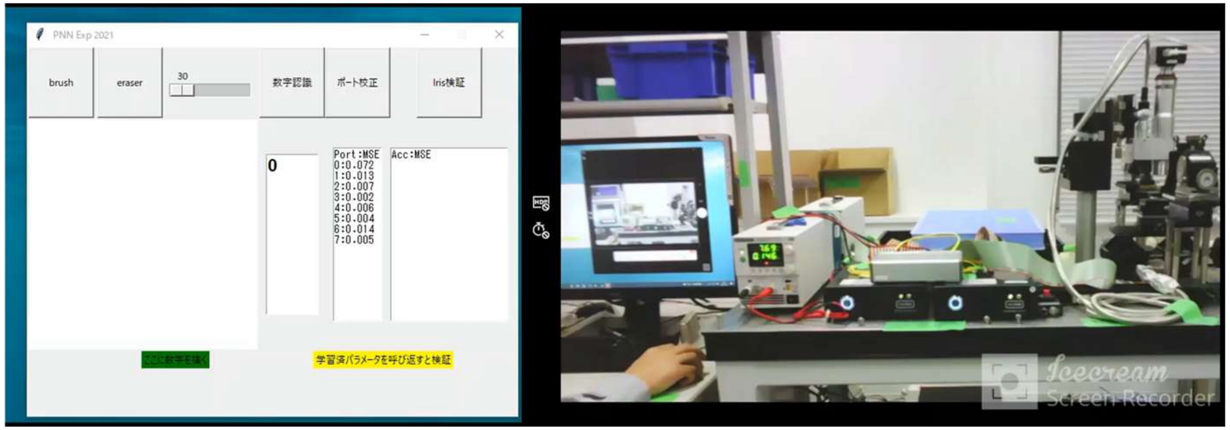


Long time stability





XOR video using BFO algorithm



Demo video

- ✓ Electronics for IO only
- ✓ Computing is done just by optical propagation
- ✓ Latency < 100 ps

(2) Bilinear projection in gate modulated PNN

$$z_1 = Wx, z_2 = Vx$$

Bilinear $\quad B = z_1 \otimes z_2$

$$B_i = \sum W_{ij} x_j \cdot \underline{V_{ik} x_k}$$

Variants

$$v_i = \sum W_{ij} x_j \cdot \underline{x_k}$$

$$v_i = \underline{\sum W_{ij} x_j} \cdot \underline{sin(x_k)}$$

arXiv:2002.05202v1,12 Feb 2020

GLU Variants Improve Transformer

Noam Shazeer
Google
noam@google.com

February 14, 2020

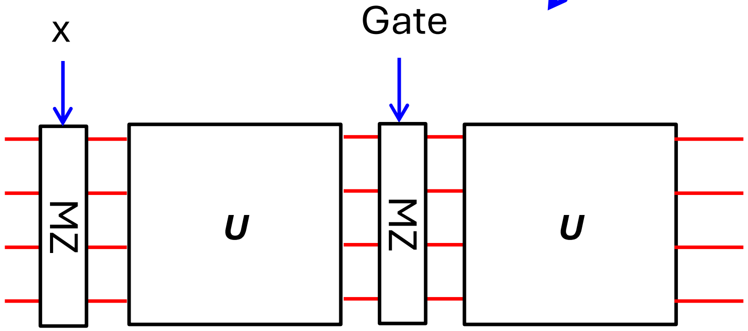
These variants have shown improvements over traditional activation functions when used in Transformer models.

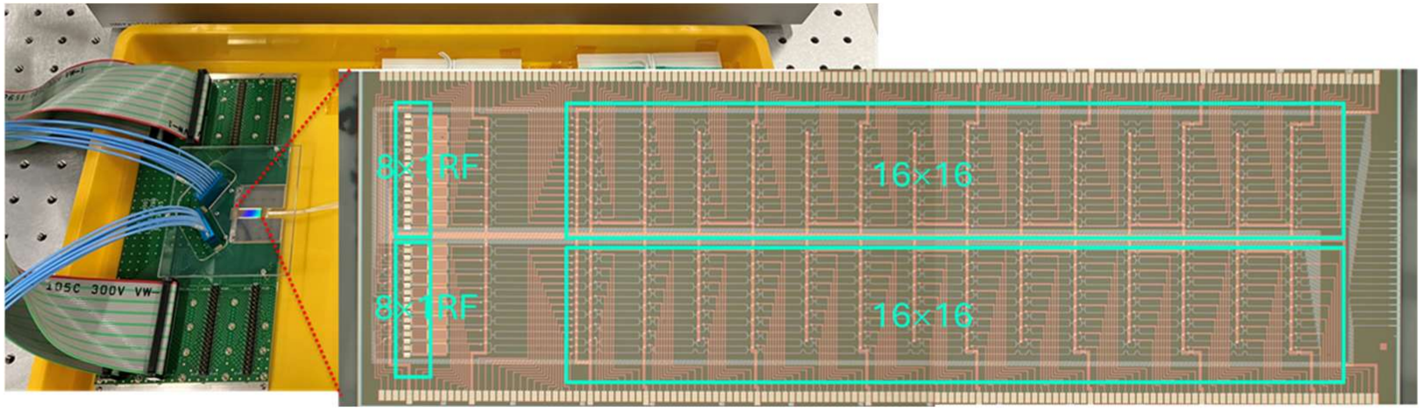
Abstract

Gated Linear Units [Dauphin et al., 2016] consist of the component-wise product of two linear projections, one of which is first passed through a sigmoid function. Variations on GLU are possible, using different nonlinear (or even linear) functions in place of sigmoid. We test these variants in the feed-forward sublayers of the Transformer [Vaswani et al., 2017] sequence-to-sequence model, and find that some of them yield quality improvements over the typically-used ReLU or GELU activations.

$$\mathrm{GLU}(x, W, V, b, c) = \sigma(xW + b) \otimes (xV + c)$$

$$\mathrm{Bilinear}(x, W, V, b, c) = (xW + b) \otimes (xV + c)$$





Acknowledgments

https://www.jst.go.jp/kisoken/crest/

JST CREST (JPMJCR21C3)

JSPS KAKENHI JP23H01885

https://www.jst.go.jp/kisoken/crest/

JST CREST (JPMJCR24R3) (2024 newly launched)

AIST SCR Platform

Thank you for kind attention!