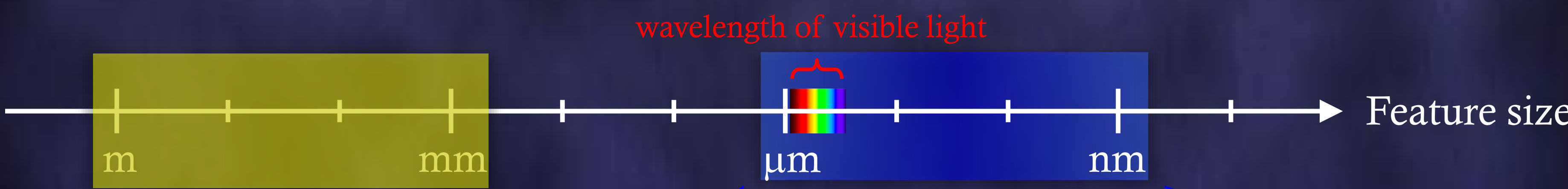
Photonics for AI

Dr. Yichen Shen
Presentation @ MIT
Apr 26th, 2018

Nanophotonics

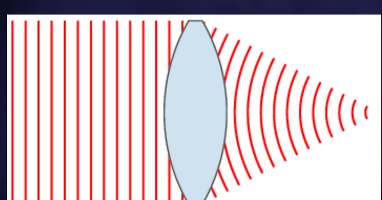
Optical structures (dielectrics) with nanometer-scale features



Macroscopic structures
(Feature size >> Wavelength)

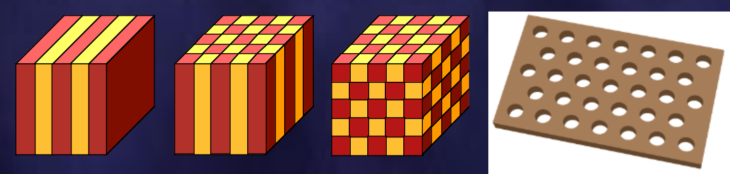
Nanophotonic structures
(Feature size <≈ Wavelength)

Ray optics

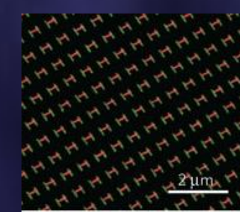


Limited ways to manipulate light

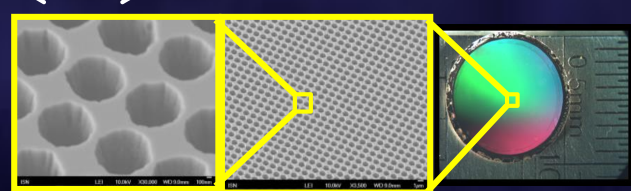
Photonic crystals



Metamaterials



500 nm



Integrated Photonics

Thanks to improved computation power & fabrication techniques

Artificial Neural Networks (ANN)

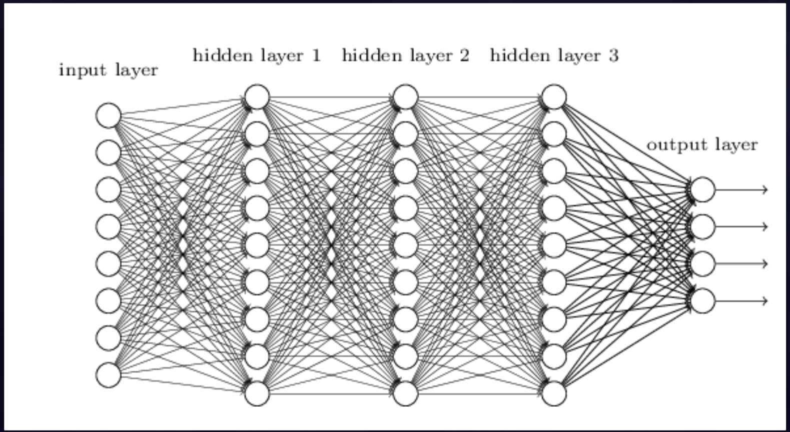
<u>Breakthroughs in deep learning</u>:

- Natural Language Processing (NLP)
- Game Playing (Go, Atari)
- Autonomous Vehicles
- Control
- Ad Placement
- Researches (drug discovery, material study)
- Etc.

Google translate

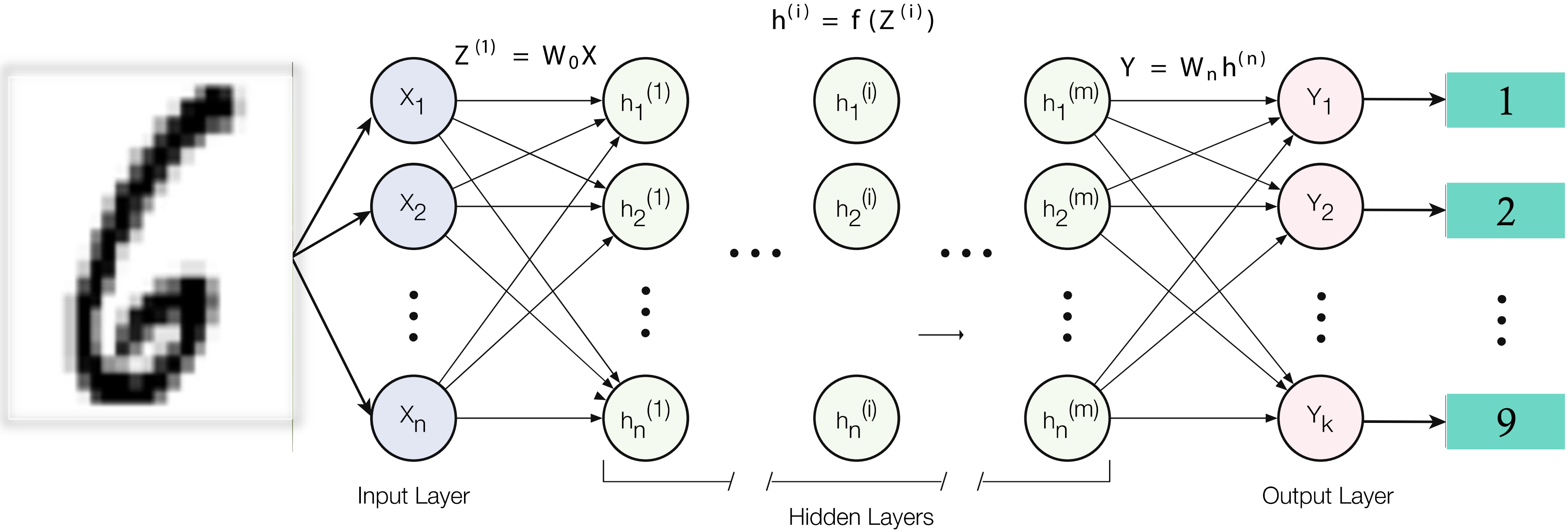
Neuromorphic Computing

Biological Neural Networks



Artificial Neural Networks

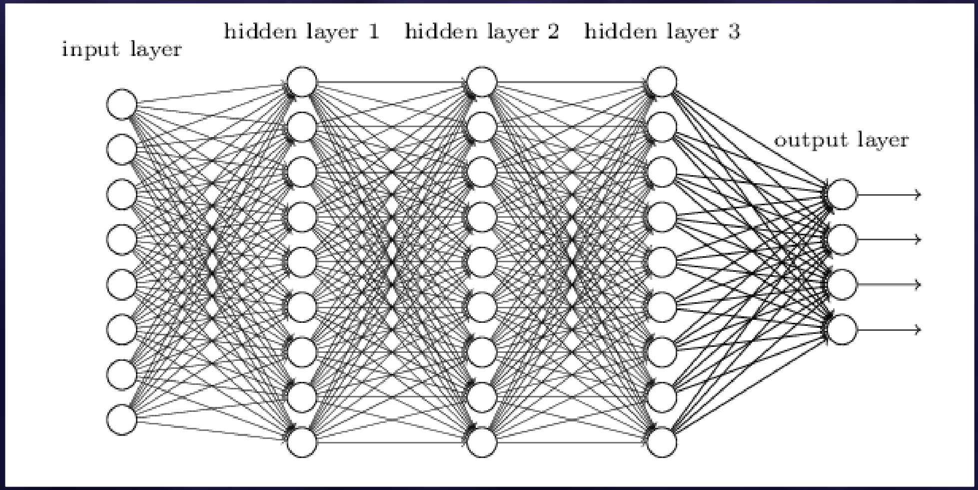
Basic Algorithm of ANN



Matrix Multiplication: $z_j^{(1)} = \sum_{i=1}^{n} w_{ji}^{(1)} x_i$ $\quad z_j^{(k)} = \sum_{i=1}^{n} w_{ji}^{(k)} h_i^{(k)}$

Nonlinear Activation: $h_j^{(1)} = f(z_j^{(1)})$ $\quad h_j^{(k)} = f(z_j^{(k)})$

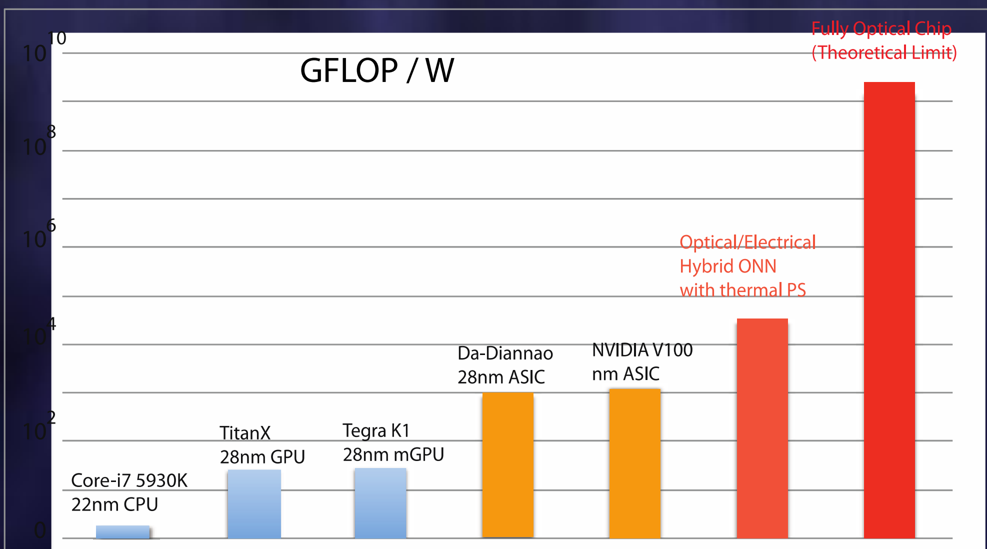
Hardware and Data Enable Deep Learning



The Need for Speed

More Data → Bigger Models → More need for Computation

But Moore's Law is no-longer providing more computation…



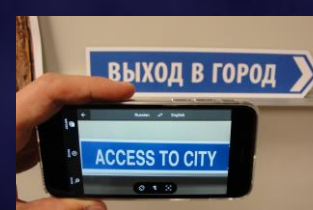
Von Neumann

ASIC/FPGA

Optical Processing

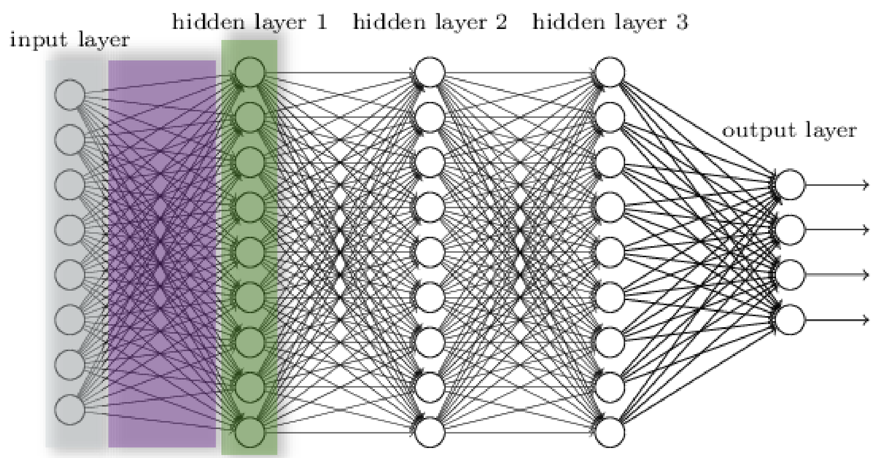
The Market:

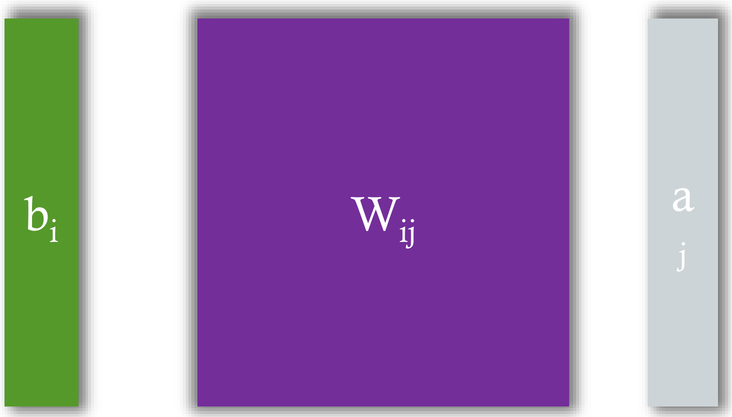
On clouds:
Millions of high power AI processors ($10,000 each) in data centers by 2020



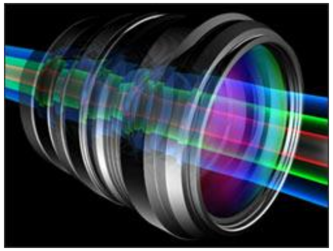
On premise:
Billions of compact AI processors needed due to the rise of autonomouse driving, AR and IoT.

In Deep Learning
Key Operation is dense M x V

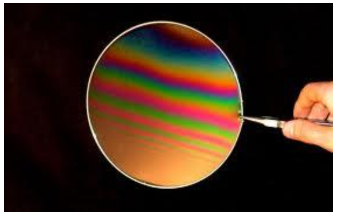


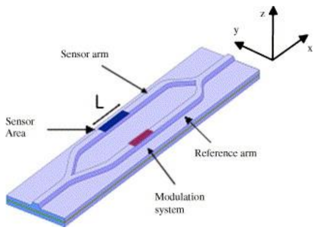


In Optics, Matrix Multiplication is very common & (usually) consumes no energy !



Convolution / FFT





Matrix Multiplication

ANN does NOT require high resolution

Category	Method	Weights (# of bits)	Activations (# of bits)	Accuracy Loss vs. 32-bit float (%)
Dynamic Fixed Point	w/o fine-tuning	8	10	0.4
	w/ fine-tuning	8	8	0.6
Reduce weight	Ternary weights Networks (TWN)	2*	32	3.7
	Trained Ternary Quantization (TTQ)	2*	32	0.6
	Binary Connect (BC)	1	32	19.2
	Binary Weight Net (BWN)	1*	32	0.8
Reduce weight and activation	Binarized Neural Net (BNN)	1	1	29.8
	XNOR-Net	1*	1	11
Non-Linear	LogNet	5(conv), 4(fc)	4	3.2
	Weight Sharing	8(conv), 4(fc)	16	0

* first and last layers are 32-bit float

Sze et al, arXiv:1703.09039 (2017)

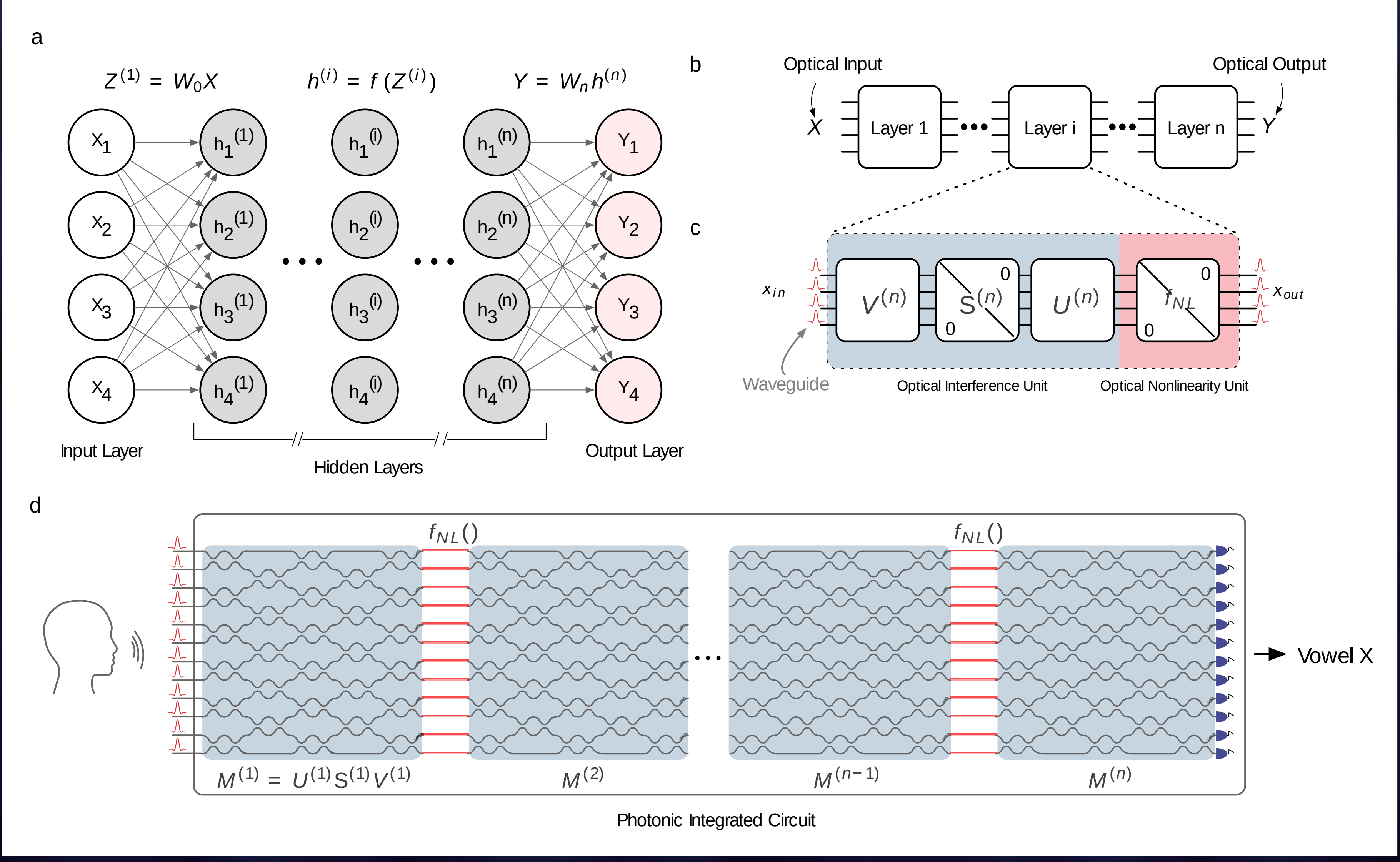
Deep Learning Inference is "Passive"

Once the Optical Neural Network is trained, no need to update the weights frequently…

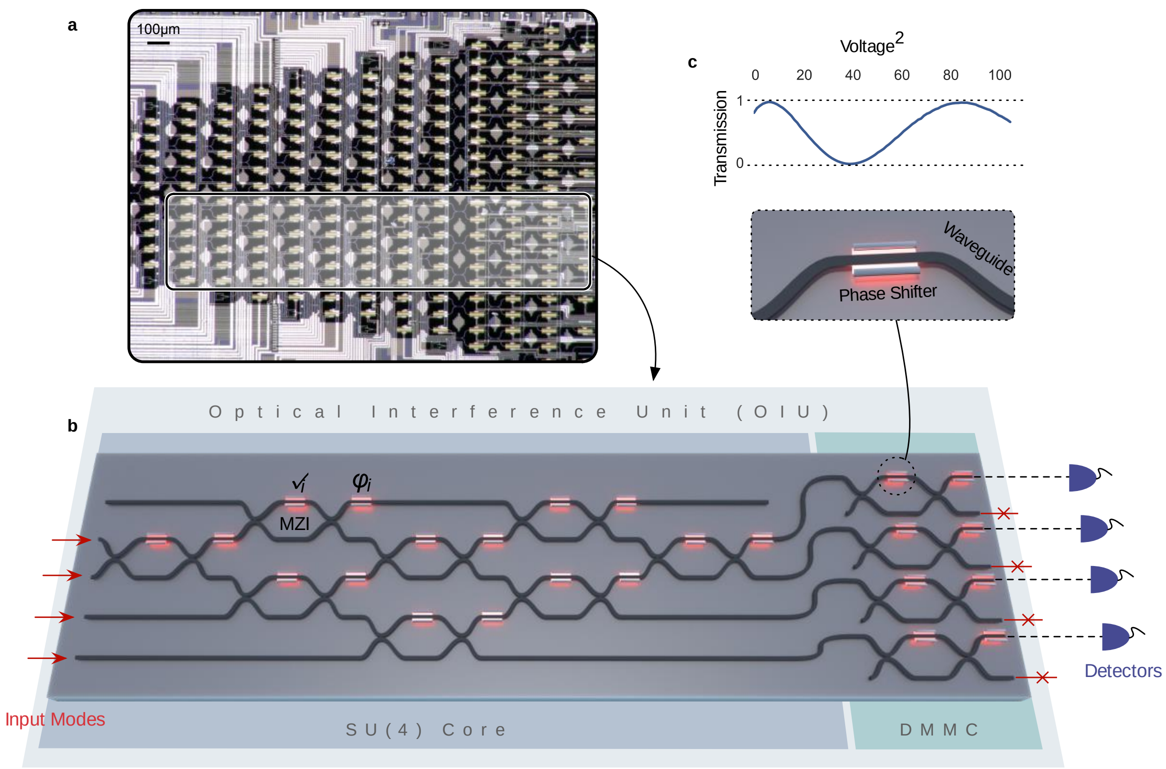
Deep Learning is very parallelizable

Multiple wavelengths can be used to simultaneously execute batch of data

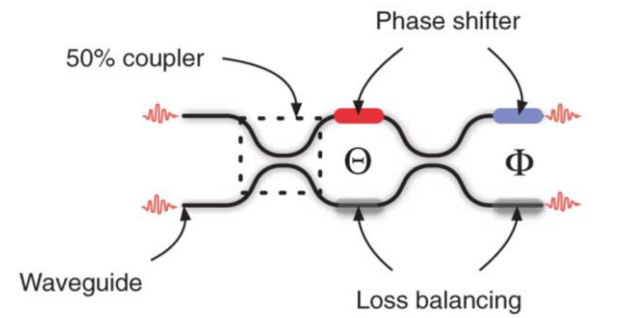
Coherent Optical Neural Networks (ONN)

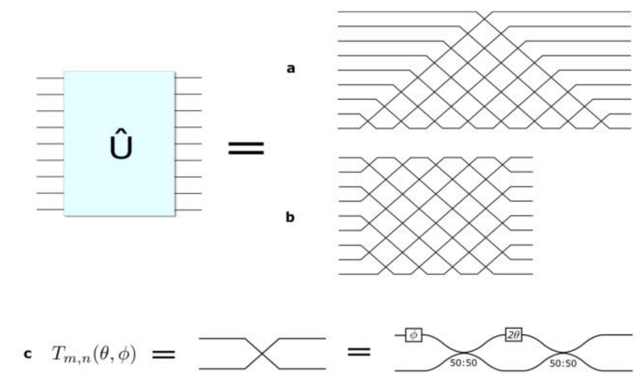


Programmable Nanophotonic Processors



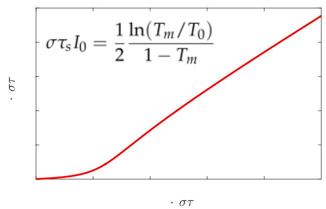
$$\begin{pmatrix} e^{i\phi}\cos\theta & -e^{i\phi}\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

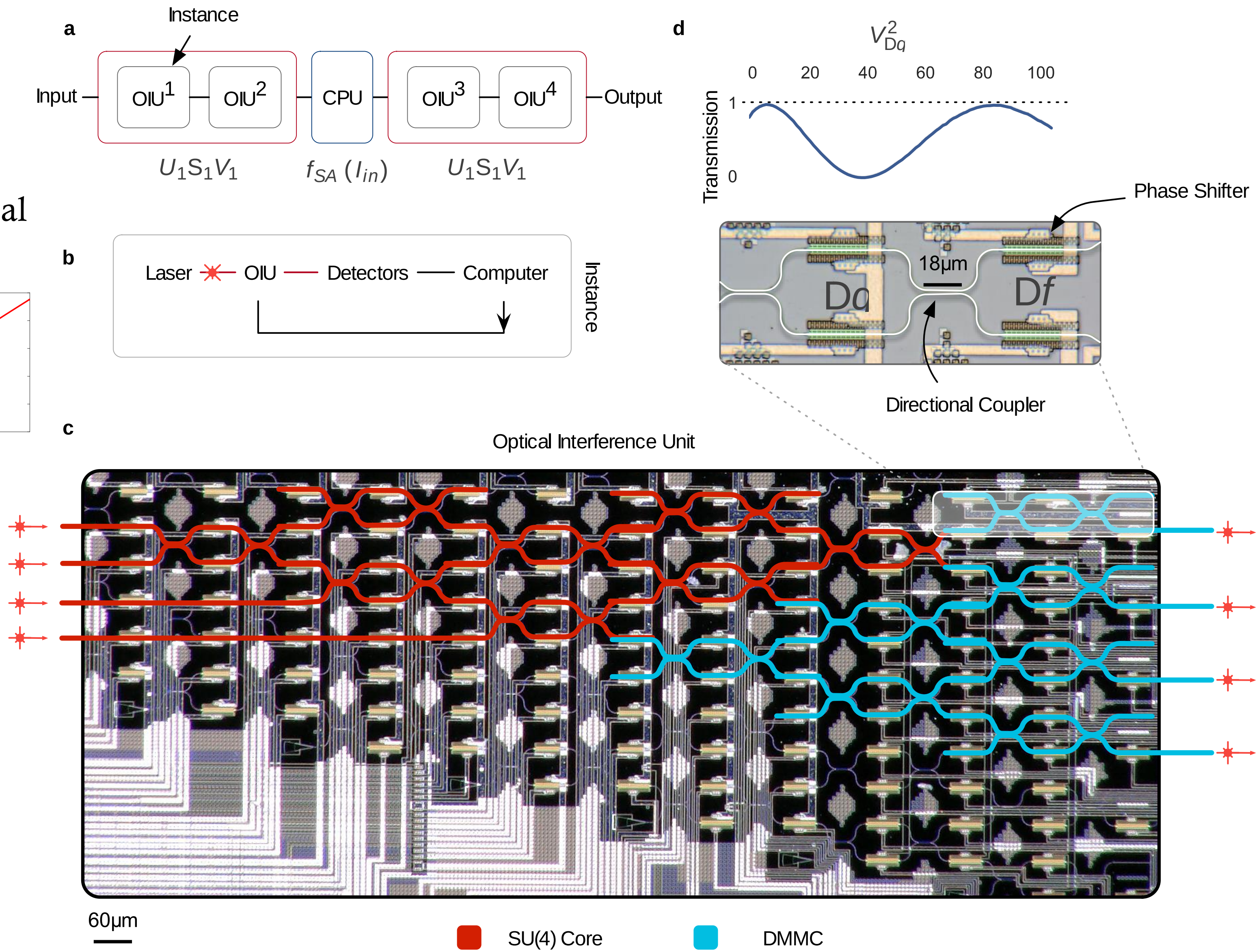




Y.Shen and N. Harris et al. *"Deep Learning with Coherent Nanophotonic Circuit" Nature Photonics* **11**, 441–446 (2017)

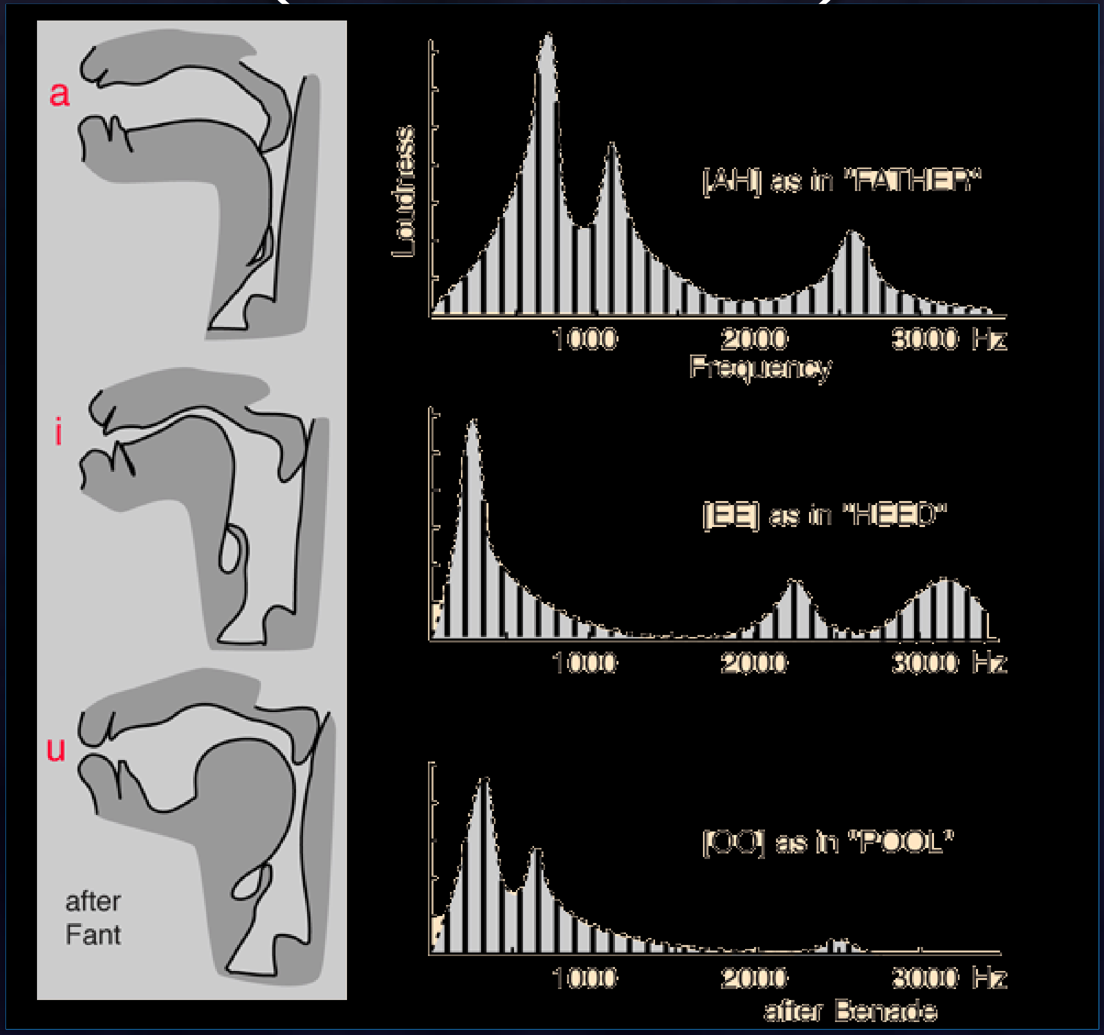
Simulated Optical Nonlinearity



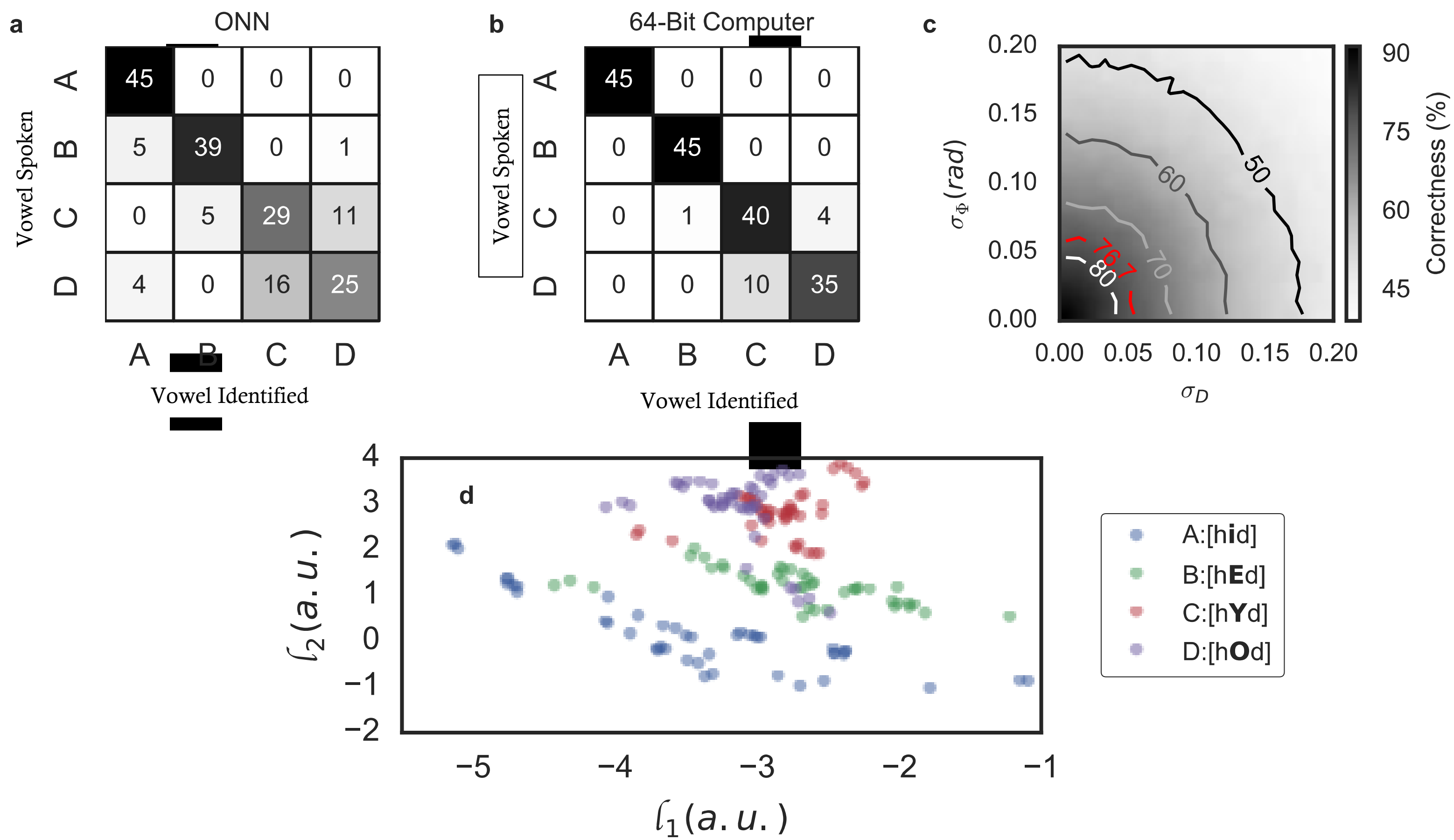


Y.Shen and N. Harris et al. *"Deep Learning with Coherent Nanophotonic Circuit" Nature Photonics* **11**, 441–446 (2017)

Optical Vowel Recognition (4d 4 classes)



Experimental Result



Simulation Result: 165/180=91.7%

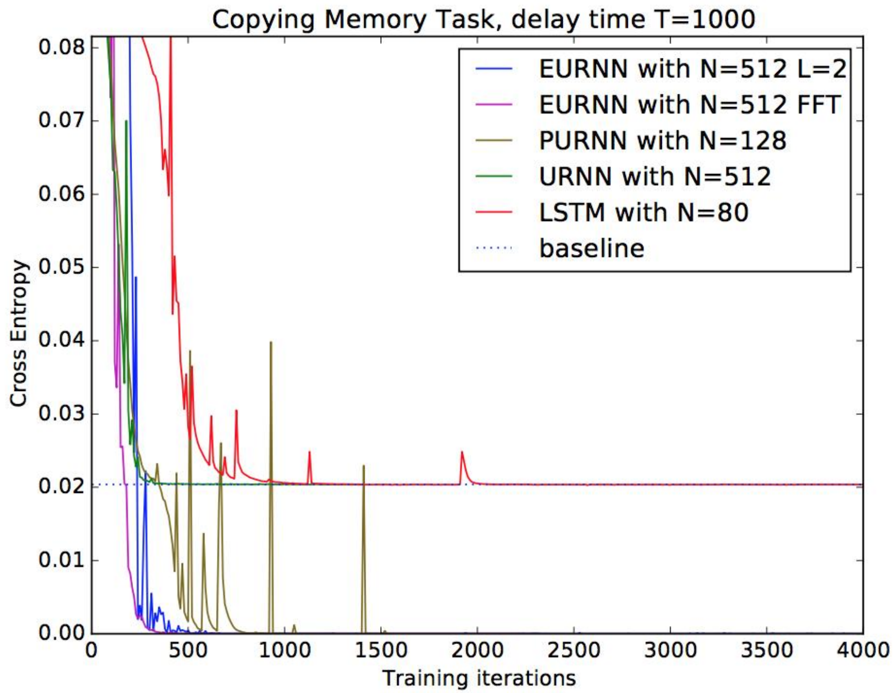
Experiment Result: 138/180=76.7%

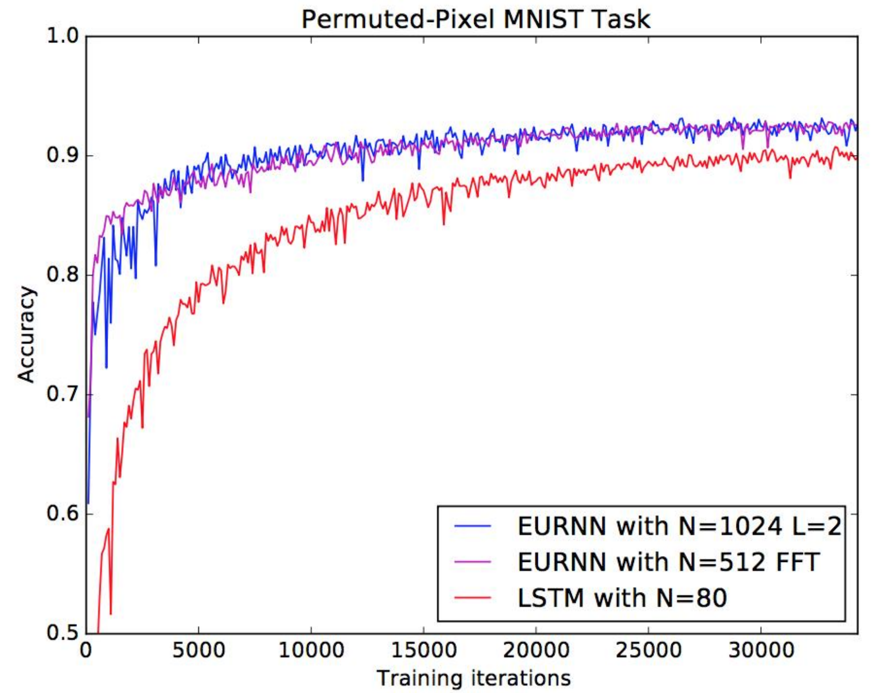
The other side of the Story…

- Immature photonics eco-system (low yield, high cost)
- Large device size
- Non-ideal PDK component design (lossy, low resolution, power hungry)
- AD/DA interface

Software & Hardware

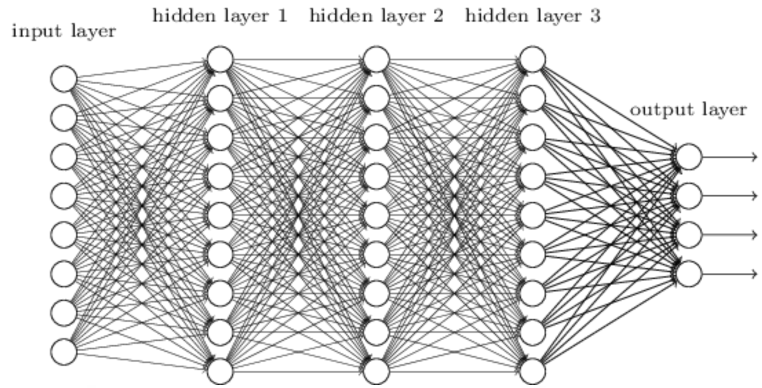
AI algorithms DESIGNED to be run on photonics chip

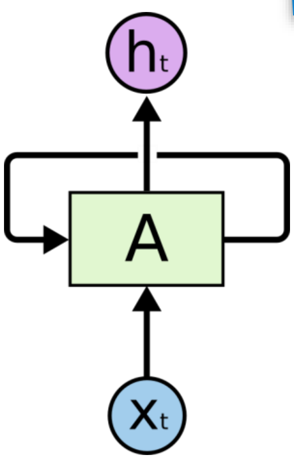




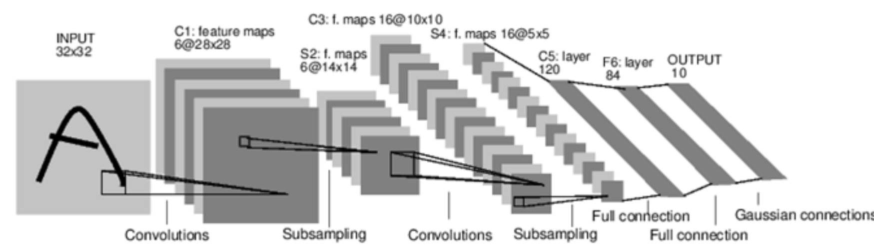
L. Jing & Y. Shen et al, International Conference for Machine Learning (ICML 2017)

Fully Connected Neural Networks





Recurrent Neural Networks

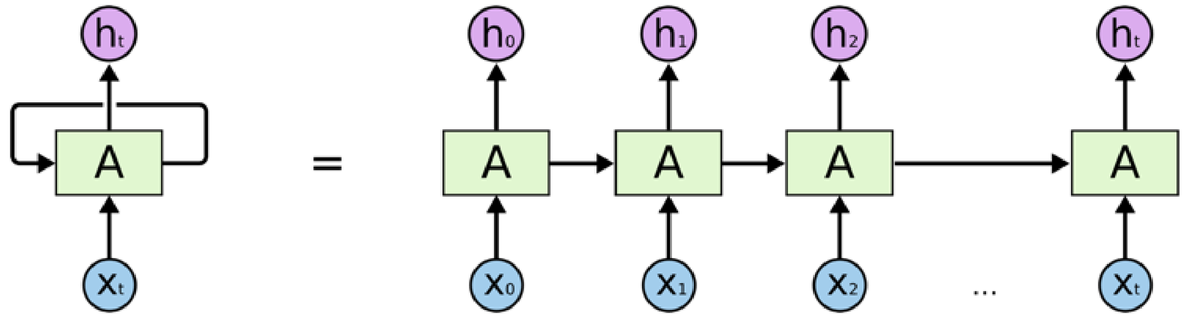


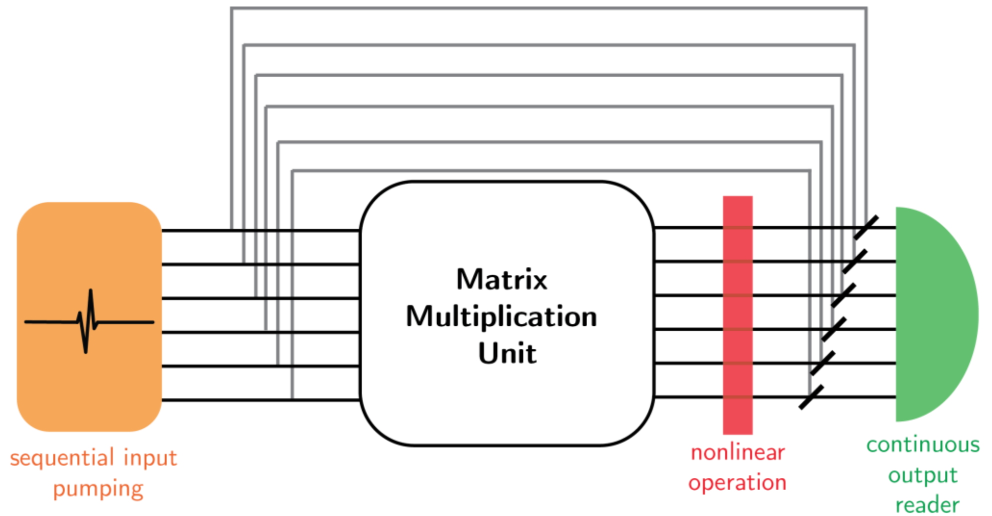
A Full Convolutional Neural Network (LeNet)

Convolutional Neural Networks

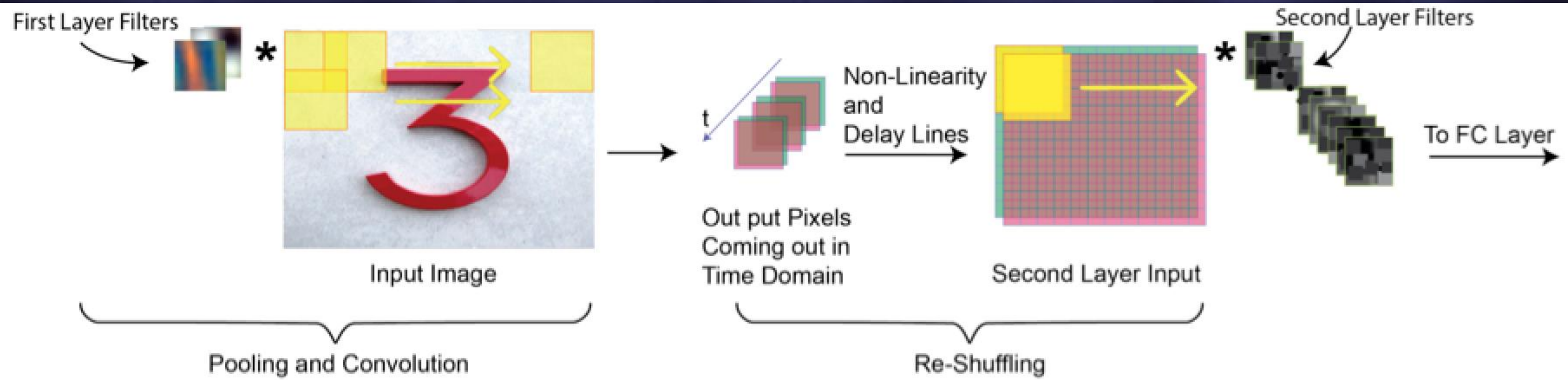
Recurrent Neural Networks

Commonly used for Speech Recognition and Language Processing





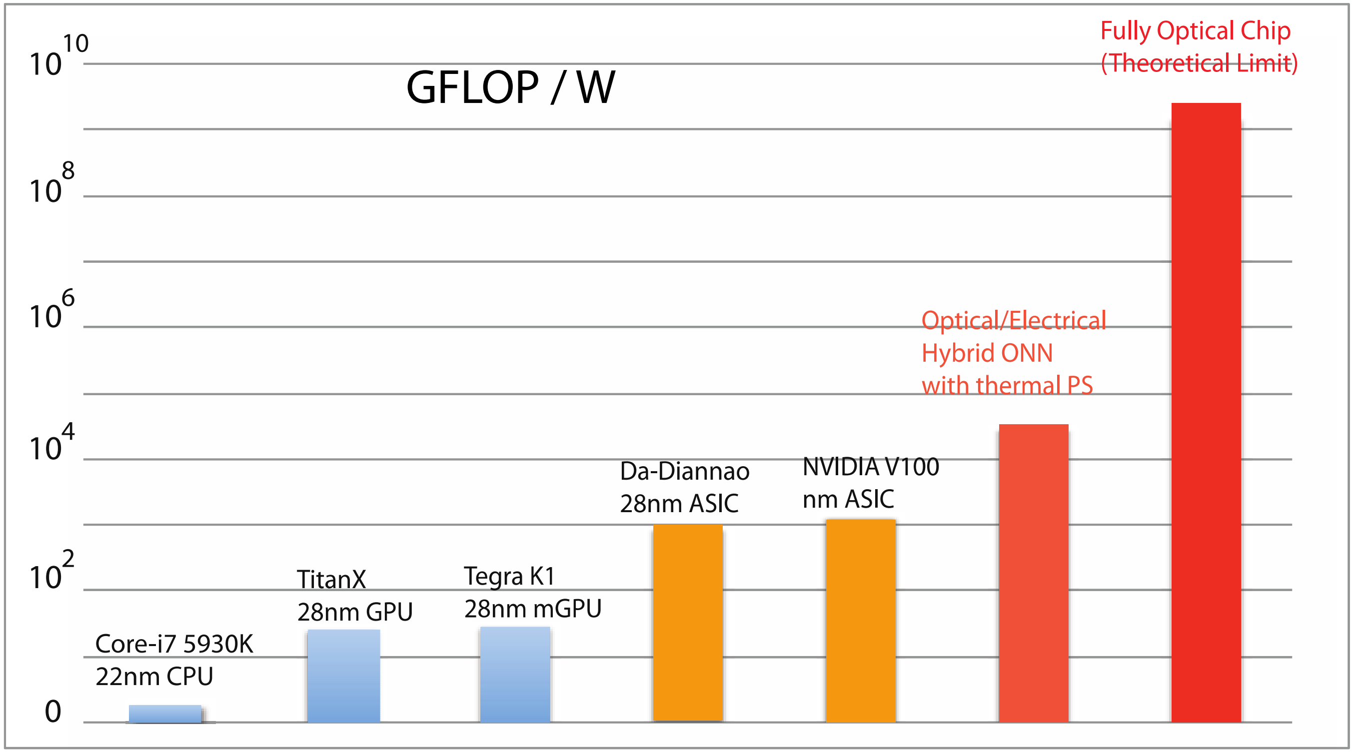
Convolution Neural Networks



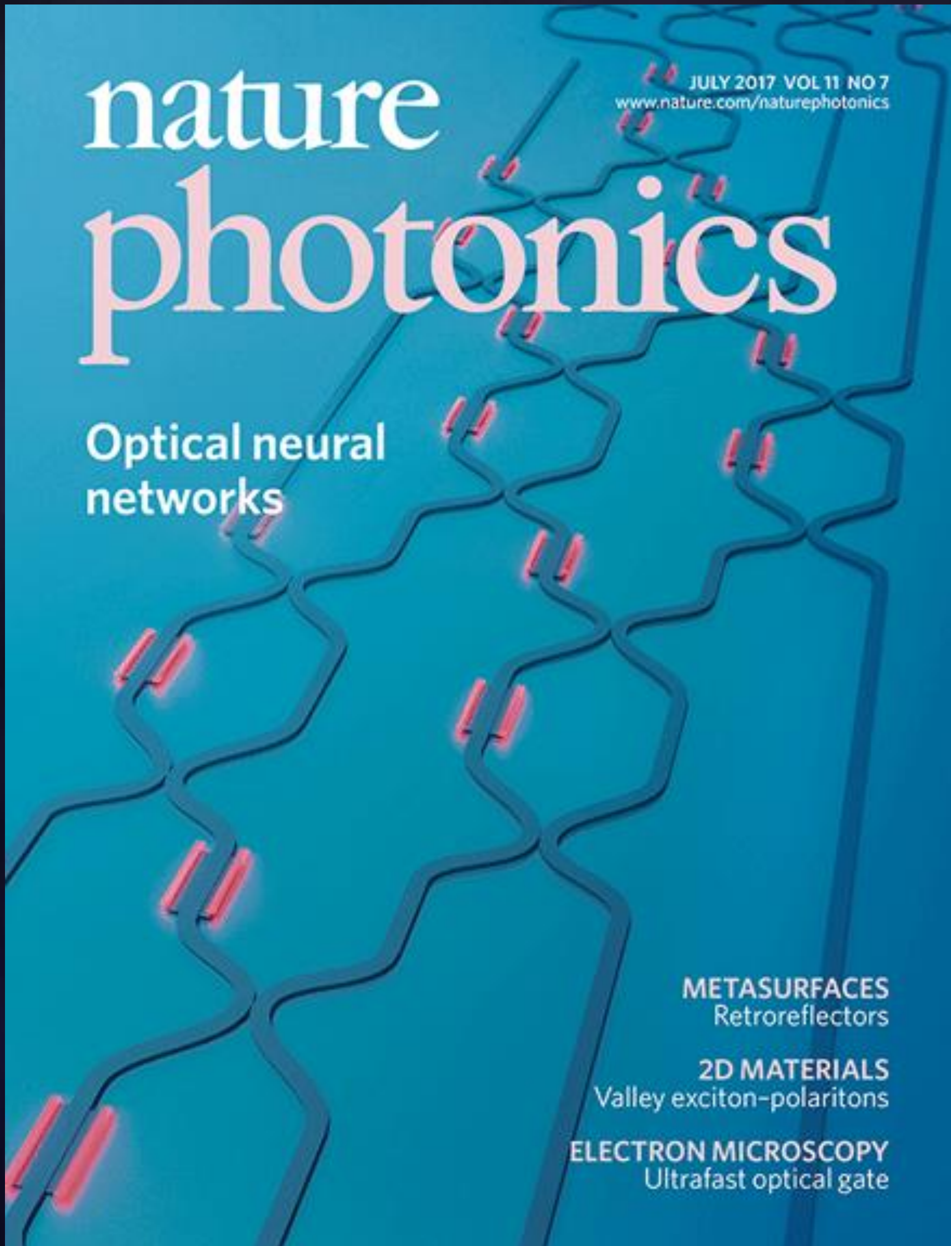
Scott Skirlo and Yichen Shen et al, Manuscript in Preparation

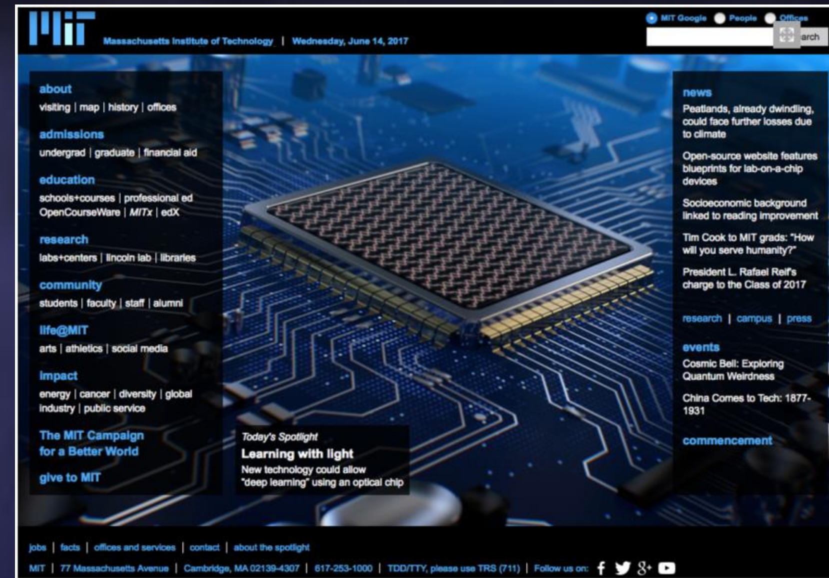
Speed and Energy Efficiency Comparison with Electrical ANN

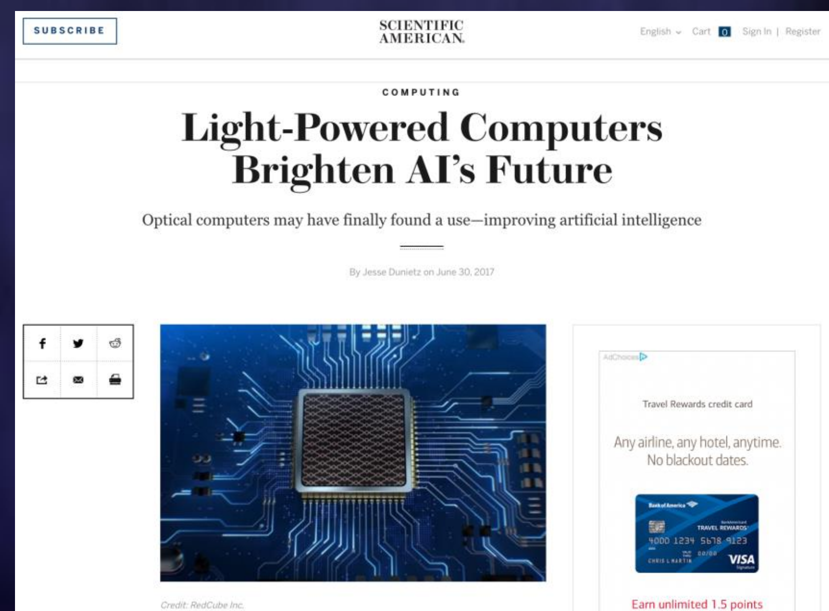
	NVIDIA TITAN X	ONN (with thermal PS)
Architecture	Von Neumann	Neuromorphic
Power Consumption	1 kW	1-2 kW
Operation Speed	10 TFLOP	10,000 TFLOP



Y. Shen and N. Harris et al, Nature Photonics 11, 441 (2017)


nature photonics
JULY 2017 VOL 11 NO 7
www.nature.com/naturephotonics
Optical neural networks
METASURFACES
Retroreflectors
2D MATERIALS
Valley exciton-polaritons
ELECTRON MICROSCOPY
Ultrafast optical gate


Massachusetts Institute of Technology | Wednesday, June 14, 2017
about
visiting | map | history | offices
admissions
undergrad | graduate | financial aid
education
schools+courses | professional ed
OpenCourseWare | MITx | edX
research
labs+centers | lincoln lab | libraries
community
students | faculty | staff | alumni
life@MIT
arts | athletics | social media
impact
energy | cancer | diversity | global
industry | public service
The MIT Campaign for a Better World
give to MIT
Today's Spotlight
Learning with light
New technology could allow "deep learning" using an optical chip
news
Peatlands, already dwindling, could face further losses due to climate
Open-source website features blueprints for lab-on-a-chip devices
Socioeconomic background linked to reading improvement
Tim Cook to MIT grads: "How will you serve humanity?"
President L. Rafael Reif's charge to the Class of 2017
research | campus | press
events
Cosmic Bell: Exploring Quantum Weirdness
China Comes to Tech: 1877-1931
commencement
jobs | facts | offices and services | contact | about the spotlight
MIT | 77 Massachusetts Avenue | Cambridge, MA 02139-4307 | 617-253-1000 | TDD/TTY, please use TRS (711) | Follow us on:


SUBSCRIBE
SCIENTIFIC AMERICAN
English Cart Sign In | Register
COMPUTING
Light-Powered Computers Brighten AI's Future
Optical computers may have finally found a use—improving artificial intelligence
By Jesse Dunietz on June 30, 2017
Travel Rewards credit card
Any airline, any hotel, anytime. No blackout dates.
Earn unlimited 1.5 points
Credit: RedCube Inc.

Optical AI Computing

The New York Times

Science

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

ENVIRONMENT SPACE & COSMOS

LIGHT MAY BE KEY TO NEW GENERATION OF FAST COMPUTERS

By WILLIAM J. BROAD
Published: October 22, 1985

SINCE its start nearly half a century ago, the computer revolution has advanced by virtue of a simple physical phenomenon: that streams of speeding electrons can start or stop the flow of other streams of electrons. In short, electrons can act as a switch.

FACEBOOK
TWITTER
GOOGLE+
EMAIL

Some History on Optical Neural Networks

2005

> *"The biggest issue with this paper is that it relies on neural networks."*
>
> Anonymous Reviewer

2016

Springtime for AI: The Rise of Deep Learning

After decades of disappointment, artificial intelligence is finally catching up to its early promise, thanks to a powerful technique called deep learning

By Yoshua Bengio on June 1, 2016

SCIENTIFIC AMERICAN