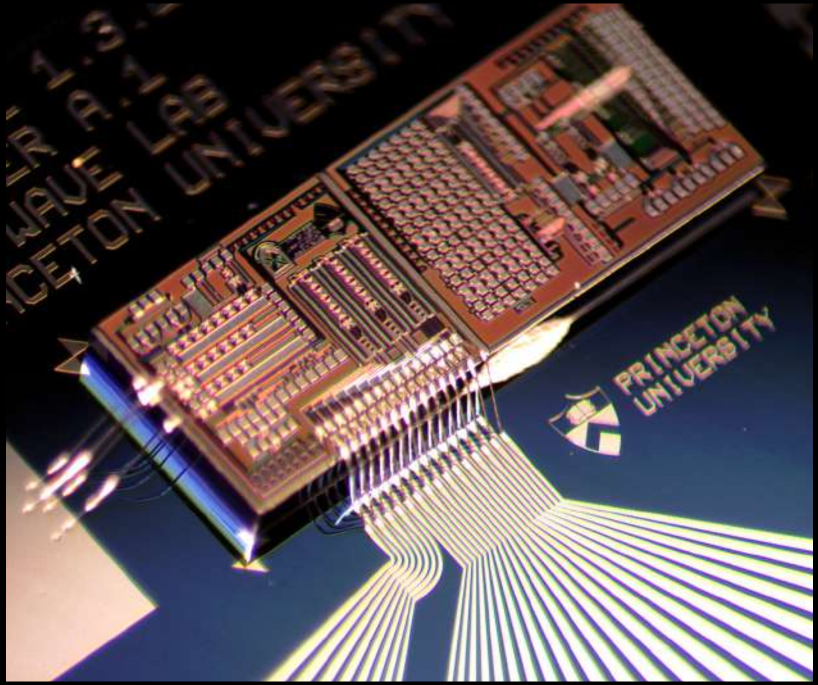
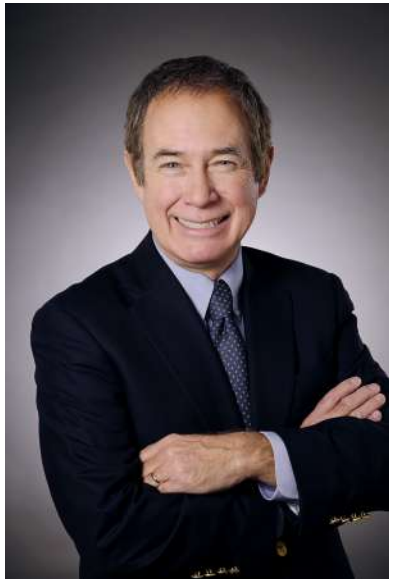
Silicon Photonics for Artificial Intelligence and Neuromorphic Computing

Bhavin J. Shastri
Queen's University, Canada
Vector Institute for Artificial Intelligence



Queen's University
VECTOR INSTITUTE
PRINCETON UNIVERSITY

Collaborators



Paul Prucnal
Princeton

Lukas Chrostowski
UBC

Sudip Shekhar
UBC

Alexander Tait
Queen’s

Chaoran Huang
CUHK

Volker Sorger
GWU

+ many others and co-authors

Fabrication Support

CMC MICROSYSTEMS

SiEPIC FAB

ANT APPLIED NANOTOOLS INC.

amf ADVANCED MICRO FOUNDRY

Students and Postdocs

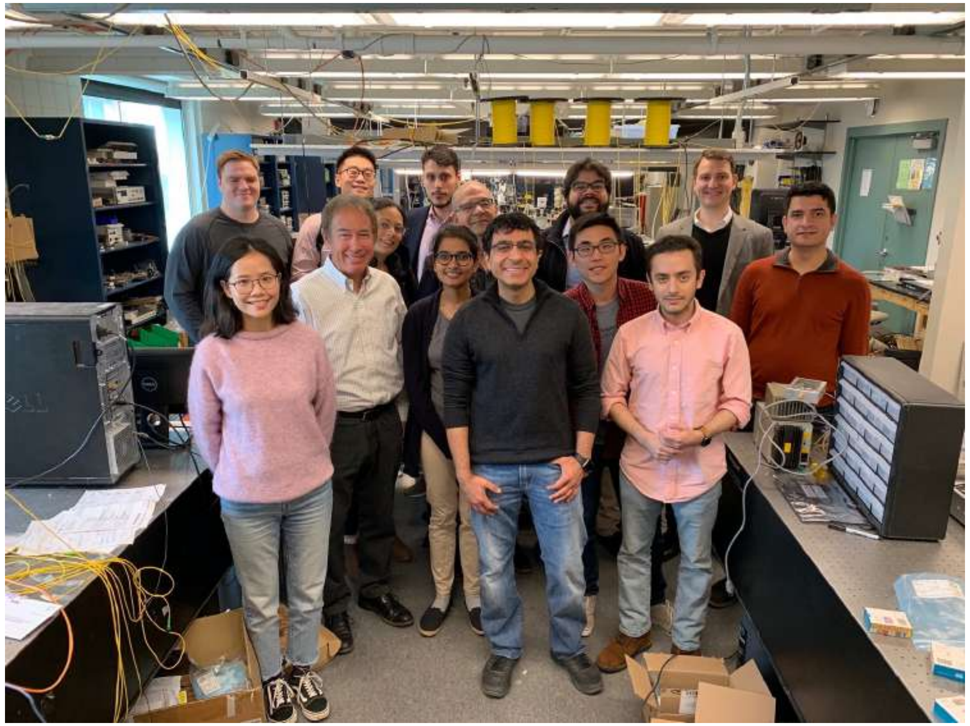
Credit: Ishana Gopaul

Shastri Lab (Queen's)

Postdoc: Bicky Marquez

Graduate Students: Matthew Filipovich; Zhimu Guo; Hugh Morison; Jagmeet Singh; Marcus Tamura; Maryam Moridsadat; Ahmed Khaled; Karanpreet Singh; Jacob Ewaniuk

+ former students



Lightwave Prucnal Lab (Princeton)

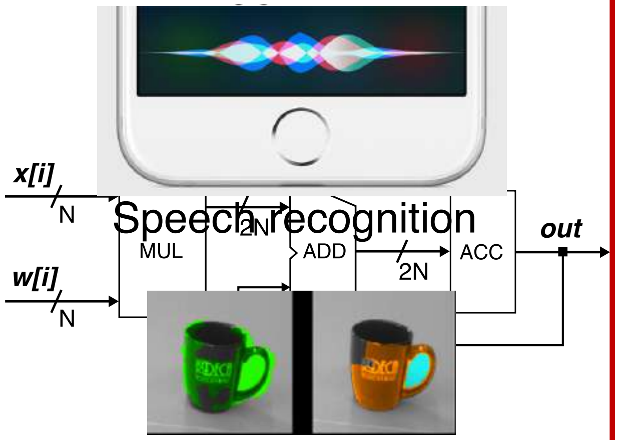
Researcher: Lei Xu

Graduate Students: Thomas Ferreira de Lima; Hsuan-Tung Peng; Eric Blow, Aashu Jha; Simon Bilodeau; Weipang Zhang; Josh Lederman

+ former students and postdocs

Need for Speed

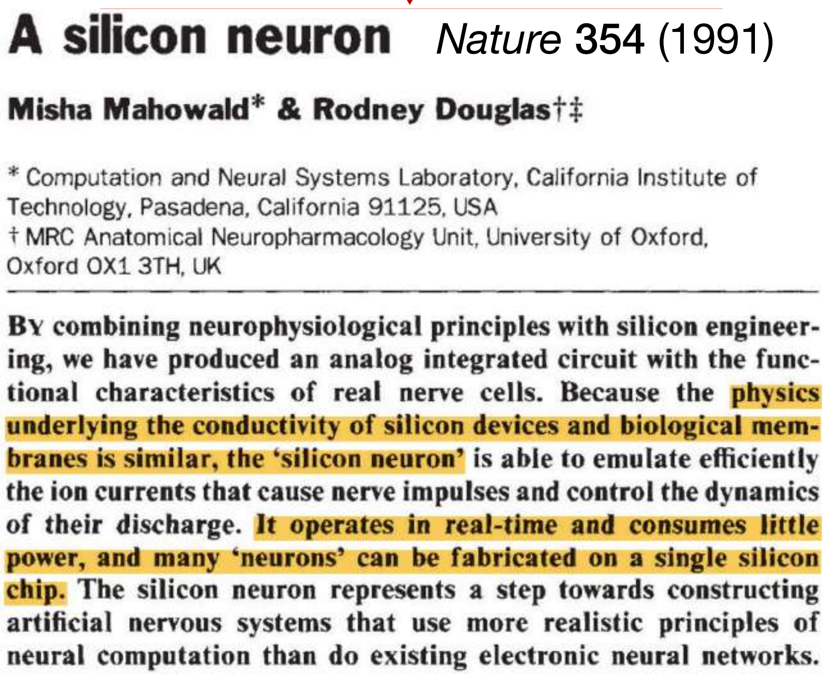
Software neural networks on **digital hardware**	Hardware neural networks **Neuromorphic Electronics**	Optical neural networks **Neuromorphic Photonics**



Speech recognition

Object recognition

Hz - kHz



A silicon neuron

Misha Mahowald* & Rodney Douglas†‡

* Computation and Neural Systems Laboratory, California Institute of Technology, Pasadena, California 91125, USA
† MRC Anatomical Neuropharmacology Unit, University of Oxford, Oxford OX1 3TH, UK

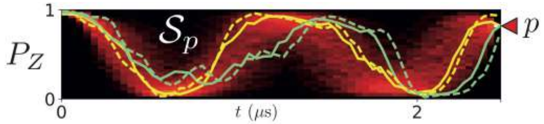
BY combining neurophysiological principles with silicon engineering, we have produced an analog integrated circuit with the functional characteristics of real nerve cells. Because the physics underlying the conductivity of silicon devices and biological membranes is similar, the 'silicon neuron' is able to emulate efficiently the ion currents that cause nerve impulses and control the dynamics of their discharge. It operates in real-time and consumes little power, and many 'neurons' can be fabricated on a single silicon chip. The silicon neuron represents a step towards constructing artificial nervous systems that use more realistic principles of neural computation than do existing electronic neural networks.

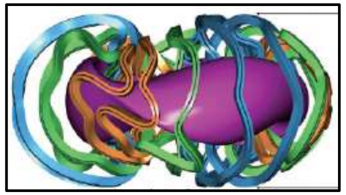
Nature **354** (1991)

Bandwidth-complexity tradeoff!

Capacitive loading, EMI, topological constraints

- *Bandwidth-complexity tradeoff!*





Nonlinear optimization
Ultrafast control
RF signal processing
Quantum tomography

GHz

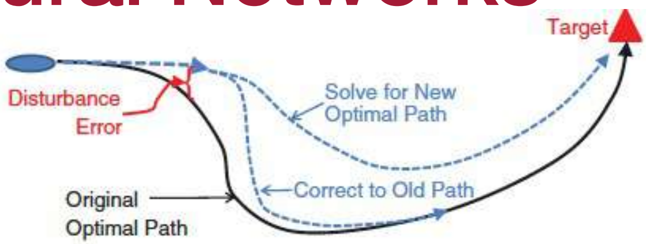
Could enable new applications that are challenging to be achieved with electronics!

Prucnal, Shastri et al. *Adv. Opt. Photonics* **8** (2016)

Applications for Photonic Neural Networks

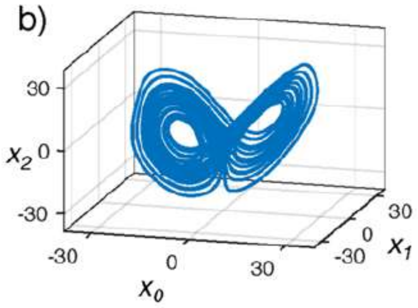
- ❑ **Nonlinear programming**
  - Nonlinear optimization problems (robotics, predictive control, autonomous vehicles)
  - Ordinary/partial differential equations
- ❑ **High-performance computing and Machine Learning**
  - Vector-matrix multiplications
  - Deep learning inference
  - Ultrafast and online learning
- ❑ **Intelligent signal processing**
  - Optical fiber communications
  - mm-wave edge processing
  - Spectral mining
- ❑ **Physics**
  - Qubit readout classification
  - High-energy particle collision experiments

Huang, Prucnal, Shastri et al. arXiv:2105.09943 (2021)
Shastri, Tait, Prucnal et al. *Nature Photonics* **15** (2021)

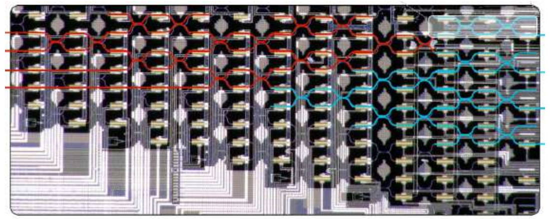


Ferreira de Lima, Prucnal et al. JLT (2019)
Model predictive control

Tait, Shastri, Prucnal et al. Sci. Rep. (2017)
Lorenz attractor



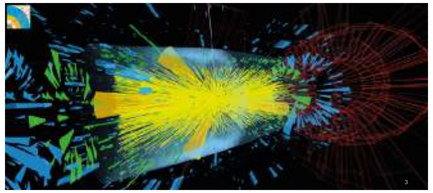
Shen et al. Nat. Photon. (2017)
Vector-matrix multiplier



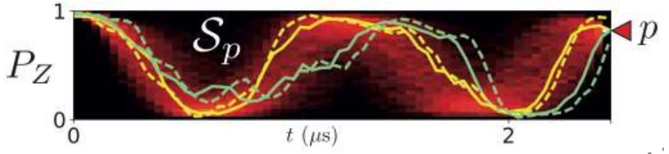
Ma, Shastri, Prucnal et al. OE (2019)
PCA, ICA, BSS

Huang, Shastri, Prucnal et al.
Nature Electron. (2021)
Fiber nonlinearity compensation

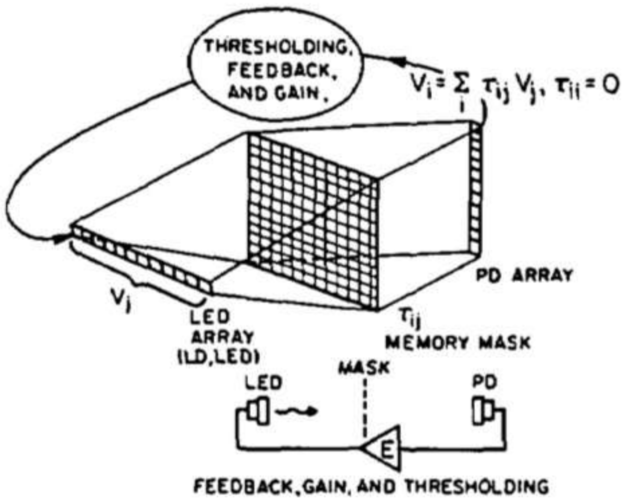
CMS Detector at CERN



Quantum tomography

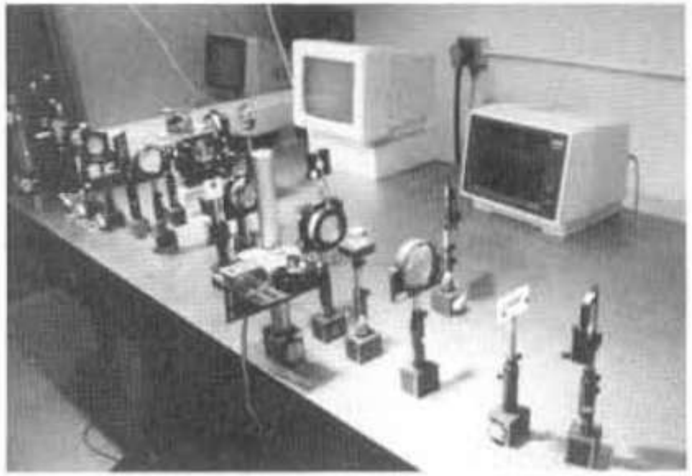


We Have Been Here Before: What Has Changed?



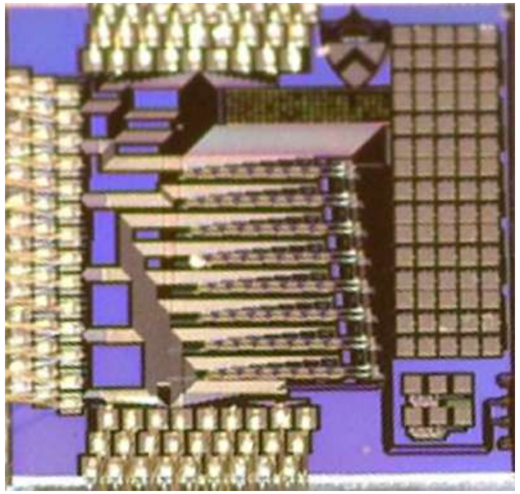
Psaltis, D., & Farhat, N. (1985)

Optical information-processing systems can have high processing power because of the large degree of parallelism as well as the interconnection capability that is achievable. Typically, more than 10^6 parallel processing channels are available in the optical system, and furthermore each of these channels can be optically interconnected (broadcasted) to 10^6 other channels. The majority of optical processors are analog systems, designed to perform linear operations. The accuracy of an analog processor is limited by the linear dynamic range of the devices used (detectors, light modulators).



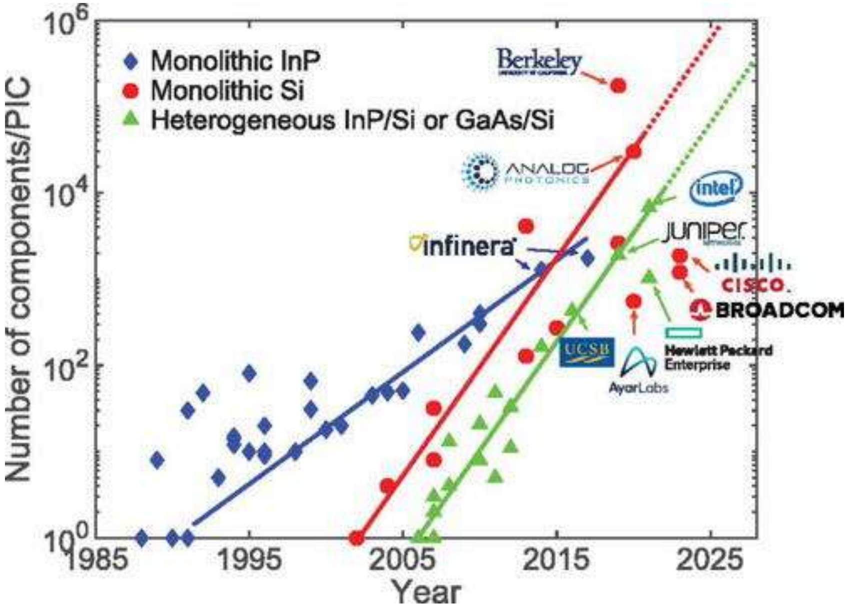
Free-space photonic neural networks

Lightwave Lab Princeton (2018)

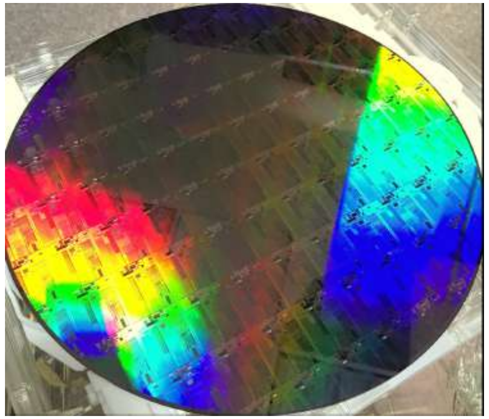


Silicon photonic neural networks (2x2 mm^2)

Photonic Moore's Law



Margalit et al. *Appl. Phys. Lett* **118** (2021)



Neuromorphic Hardware for AI

Robert Keyes' main criticisms of optical computing in the 1980's

R. W. Keyes, Optical Logic: In the light of computer technology *Optica Acta*, **32** (5), 1985
R. W. Keyes, What makes a good computer device? *Science*, **230**, pp.138–144, 1985

		Electronics	Photonics*
Computing	Nonlinearity	Easy (transistors)	Hard (but O-E-O)**
	Memory	Easy (DRAM, flip-flop)	Hard
	Gain	Easy	Easy (O-E-O)
Communication	Communication	~(1/2)CV^2 energy cost	Free (waveguide)
	Fan-out	~(1/2)CV^2 energy cost	Free (beamsplitter, WDM)
Linear Algebra	MAC (matrix-vector)	Hard (for digital)	Easy (EO Components)
System	Domain Crossings	No	Yes** (O-E-O)

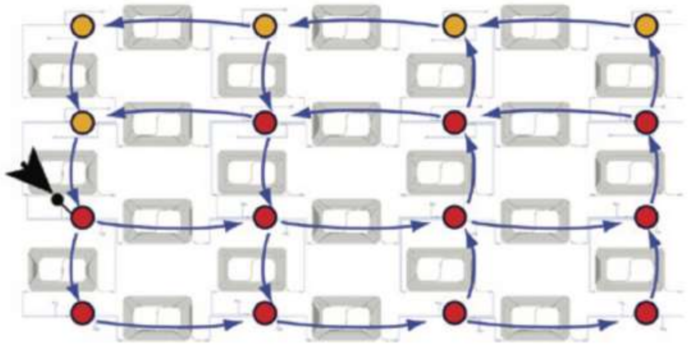
* State-of-art, e.g. Review in Shastri et al. *Nature Phot*. (2021)
Feldman *Nature* (2021)

** Tait et al. *Sci. Rep.* (2017)
Tait et al. *Phys. Rev. Appl.* (2018)

Credit: Tait (NIST), Englund (MIT) Sorger (GWU)

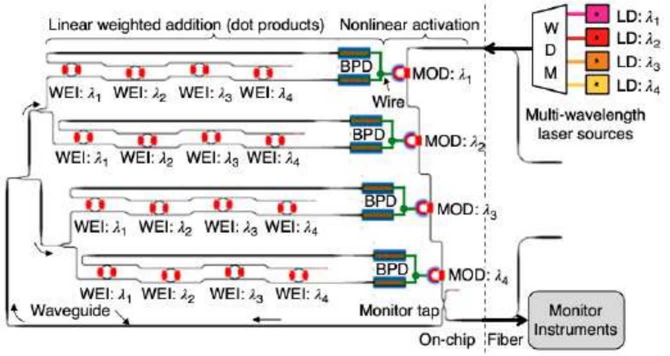
Neuromorphic Photonic Architectures

Reservoir computing
(UIB, Ghent, Femto-St)



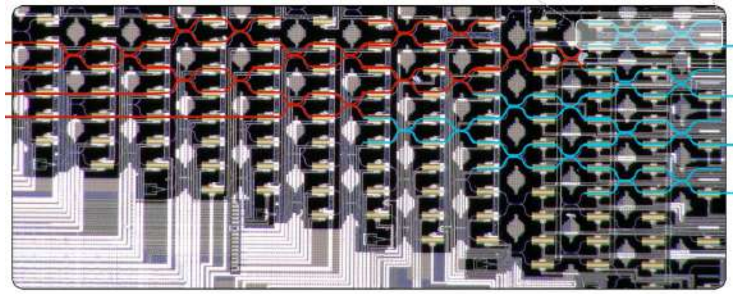
Brunner et al. *Nat. Commun.* **4** (2012)
Vandoorne et al. *Nat. Commun.* **5** (2014)

Multiwavelength Networks
(Princeton, Queen's, GWU)



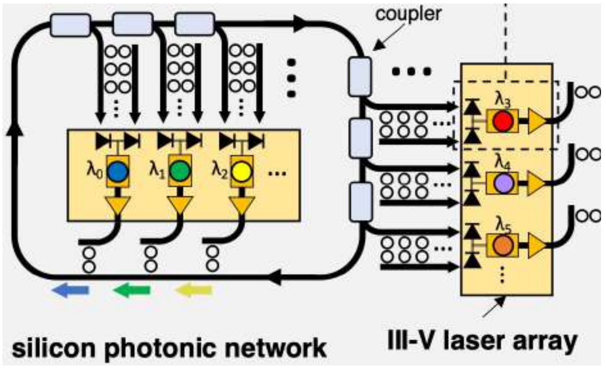
Tait et al *Sci. Rep.* **4** (2017)
Feldman et al. *Nature* **589** (2021)

Coherent networks
(MIT, Stanford)



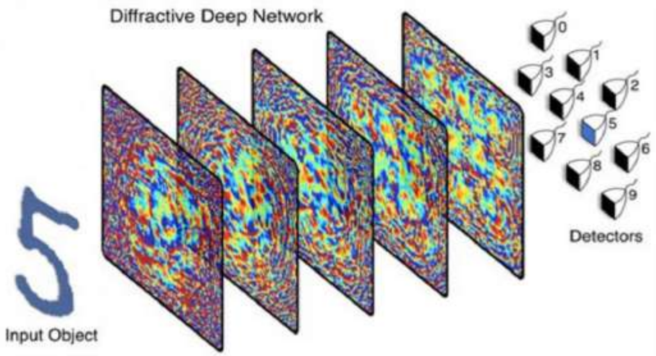
Shen et al. *Nat. Photon.* **11** (2017)
Hughes et al. *Optica* **5** (2018)

Spiking networks
(Princeton, Oxford, Strathclyde)



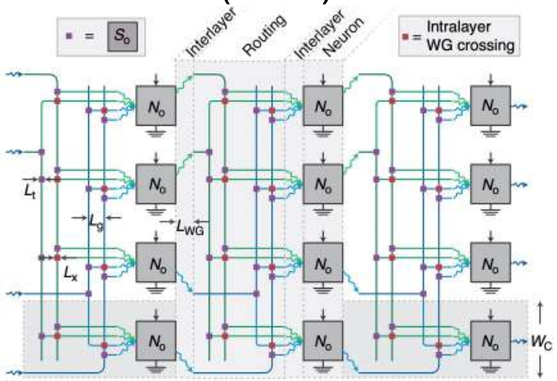
Shastri et al. *Sci. Rep.* **6** (2016)
Feldman et al. *Nature* **569** (2019)

Diffractive Optics
(UCLA, Femto-St)



Bueno et al. *Optica* **5** (2018)
Lin et al. *Science* **361** (2019)

Superconducting
(NIST)



Shainline et al. *Phys. Rev. Appl.* **7** (2017)

Shastri, Tait, Prucnal et al. *Nature Photonics* **15** (2021)

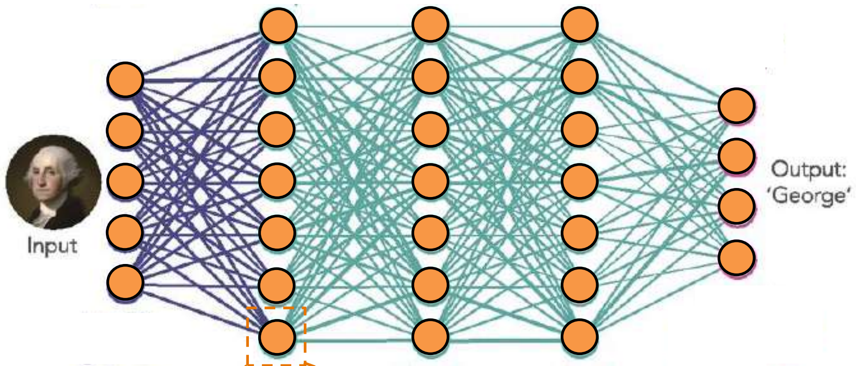
Artificial Neuron Model that Emulates Biology

Neurons have three key functions

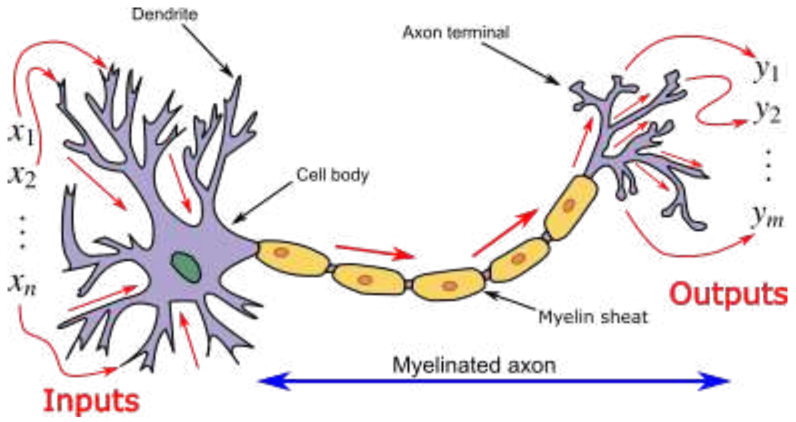
Pattern Matching
1) **Weight** multiple incoming input signals
2) **Sum** weighted inputs

Decision
3) **Threshold** the weighted sum to determine output

Neural network structure

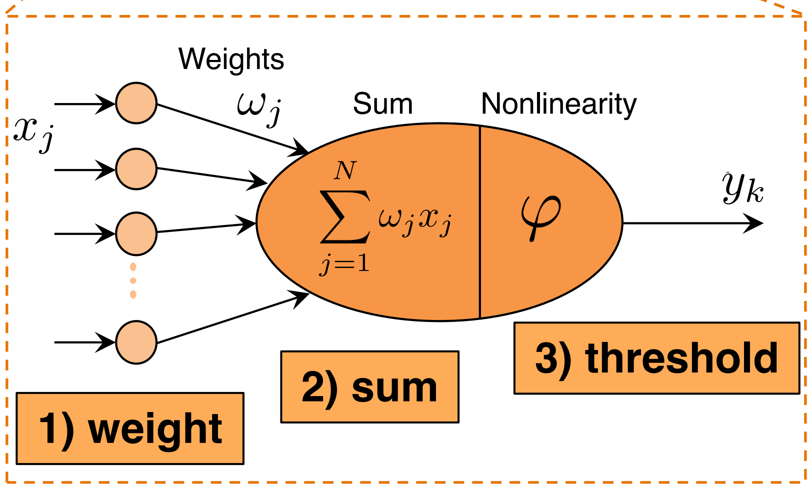


Biological neuron

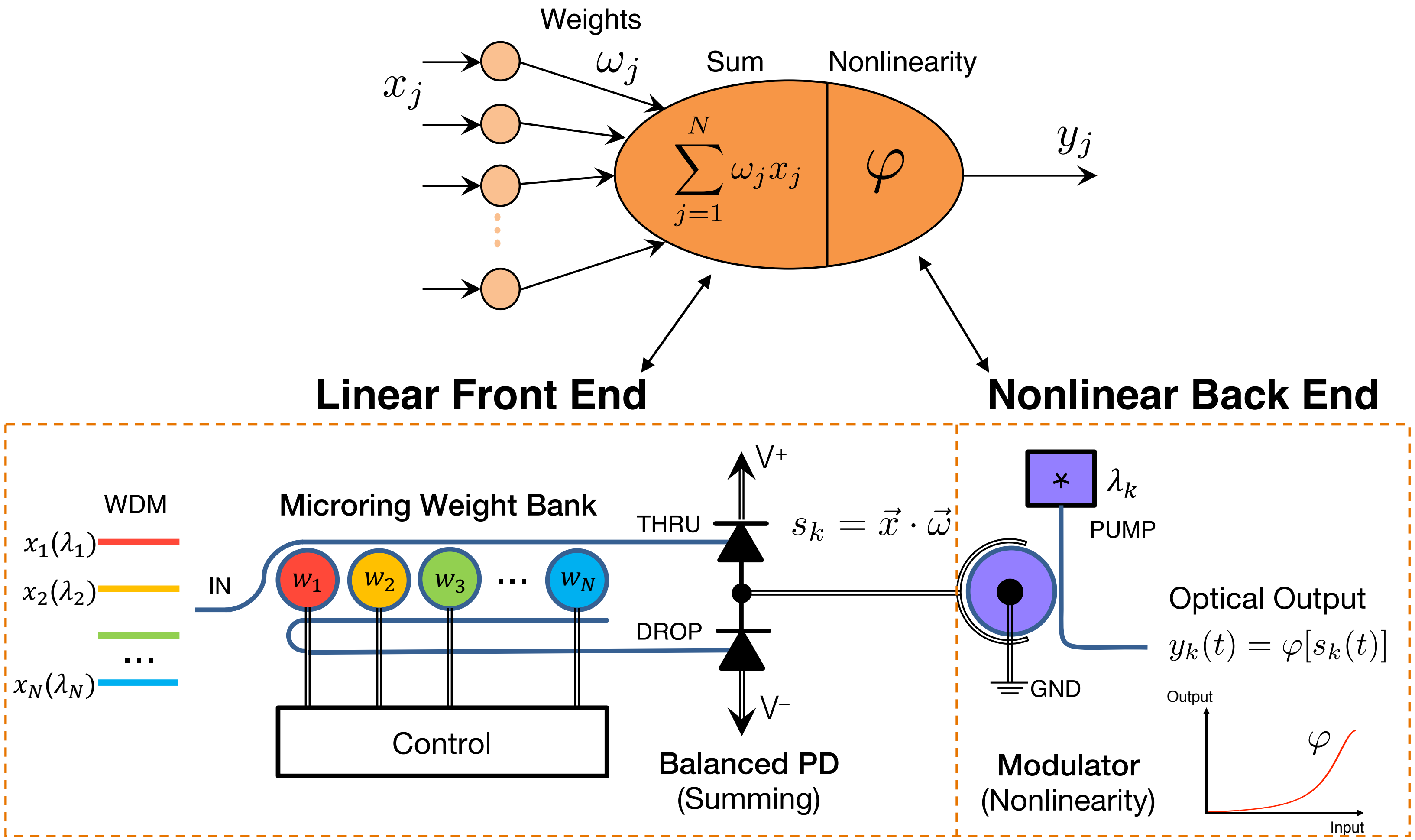


inspired

Artificial neuron

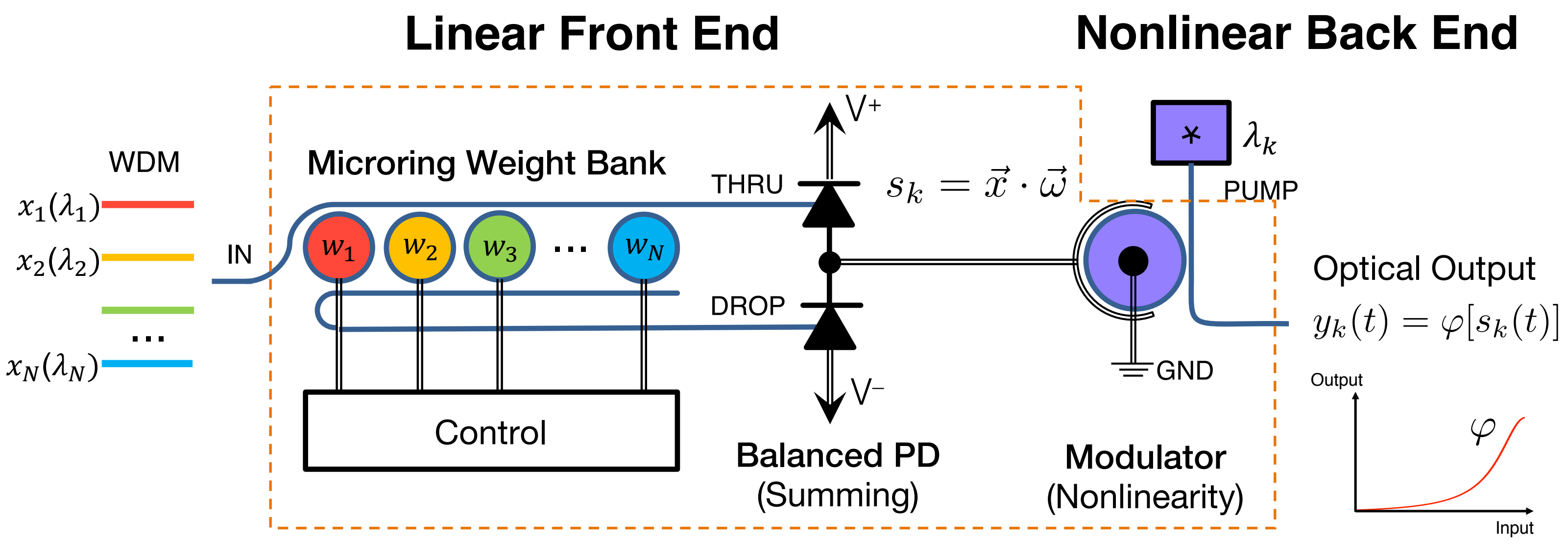


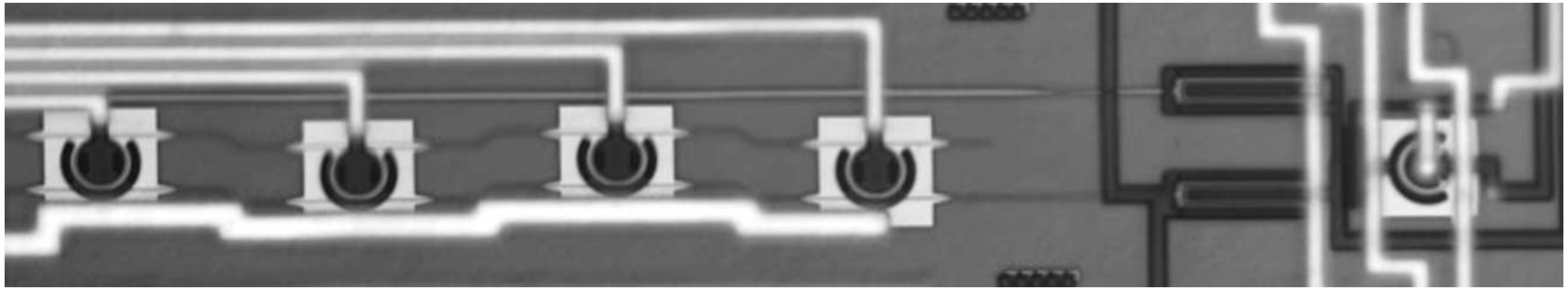
Scalable, Cascadable & Localized Multiwavelength Photonic Neuron



- ❑ All key functions are physically localized in each neuron, enabling distributed processing
- ❑ OEO enables gain, thresholding and cascadability/recurrent operation

Monolithically-Integrated Photonic Neuron



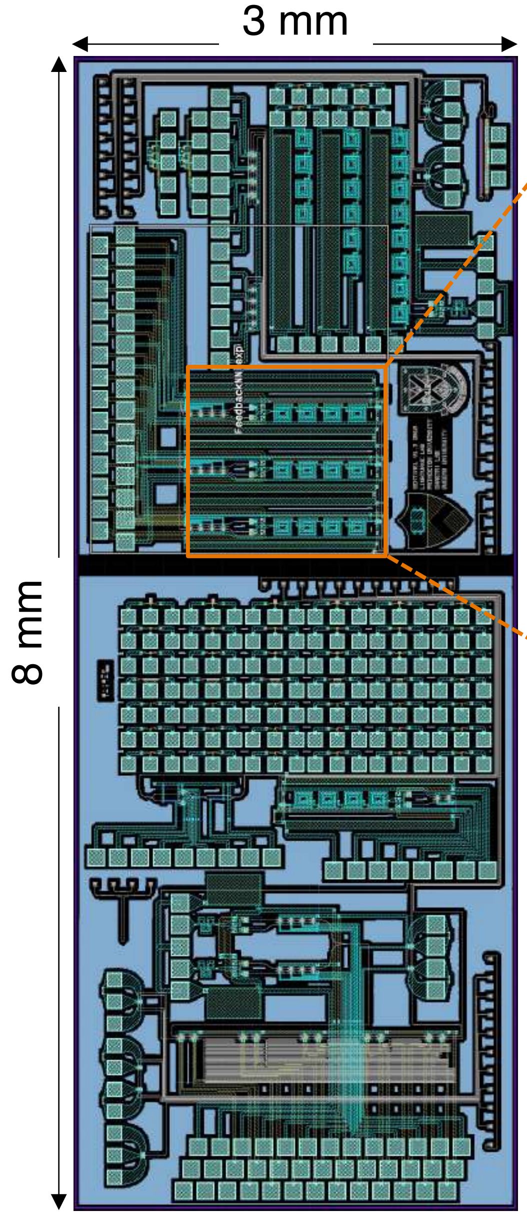


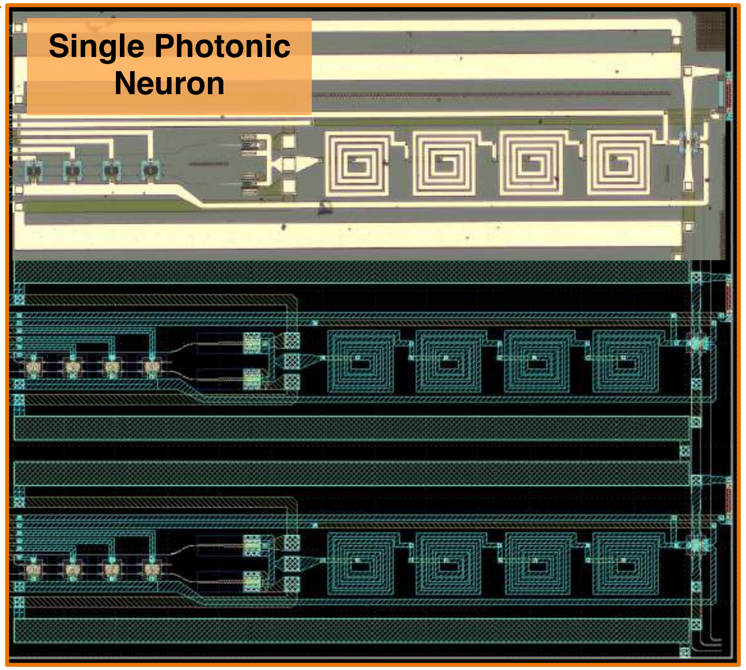
- ❑ **Integrates both linear and nonlinear neuron functionality**
- ❑ **Energy efficiency today: 500fJ/MAC; foreseeable: 1.1 fJ/MAC***
- ❑ **Operational speed: GHz**

*Nozaki, K. *et al.* *Nat. Photonics* **13**, 454–459 (2019)

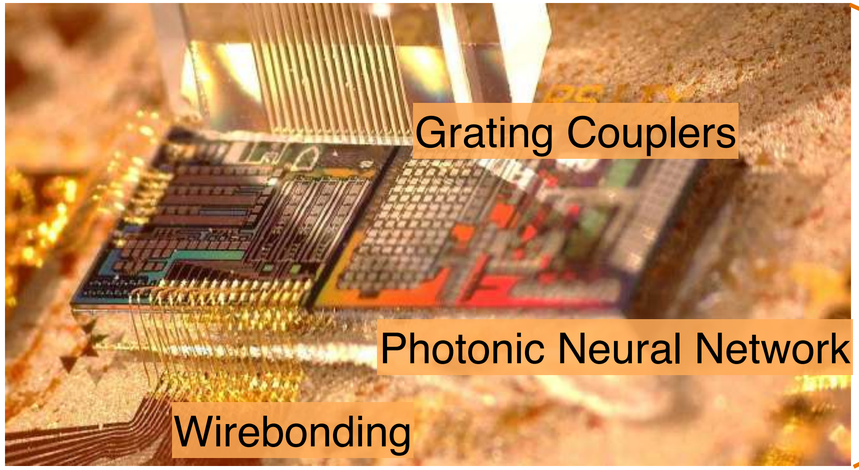
Huang, Shastri, Prucnal et al. *Nature Electronics* 4 (2021)

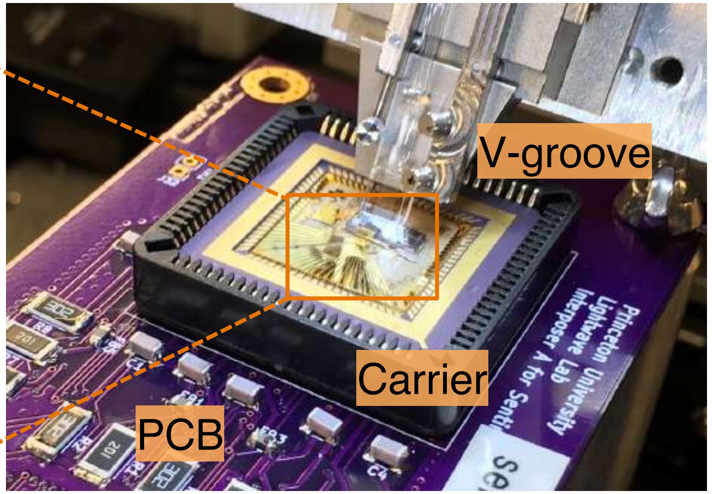
Integrated WDM Silicon Photonic Neural Network Chips





- We have demonstrated complete system integration and applications
- All neuron components are foundry compatible
- Utilizes open-source design tools created by our lab (https://github.com/lightwave-lab)
- Optical and electrical packaging are simple and can be done in-house
- Neural networks can be trained via mature software tools such as TensorFlow
- Experimentation is fully automated





Huang, Shastri, Prucnal et al. *Nature Electronics* 4 (2021)

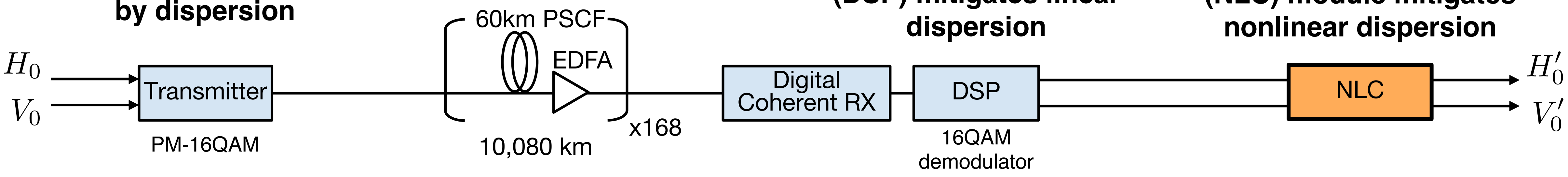
Nonlinear Dispersion Compensation

Transmitter bitrate is limited by dispersion

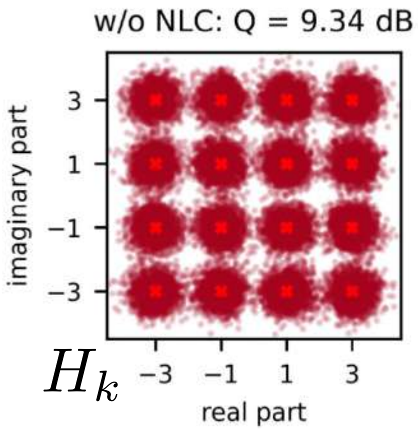
Fiber optic link introduces both linear and nonlinear dispersion

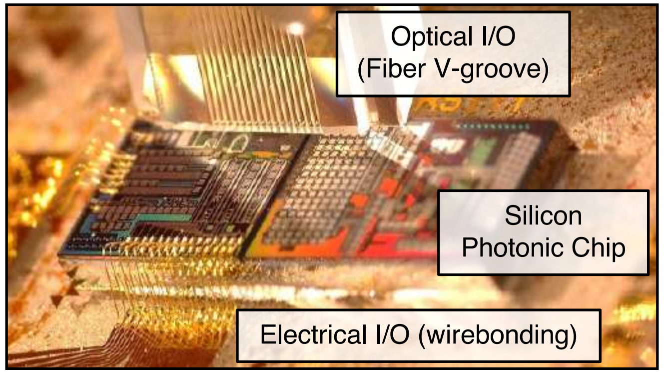
Digital signal processing (DSP) mitigates linear dispersion

Nonlinear compensation (NLC) module mitigates nonlinear dispersion

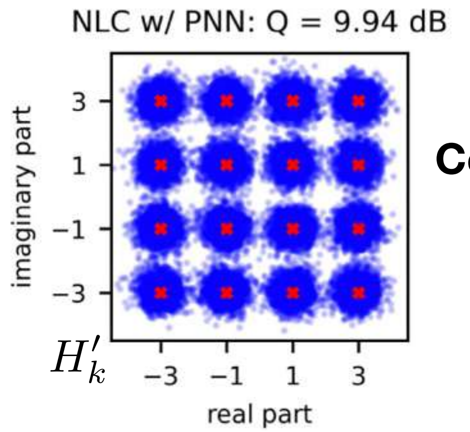


Uncompensated input





Princeton packaged silicon photonic NN



Compensated output

- ❑ **Fully-integrated photonic neural network achieved real-time nonlinearity compensation using captured long-haul fiber-optic communication data**
- ❑ **Simulated NN improvement: 0.65 dB. Real photonic NN improvement: 0.60 dB**

NEC

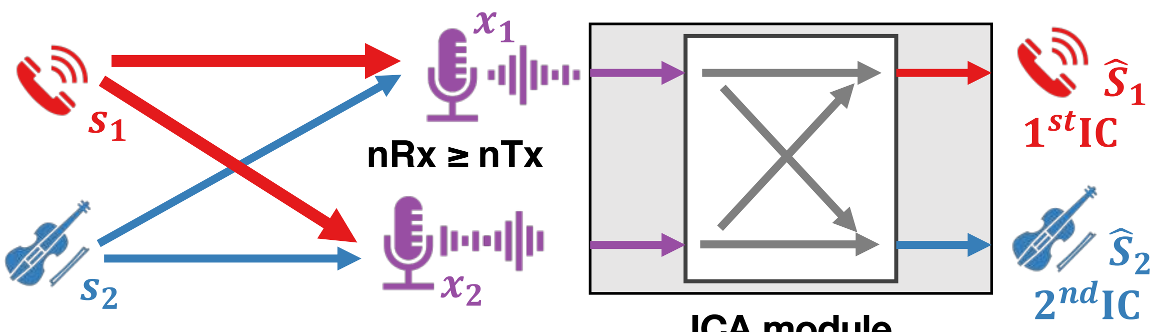
Huang, Shastri, Prucnal et al. *Nature Electronics* 4 (2021)

Blind Source Separation (BSS)

BSS: Recovery of unknown signals from arbitrary mixtures using multiple receivers

- ❑ Employs independent component analysis (ICA), adjusting network weights to maximize kurtosis
  - ❑ Outputs signals that are maximally non-Gaussian
  - ❑ Requires as many (or more) receivers as transmitters
- ❑ Applications:
  - ❑ Image denoising, object detection
  - ❑ RF signal separation, speech isolation

BSS using ICA

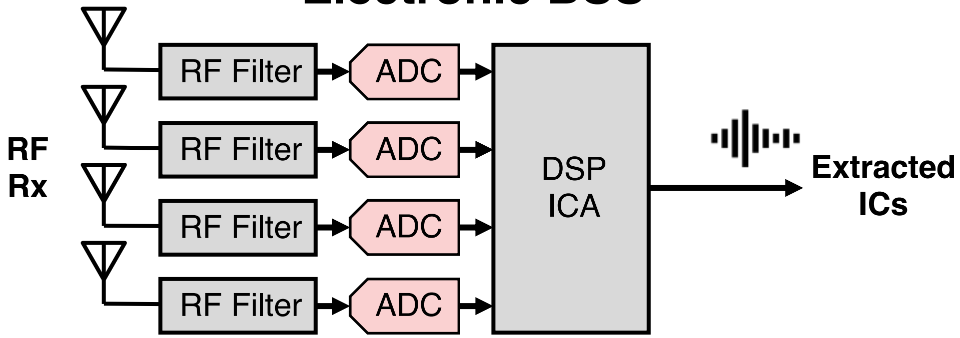


Independent sources (Tx) are linearly mixed on their way to receivers (Rx)

ICA module performs linear demixing

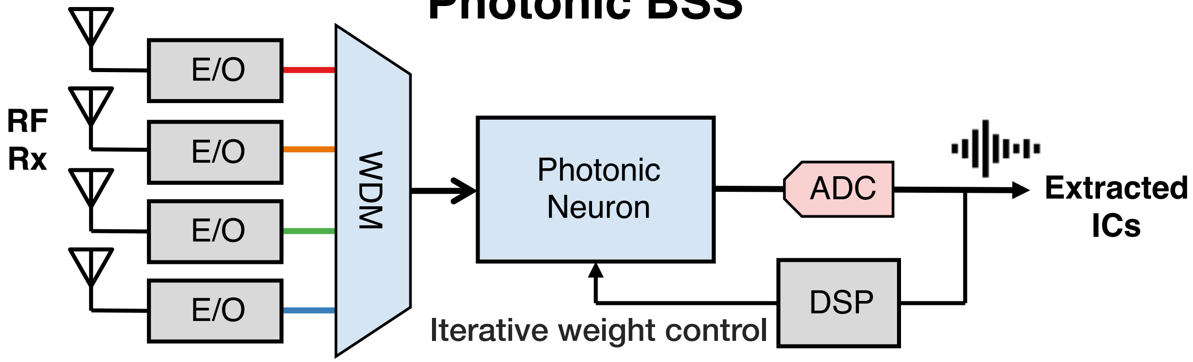
Original sources are recovered

Electronic BSS



RF components are limited in bandwidth, requiring many different receivers to cover the RF spectrum

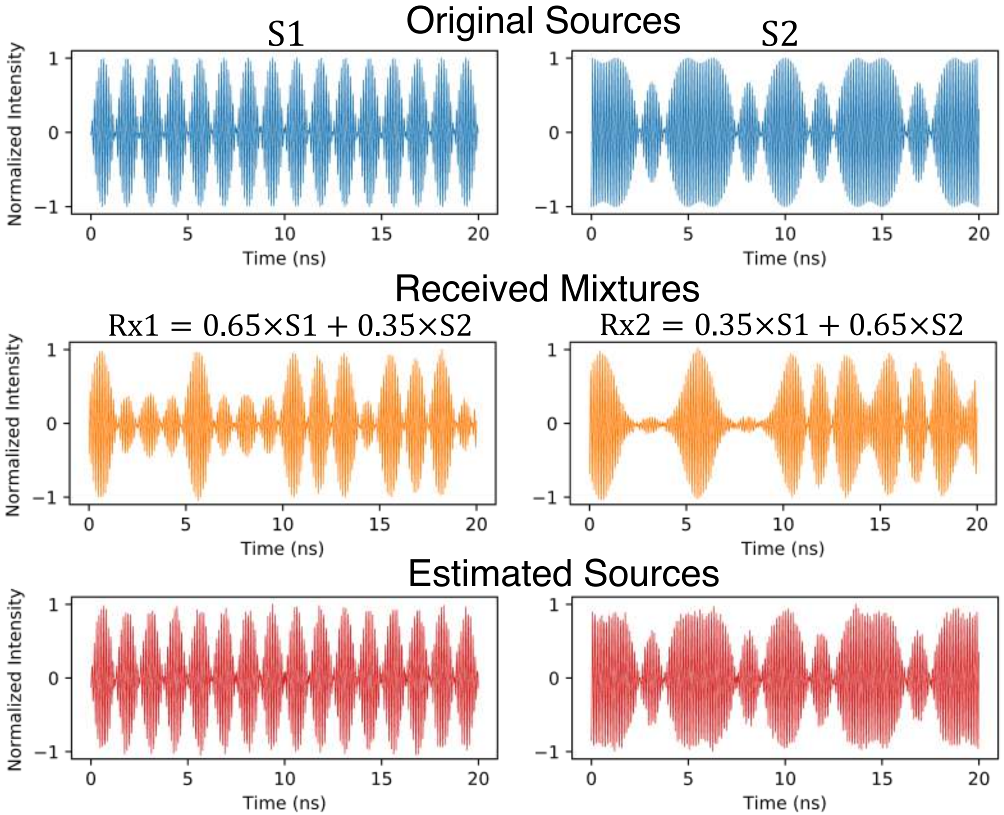
Photonic BSS

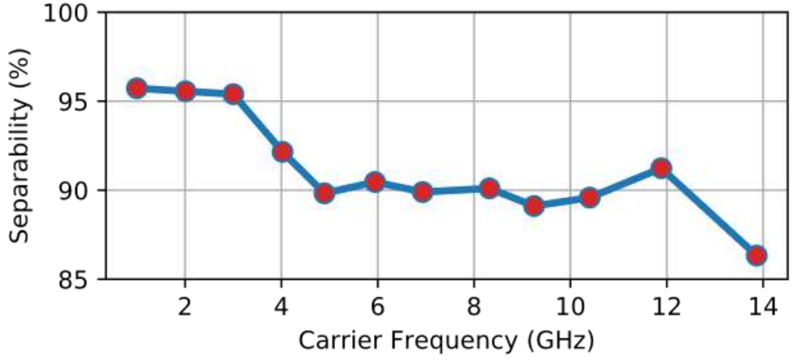


Linear photonic front-end allows ICA to be performed on the entire RF spectrum simultaneously (i.e., collective processing)

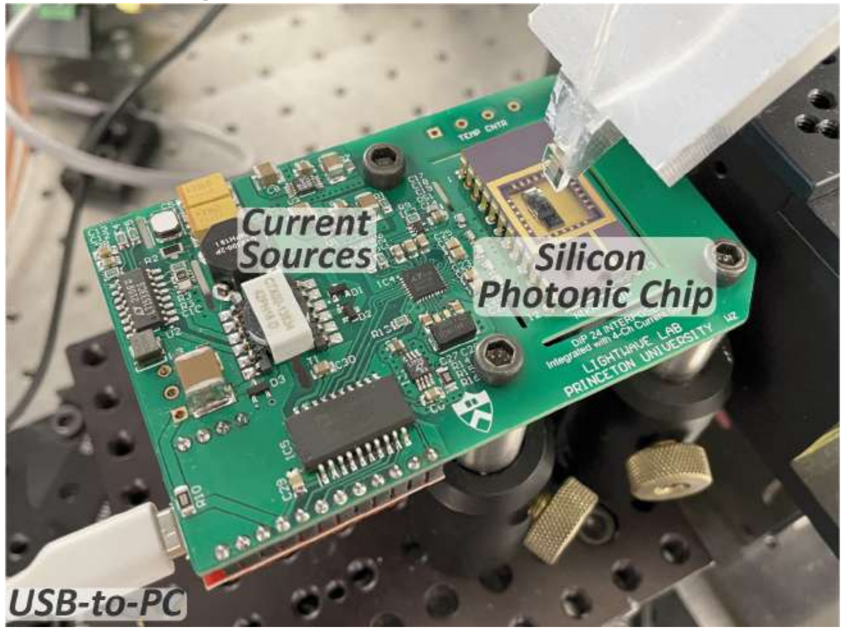
Overview of BSS see: Choi, A. Cichocki, H. M. Park, and S. Y. Lee, Neural Inf. Process. Lett. Rev. 6, 1 (2005).

RF Blind Source Separation using Photonic Neurons





Integrated Photonic Neuron



- ❑ **Fully-integrated neuron linear front end used to demonstrate broadband BSS (ICA) from 1 GHz to 13.8 GHz**
- ❑ **Separability of over 85% achieved across the entire operating bandwidth**

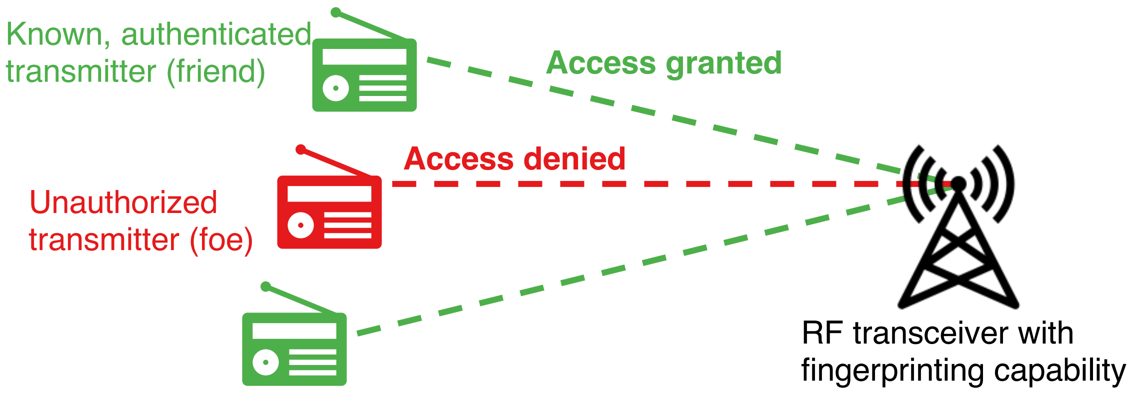
Zhang, Huang, Shastri, Prucnal, et al, arXiv:2104.01164 (2021)

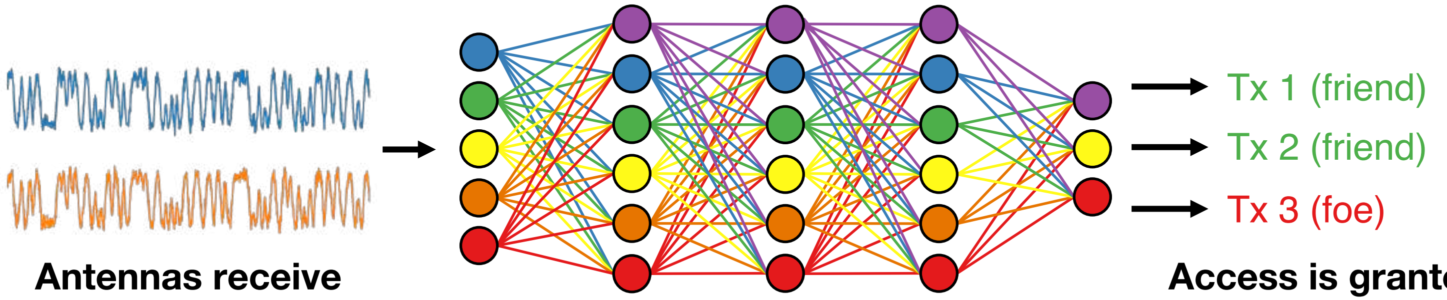
RF Fingerprinting using Neural Networks

RF fingerprinting discriminates between nominally identical RF transmitters using their nearly unidentifiable idiosyncrasies

- ❑ Exploits minor physical variances from manufacturing and environmental effects
- ❑ Capable of distinguishing nominally “identical” transmitters (i.e., same make and model)

Use Case: Identify Friend or Foe





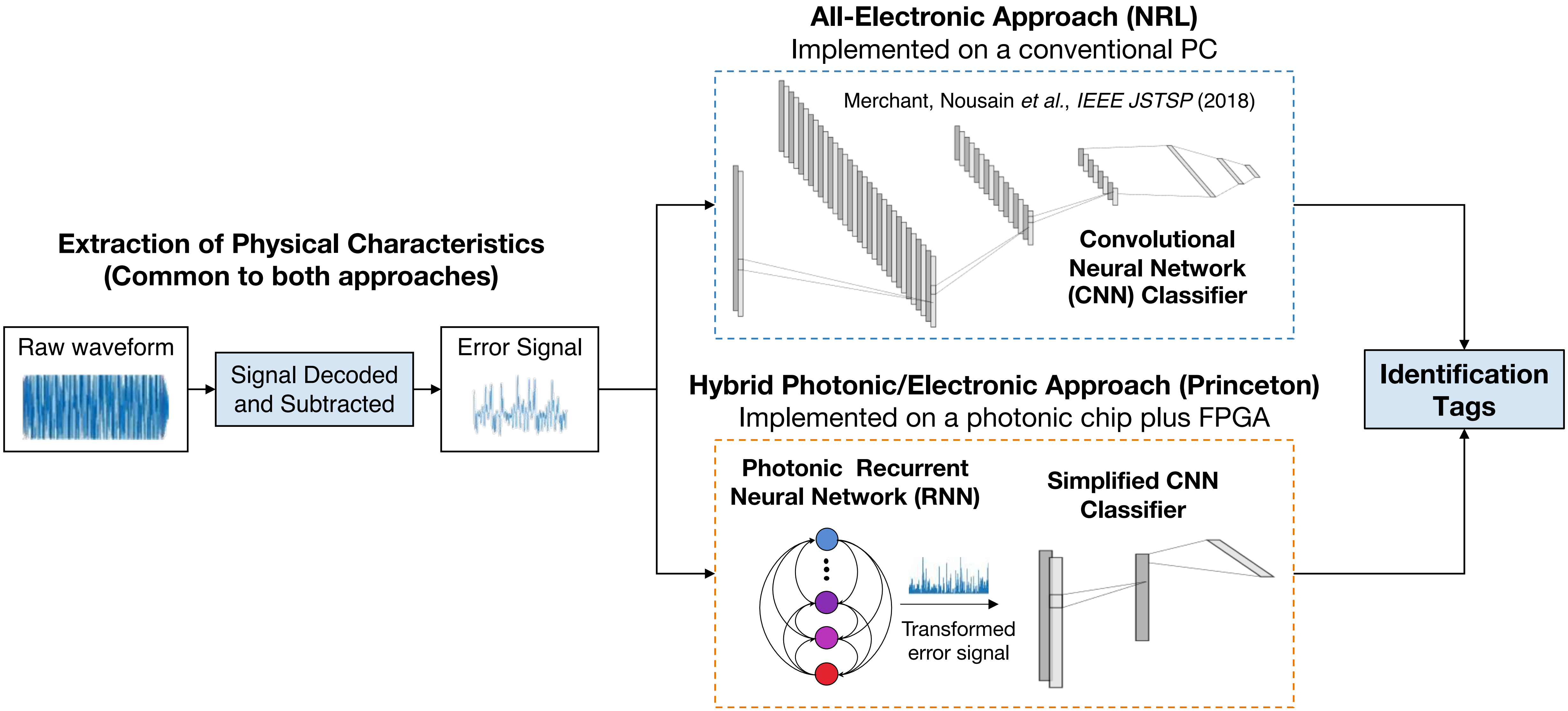
Antennas receive RF waveforms

Network is trained to recognize individual transmitters from a set of nominally identical transmissions

Access is granted or denied based on transmitter identity

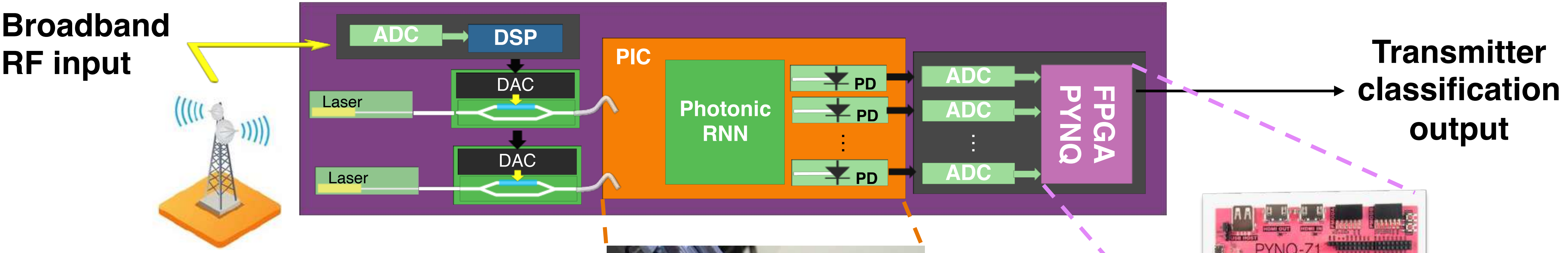
The photonic neural network can enable real-time transmitter identification across 10’s of GHZ of RF spectrum

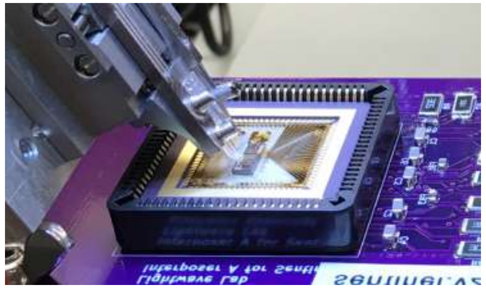
RF Fingerprinting: All-Electronic vs. Hybrid Approach



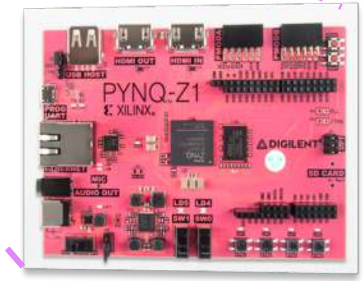
Peng, Shastri, Prucnal, et al, arXiv:2106.13865 (2021)

RF Fingerprinting Hybrid Architecture





Photonic RNN at frontend processes GHz RF signals in real time



Reduced CNN is implemented in a small form-factor FPGA: Xilinx Pynq-Z1

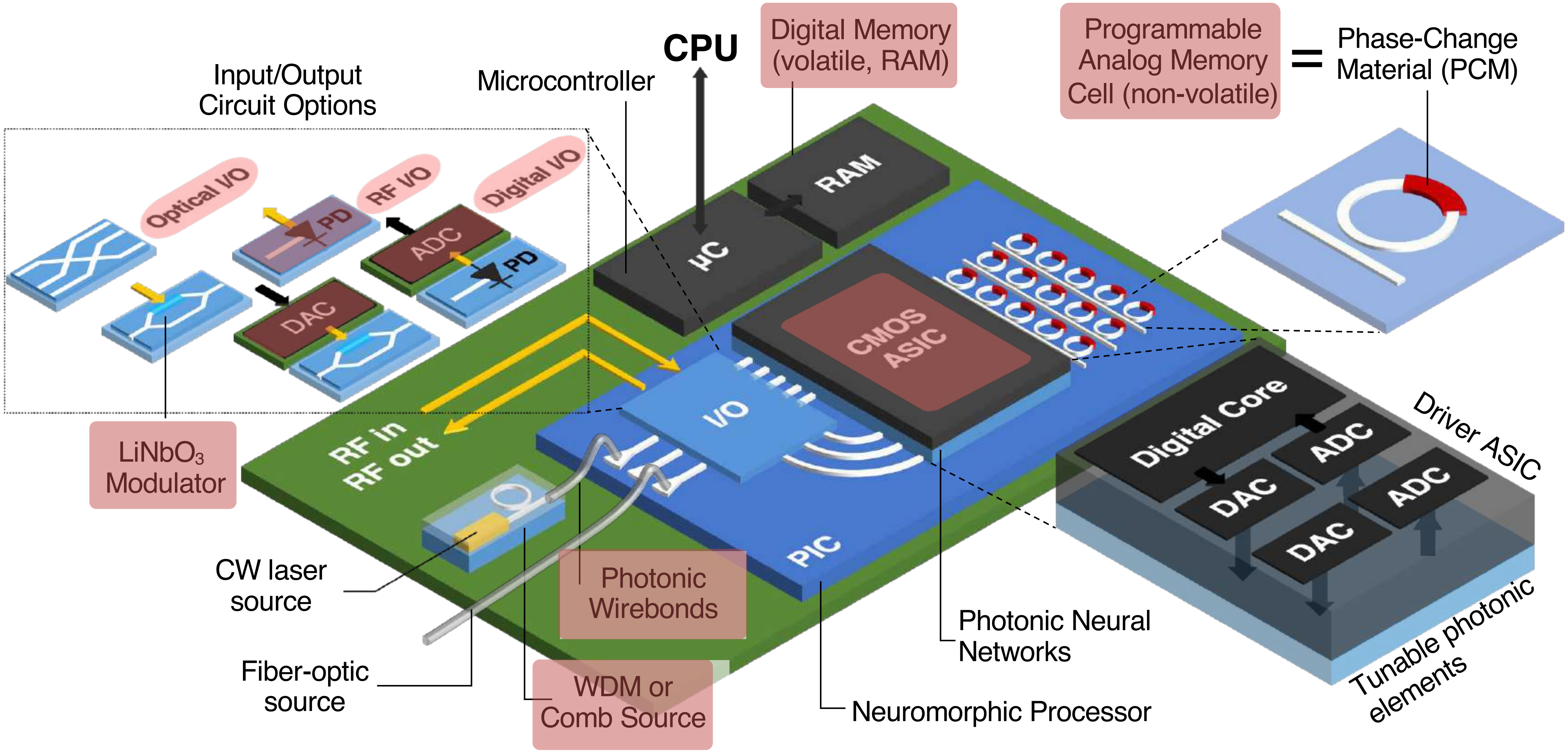
Performance Summary

Approach	Estimated Energy per classification (µJ)	Latency (ms)	Throughput (Classifications/sec)
NRL Convolutional NN implemented on Larger FPGA	6,194	26.2	50
Princeton/Queen's Hybrid Approach	15.11	0.154	12,190
Improvement factor	**410x**	**169x**	**244x**

- ❑ Classifies 30 identical transmitters in real time with over 96% accuracy
- ❑ **Simultaneous improvement of power, latency, throughput, and SWaP over purely electronic approach**

Peng, Shastri, Prucnal, et al, arXiv:2106.13865 (2021)

Roadmap: Emerging Ideas and Challenges



Shastri, Tait, Prucnal et al. *Nature Photonics* **15** (2021)

REVIEW ARTICLE | FOCUS
https://doi.org/10.1038/s41566-020-00754-y
nature photonics

Check for updates

Photonics for artificial intelligence and neuromorphic computing

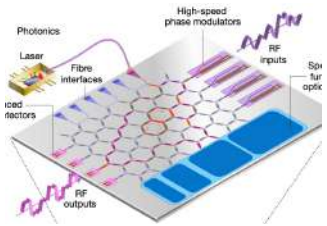
Bhavin J. Shastri[1,2,7] ✉, Alexander N. Tait[2,3,7] ✉, T. Ferreira de Lima[2], Wolfram H. P. Pernice[4], Harish Bhaskaran[5], C. D. Wright[6] and Paul R. Prucnal[2]

Research in photonic computing has flourished due to the proliferation of optoelectronic components on photonic integration platforms. Photonic integrated circuits have enabled ultrafast artificial neural networks, providing a framework for a new class of information processing machines. Algorithms running on such hardware have the potential to address the growing demand for machine learning and artificial intelligence in areas such as medical diagnosis, telecommunications, and high-performance and scientific computing. In parallel, the development of neuromorphic electronics has highlighted challenges in that domain, particularly related to processor latency. Neuromorphic photonics offers sub-nanosecond latencies, providing a complementary opportunity to extend the domain of artificial intelligence. Here, we review recent advances in integrated photonic neuromorphic systems, discuss current and future challenges, and outline the advances in science and technology needed to meet those challenges.

Shastri, Tait, Prucnal et al. *Nature Photonics* **15** (2021)

Optics & Computing: A 2022 Perspective

Programmable photonics



Bogaerts et al. *Nature* (2020)

Parallel processing, passive waveguides, matrix multipliers

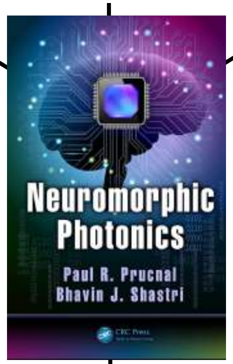
Analog Photonics

Silicon Industry

Silicon photonics, electronics + photonics

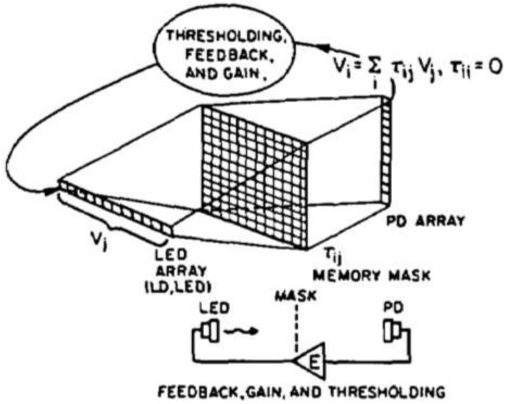
Neuromorphic Photonics

Shastri, *Nat. Photon.* (2021)



Free-space Linear NNs

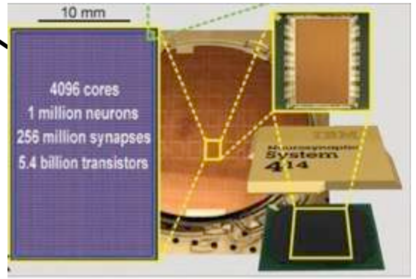
Challenge: Integration, nonlinearity



Psaltis & Farhat, *Opt. Lett.* (1985)

Neuromorphic Electronics

Challenge: bandwidth



Merolla et al. *Science* (2014)

Scalable Computing Models

Deep learning, spiking, equilibrium propagation

Prucnal & Shastri, *Neuromorphic Photonics* CRC Press (2017) *fin*

shastri@ieee.org