nature photonics

ARTICLES

PUBLISHED ONLINE: 12 JUNE 2017 | DOI: 10.1038/NPHOTON.2017.93

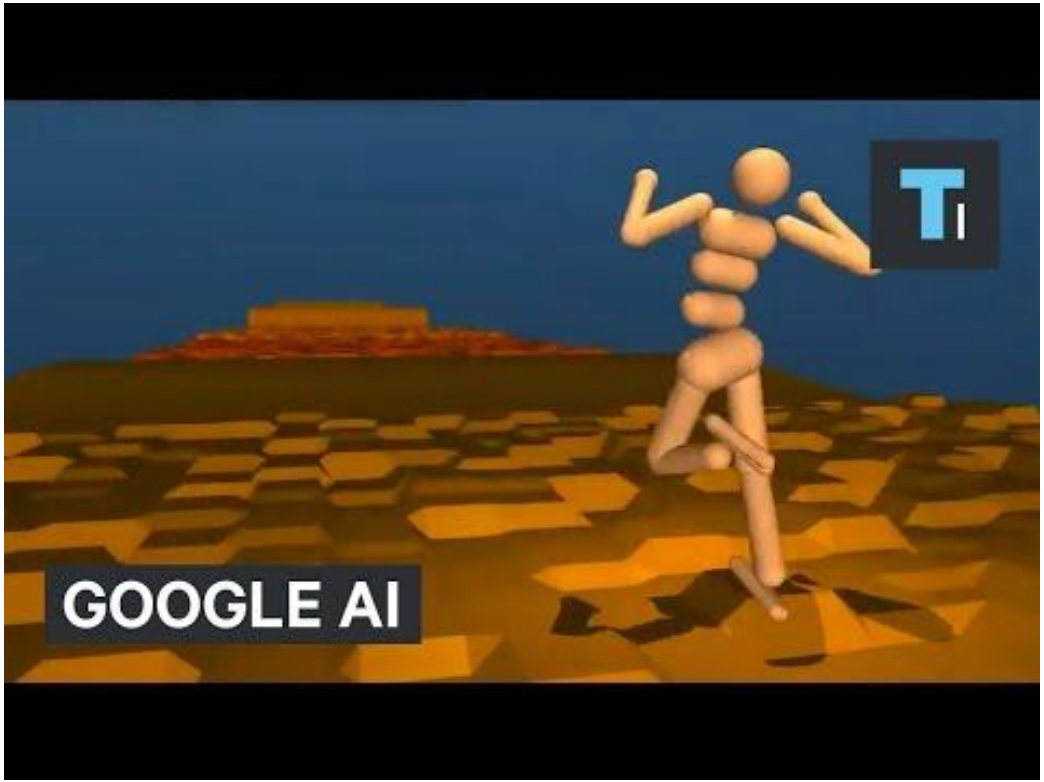
Deep learning with coherent nanophotonic circuits

Yichen Shen[1*†], Nicholas C. Harris[1*†], Scott Skirlo[1], Mihika Prabhu[1], Tom Baehr-Jones[2], Michael Hochberg[2], Xin Sun[3], Shijie Zhao[4], Hugo Larochelle[5], Dirk Englund[1] and Marin Soljačić[1]

Journal Club - Loris Marini
18th August 2017

Google Deepmind

Towards General Purpose AI

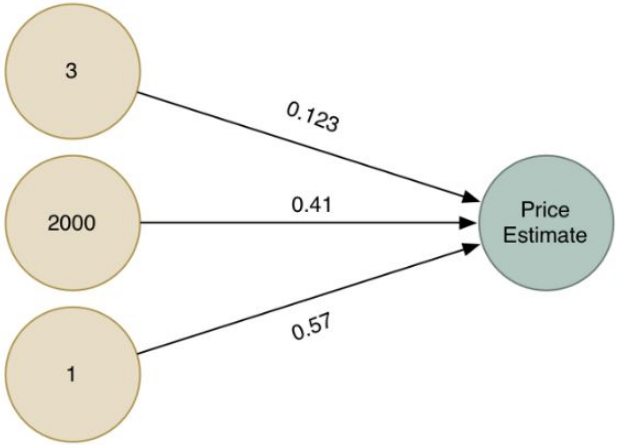


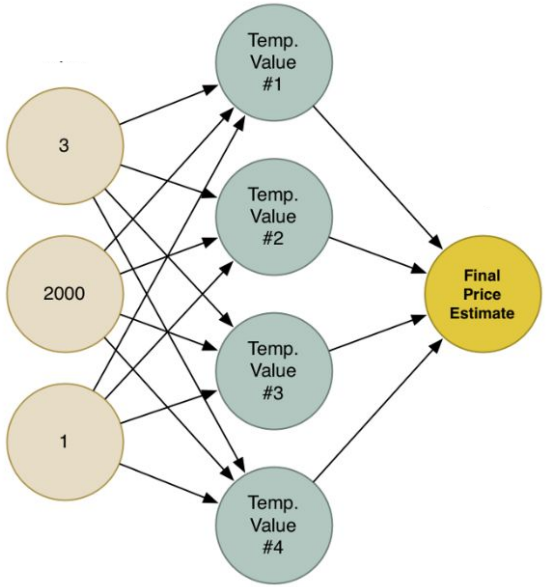
Wy Machine Learning?

We do not want to write a piece of code for every single task.
Rather we want to teach a machine the “principle’ and let it solve problems for us.

Neural Networks: Intro

Useful Architectures to model nonlinear systems

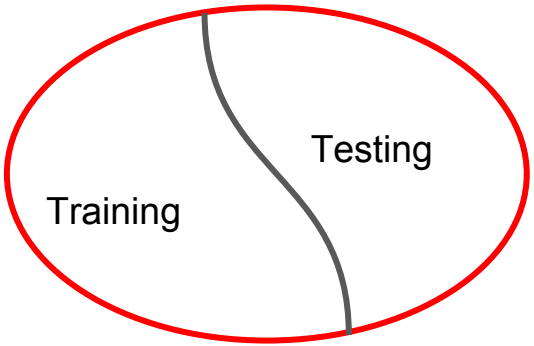




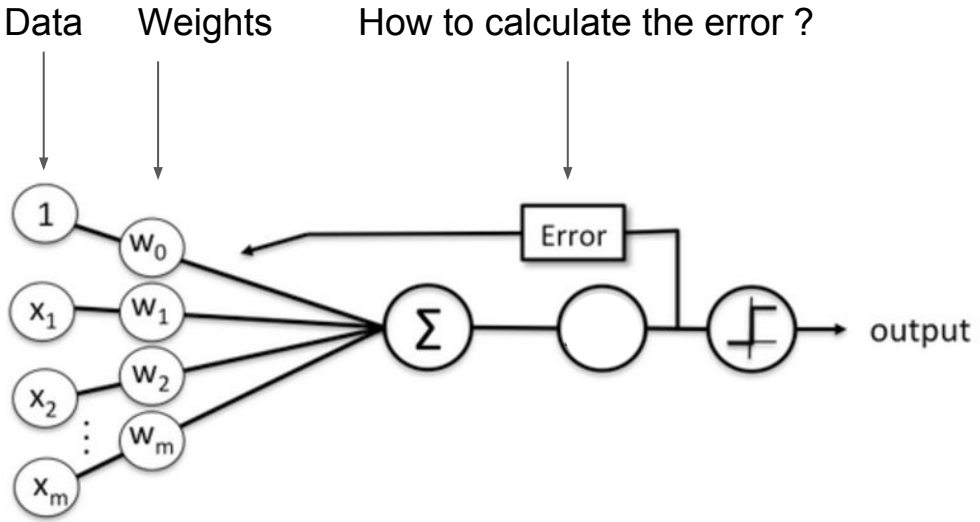
Source: http://bit.ly/237saIB

Training a Neural Network

Find the right sets of weights by feeding back the error.

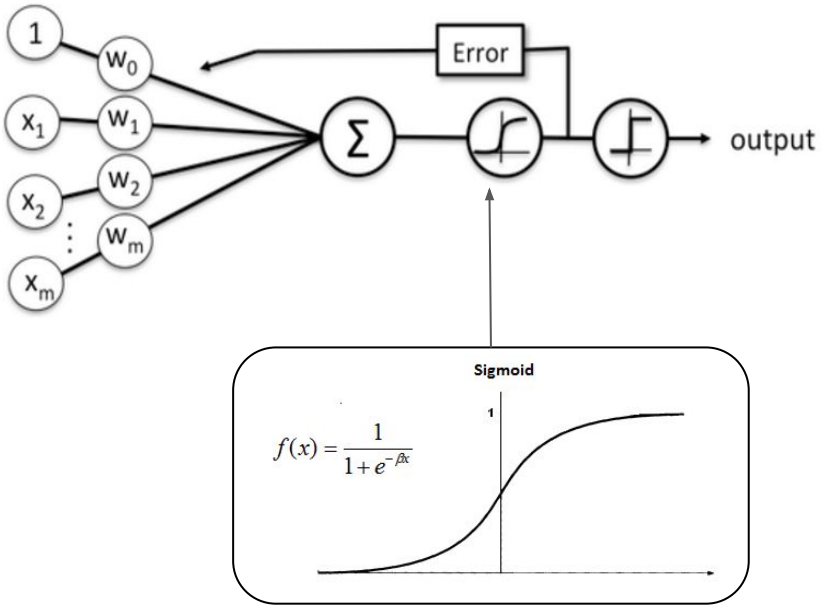


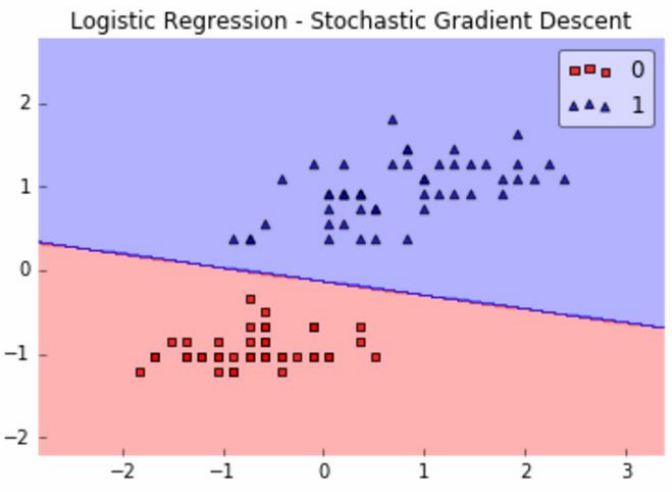
Dataset



http://bit.ly/2c7wjZ9

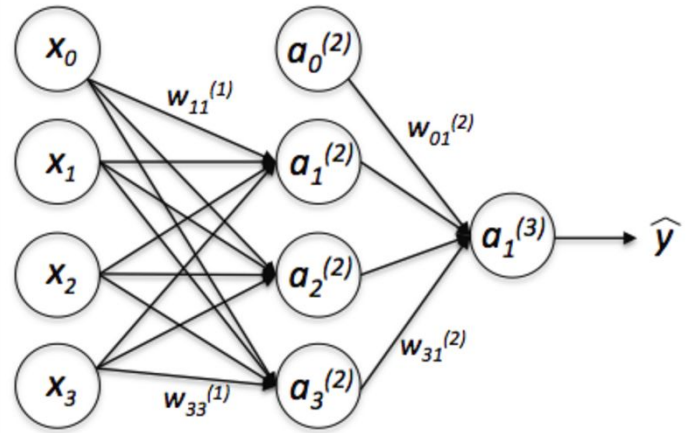
For a nonlinear behavior a single layer is not enough

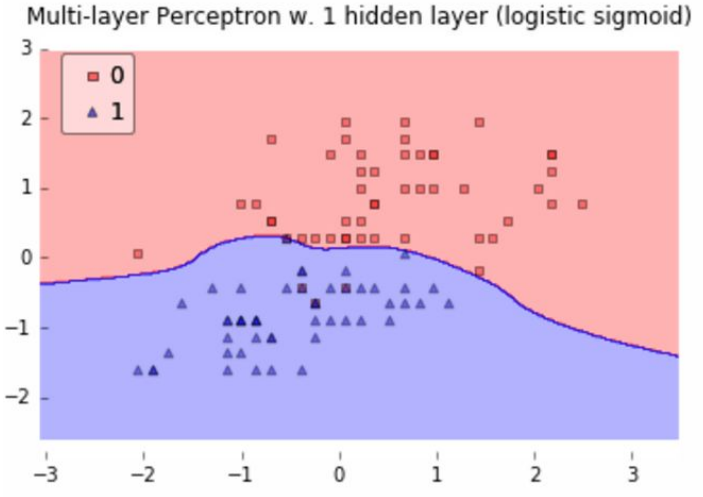




http://bit.ly/2c7wjZ9

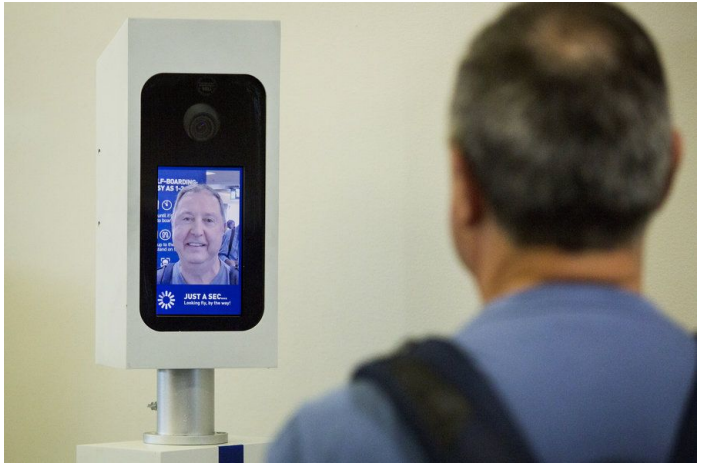
Multilayer + Nonlinear activation function = Nonlinear Response



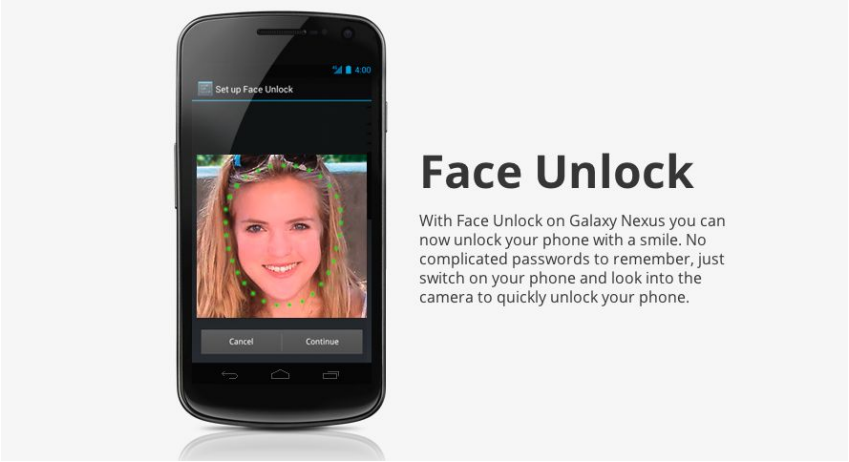


http://bit.ly/2c7wjZ9

If you do it right…



Airport safety



Cryptography

If you do it right…

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

Google AlphaGo

If you do it right…

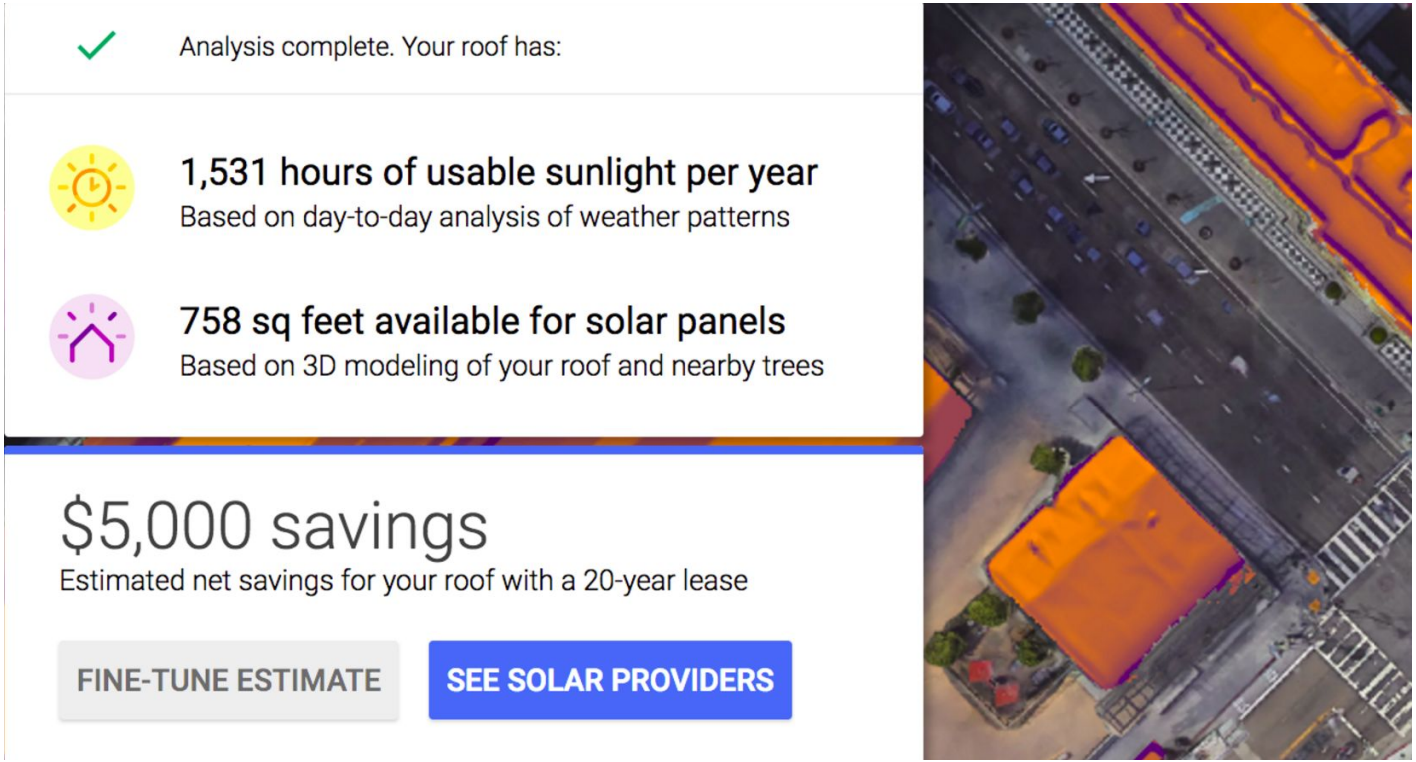
Autonomous Vehicles

http://bit.ly/2i9z3fl

If you do it right…

Tracking and logging wildlife

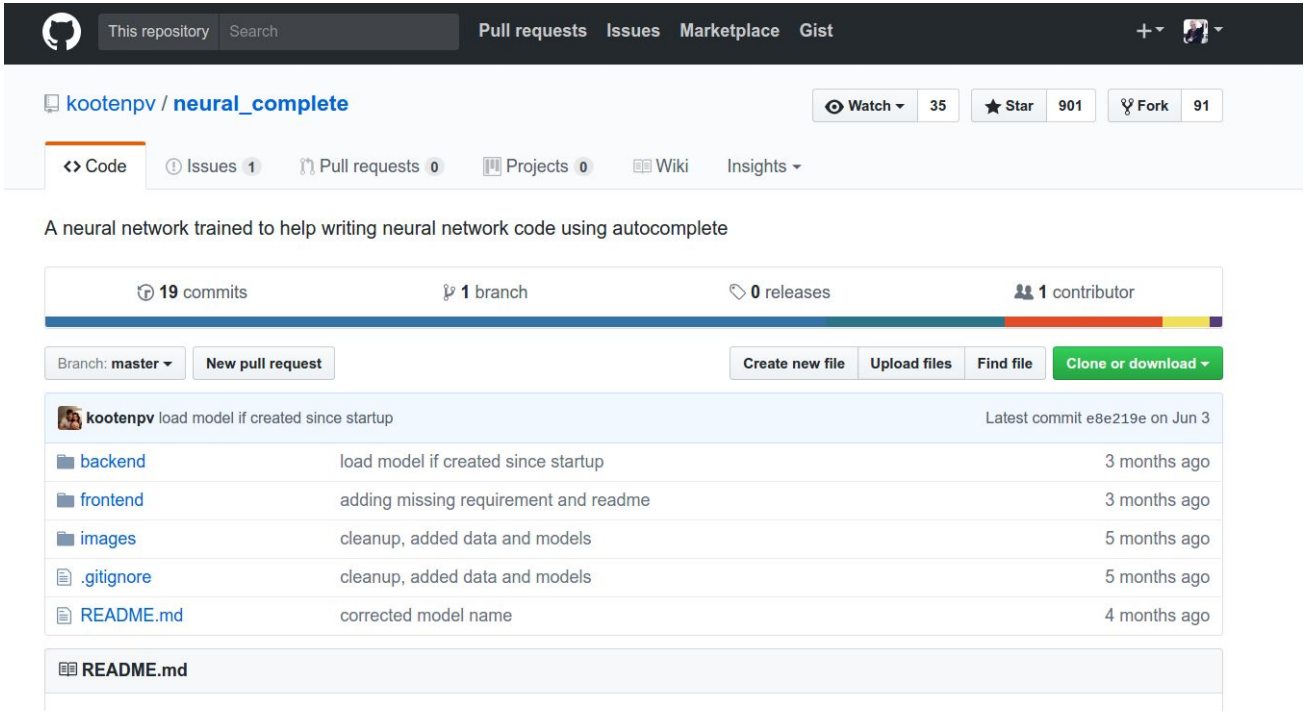
If you do it right…



Predicting real estate for solar

http://bit.ly/1Pw78gb

If you do it right…



Using neural nets to create other neural nets

http://bit.ly/2oB89hQ

Why Deep Learning Works?

Thanks to the optimization algorithms used.

The most common optimization algorithm used in machine learning

Stochastic Gradient Descent as Approximate Bayesian Inference

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

Stephan Mandt
Data Science Institute, Department of Computer Science
Columbia University

Matthew D. Hoffman
Adobe Research
Adobe Systems Incorporated

David M. Blei
Data Science Institute, Departments of Computer Science and Statistics
Columbia University

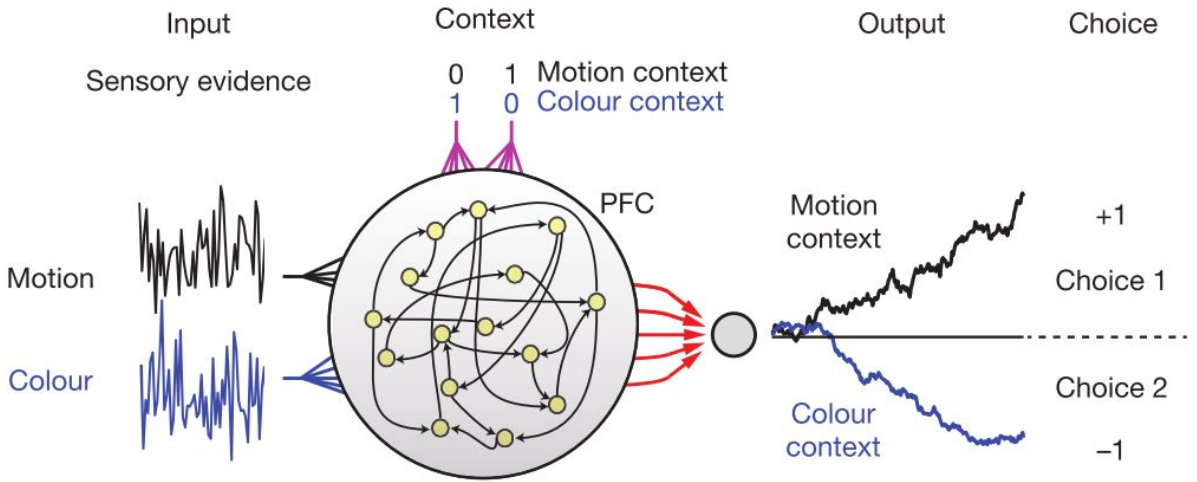
April 17, 2017

inFERENCe

http://bit.ly/2qnAA0j

Recurrent Dynamic Systems

Can reproduce dynamics of neuron populations in the prefrontal cortex.

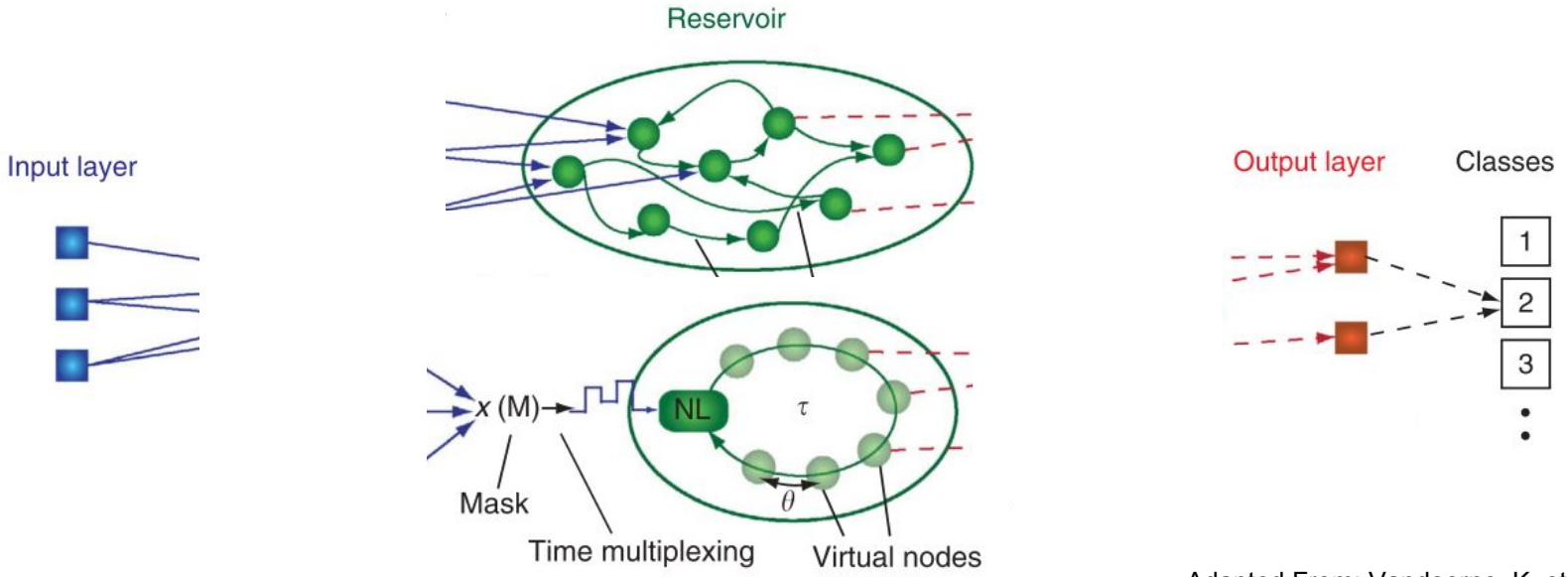


Mante, V, et. al. (2013).
https://doi.org/10.1038/nature12742

Recurrent Dynamic Systems

What happens if you allow information transfer backwards?

- Recurrent networks manifest transient dynamics.
- The transient dynamical states are fundamental for computation.

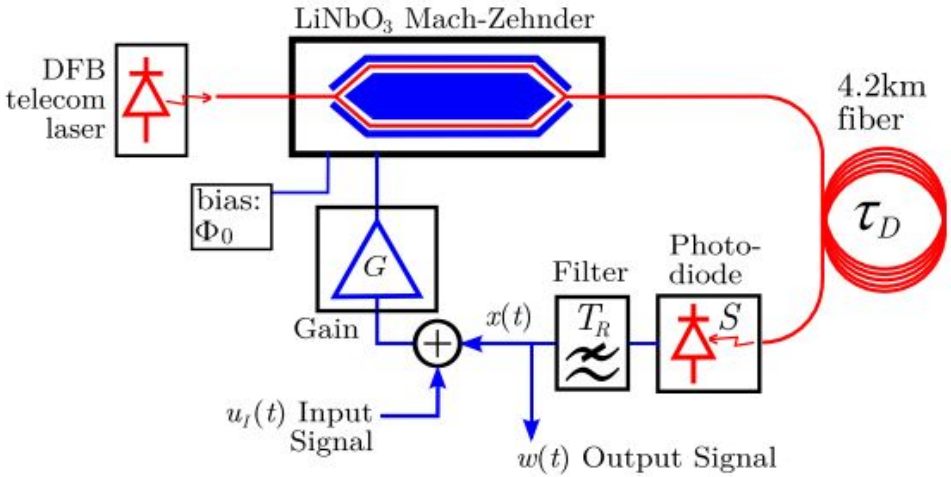


Adapted From: Vandoorne, K. et. al. (2014). *Nature Communications*, *5*, 3541. Read not recommended.

Optical
Artificial Neural Networks

Optical Reservoir Computing

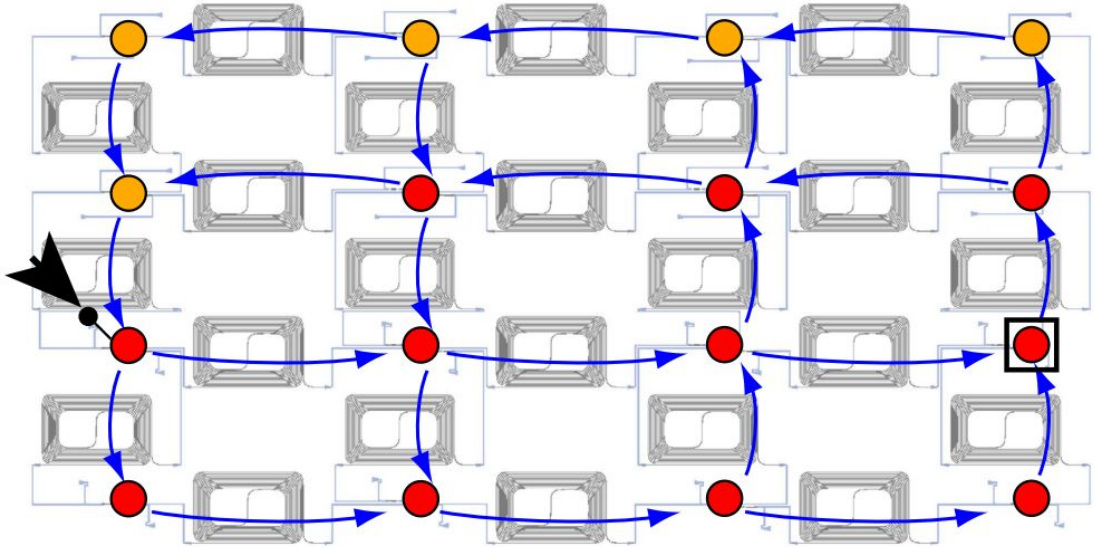
Optoelectronic version with time multiplexing.



Larger, L. et. al. (2012). *Optics Express*, *20*(3), 3241–9.

Optical Reservoir Computing

All optical with ring resonators.



Mesaritakis, C. et. al. J. Opt. Soc. Am. B 30, 3048–3055 (2013)
Reading not recommended.

The paper….

nature photonics

ARTICLES

PUBLISHED ONLINE: 12 JUNE 2017 | DOI: 10.1038/NPHOTON.2017.93

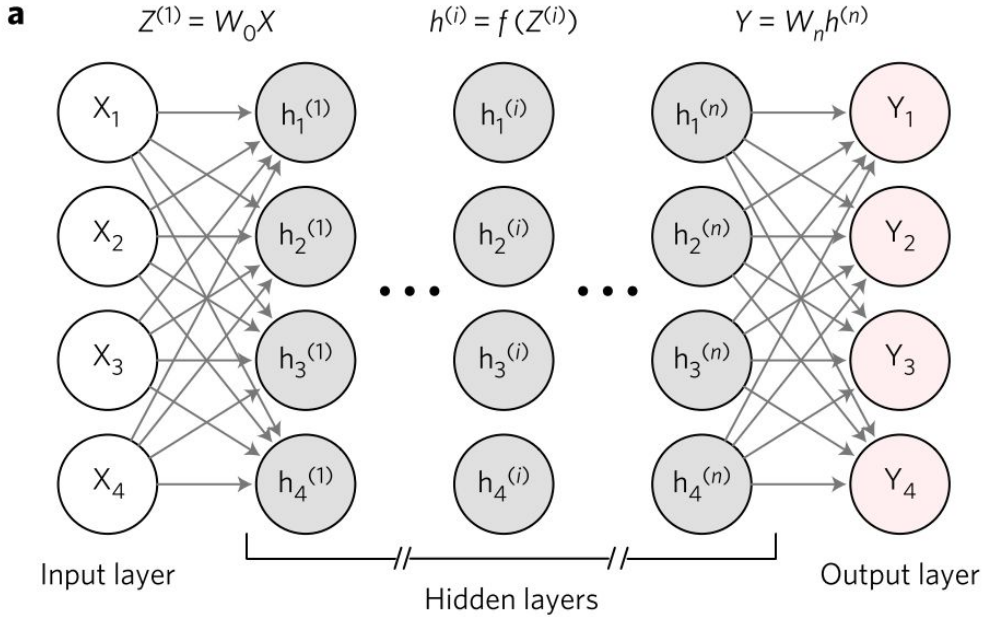
Deep learning with coherent nanophotonic circuits

“There is a need for a better architecture for computing that the Von Neumann.”

“ We propose a new architecture for fully optical neural network that is **energy efficient** and **faster** than conventional electronic systems.”

“Once a neural network is trained, the architecture can be passive, and computation on the optical signals will be performed without additional energy input.”

Architecture: a beautiful drawing...



Nonlinear Activation Function

Saturable Absorbers based on Graphene

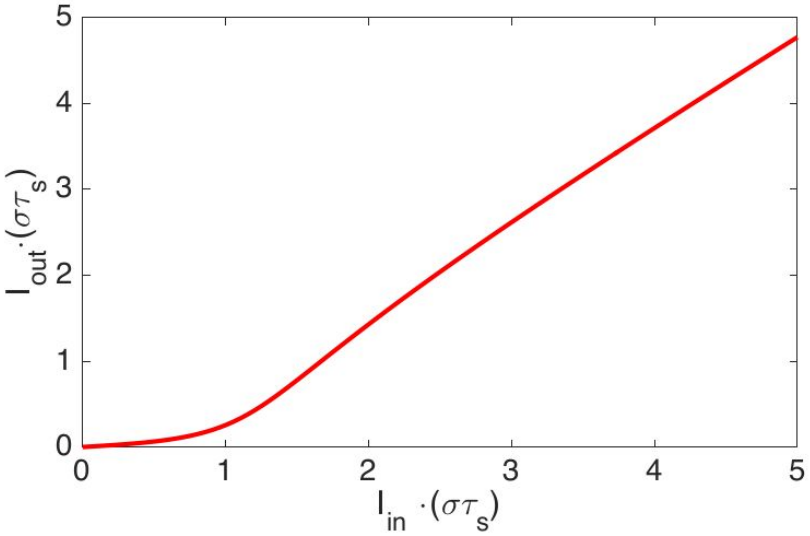
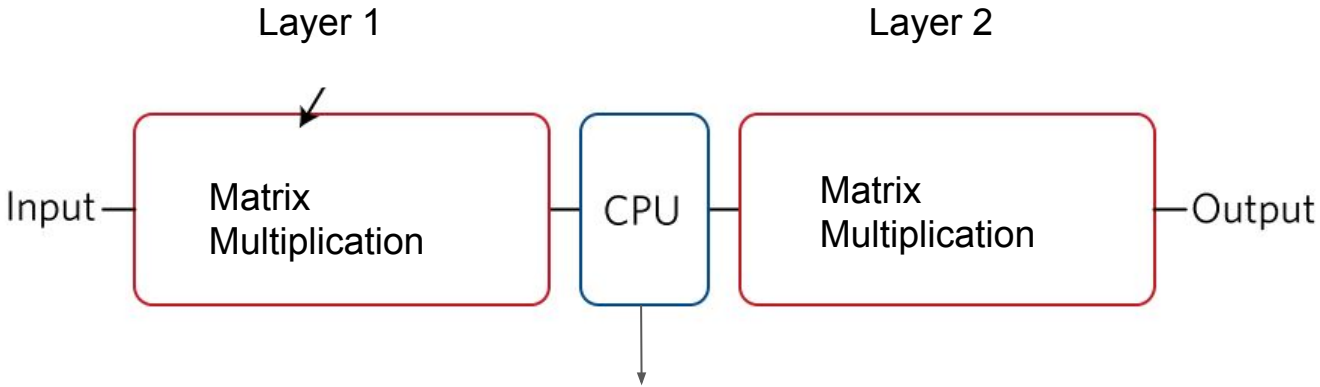


Fig. S3: Nonlinear response function plot of Optical Saturable Absorption Effect

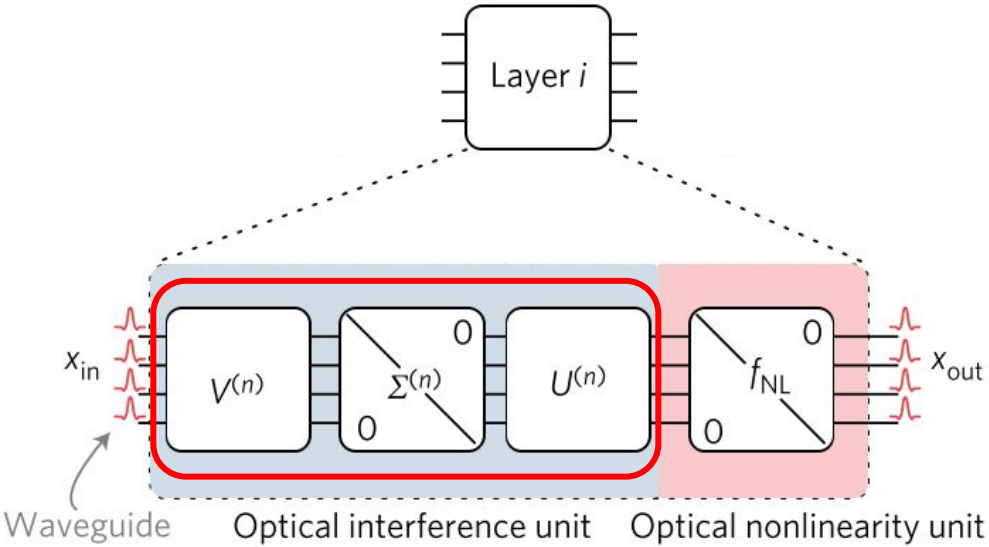
Architecture: In reality

4 neurons - 2 layers



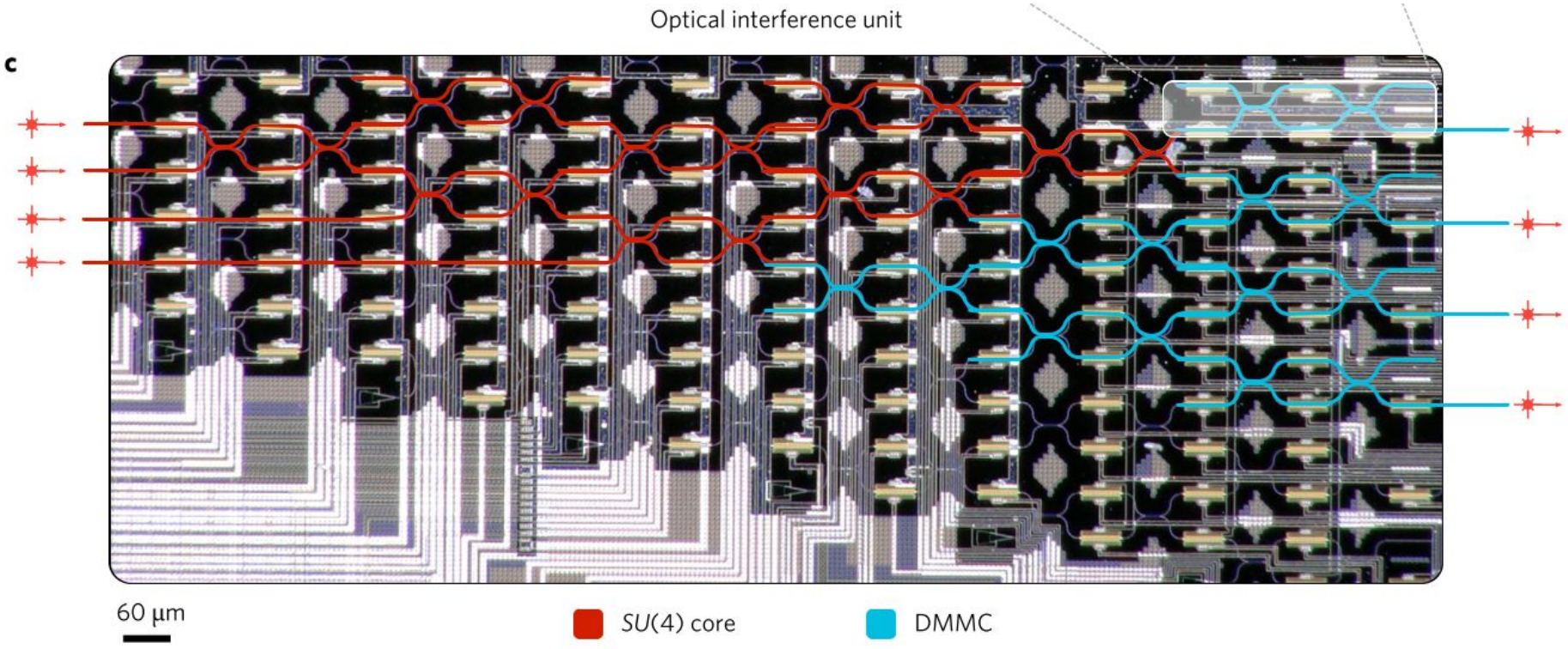
Nonlinear Activation Functions
The central part of a Neural Network.

How a single layer looks like



Linear Matrix Multiplication

Nonlinear Activation Function


c
Optical interference unit
60 μm
SU(4) core
DMMC

Experiment:

Dataset: 90 people recorded the sound of letter ‘A’, ‘B’, ‘C’, ‘D’.

1. Multiply sound by Hamming Function
2. Fourier Transfor
3. Find power spectral density
4. Use a normal CPU to train the network

Electron

5. Input lasers inside photonic chip

Photon

6. Get off-chip to introduce some nonlinearity

Electron

7. Go back on chip with new lasers
8. Propagate through network

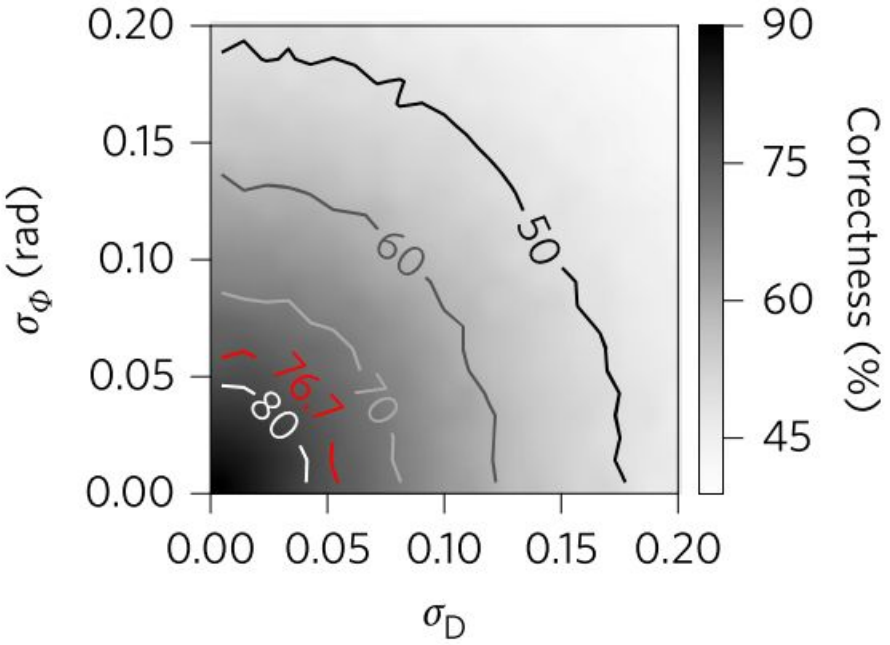
Photon

9. Detect result

Electron

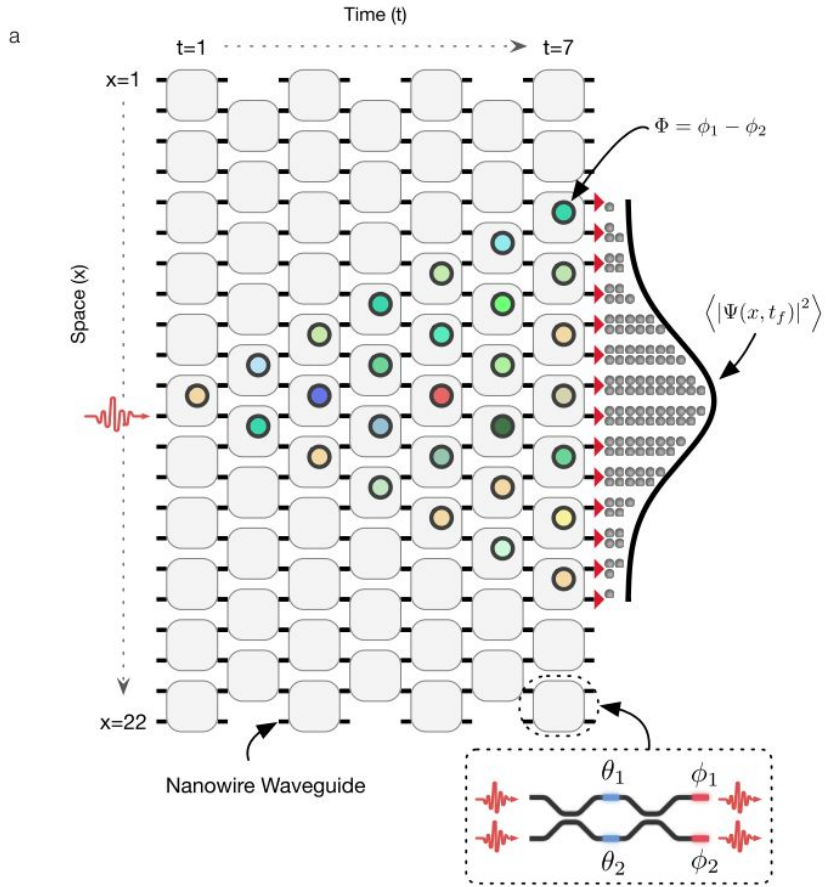
Results

76.7 % vs 91.7 % accuracy for a 64 bit CPU



Claim on the architecture #1

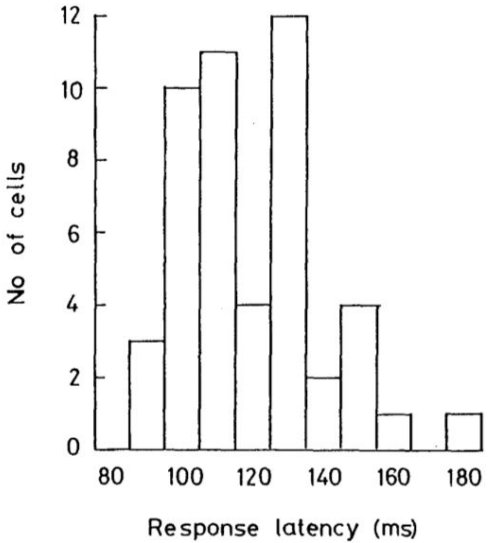
“Integrated photonics provides a scalable solution to large, phase-stable optical transformations”



Harris, N. C. et al. Preprint at http://arXiv.org/abs/1507.03406 (2015).

Claim on the architecture #2

“... would enable forward propagation 100 + times faster than electronic / optoelectronic systems”



D. I. Perrett et. al. Exp. Brain Res. 47, 329–342 (1982).

Google Tensor Flow Unit
Built on a 28nm process, runs at 700MHz

Claim on the architecture #3

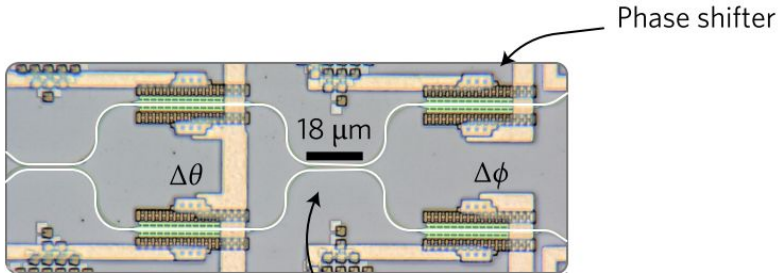
"… significantly lower latency than electronic digital computers. Useful for real time applications (autonomous driving or missile tracking)."

1. This is a physical truth. Does it really matter ?

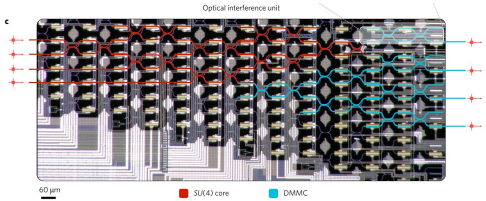
Claim on the architecture #4

"...after training, forward propagation computing is performed optically **on a passive system.**"

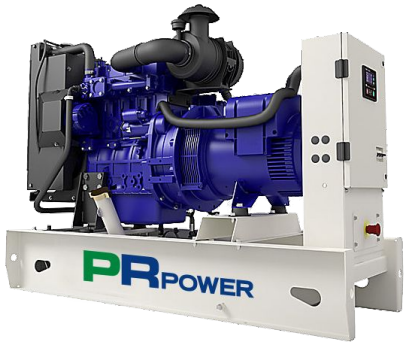
1. 10mW per phase shifter (40 mW per neuron).
2. In a 10^6 neurons (assuming you can ever make a photonic chip with 10^6 interconnects) would be 40 kW.
3. This is 10^6 times less efficient than IBM's TrueNorth: < 100 mW of power for 10^6 neurons.
4. You still need the nonlinear response for the net to work well… pump your saturable absorber!



Claim on the architecture #4 (Continued)



vs

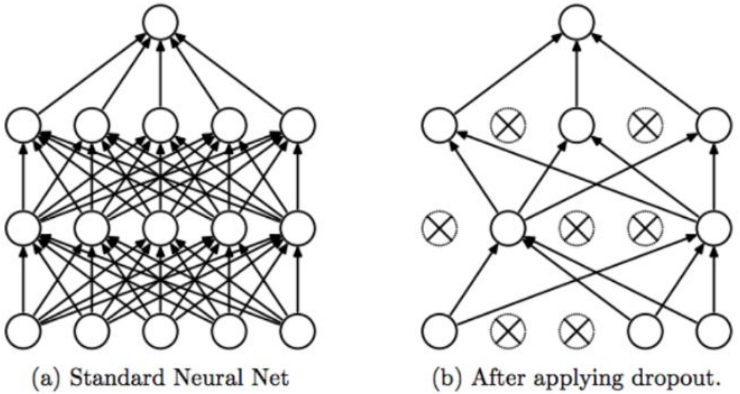


vs

Claim on the architecture #5

“ ..in optical architecture power consumption is nearly proportional to the number of neurons (instead of quadratic, as in electronics).”

Spiking Networks: Not ON all the time.



Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting", JMLR 2014

Drop-out in neural Networks: http://bit.ly/2wdnXvB

Claim on the architecture #6

"...time multiplexing = bigger effective networks with less neurons."

1. This is true for both optical and electronic nets.
2. It does not mean necessarily that the energy efficiency increases.
3. learning not as good (connectivity between virtual nodes is limited to local couplings including few nearest neighbors See: Larger, L. et. al. (2012). *Optics Express*, *20*(3), 3241–9.

A good reference from the paper!

Neural Networks 61 (2015) 85–117

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Review

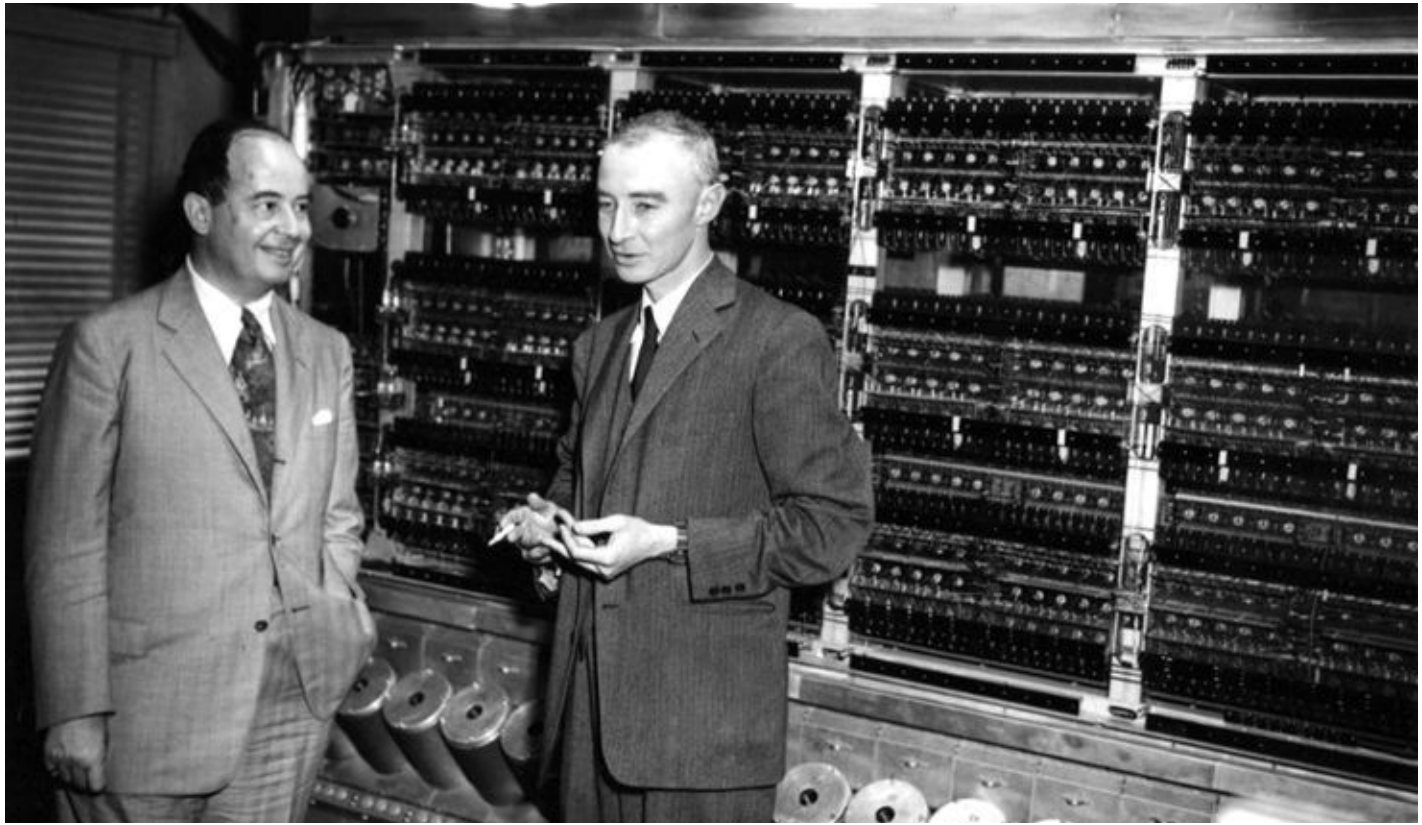
Deep learning in neural networks: An overview

Jürgen Schmidhuber

The Swiss AI Lab IDSIA, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, University of Lugano & SUPSI, Galleria 2, 6928 Manno-Lugano, Switzerland

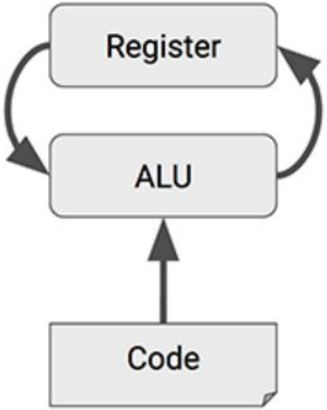
Hardware for Artificial Neural Networks

65 Years from the first Turing-complete Machine...

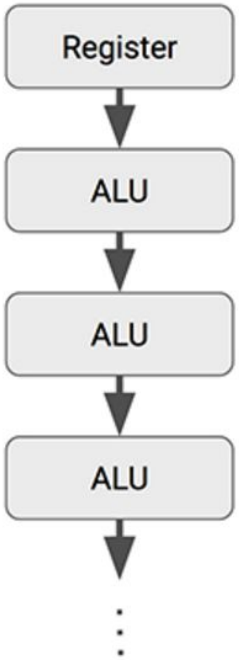


The Problem with the Von Neumann Architecture

Google TPU (Tensor Flow Unit) new architecture



Conventional CPUs



Systolic Array - MUX
(Google Designed) CPUs

Neuromorphic chips inspired by our brain

Neurons are not always ON!

Spike Processing

It’s all in the architecture.

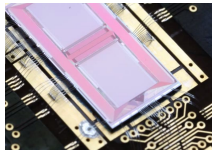
DeepMind

Google TPU

intel nervana SYSTEMS

Intel

IBM’s TrueNorth



Human Brain Project
HICANN

Architecture - Google TPU (Tensor Flow Unit)

Machine Learning accelerators in Google Data centers since 2015

Source: http://bit.ly/2se11Xu

# of network layers	# of weights
5	20M
4	5M
58	52M
56	34M
16	8M
89	100M

Source: http://bit.ly/2se11Xu

Neurosynaptic Chips - IBM’s TrueNorth

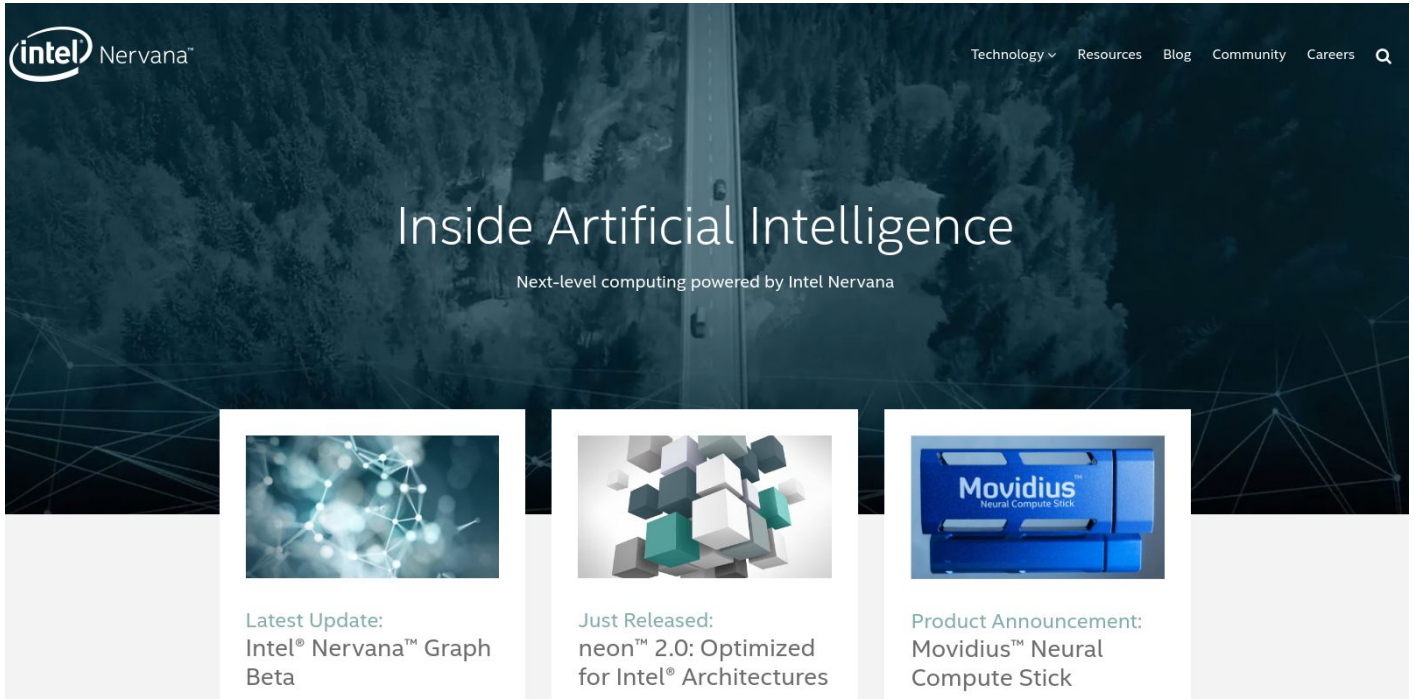
5.4 billion transistors, no clock.

Programmable neurons	1 million
Programmable synapses	256 million
Neurosynaptic cores	4,096

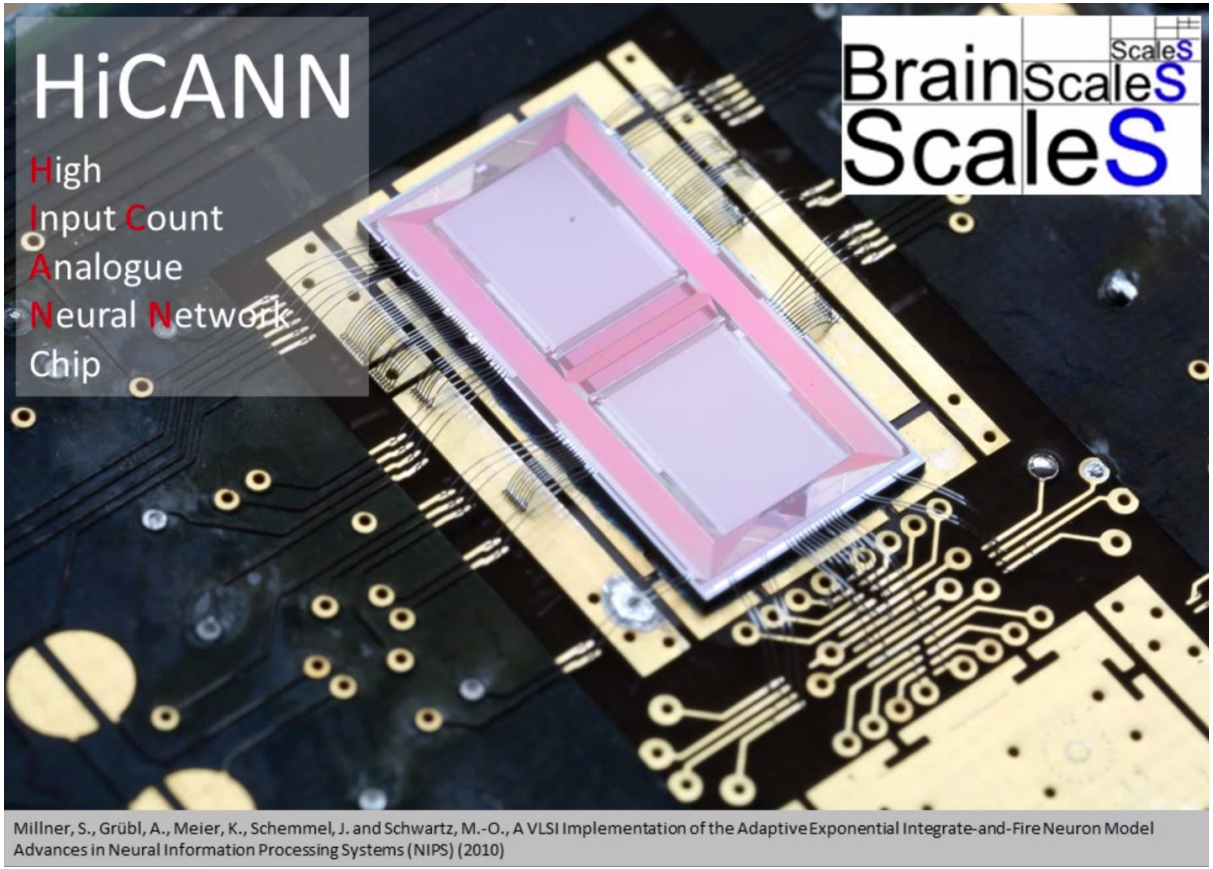
Source: https://ibm.co/1r7vKPq

Architecture - Google TPU (Tensor Flow Unit)

Machine Learning accelerators in Google Data centers since 2015



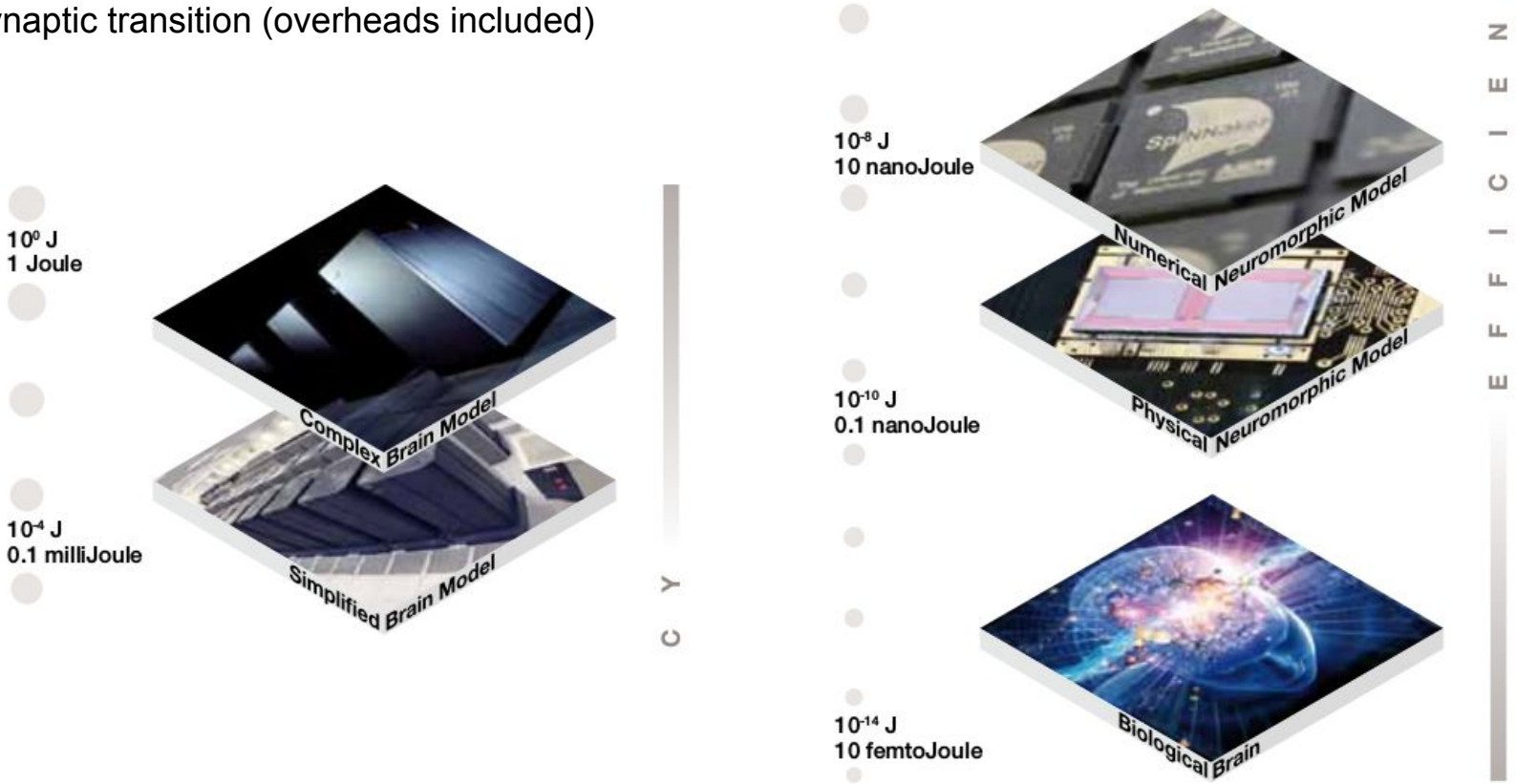
Intel Fact Sheet: http://intel.ly/2usZTSy



Extracted from a Talk by Prof Karlheinz Meier,Univ. Heidelberg (DE): http://bit.ly/2x25jE5

Energy Efficiency HICANN

J/synaptic transition (overheads included)



Source: http://bit.ly/2uZI77R

Human Brain Project (European 2020)

HICANN, University of Heidelberg, Germany

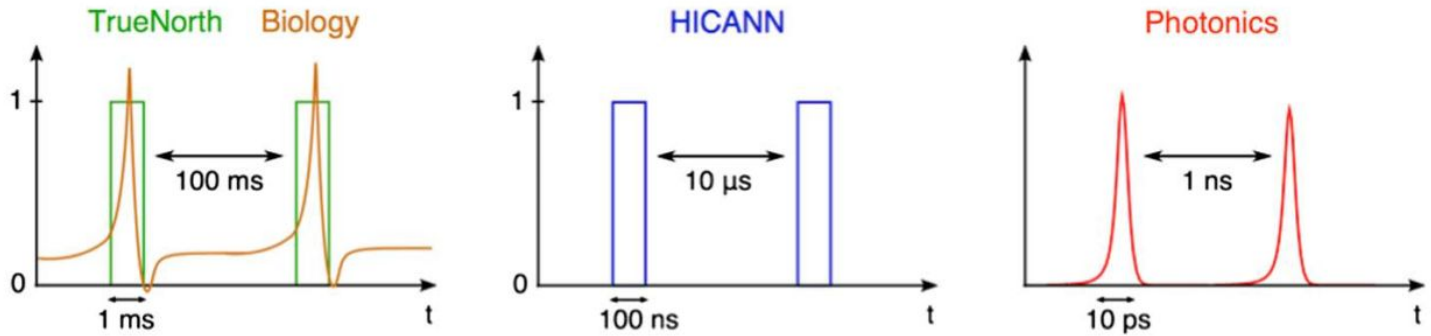
TimeScales	Nature	Simulation	Accelerated Model
Learning	Day	1000 Days	10 s

BrainScaleS

Extracted from a Talk by Prof Karlheinz Meier,Univ. Heidelberg (DE): http://bit.ly/2x25jE5

Neuromorphic chips and spike processing

Time scales: biological neurons vs electronic spiking neurons vs photonic neurons.



Prucnal, P. R. et. al. (2016). *Advances in Optics and Photonics*, *8*(2), 228.

nature photonics

ARTICLES

PUBLISHED ONLINE: 12 JUNE 2017 | DOI: 10.1038/NPHOTON.2017.93

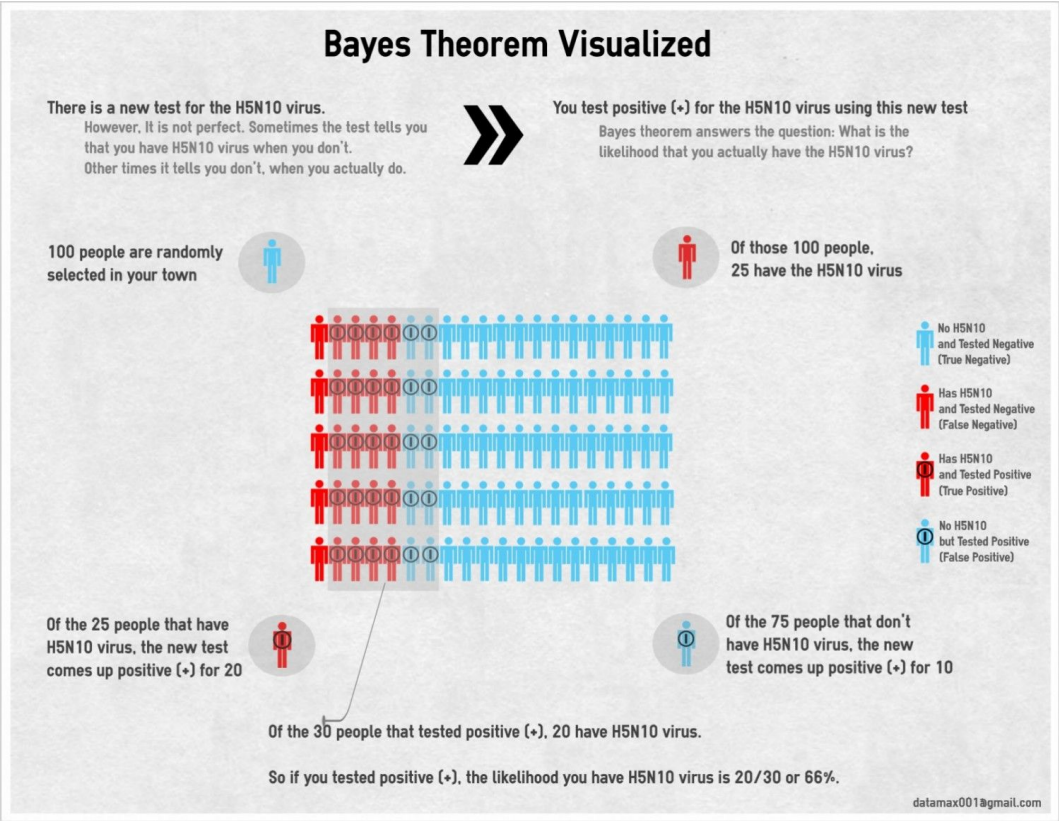
Deep learning with coherent nanophotonic circuits

Yichen Shen[1*†], Nicholas C. Harris[1*†], Scott Skirlo[1], Mihika Prabhu[1], Tom Baehr-Jones[2], Michael Hochberg[2], Xin Sun[3], Shijie Zhao[4], Hugo Larochelle[5], Dirk Englund[1] and Marin Soljačić[1]

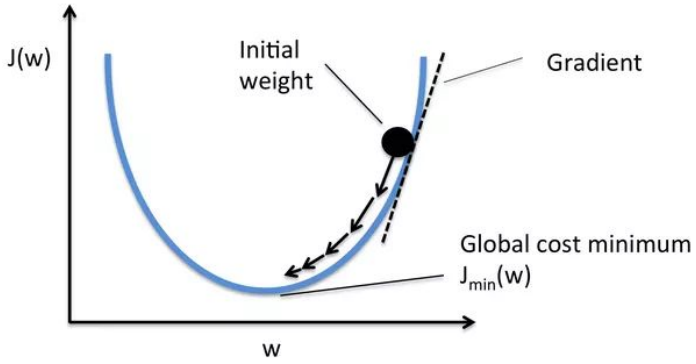
A solution looking for a problem?

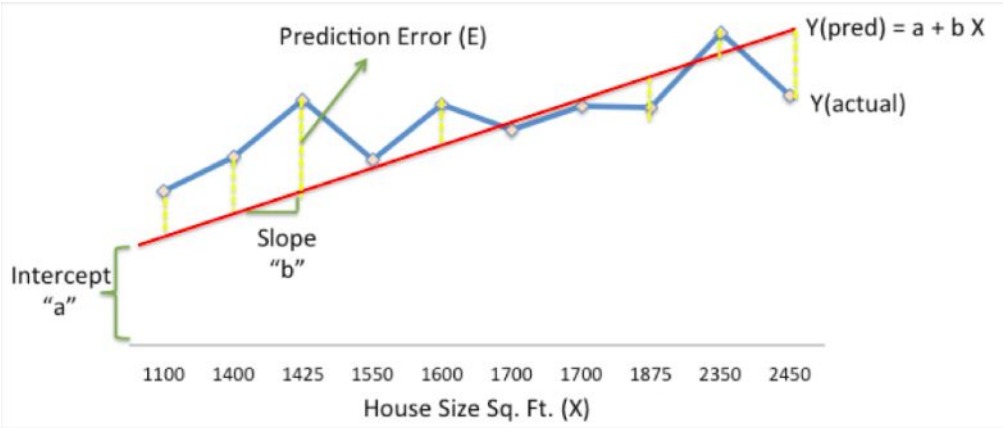
Extras

Bayes Theorem Visualized



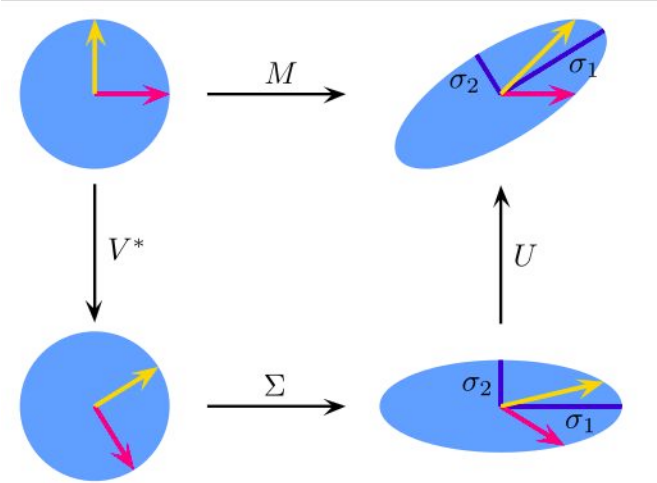
Stochastic Gradient Descent





Matrix Multiplication with linear optics

Singular Value Decomposition



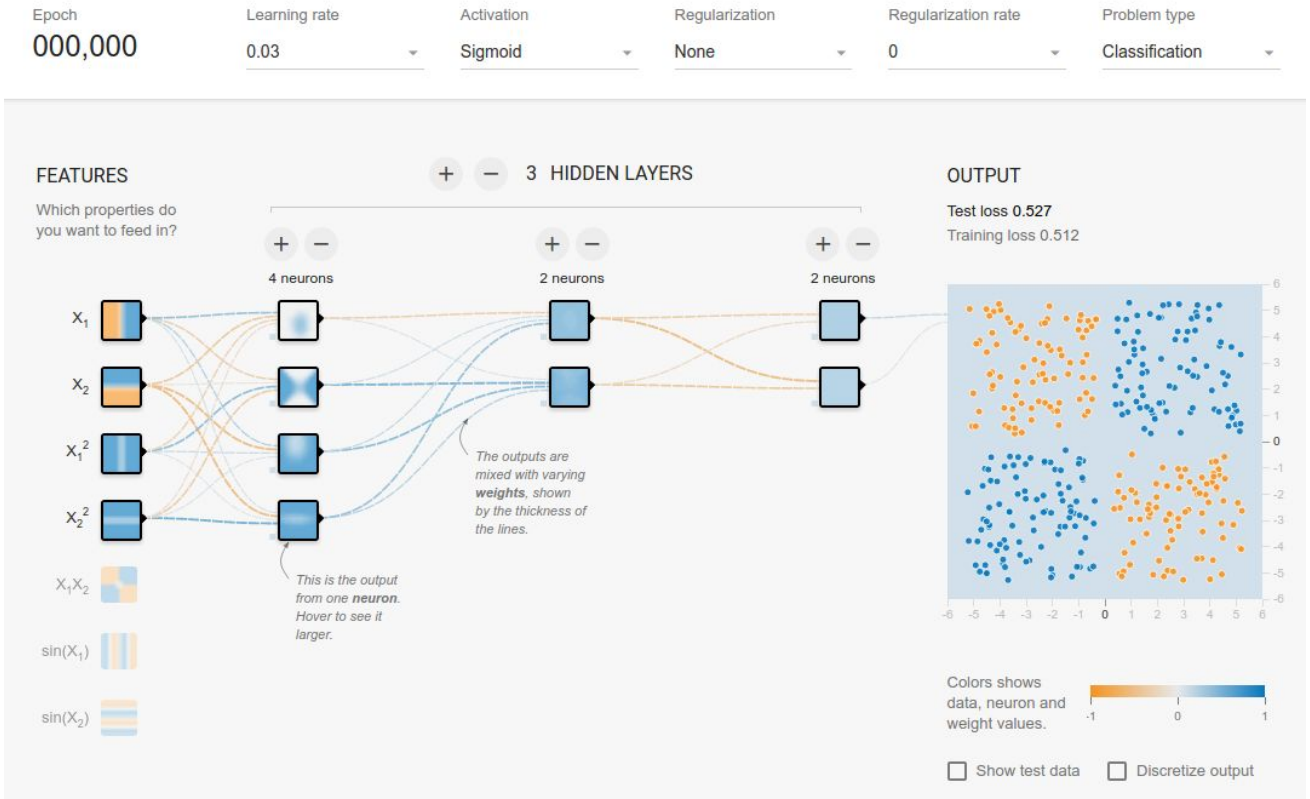
Unitary Matrix
(Beam splitters and phase shifters)

$$M = U \cdot \Sigma \cdot V^*$$

Optical attenuation
Optical amplification

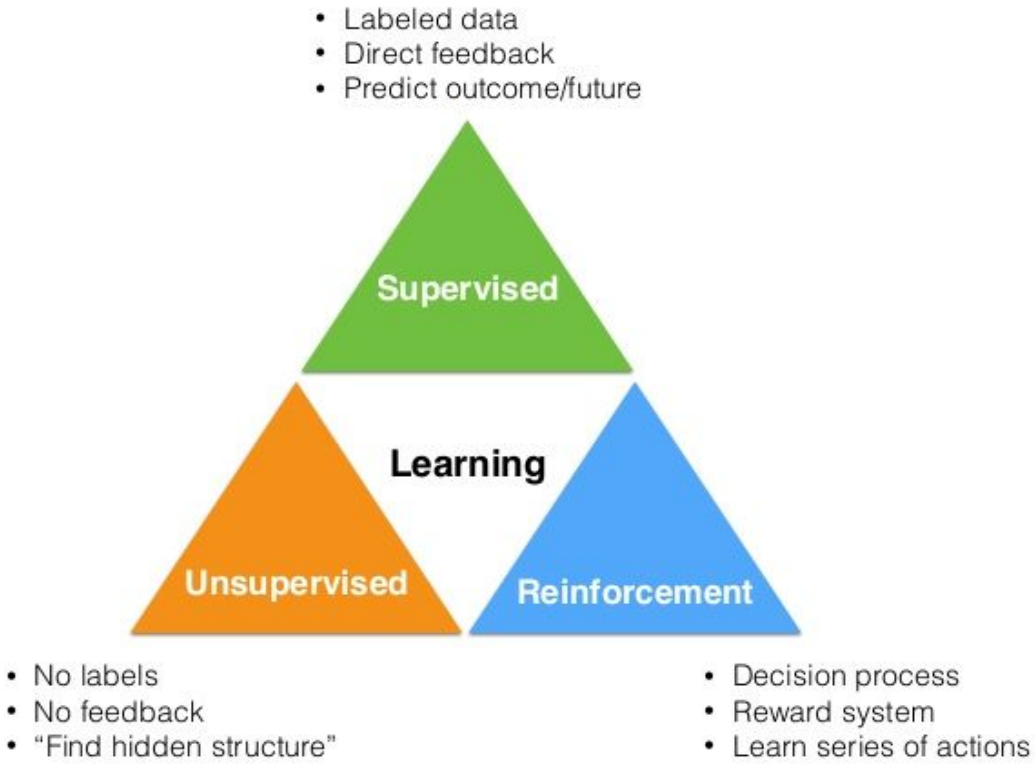
“Matrix multiplication with unitary matrices implemented in the manner above consumes, in principle, no power.”

A playground for Neural Networks

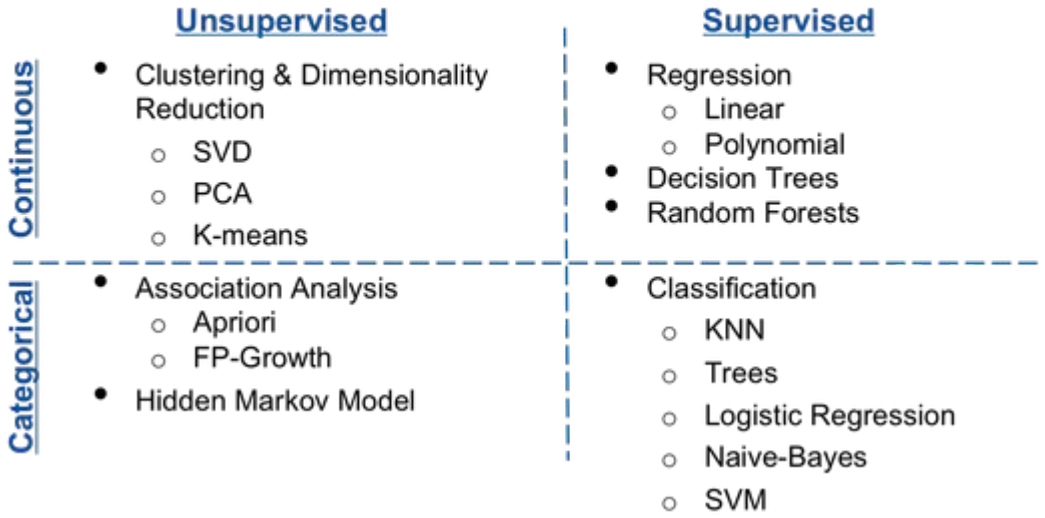


Source: http://bit.ly/2w8RJ4F

Types of Machine Learning



Neural Nets, Deep Learning and Machine Learning



Source: http://bit.ly/2wrkGGz

Libraries for Deep Learning

Lua

- Torch & Pytorch

Python Frameworks

- Theano & Ecosystem
- TensorFlow
- Caffe
- Caffe2
- Chainer
- CNTK
- DSSTNE
- DyNet
- Keras
- Mxnet
- Paddle
- BigDL
- Licensing

JVM Considerations

- Speed
- DL4J: Why the JVM?
- DL4J: Ecosystem
- DL4S: Deep Learning in Scala
- Machine-Learning Frameworks
- Further Reading

Source: http://bit.ly/2ro1t9O

Human Brain Project: Learning

HBP Science for non-speci...

Neurobiology for non-specialists

HBP Curriculum online workshop - neurobiology for non-specialists

Brain medicine for non-specialists

HBP Curriculum online course - Brain medicine for non-specialists

ICT for non-specialists

HBP Curriculum online course - ICT for non-specialists

HBP Complementary Topics

Research, ethics and societal impact

HBP Curriculum online course - Research, ethics and societal impact

IPR, translation and exploitation of research

HBP Curriculum online course - Intellectual property rights, translation and exploitation of research

Source: http://bit.ly/2w853XJ

7 examples of machine learning (witch code)

Medium

Sign in / Sign up

Adam Geitgey Follow

Interested in computers and machine learning. Likes to write about it.

May 6, 2014 · 15 min read