The Science and the Engineering of Intelligence

Tomaso Poggio,
The MIT Quest
Center for Brains, Minds & Machines,
McGovern Institute for Brain Research,
Computer Science and Artificial Intelligence Agency,
Brain and Cognitive Sciences, MIT

The CBMM-FLAB partnership

http://www.fujitsu.com/global/about/resources/news/press-releases/2018/1005-01.html

http://pr.fujitsu.com/jp/news/2018/10/5.html

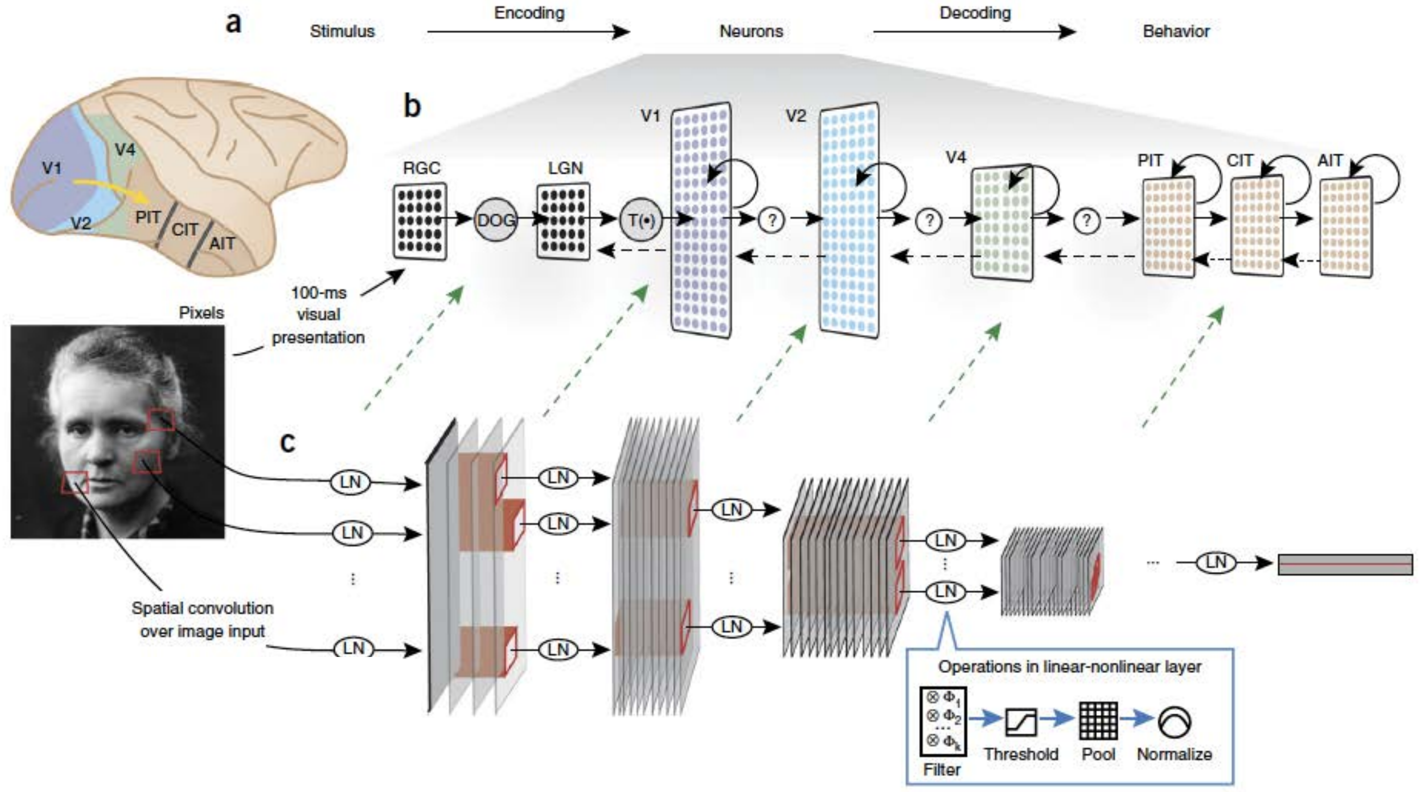

HIGHLAND PARK
POLAR

Overview

- Motivations: the greatest problem in science, CBMM, the MIT Quest
- A bit of history: Neuroscience and AI, Science and Engineering
- CBMM and the Quest
- AI ethics and its neural bases
- Theory: explaining how deep networks work and what are their properties and limitations.

CBMM’s focus is the Science and the Engineering of Intelligence

We aim to make progress in understanding intelligence, that is in understanding how the brain makes the mind, how the brain works and how to build intelligent machines.



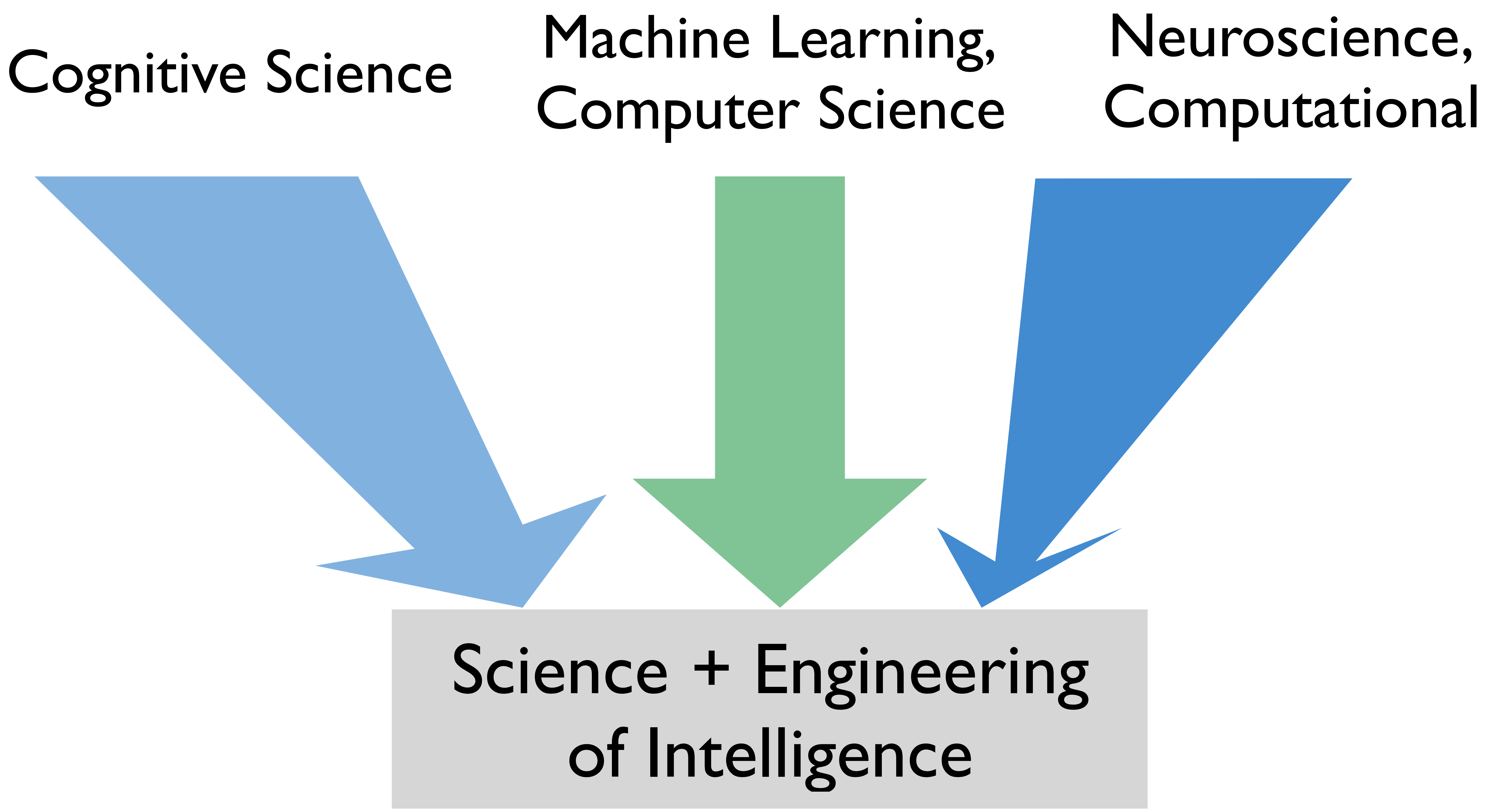
CENTER FOR
Brains
Minds+
Machines

CBMM Overview

CENTER FOR Brains Minds+ Machines

The Center for Brains, Minds and Machines (CBMM) is a multi-institutional NSF Science and Technology Center dedicated to the study of intelligence - how the brain produces intelligent behavior and how we may be able to replicate intelligence in machines. We believe in the synergy between the science and the engineering of intelligence.

Funding 2013-2023	**~$50M**
Research Institutions	**~4**
Educational Institutions	**12**
Faculty (CS+BCS+…)	**~23**
Researchers	**223**
Publications	**397**



Research, Education & Diversity Partners

MIT
Boyden, Desimone, DiCarlo, Kanwisher, Katz, McDermott, Poggio, Rosasco, Sassanfar, Saxe, Schulz, Tegmark, Tenenbaum, Ullman, Wilson, Winston

Harvard
Blum, Gershman, Kreiman, Livingstone, Nakayama, Sompolinsky, Spelke

Allen Institute
Koch

Howard U.
Chouika, Manaye, Rwebangira, Salmani

Hunter College
Chodorow, Epstein, Sakas, Zeigler

Johns Hopkins U.
Yuille

Queens College
Brumberg

Rockefeller U.
Freiwald

Stanford U.
Goodman

Universidad Central del Caribe (UCC)
Jorquera

University of Central Florida
McNair Program

UMass Boston
Blaser, Ciaramitaro, Pomplun, Shukla

UPR – Mayagüez
Santiago, Vega-Riveros

UPR– Río Piedras
Garcia-Arraras, Maldonado-Vlaar, Megret, Ordóñez, Ortiz-Zuazaga

Wellesley College
Hildreth, Wiest, Wilmer

Academic and Corporate Partners

IIT Cingolani	**A*star** Chuan Poh Lim	**Hebrew U.** Weiss	**MPI** Bülthoff
Genoa U. Verri, Rosasco	**Weizmann** Ullman	**City U. HK** Smale	

Google	**IBM**	**Microsoft**	**Orcam**	**Siemens**	**Fujitsu**	
DeepMind	**Honda**	**Boston Dynamics**	**NVIDIA**	**Schlumberger**	**Mobileye**	**Intel**

EAC Members & Meetings

Demis Hassabis, *DeepMind*

Charles Isbell, Jr., *Georgia Tech*

Christof Koch, *Allen Institute*

Fei-Fei Li, *Stanford*

Lore McGovern, *MIBR, MIT*

Joel Oppenheim, *NYU*

Pietro Perona, *Caltech*

Marc Raibert, *Boston Dynamics*

Judith Richter, *Medinol*

Kobi Richter, *Medinol*

Dan Rockmore, *Dartmouth*

Amnon Shashua, *Mobileye*

David Siegel, *Two Sigma*

Susan Whitehead, *MIT Corporation*

Summer Course at Woods Hole: Our flagship initiative led by G. Kreiman

Brains, Minds & Machines Summer Course

An intensive three-week course gives advanced students a “deep” introduction to the problem of intelligence

A community of scholars is being formed

CENTER FOR
Brains
Minds+
Machines

The MIT Quest for Intelligence

Forging connections between human and machine intelligence research, its applications, and its bearing on society.

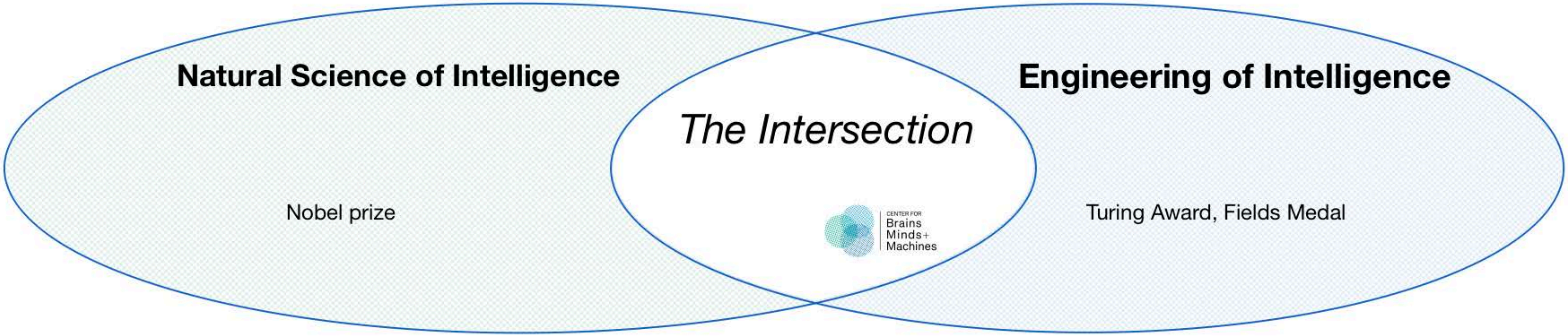
The MIT Intelligence Quest will advance the science and engineering of both human and machine intelligence. Launched on February 1, 2018, this effort seeks to discover the foundations of human intelligence and drive the development of technological tools that can positively influence virtually every aspect of society.

The Institute's culture of collaboration …we seek to answer the deepest questions about intelligence.

CENTER FOR Brains Minds+ Machines

Intelligence: The MIT Quest

CORE: Cutting-Edge Research on the Science + Engineering of Intelligence

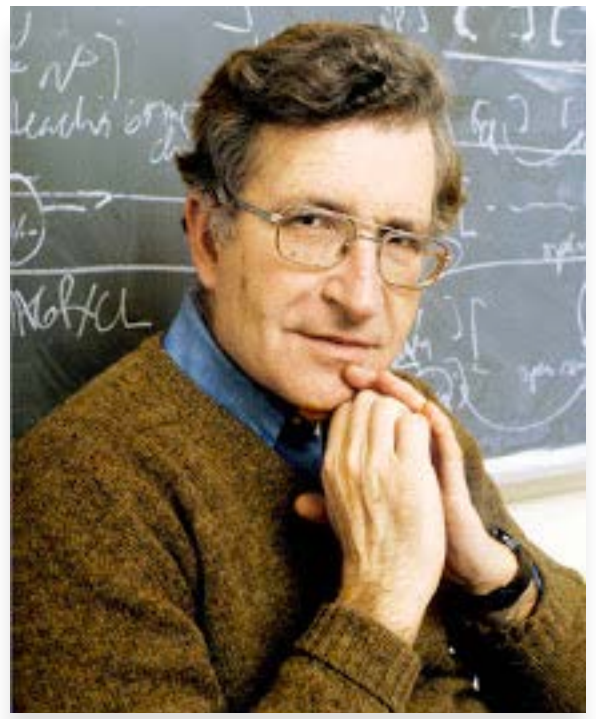


Overview

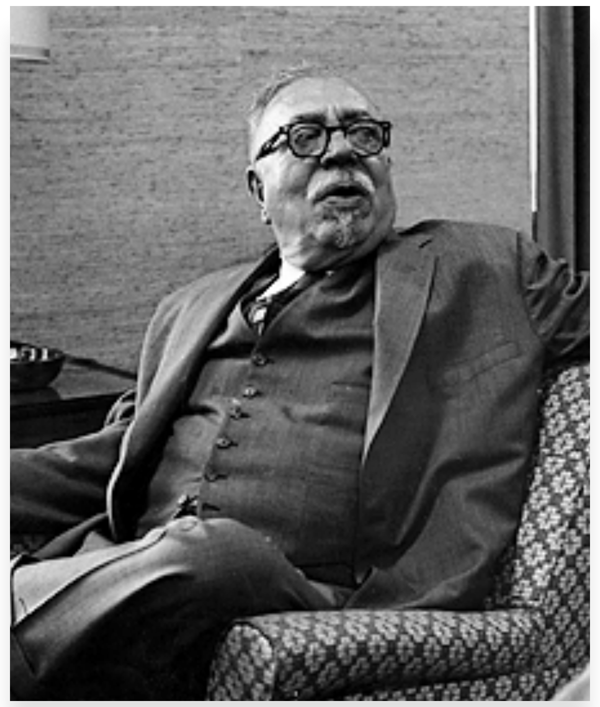
- Motivations: the greatest problem in science, CBMM, the MIT Quest
- A bit of history: Neuroscience and AI, Science and Engineering
- CBMM and the Quest
- AI ethics and its neural bases
- Theory: explaining how deep networks work and what are their properties and limitations.

Logical for MIT…

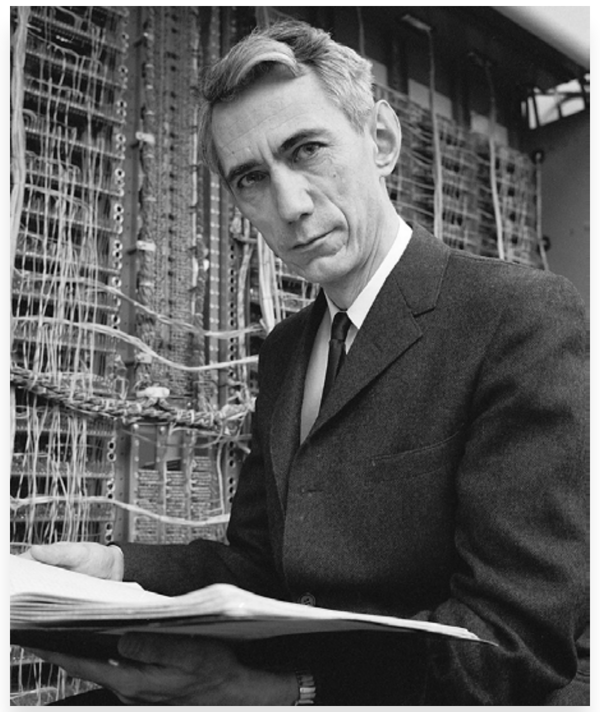
Marvin Minsky



Noam Chomsky



Norbert Wiener



Claude Shannon

“The Golden Age” 1950 - 1970

MIT Intelligence Initiative

CENTER FOR
Brains
Minds+
Machines

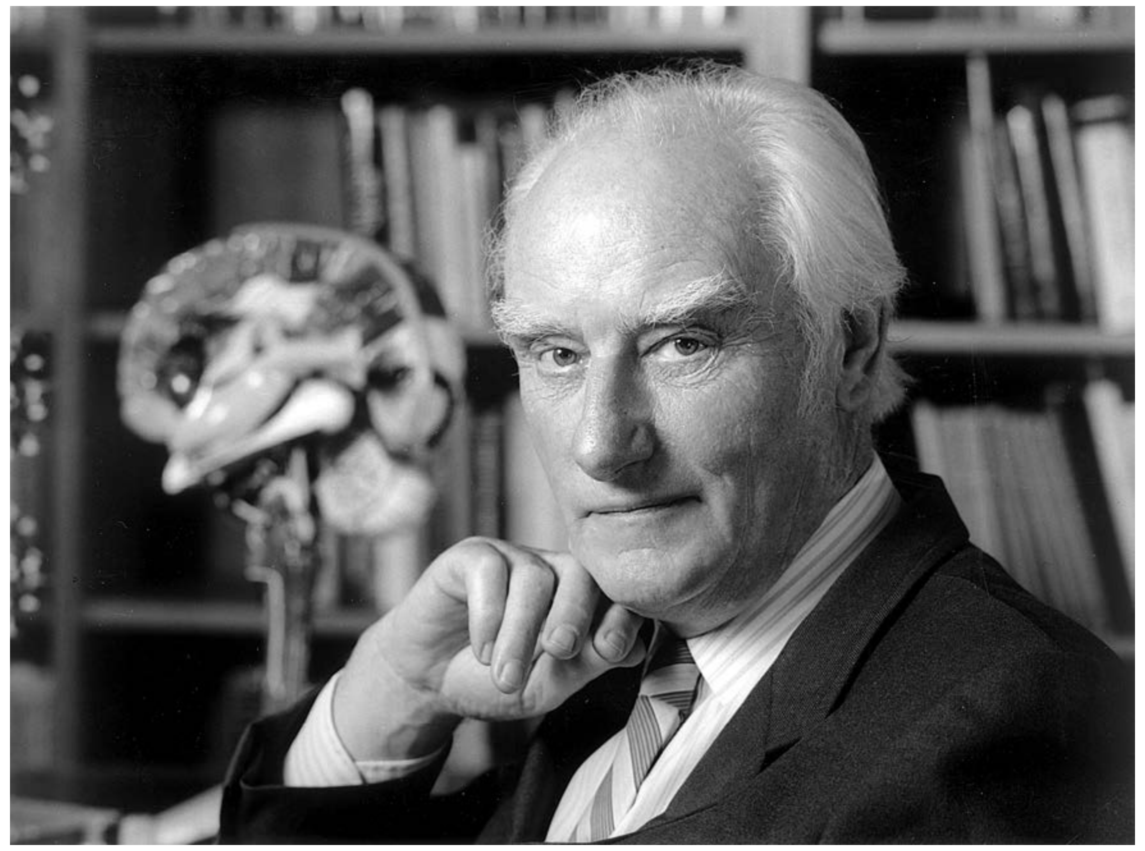
**Intelligence:
The MIT Quest**

2008

2012 - 2013

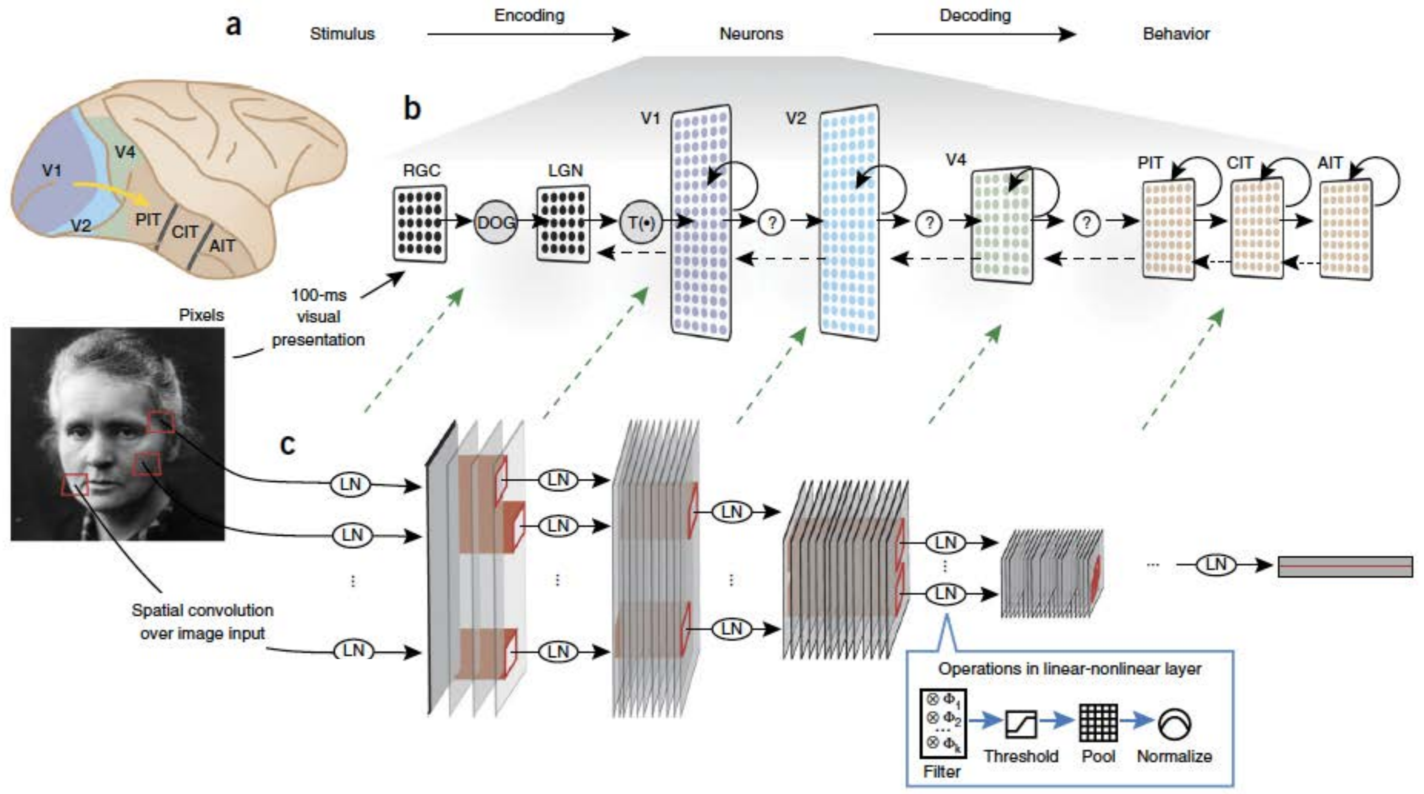
2018

Just a definition: I use the word *science* to mean *natural science*



CBMM's focus is the natural Science and the Engineering of Intelligence

We aim to make progress in understanding intelligence, that is in understanding how the brain makes the mind, how the brain works and how to build intelligent machines. We believe that the science of intelligence will enable better engineering of intelligence.



CENTER FOR
Brains
Minds+
Machines

Neuron
Review

Neuroscience-Inspired Artificial Intelligence

Demis Hassabis,[1,2,*] Dharshan Kumaran,[1,3] Christopher Summerfield,[1,4] and Matthew Botvinick[1,2]
[1]DeepMind, 5 New Street Square, London, UK
[2]Gatsby Computational Neuroscience Unit, 25 Howland Street, London, UK
[3]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, UK
[4]Department of Experimental Psychology, University of Oxford, Oxford, UK
*Correspondence: dhcontact@google.com
http://dx.doi.org/10.1016/j.neuron.2017.06.011

The fields of neuroscience and artificial intelligence (AI) have a long and intertwined history. In more recent times, however, communication and collaboration between the two fields has become less commonplace. In this article, we argue that better understanding biological brains could play a vital role in building intelligent machines. We survey historical interactions between the AI and neuroscience fields and emphasize current advances in AI that have been inspired by the study of neural computation in humans and other animals. We conclude by highlighting shared themes that may be key for advancing future research in both fields.

The successful transfer of insights gained from neuroscience to the development of AI algorithms is critically dependent on the interaction between researchers working in both these fields, with insights often developing through a continual handing back and forth of ideas between fields. In the future, we

Two Main Recent Success Stories in AI

FINANCIAL TIMES

ft.com/comment

You are signed in

Search for...

PERSON IN THE NEWS

March 11, 2016 3:14 pm

Demis Hassabis, master of the new machine age

Murad Ahmed

Share | Author alerts | Print | Clip | Comments

The creator of the AI game-playing program makes all the right moves, writes Murad Ahmed

The victories have a human mastermind in Demis Hassabis, co-founder and chief executive of DeepMind. He describes Mr Lee as the "Roger Federer of Go", and for some the computer program's achievement is akin to a robot taking to the lawns of Wimbledon and beating the legendary tennis champion.

"I think it is pretty huge but, ultimately, it will be for

More

PERSON IN THE NEWS

James Comey

Ali al-Naimi

Kyle Bass

THE BIG READ

EDF

TUNISIA


Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
KBA

Real Engineering: Mobileye

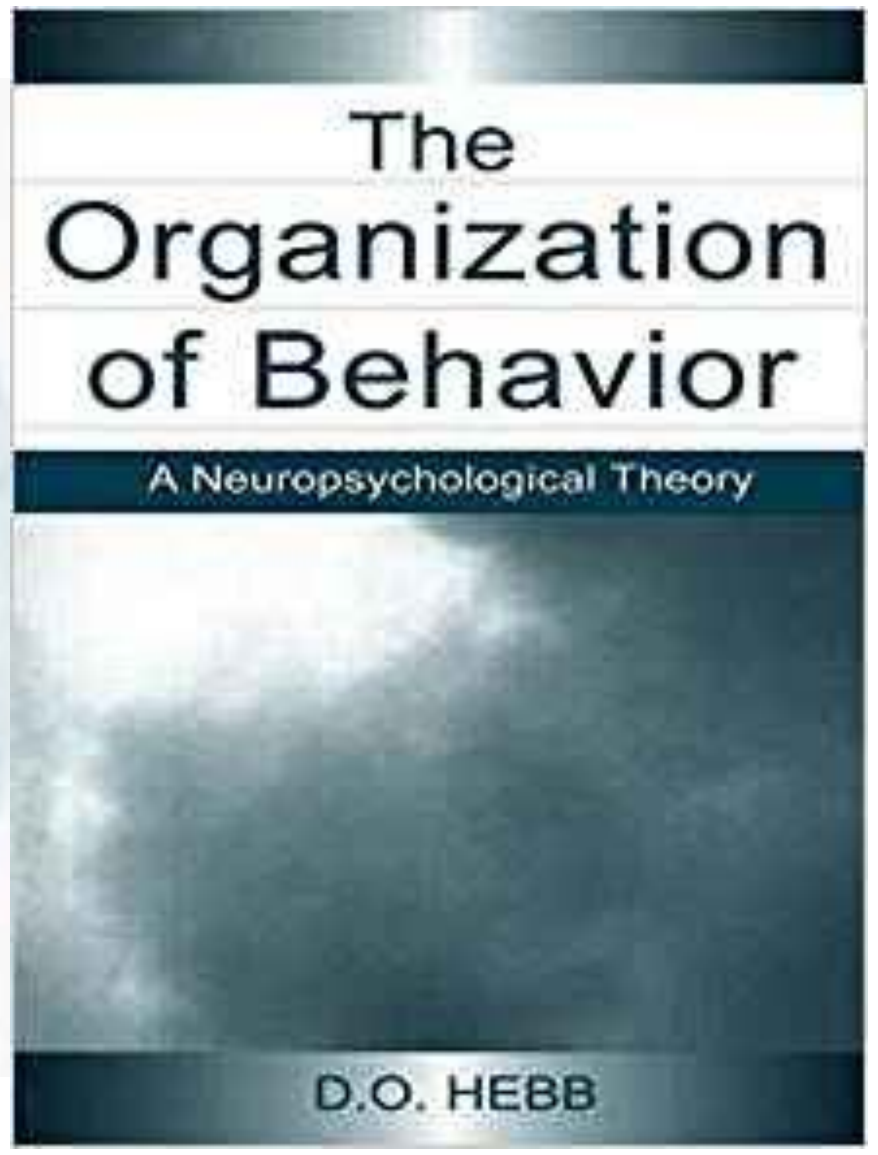
CENTER FOR
Brains
Minds+
Machines

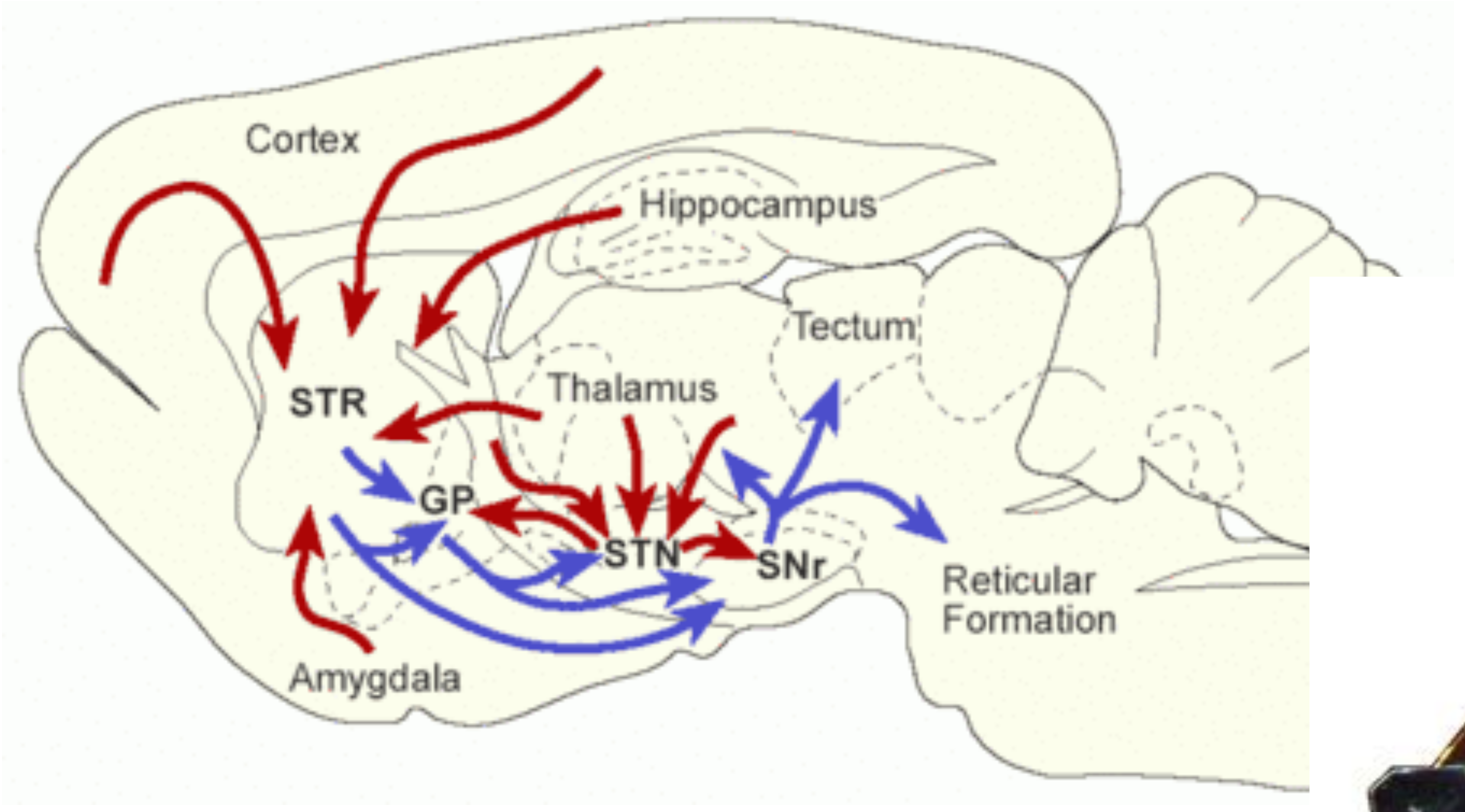
Moore-like law for ML (1995-2018)

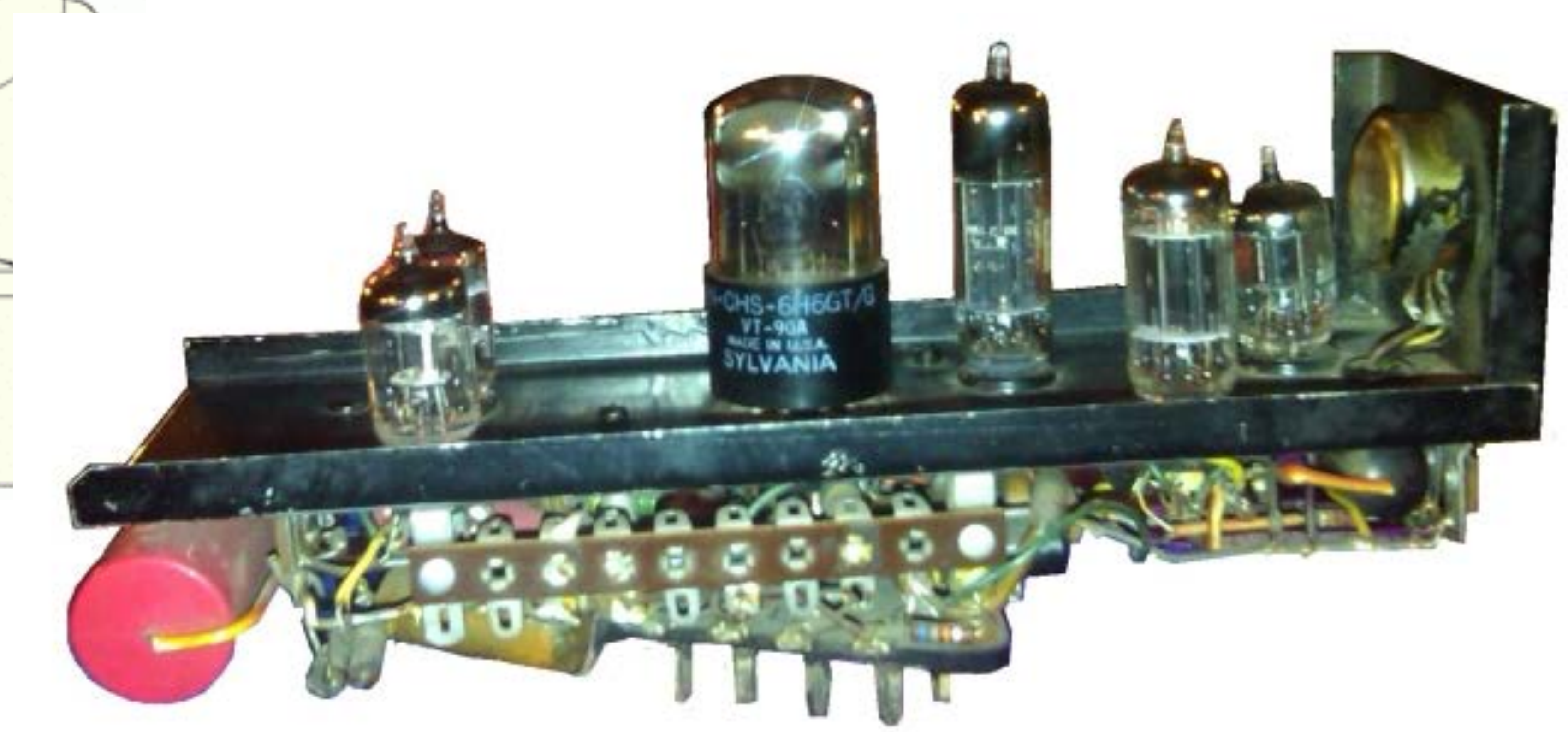
Two Main Recent Success Stories in AI

DL and RL come from neuroscience

RL







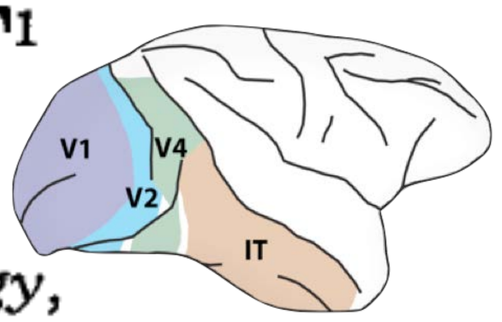
Minsky's SNARC

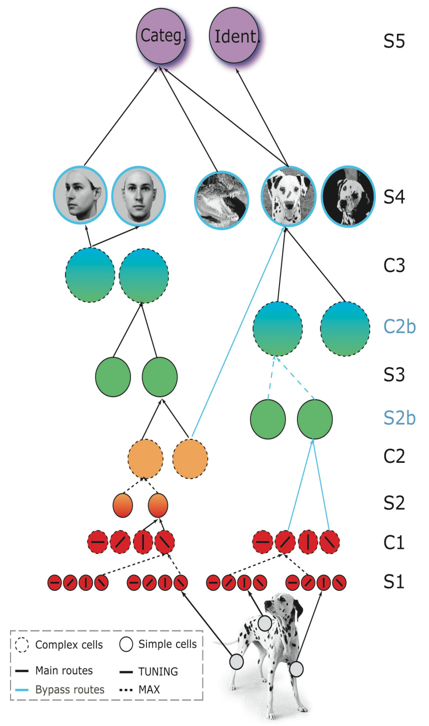
DL

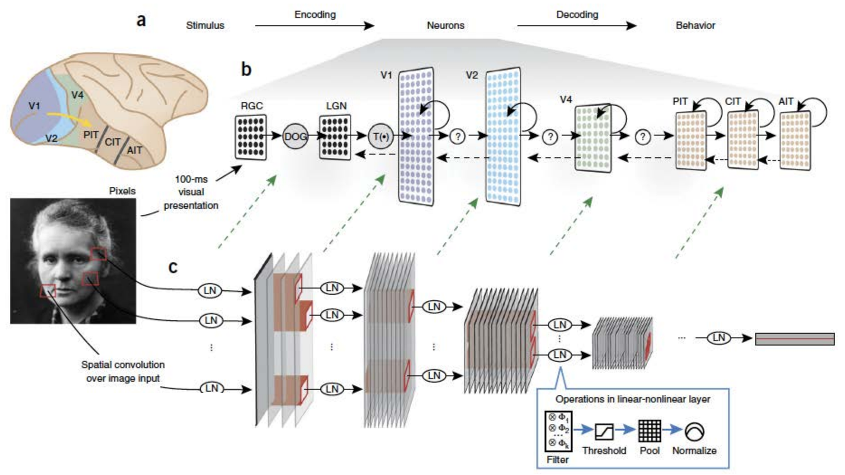
RECEPTIVE FIELDS AND FUNCTIONAL ARCHITECTURE IN TWO NONSTRIATE VISUAL AREAS (18 AND 19) OF THE CAT[1]

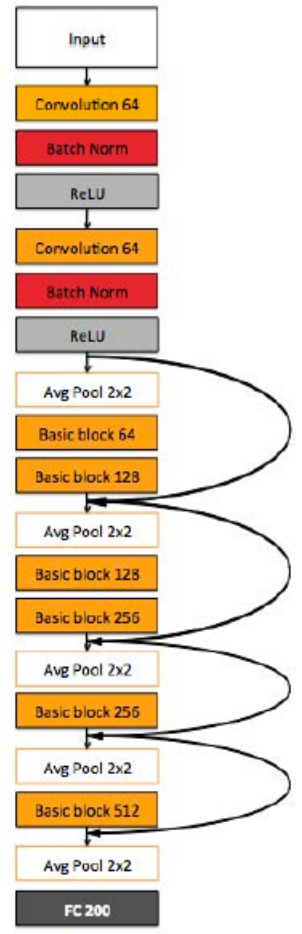
DAVID H. HUBEL AND TORSTEN N. WIESEL
Neurophysiology Laboratory, Department of Pharmacology, Harvard Medical School, Boston, Massachusetts

(Received for publication August 24, 1964)





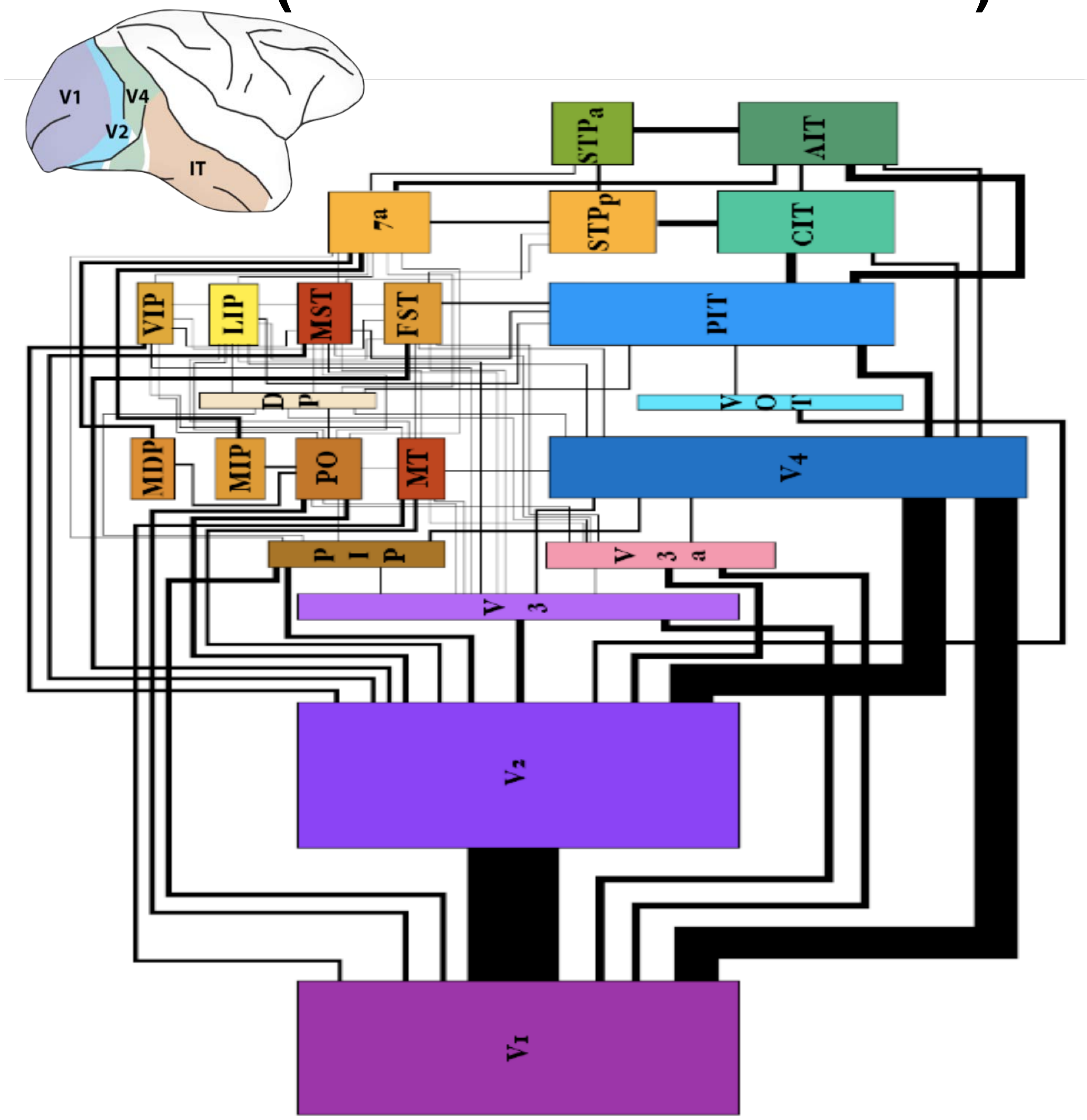




Background: State-of-the-art Machines (“Deep Learning”) Have Emerged From the Brain’s Visual Processing Architecture

Brains / Minds

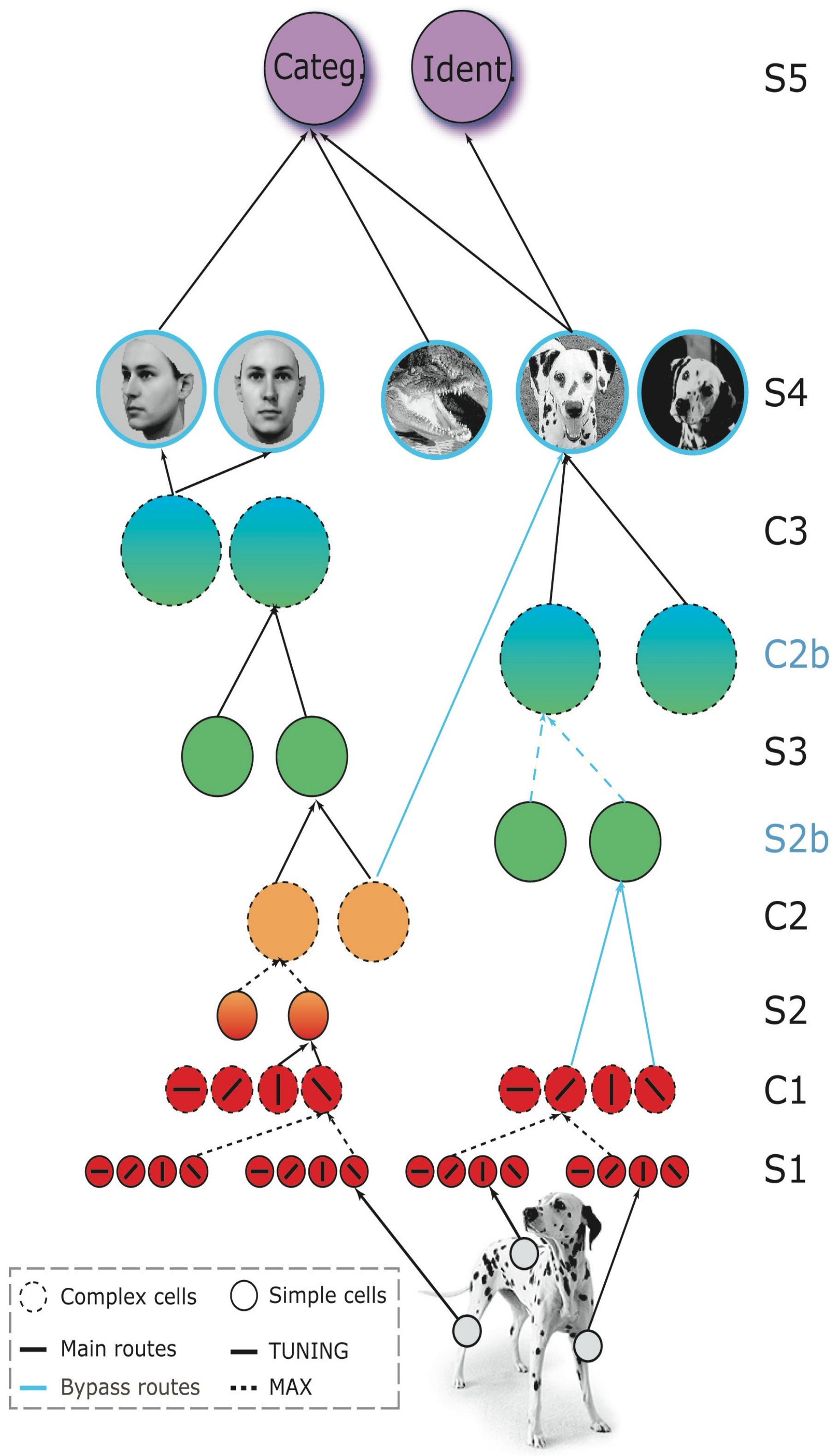
(ventral visual stream)



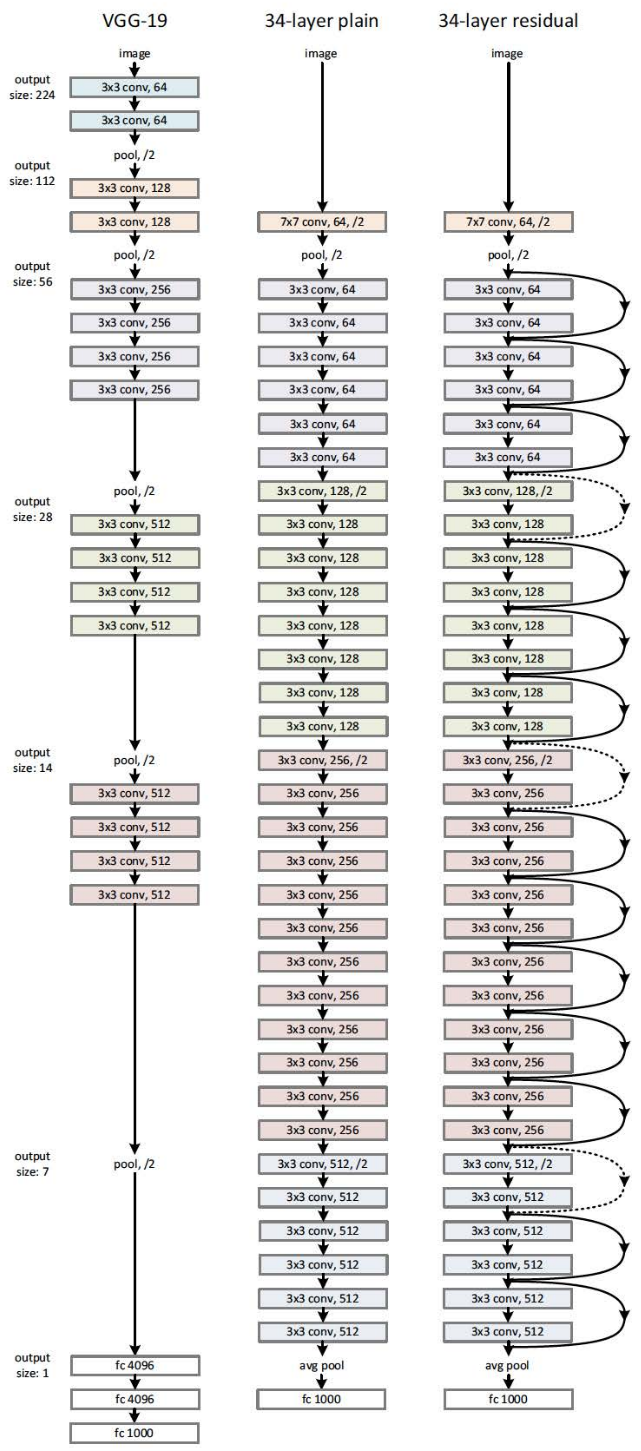
Desimone & Ungerleider 1989; vanEssen+Movshon

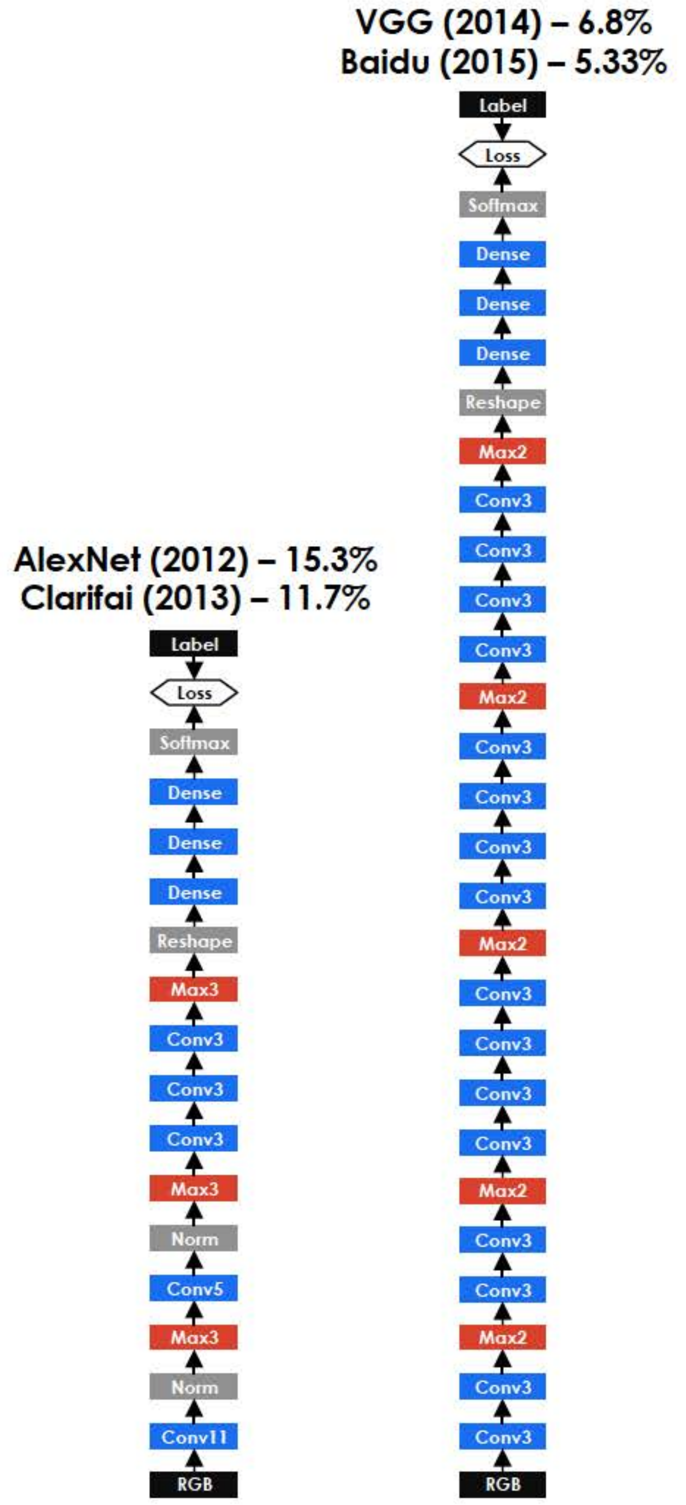
What’s the engineering of the future?

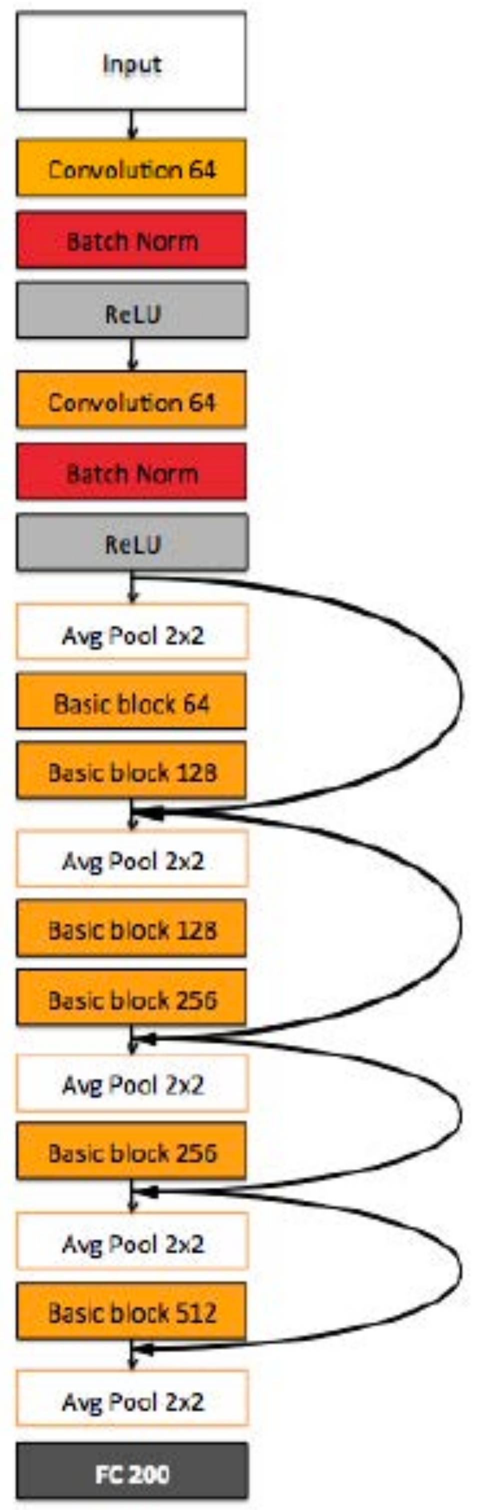
Machines → …



State of the Art ResNets

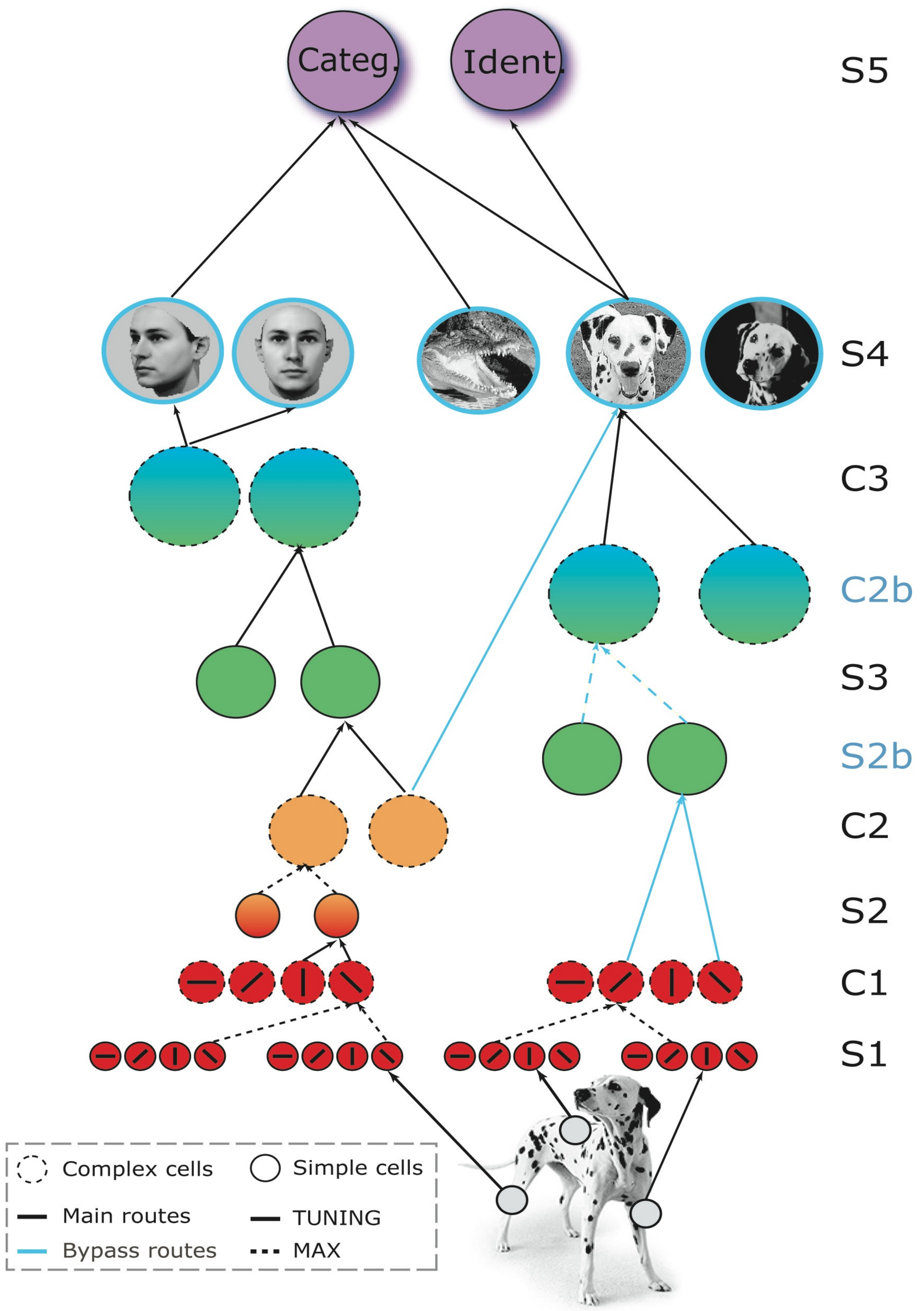






CENTER FOR
Brains
Minds+
Machines

Convolutional networks



"Hubel-Wiesel" models include

Hubel & Wiesel, 1959: Fukushima, 1980, Wallis & Rolls, 1997; Mel, 1997; LeCun et al 1998; Riesenhuber & Poggio, 1999; Thorpe, 2002; Ullman et al., 2002; Wersing and Koerner, 2003; Serre et al., 2007; Freeman and Simoncelli, 2011….

Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007

Overview

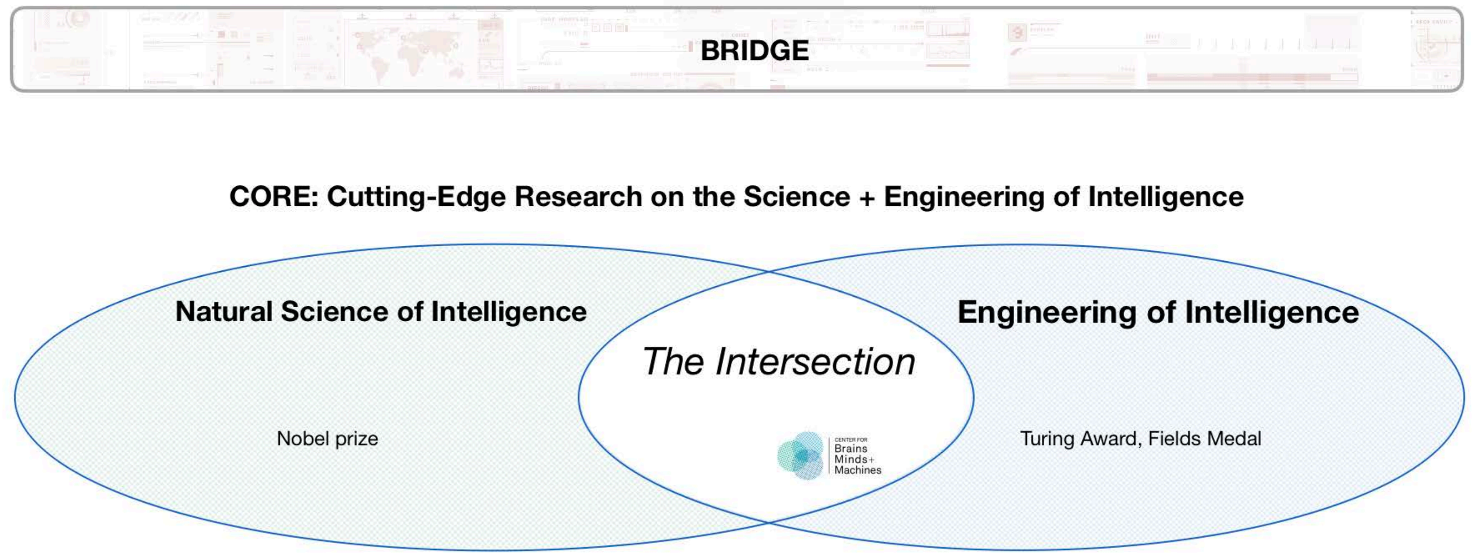
- Motivations: the greatest problem in science, CBMM, the MIT Quest
- A bit of history: Neuroscience and AI, Science and Engineering
- CBMM and the Quest
- AI ethics and its neural bases
- Theory: explaining how deep networks work and what are their properties and limitations.

Visual intelligence, *video ergo sum*

MODULE FOUR

Visual Intelligence

Within The CORE Intersection:
CBMM + additional "moonshot" projects



- *Visual Intelligence (CBMM)*
- *Development of Intelligence*
- *New circuits for deep nets in counter streams in cortical areas*
- *Planning and imagination*
- *Emotional Intelligence*
- *Language*

Overview

- Motivations: the greatest problem in science, CBMM, the MIT Quest
- A bit of history: Neuroscience and AI, Science and Engineering
- CBMM and the Quest
- AI ethics and its neural bases
- Theory: explaining how deep networks work and what are their properties and limitations.

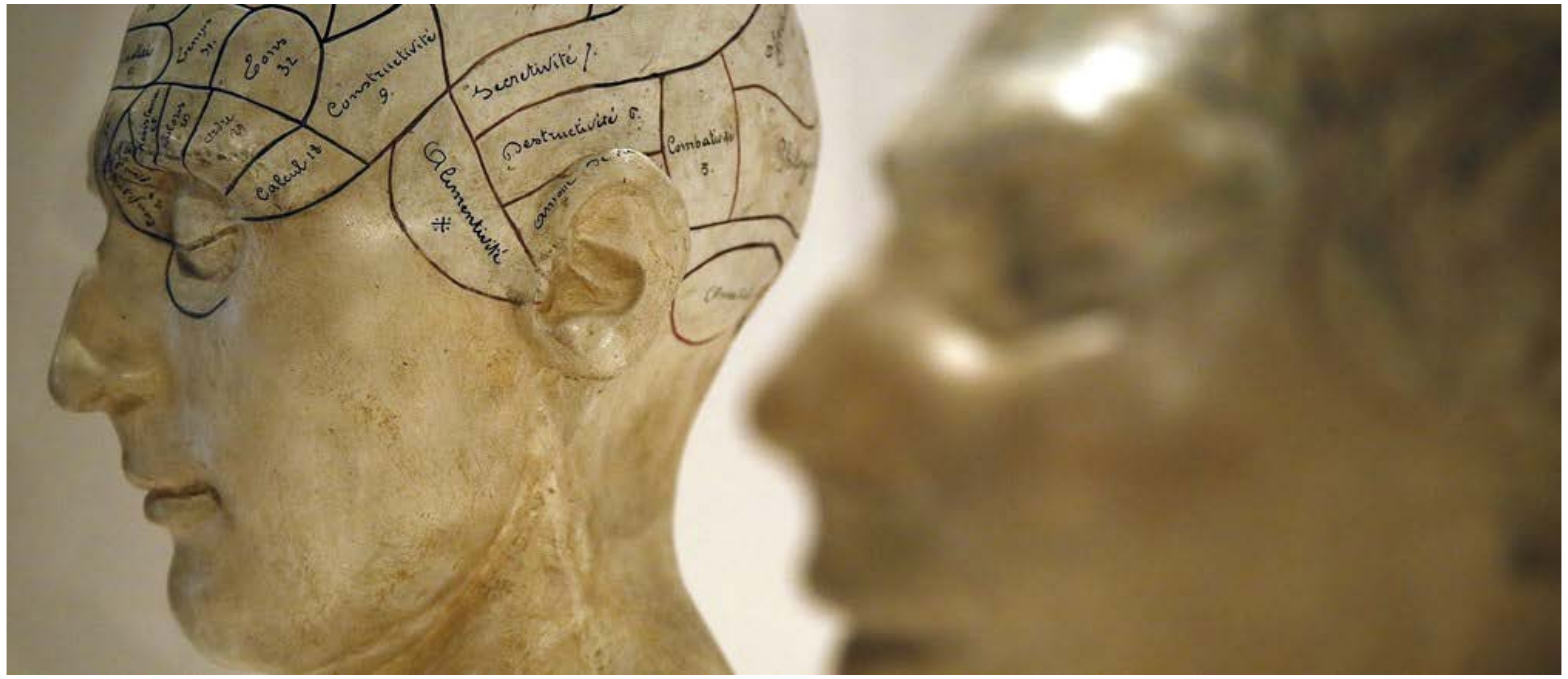
AI and ethics

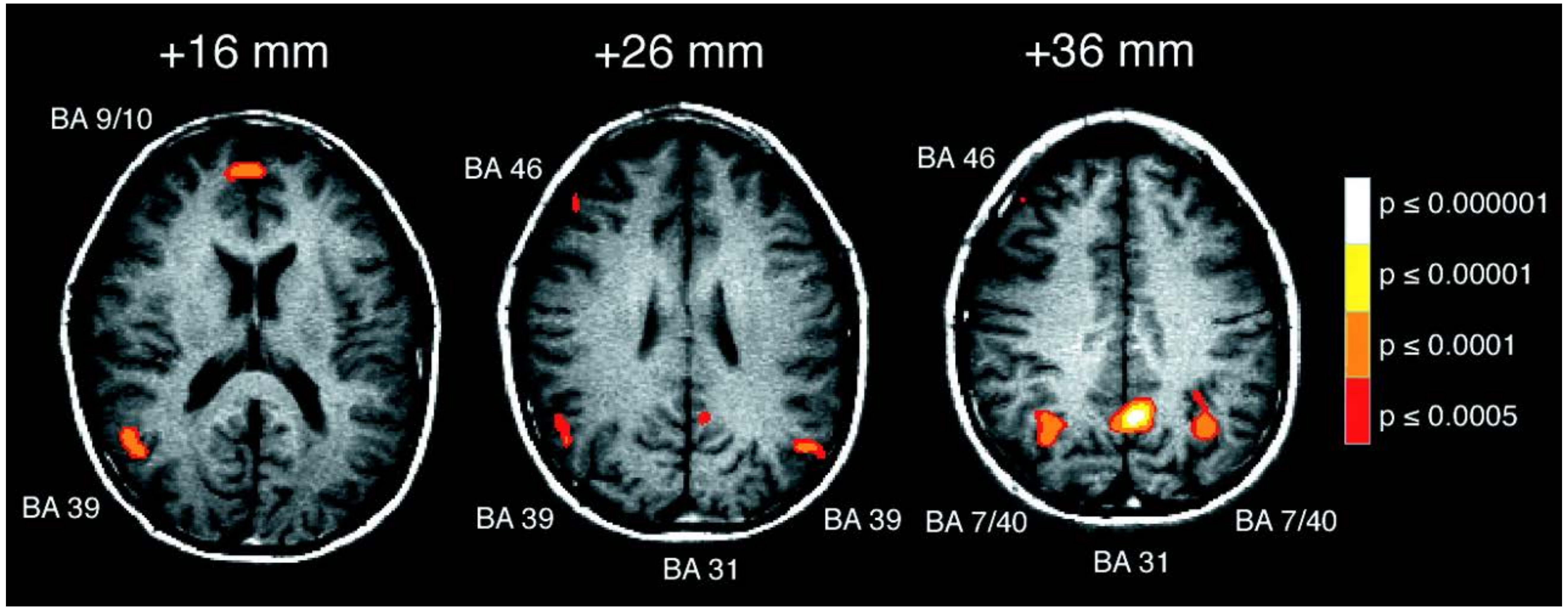
- Too much about
  - AI more dangerous than nuclear bombs
  - the trolley problem

- More pressing issues:
  - What to publish/not publish
  - Jobs lost to machines

- Future:
  - how to build ethical machines
  - can the brain teach us how?

Neuroscience of ethics



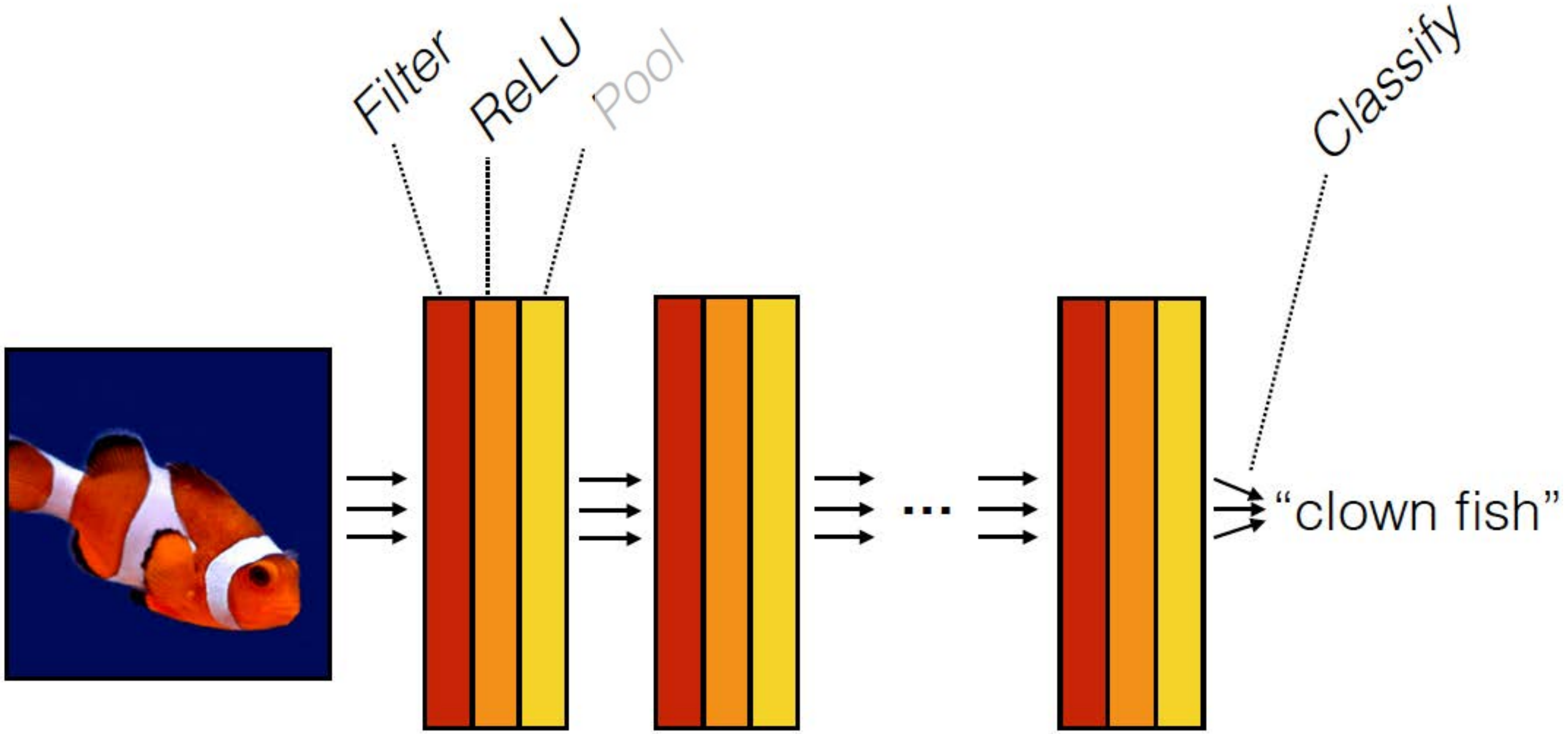


Studies with fMRI revealed that particular areas of the brain are associated with particular cognitive events such as our moral emotions and ethical reasoning.

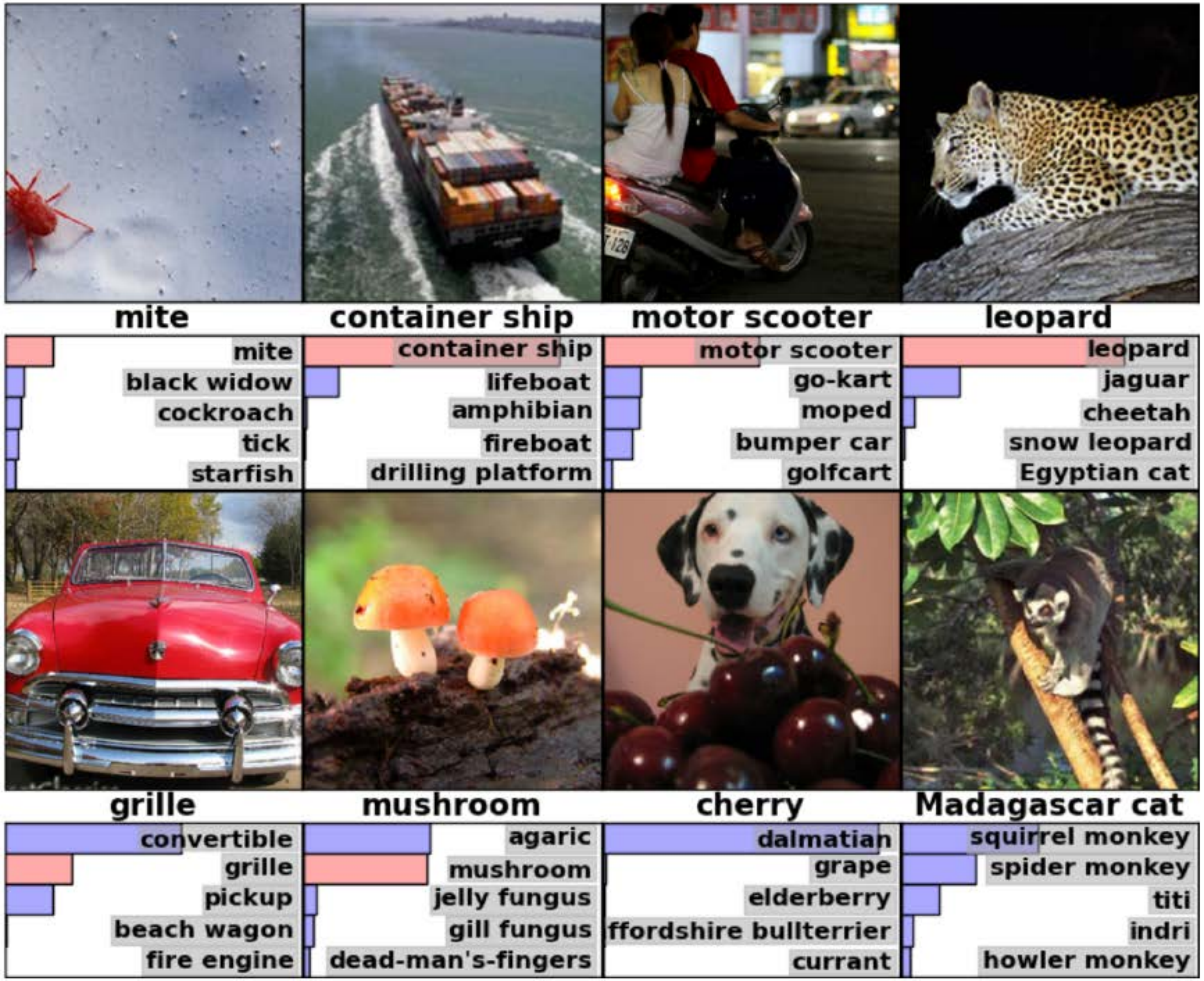
Overview

- Motivations: the greatest problem in science, CBMM, the MIT Quest
- A bit of history: Neuroscience and AI, Science and Engineering
- CBMM and the Quest
- AI ethics and its neural bases
- Theory: explaining how deep networks work and what are their properties and limitations.

Computation in a neural net



$$f(\mathbf{x}) = f_L(\dots f_2(f_1(\mathbf{x})))$$



Krizhevsky et al. NIPS 2012

WGBH | 1708.07120.pdf | Center for Brains, Minds and M | Theories of Deep Learning (ST

tp@ai.mit.edu

Secure | https://stats385.github.io

Theories of Deep Learning (STATS 385)

Stanford University, Fall 2017

The spectacular recent successes of deep learning are purely empirical. Nevertheless intellectuals always try to explain important developments theoretically. In this literature course we will review recent work of Bruna and Mallat, Mhaskar and Poggio, Papyan and Elad, Bolcskei and co-authors, Baraniuk and co-authors, and others, seeking to build theoretical frameworks deriving deep networks as consequences. After initial background lectures, we will have some of the authors presenting lectures on specific papers. This course meets once weekly.

Instructors:

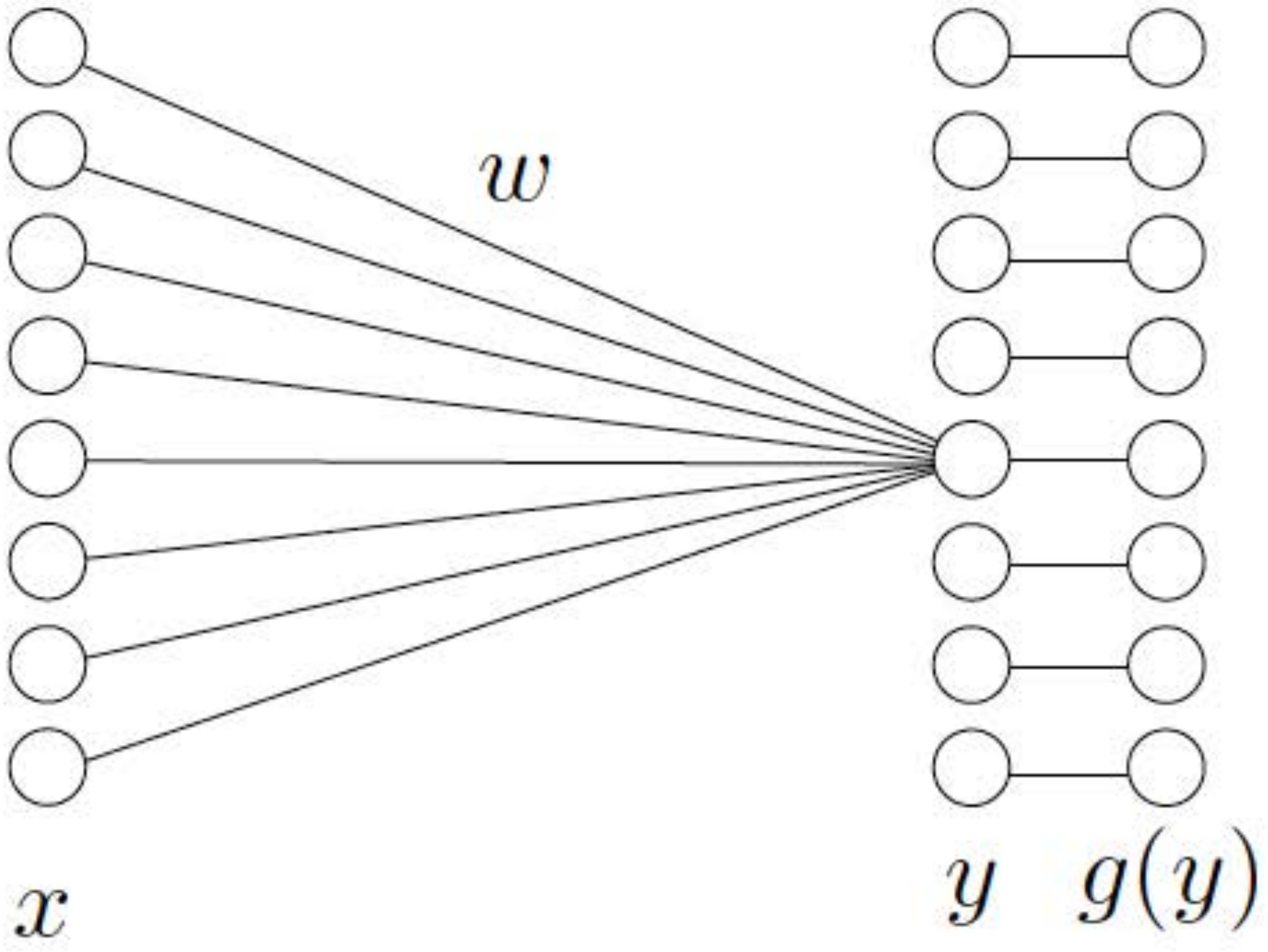
David Donoho

Hatef Monajemi

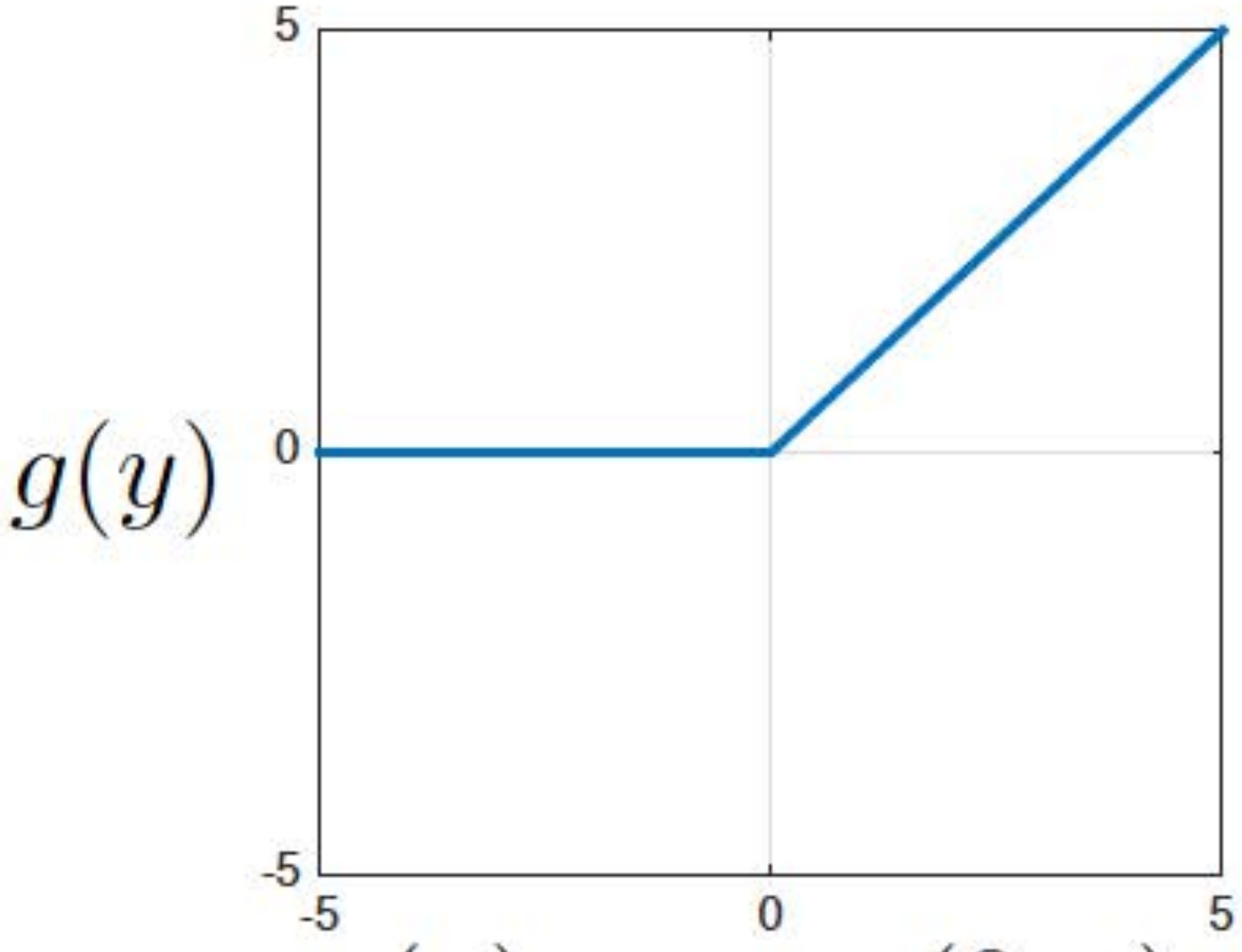
Vardan Papyan

Deep nets : a theory is needed (after alchemy, chemistry)

Computation in a neural net

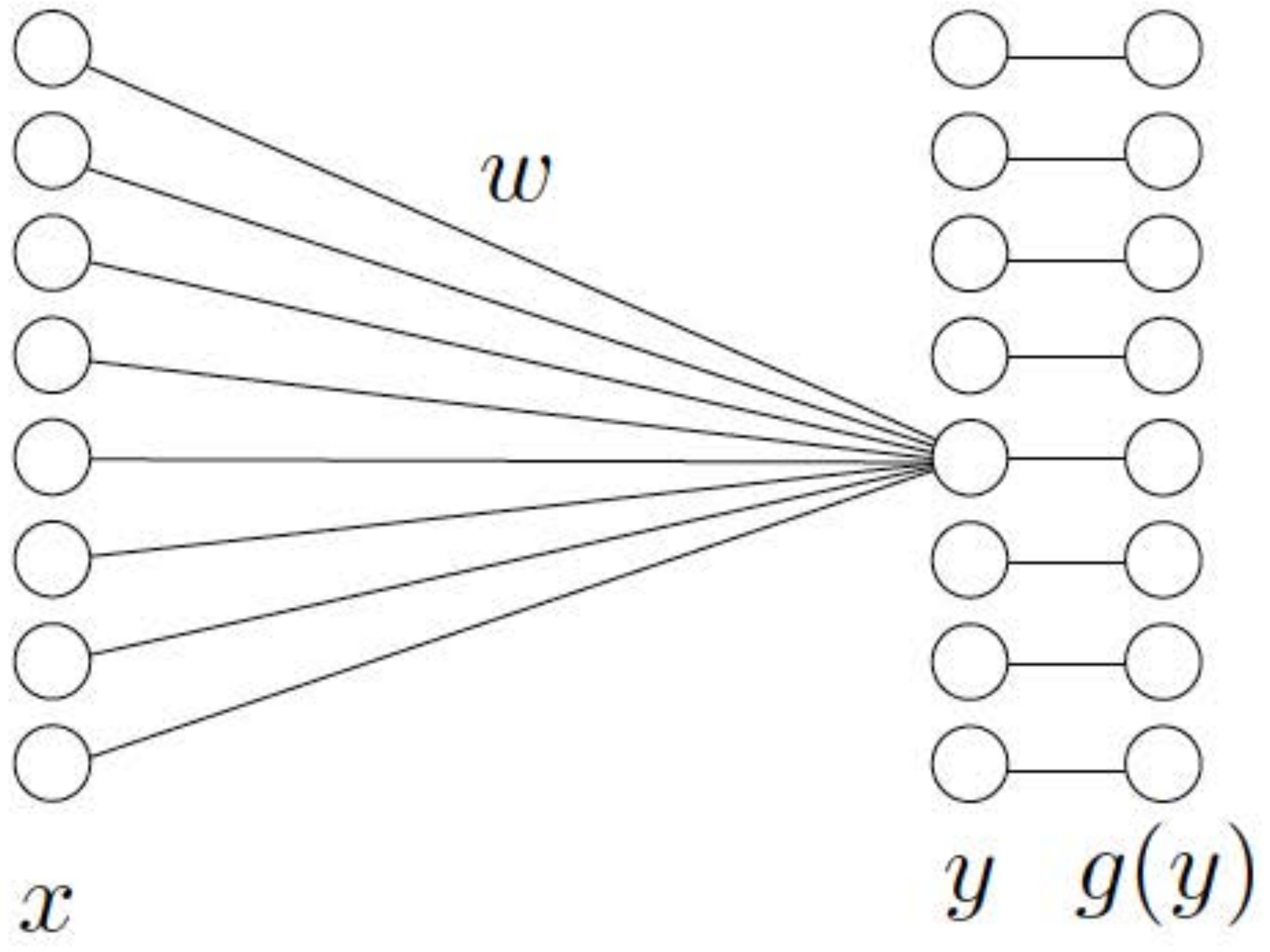


Rectified linear unit (ReLU)

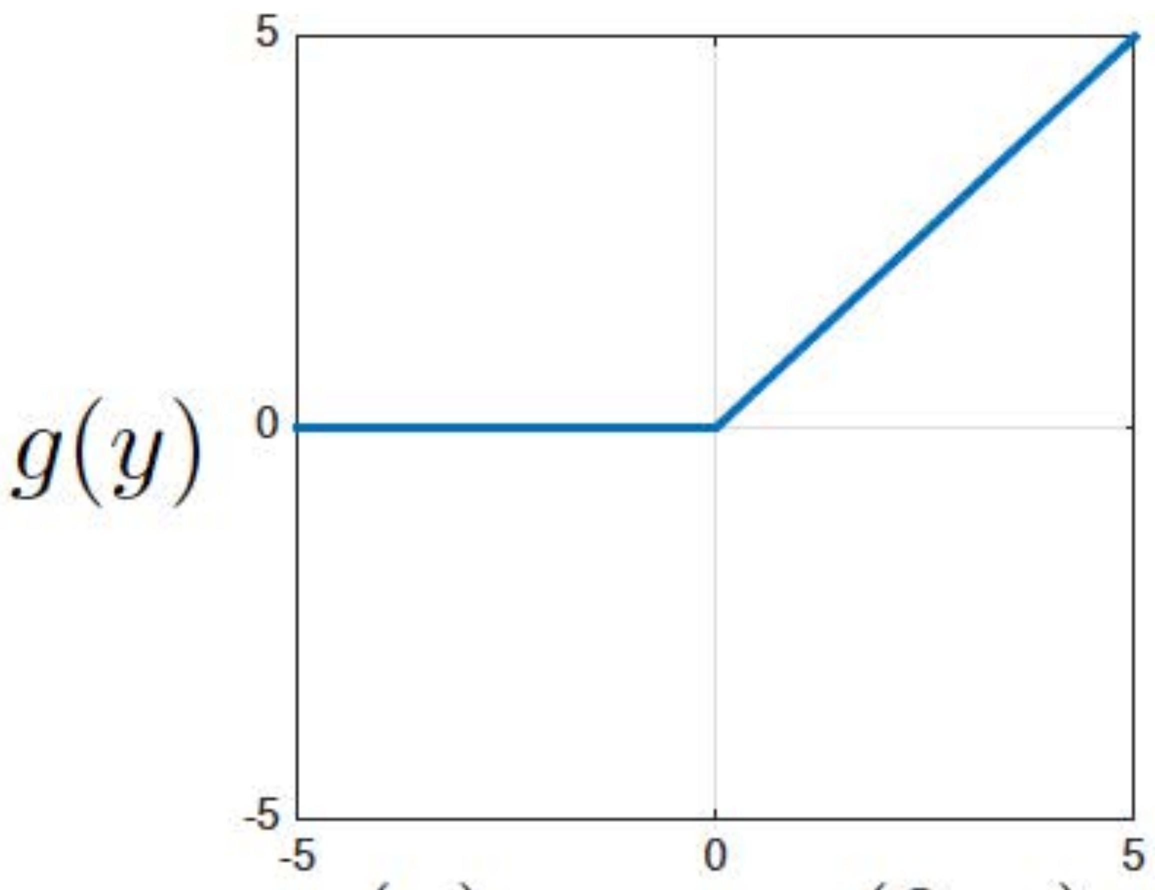


$$g(y) = \max(0, y)$$

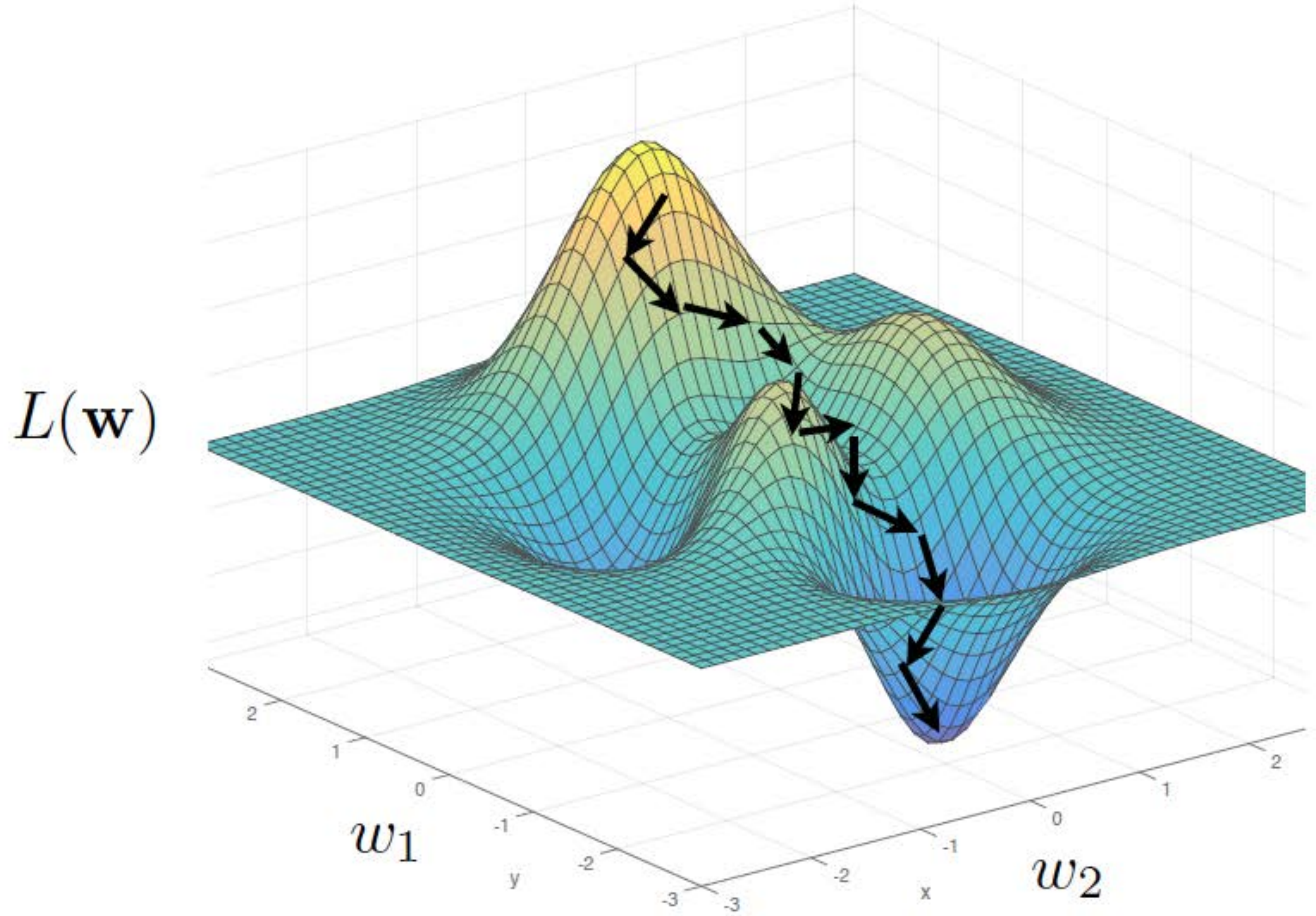
Deep nets architecture and SGD training



Rectified linear unit (ReLU)



$$g(y) = \max(0, y)$$



CENTER FOR
Brains
Minds+
Machines

Gradient descent

$$\boxed{\operatorname*{argmin}_{\mathbf{w}} \quad \sum_i \ell(\mathbf{z}_i, f(\mathbf{x}_i; \mathbf{w})) = L(\mathbf{w})}$$

One iteration of gradient descent:

$$\mathbf{w}^{t+1} = \mathbf{w}^{t} - \eta_t \frac{\partial L(\mathbf{w}^{\mathbf{t}})}{\partial \mathbf{w}}$$

learning rate

DLNNs: three main scientific questions

Approximation theory: when and why are deep networks better - no curse of dimensionality — than shallow networks?

Optimization: what is the landscape of the empirical risk?

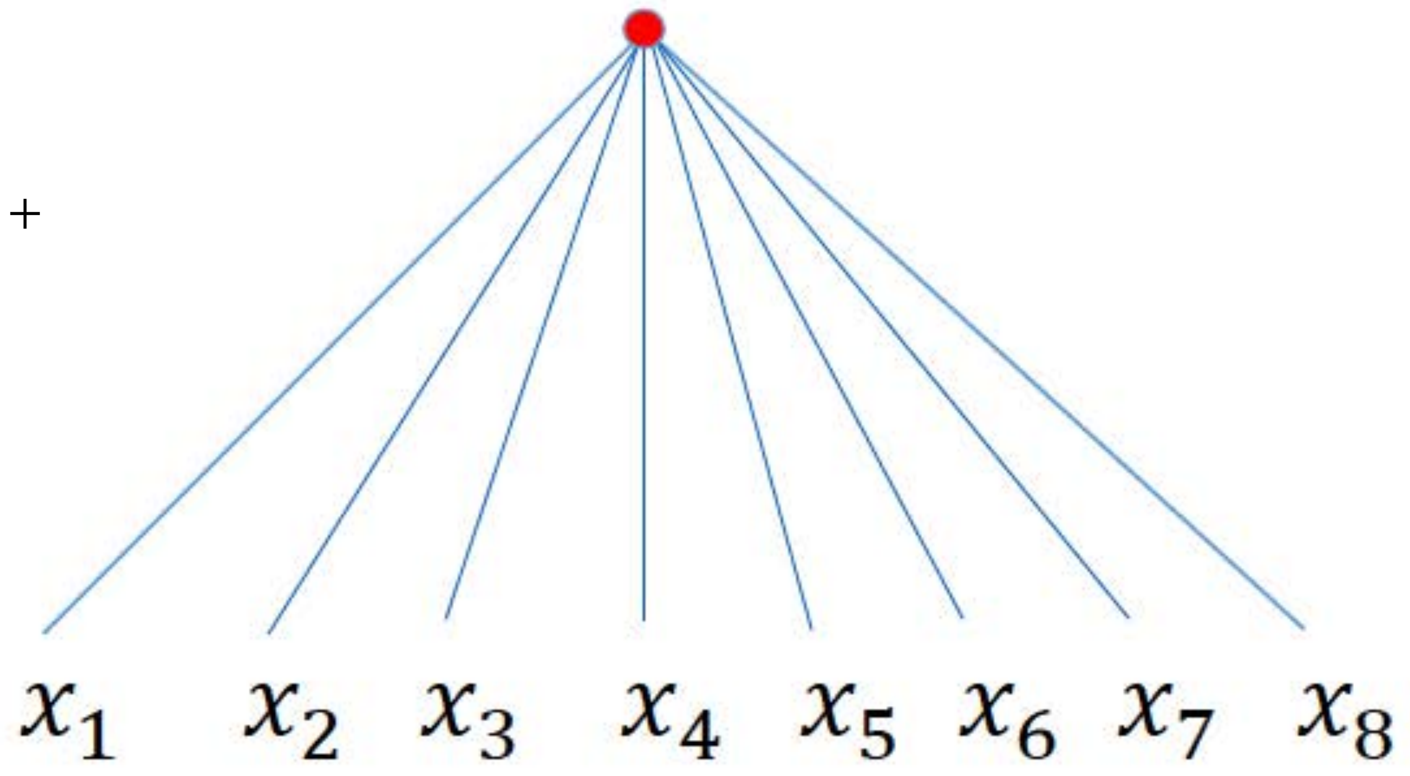
Generalization by SGD: how can overparametrized networks generalize?

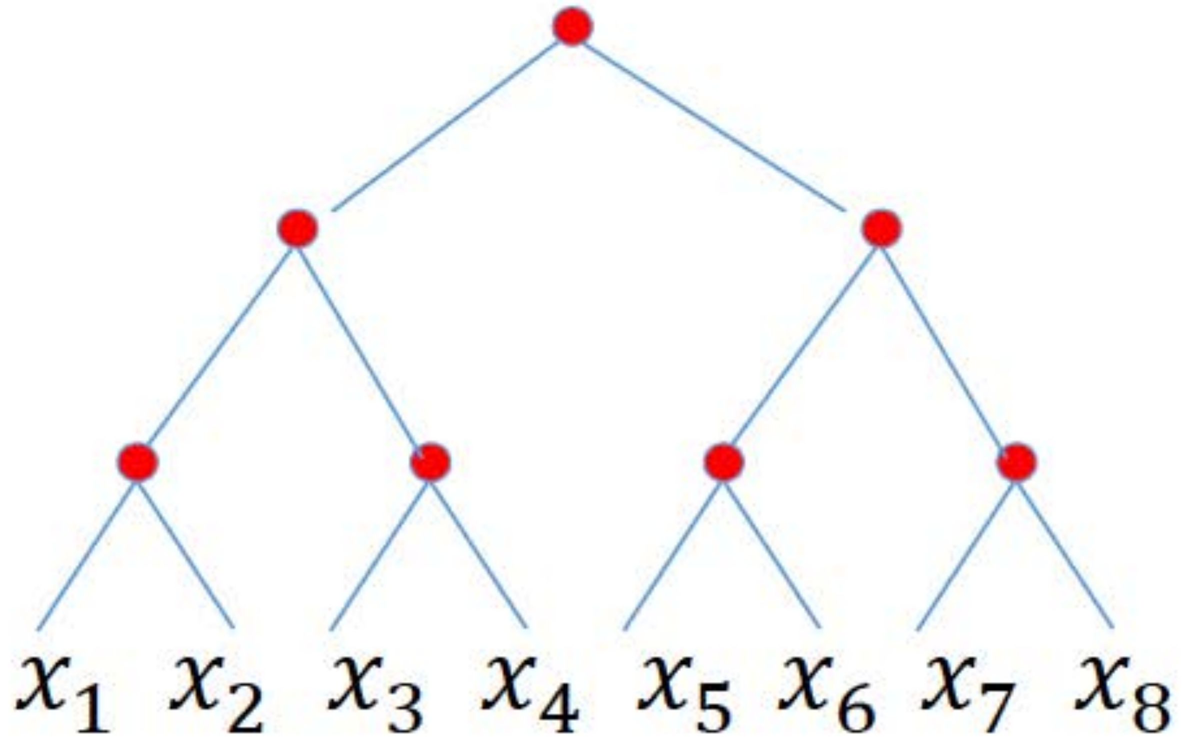
Work with Hrushikeshl Mhaskar+Lorenzo Rosasco+Fabio Anselmi+Chiyuan Zhang+Qianli Liao +Sasha Rakhlin + Noah G + Xavier B

Theory I:
Why and when are deep networks better than shallow networks?

$$f(x_1,x_2,\ldots,x_8) = g_3(g_{21}(g_{11}(x_1,x_2),g_{12}(x_3,x_4))g_{22}(g_{11}(x_5,x_6),g_{12}(x_7,x_8)))$$

$$g(x) = \sum_{i=1}^{r} c_i \left| < w_i , x > + b_i \right|_+$$





Theorem (informal statement)

Suppose that a function of d variables is compositional . Both shallow and deep network can approximate f equally well. The number of parameters of the shallow network depends exponentially on d as $O(\varepsilon^{-d})$with the dimension whereas for the deep network dance is dimension independent, i.e. $O(\varepsilon^{-2})$

CENTER FOR
Brains
Minds+
Machines

Mhaskar, Poggio, Liao, 2016

When can the curse of dimensionality be avoided

Generic functions

$$f(x_1, x_2, \ldots, x_8)$$

Compositional functions

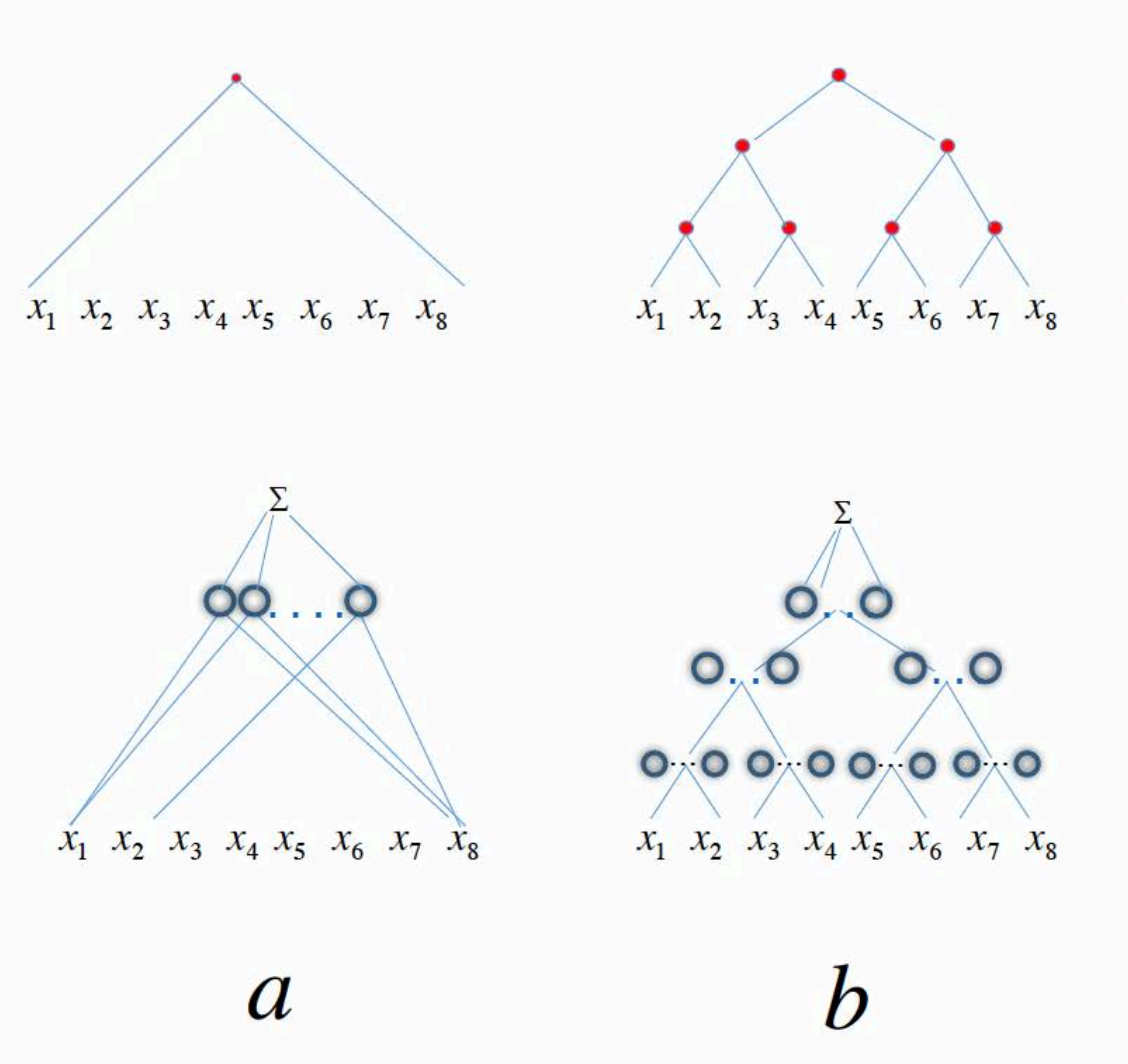
$$f(x_1, x_2, \ldots, x_8) = g_3(g_{21}(g_{11}(x_1, x_2), g_{12}(x_3, x_4))g_{22}(g_{11}(x_5, x_6), g_{12}(x_7, x_8)))$$

Mhaskar, Poggio, Liao, 2016

Microstructure of compositionality

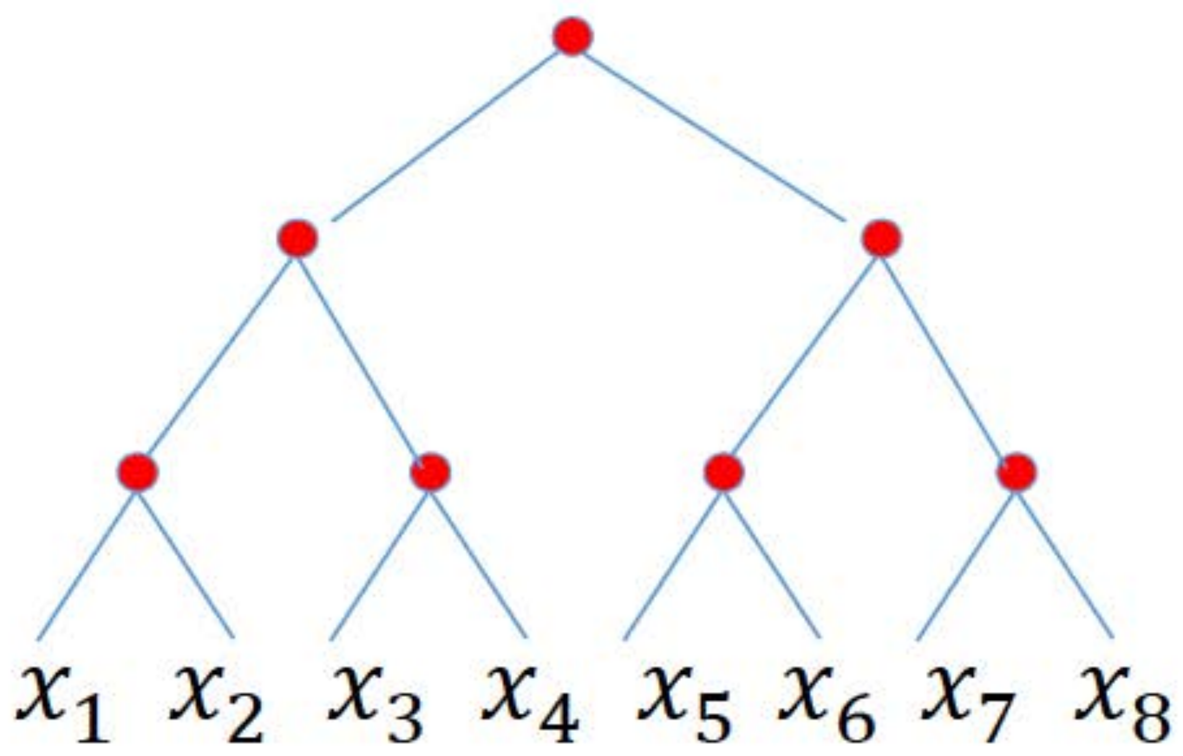
target function

approximating
function/network



Hierarchically local compositionality

$$f(x_1,x_2,...,x_8)=g_3(g_{21}(g_{11}(x_1,x_2),g_{12}(x_3,x_4))g_{22}(g_{11}(x_5,x_6),g_{12}(x_7,x_8)))$$



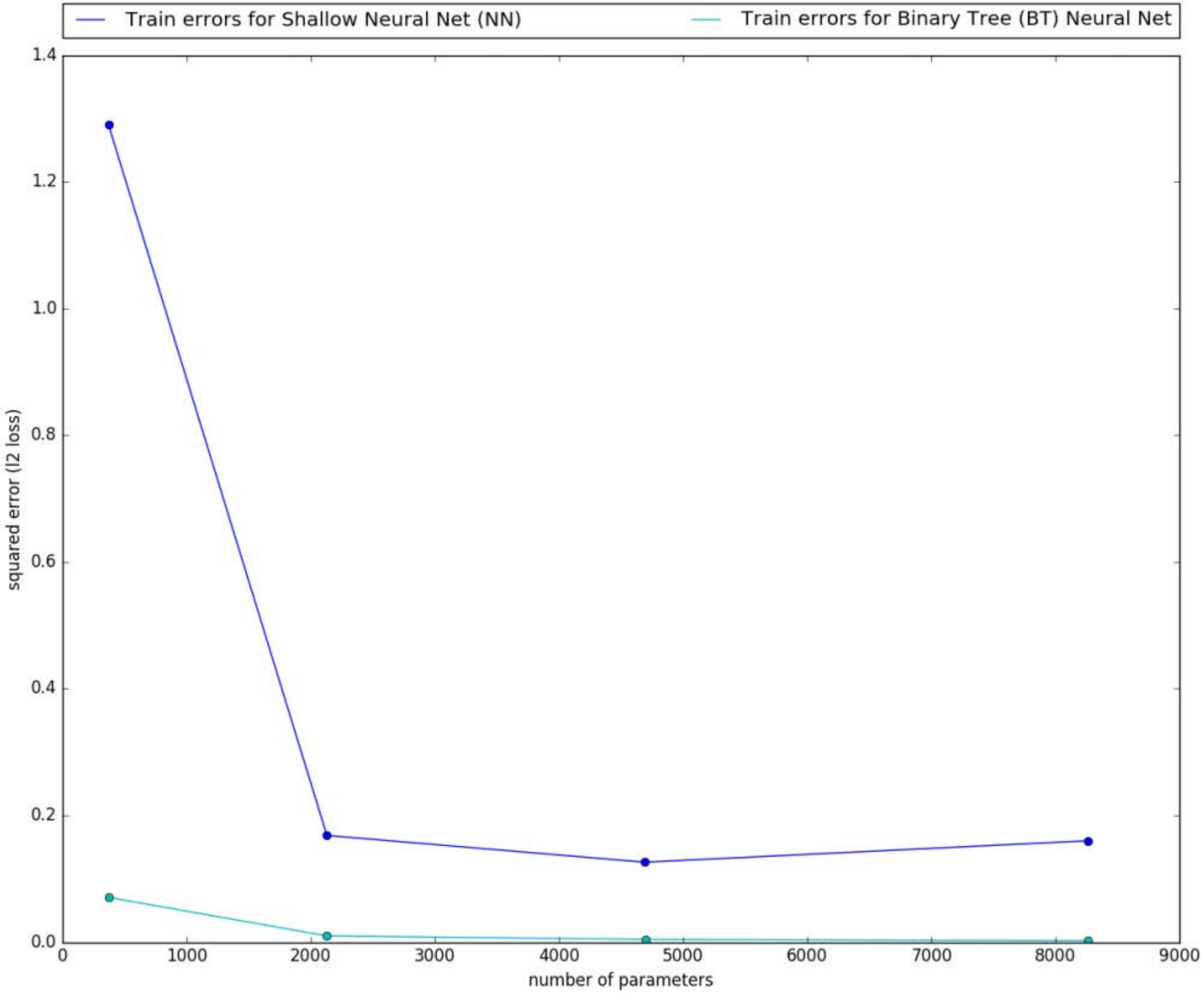
Theorem (informal statement)

Suppose that a function of d variables is hierarchically, locally, compositional . Both shallow and deep network can approximate f equally well. The number of parameters of the shallow network depends exponentially on d as $O(\varepsilon^{-d})$ with the dimension whereas for the deep network dance is $O(d\varepsilon^{-2})$

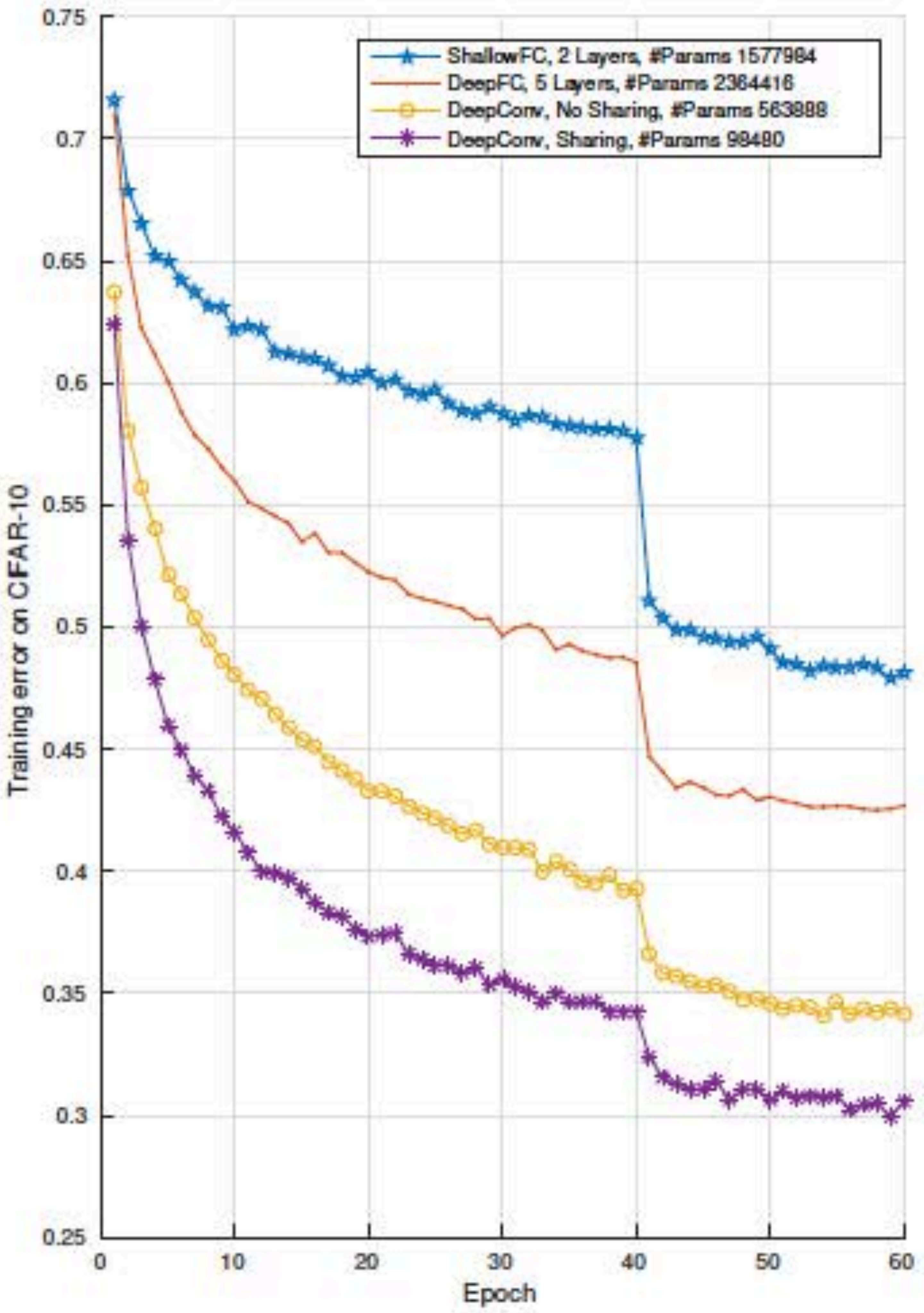
CENTER FOR Brains Minds+ Machines

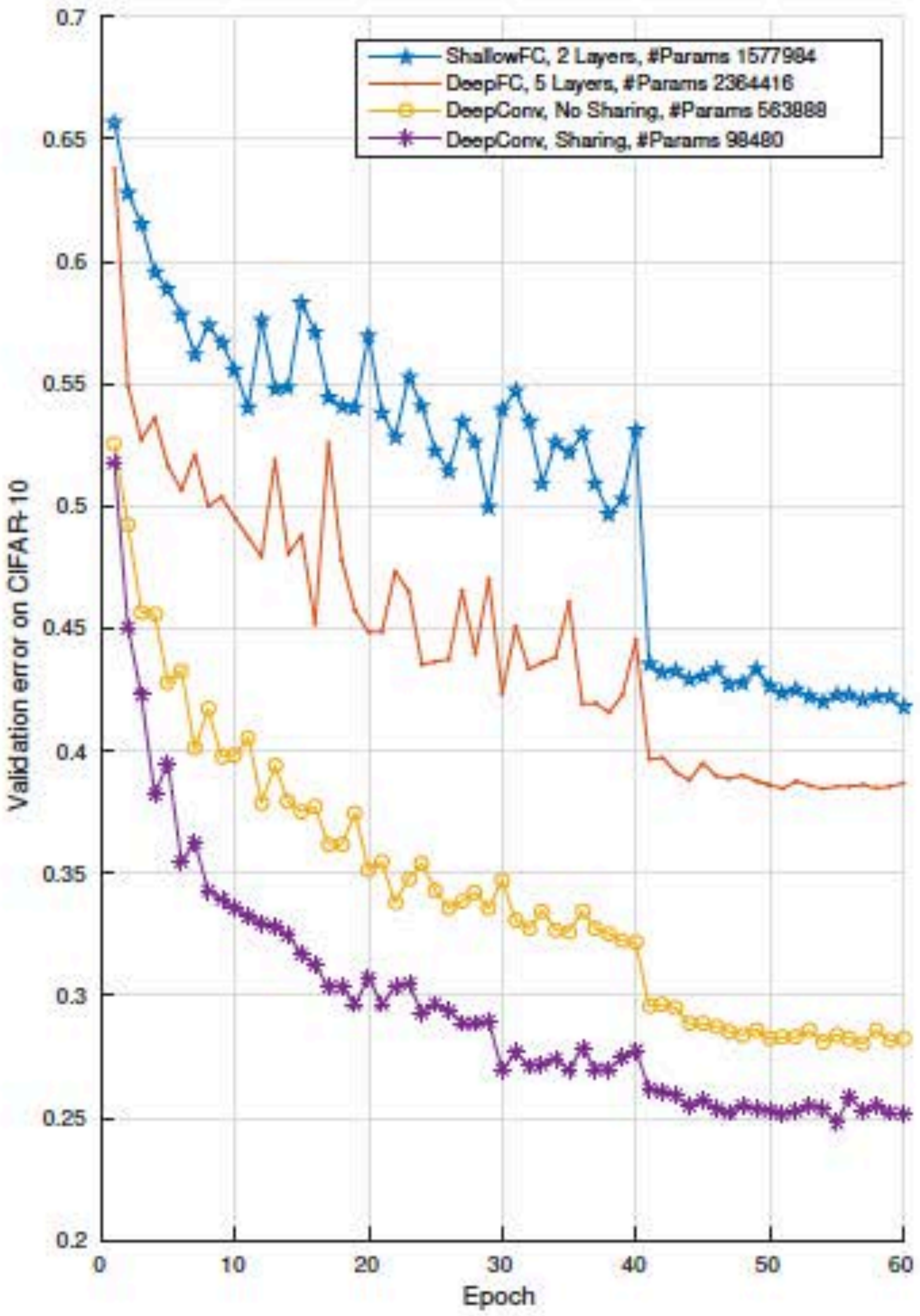
Mhaskar, Poggio, Liao, 2016

Binary Tree NN vs Shallow NN 8D



Locality of constituent functions is key **not** weight sharing: CIFAR





Why are compositional functions important?

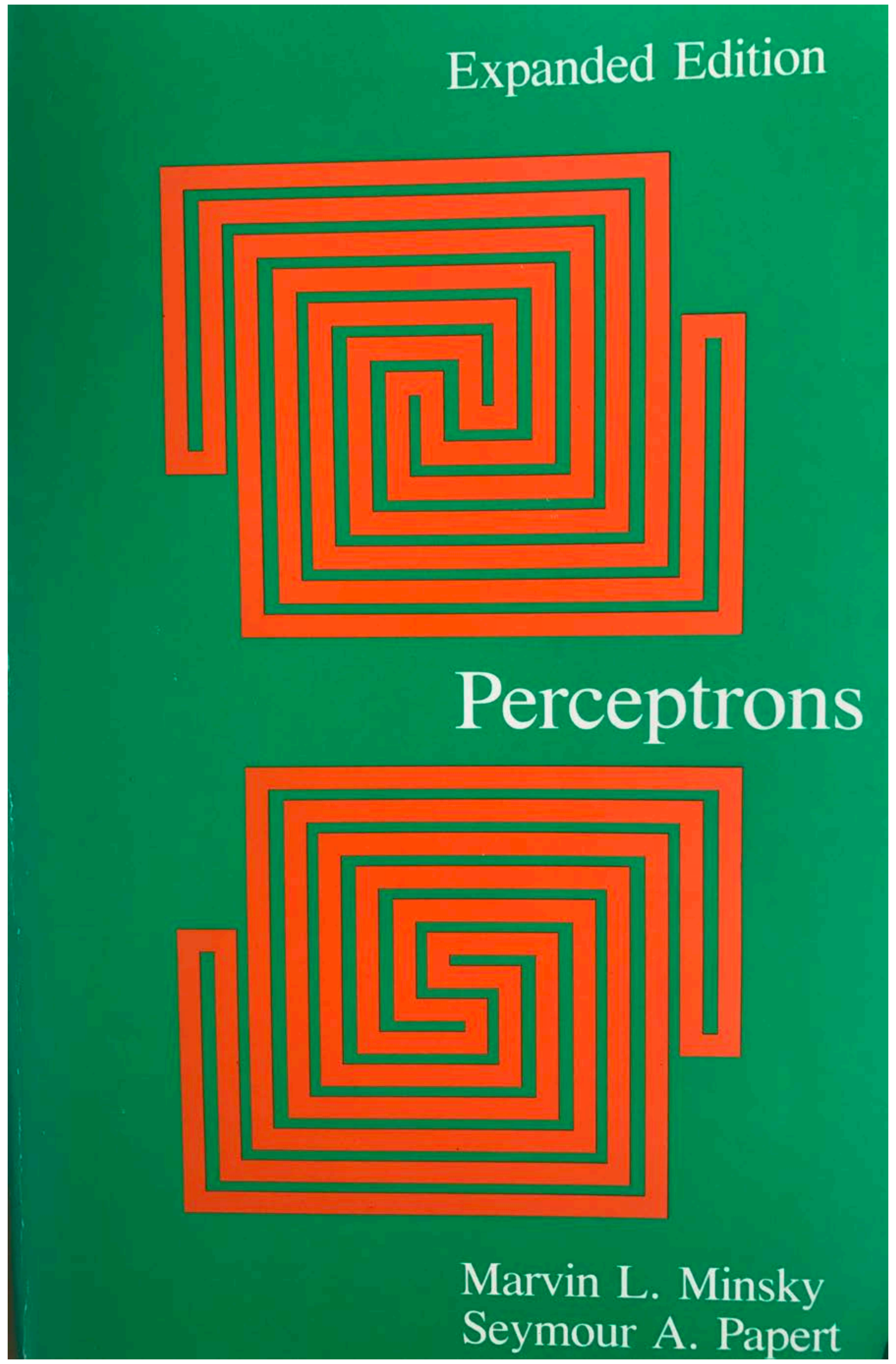
Which one of these reasons:

Physics (Max Tegmark)?
Neuroscience? <=== tp
Evolution?

Locality of Computation

What is special about locality of computation?

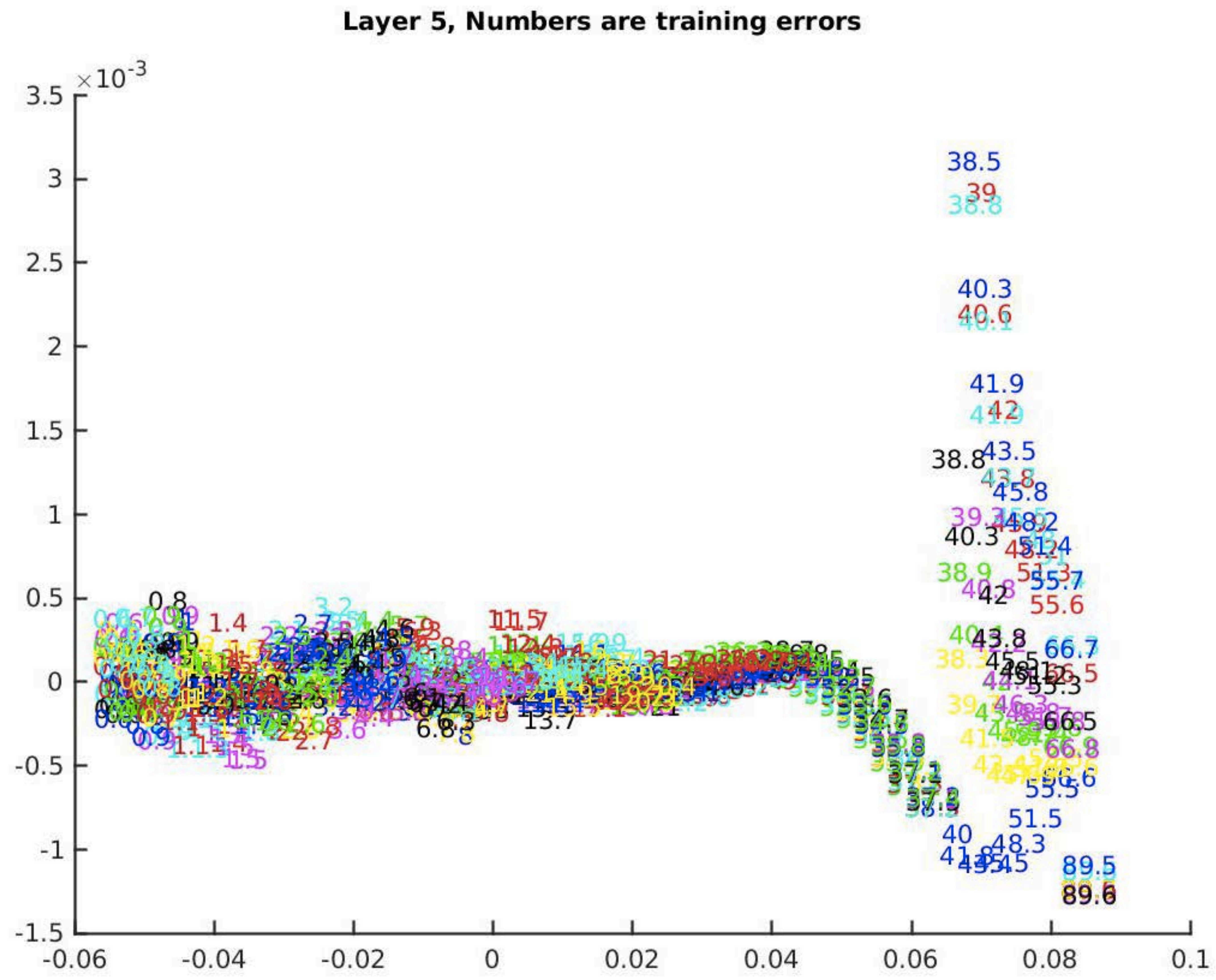
Locality in “space”?
Locality in “time”?



Theory II: What is the Landscape of the empirical risk?

Theorem (informal statement)

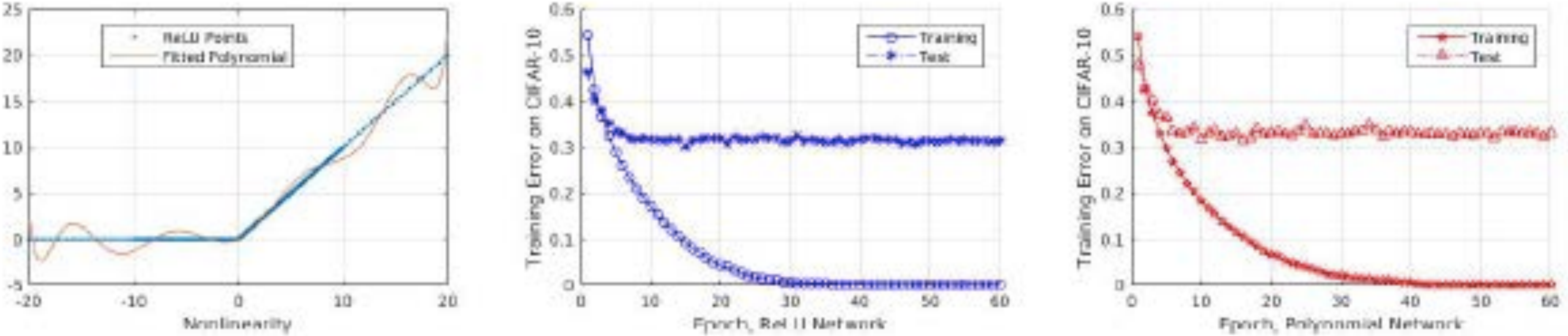
Replacing the RELUs with univariate polynomial approximation, Bezout theorem implies that the system of polynomial equations corresponding to zero empirical error has a very large number of degenerate solutions. The global zero-minimizers correspond to flat minima in many dimensions (generically unlike local minima). Thus SGD is biased towards finding global minima of the empirical risk.



CENTER FOR Brains Minds+ Machines

Liao, Poggio, 2017

Observation (theory and experiment): deep polynomial networks show same puzzles as RELU nets



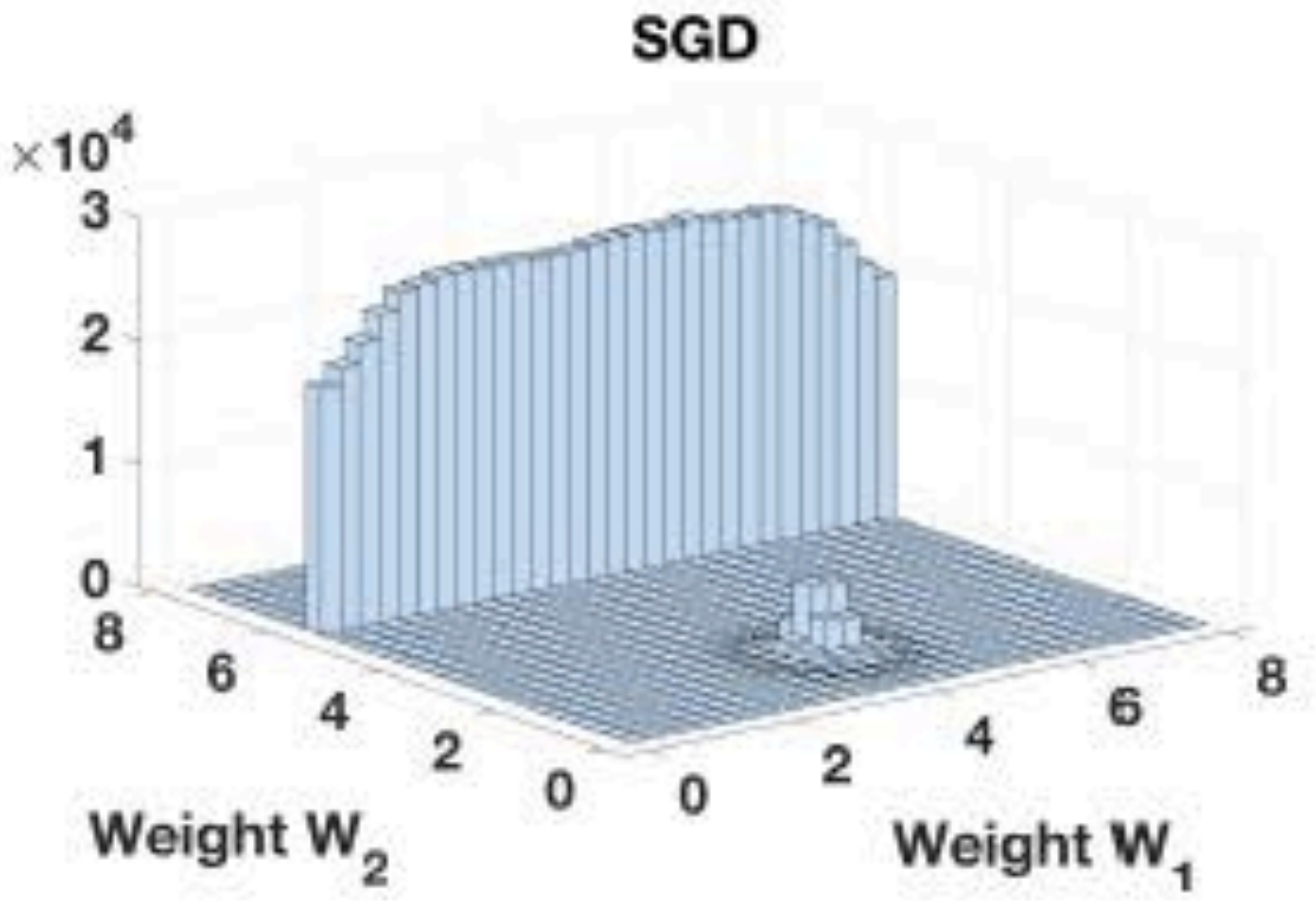
Poggio et al., 2017

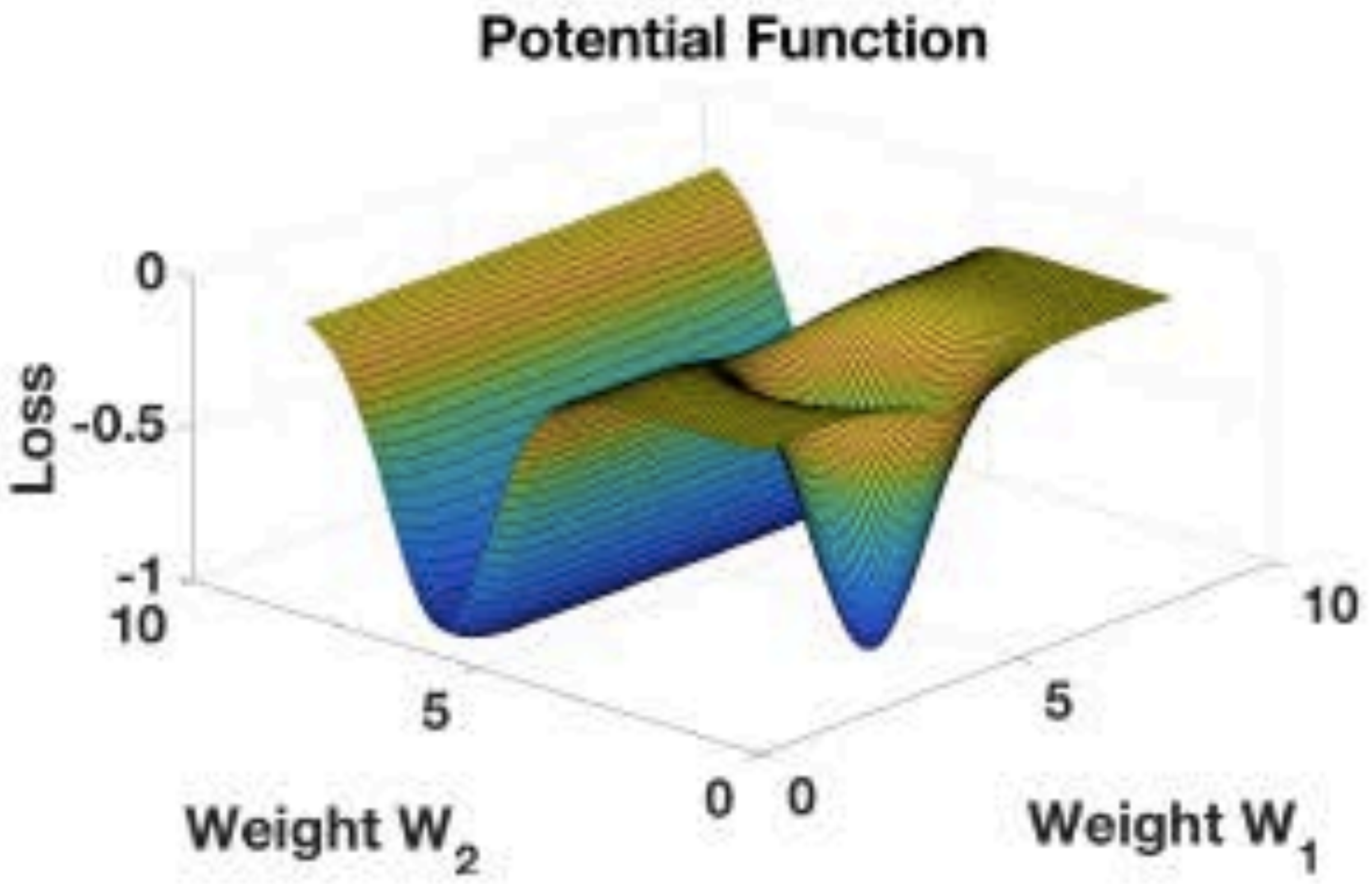
Theory of Deep Learning IIb: Optimization Properties of SGD

by

Chiyuan Zhang[1] Qianli Liao[1] Alexander Rakhlin[2] Brando Miranda[1] Noah Golowich[1] Tomaso Poggio[1]

[1]Center for Brains, Minds, and Machines, McGovern Institute for Brain Research,
Massachusetts Institute of Technology, Cambridge, MA, 02139.
[2] University of Pennsylvania





Bezout theorem

$$p(x_i) - y_i = 0 \quad \text{for } i = 1,\ldots,n$$

The set of polynomial equations above with *k= degree of p(x)* has a number of distinct zeros (counting points at infinity, using projective space, assigning an appropriate multiplicity to each intersection point, and excluding degenerate cases) equal to

$$Z = k^n$$

the product of the degrees of each of the equations. As in the linear case, when the system of equations is underdetermined – as many equations as data points but more unknowns (the weights) – the theorem says that there are an infinite number of global minima, under the form of Z regions of zero empirical error.

Global and local zeros

$f(x_i) - y_i = 0 \quad \text{for } i = 1, \ldots, n$

n equations in W unknowns with $W >> n$

$$\nabla_w \sum_{i=1}^{N} (f(x_i) - y_i)^2) = 0$$

W equations in W unknowns

There are a very large number of zero-error minima which are highly degenerate unlike the local non-zero minima.

Langevin equation

$$\frac{dw}{dt} = -\gamma_t \nabla V(w(t), z(t)) + \gamma_t' \, dB(t)$$

with the Boltzmann equation as asymptotic "solution"

$$p(w) \sim \frac{1}{Z} = e^{-\frac{V(w)}{T}}$$

SGD

$$f_{t+1} = f_t - \gamma_t \nabla V(f_t, z_t),$$

$$\nabla V(f_t, z_t) = \frac{1}{|z_t|} \sum_{z \in z_t} \nabla V(f_t, z).$$

We define a noise "equivalent quantity"

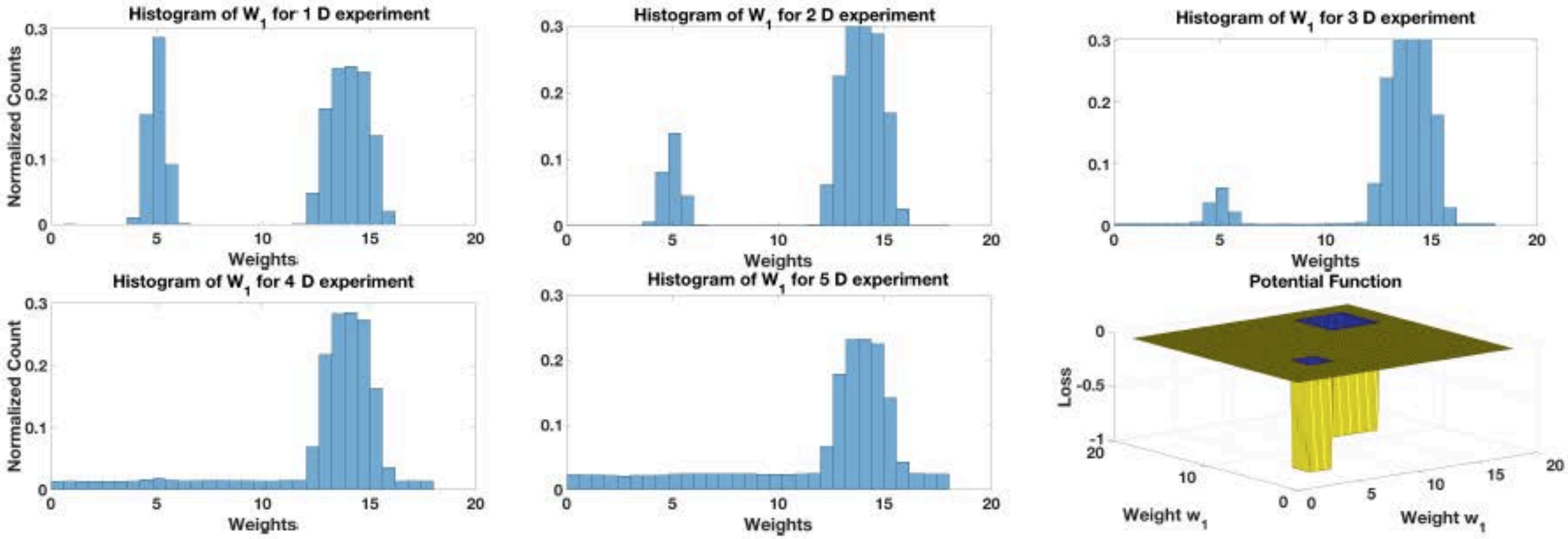
$$\xi_t = \nabla V(f_t, z_t) - \nabla I_{S_n}(f_t),$$

and it is clear that $\mathbb{E}\xi_t = 0$.

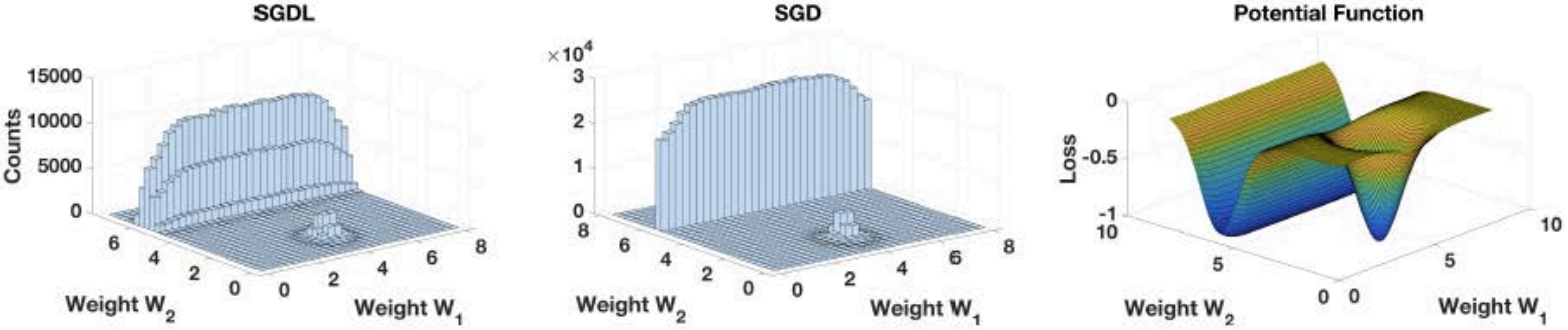
We write Equation 6 as

$$f_{t+1} = f_t - \gamma_t(\nabla I_{S_n}(f_t) + \xi_t).$$

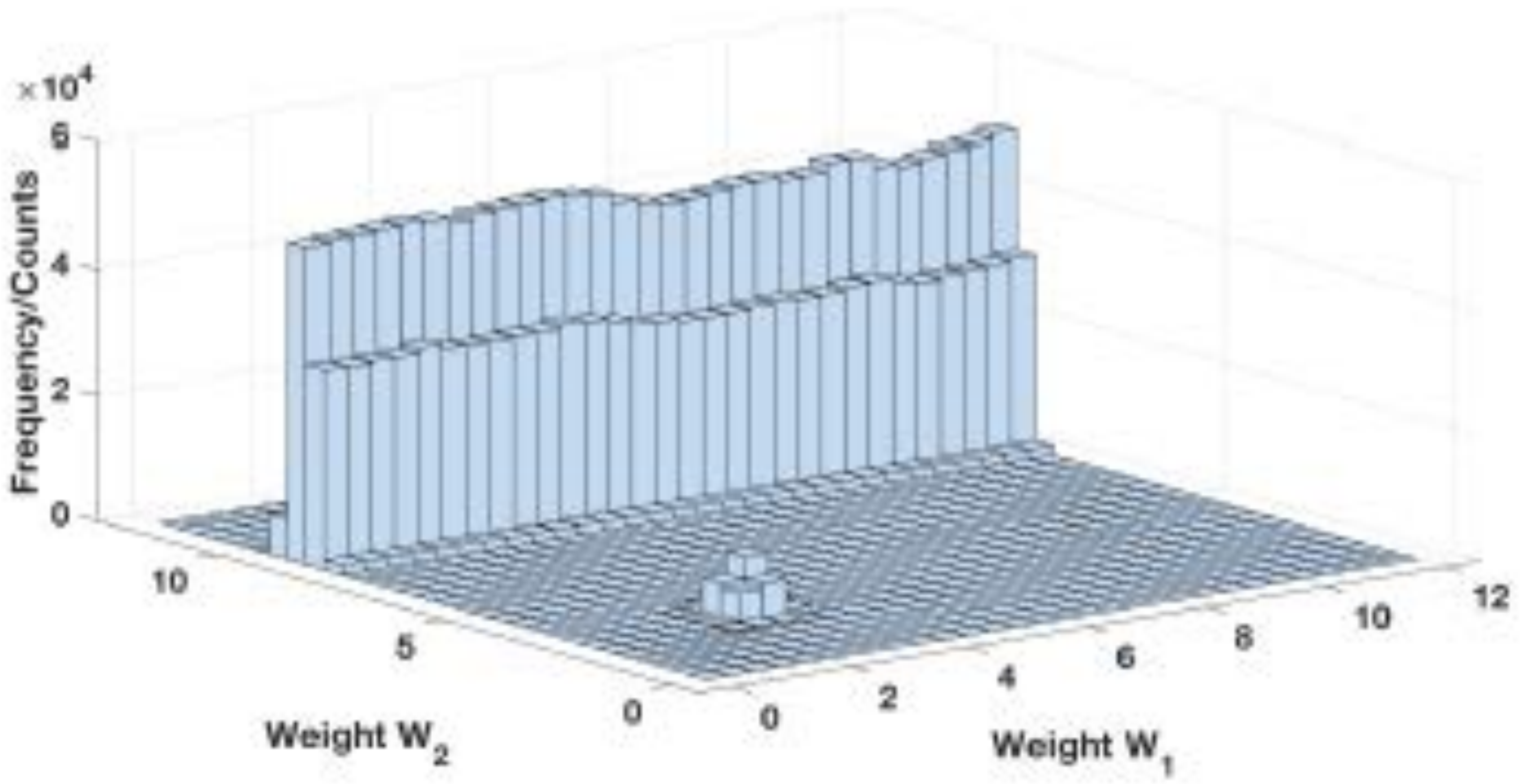
GDL selects larger volume minima

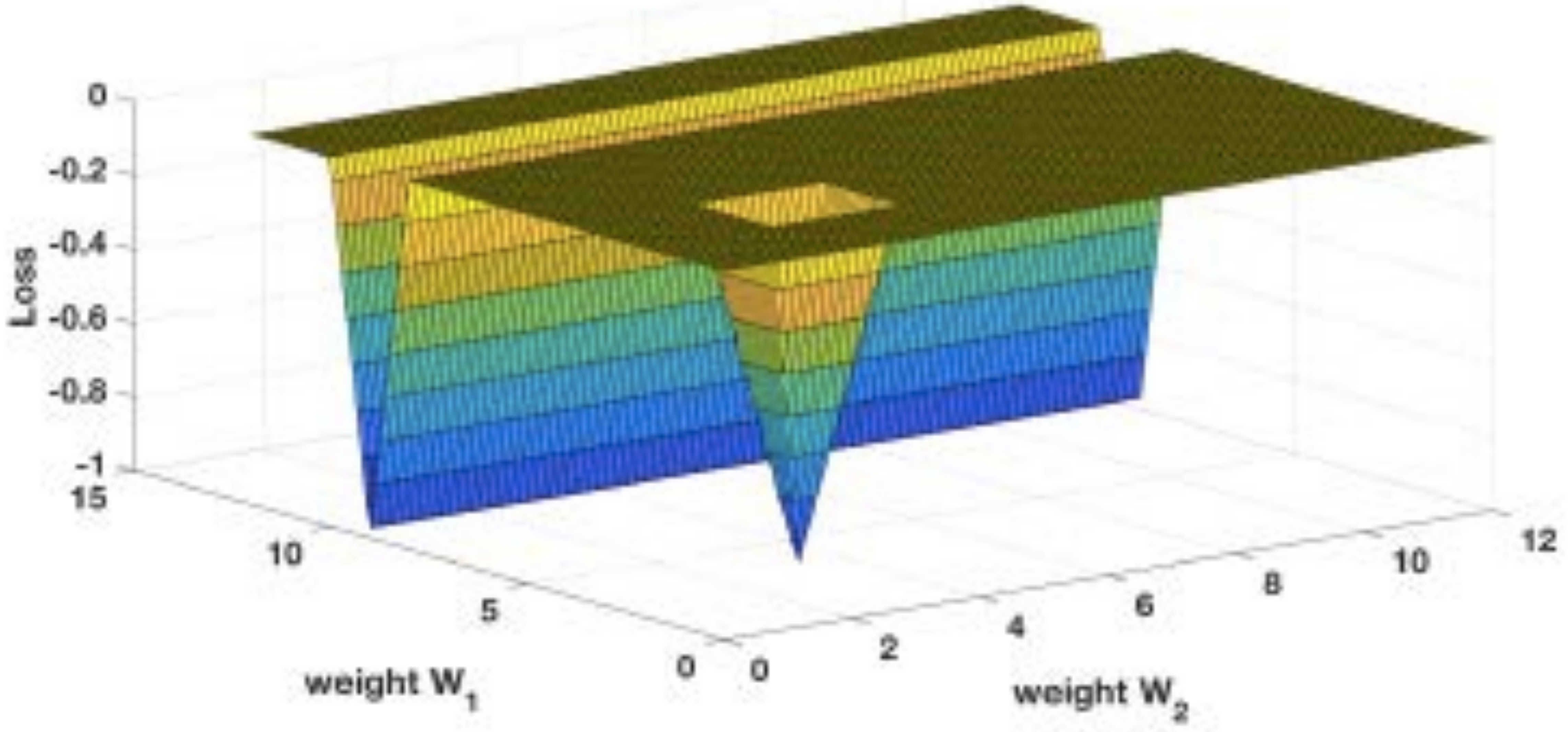


GDL ~ SGD (empirically)

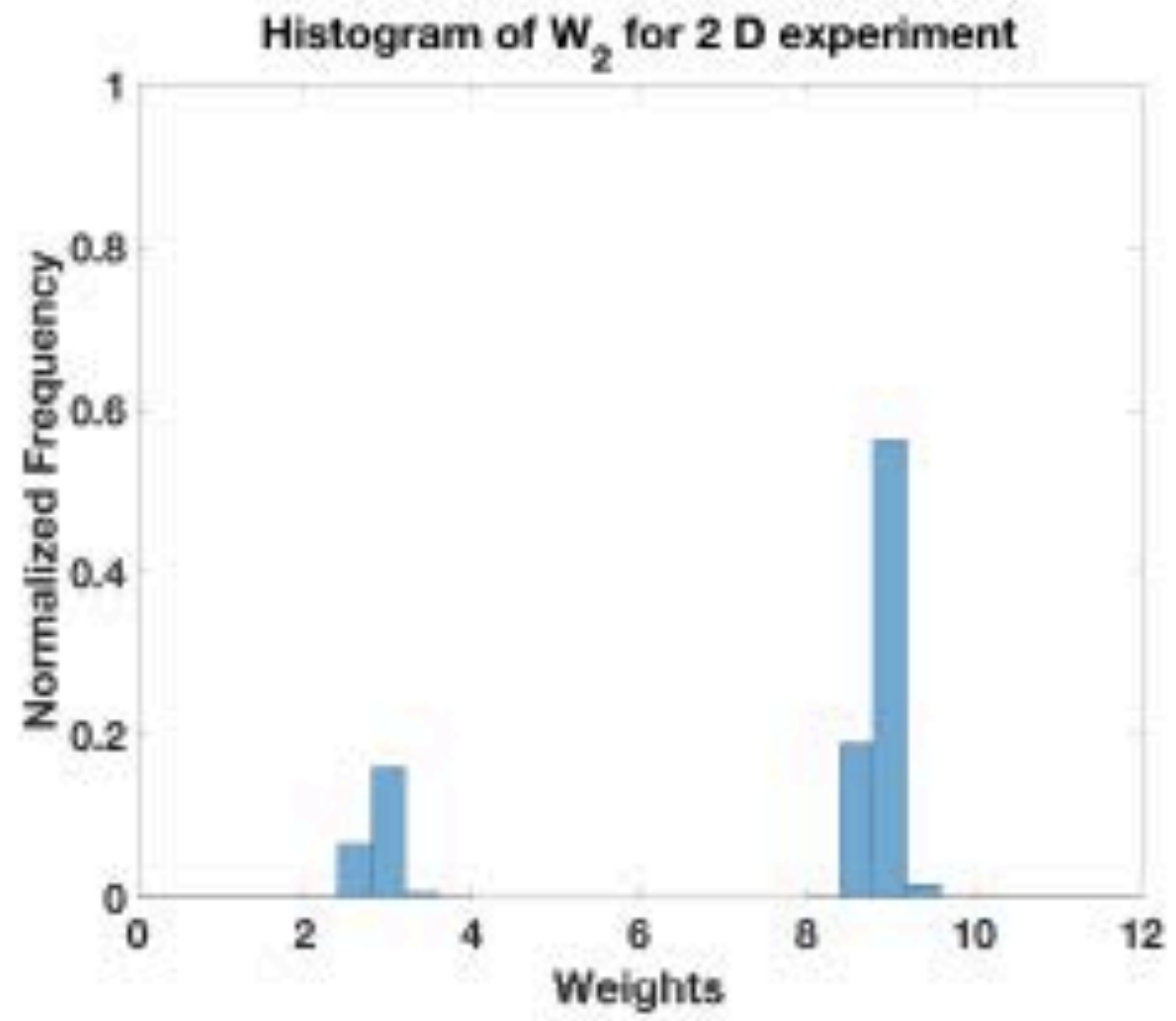


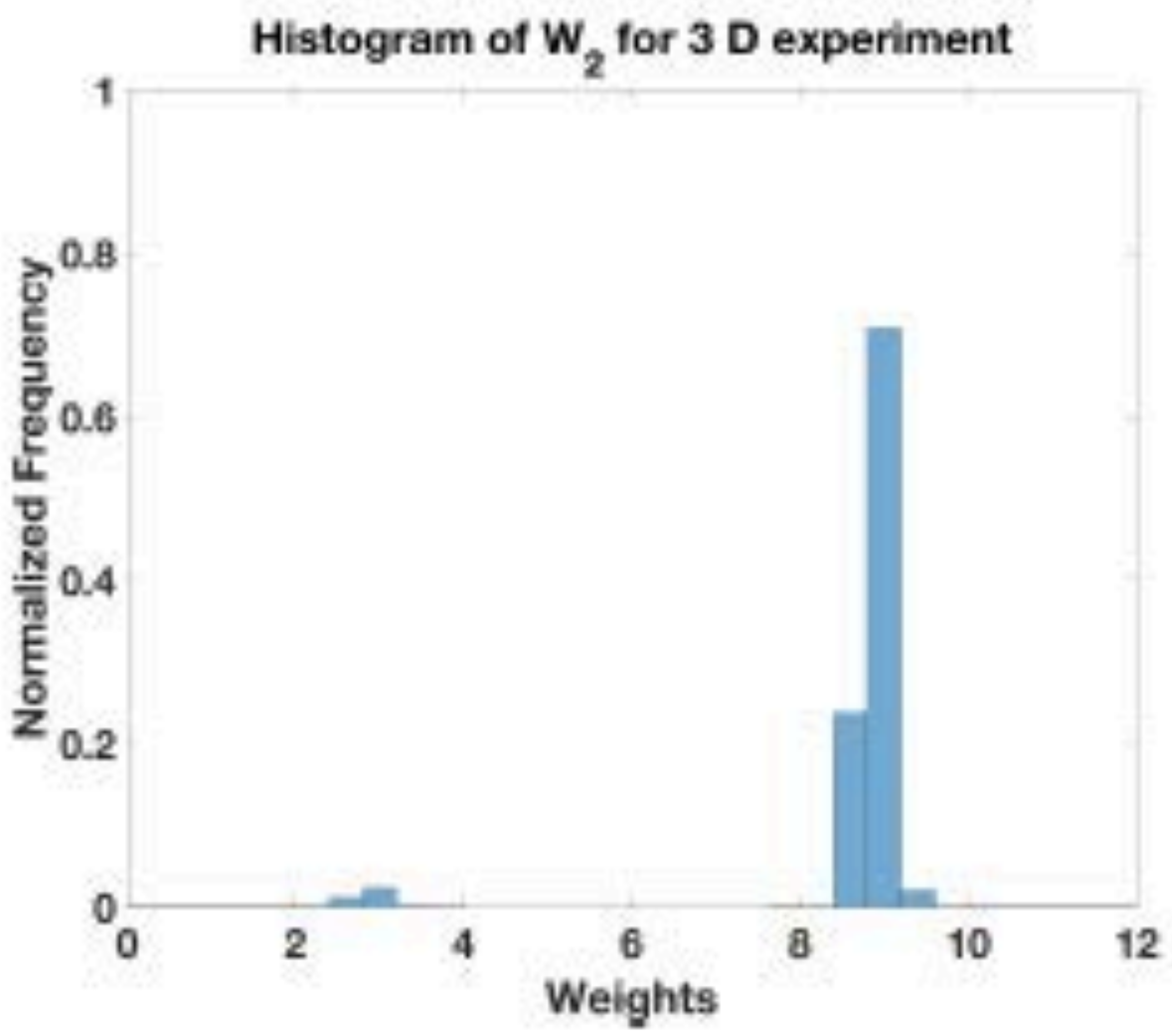
GDL selects degenerate minima

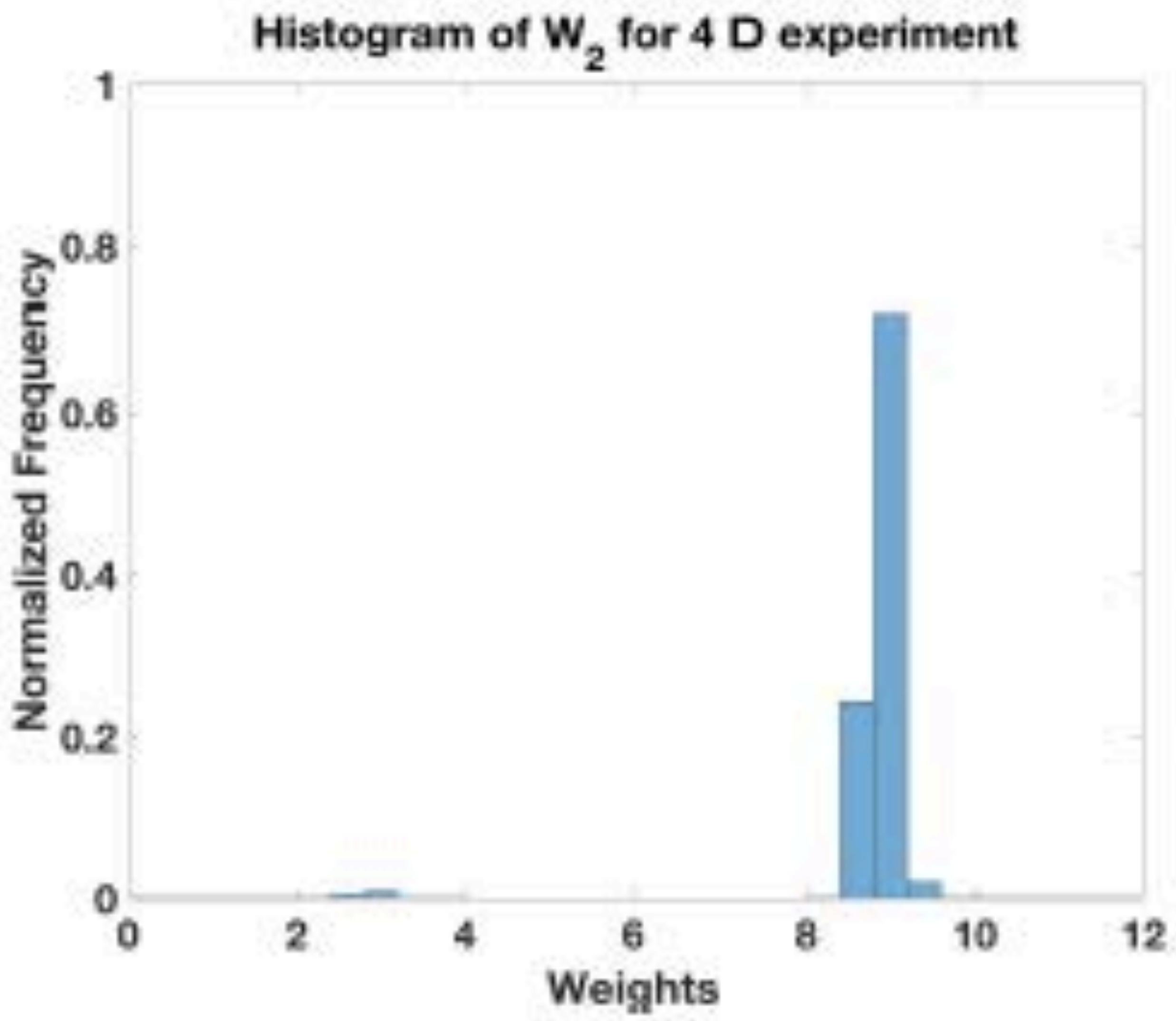


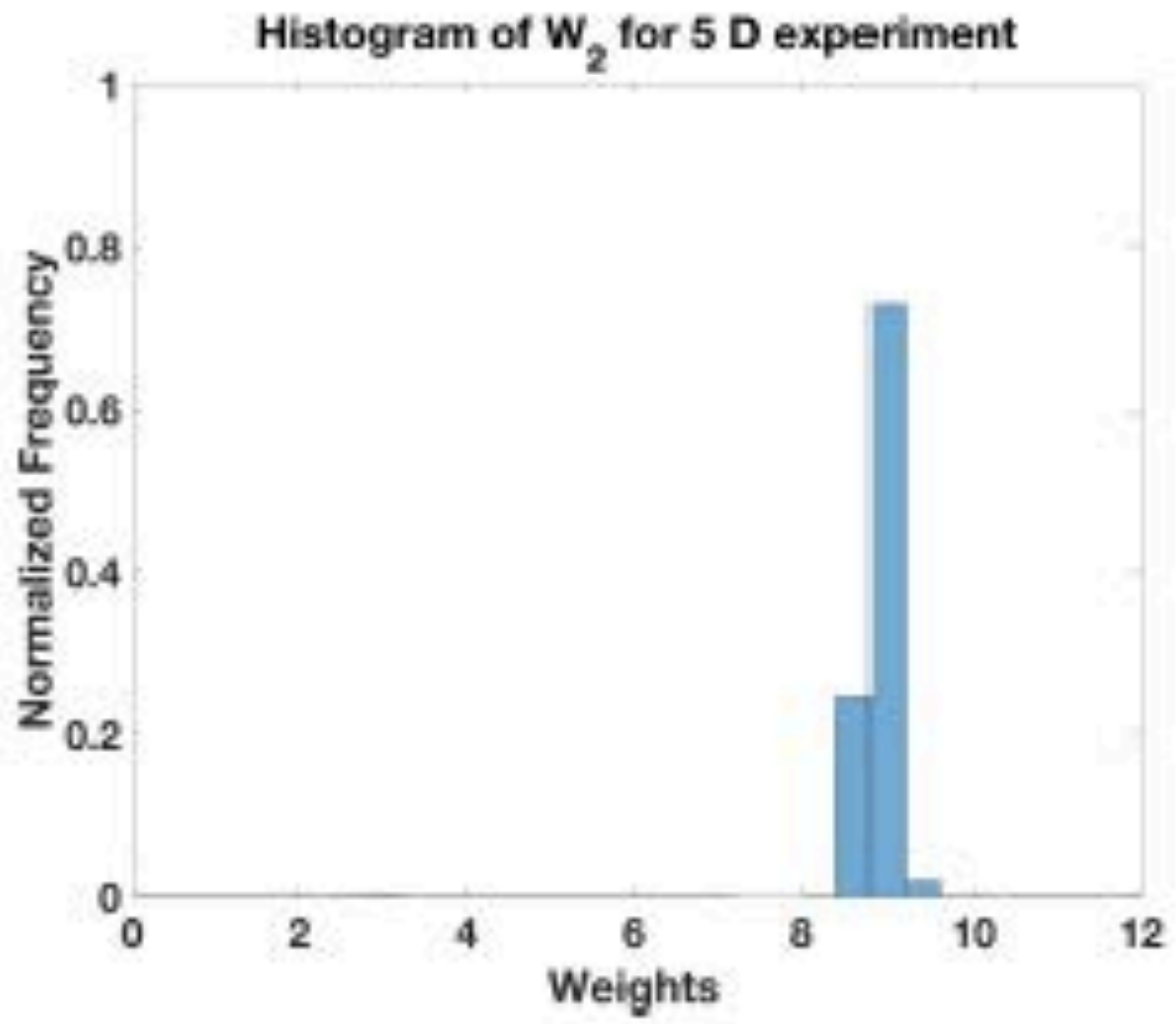


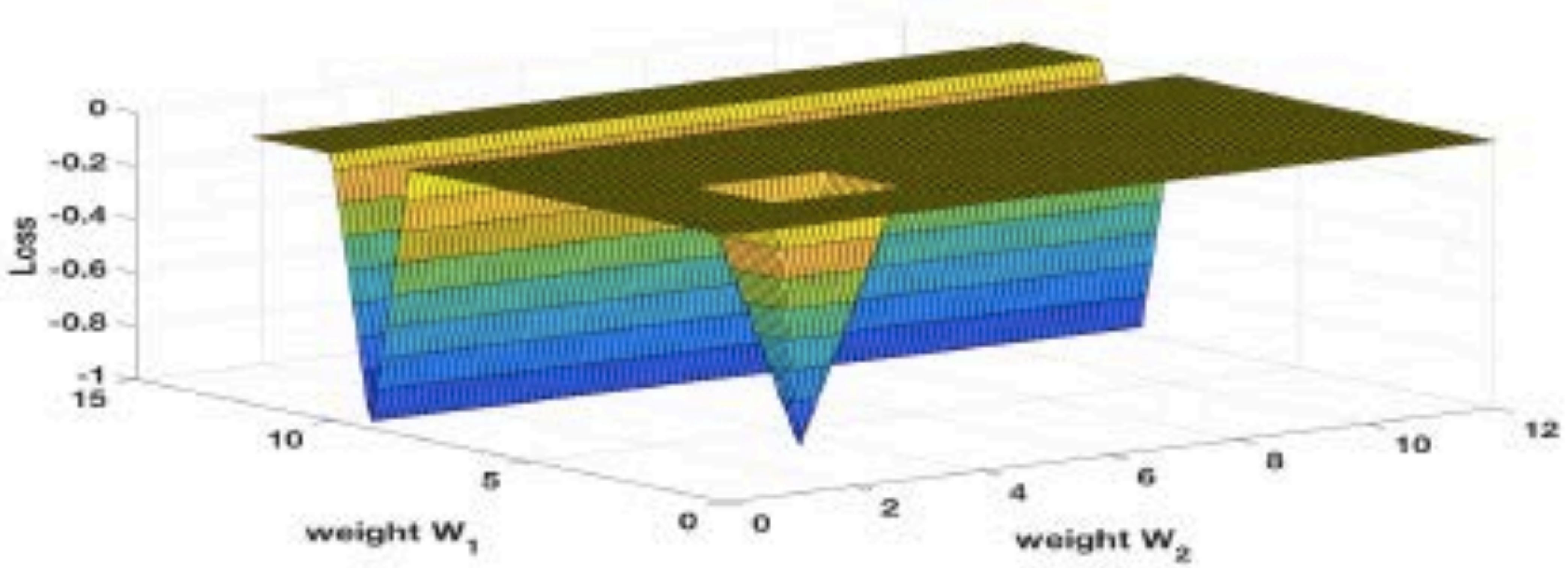
Concentration because of high dimensionality











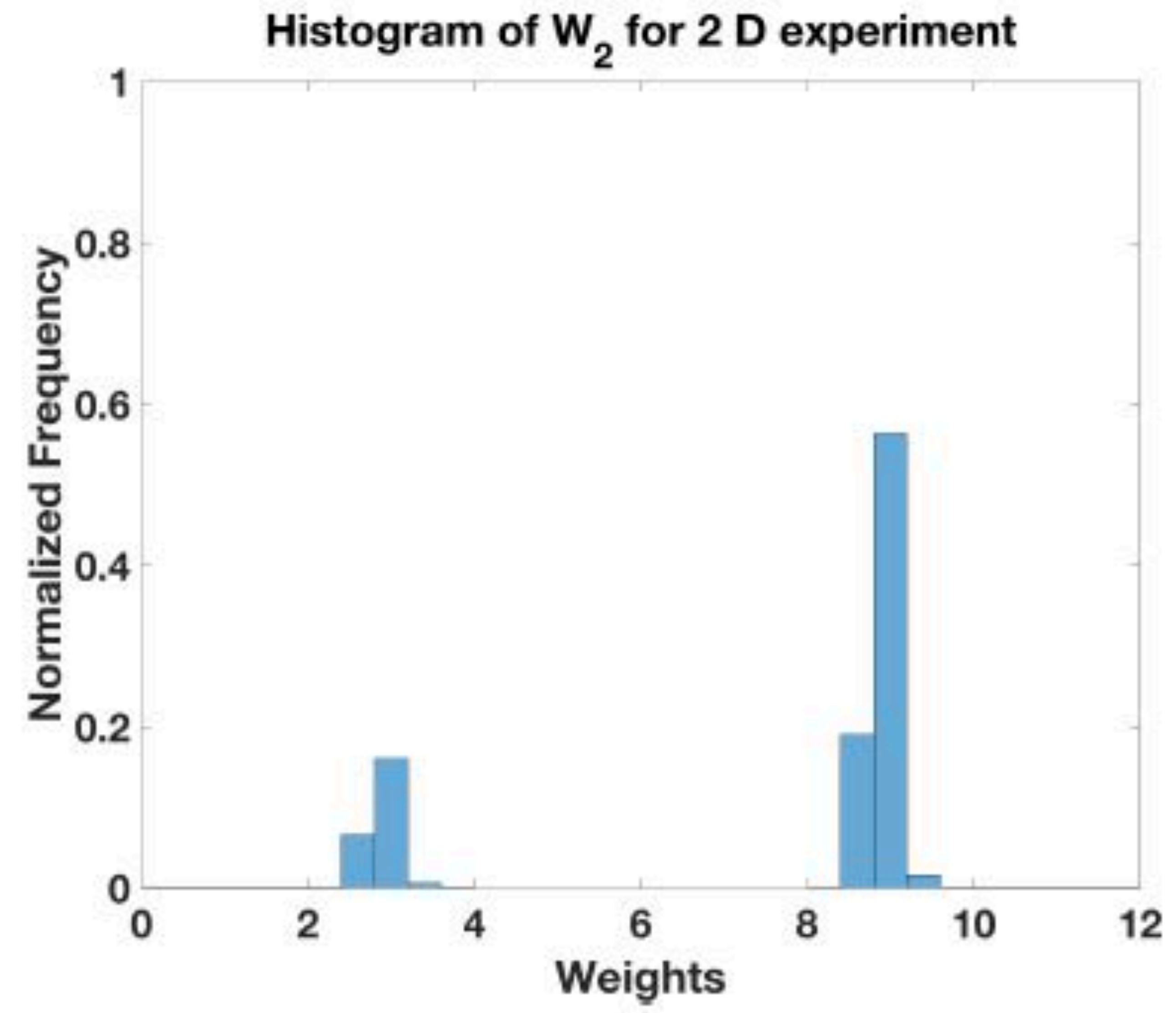
Poggio, Rakhlin, Golovitc, Zhang, Liao, 2017

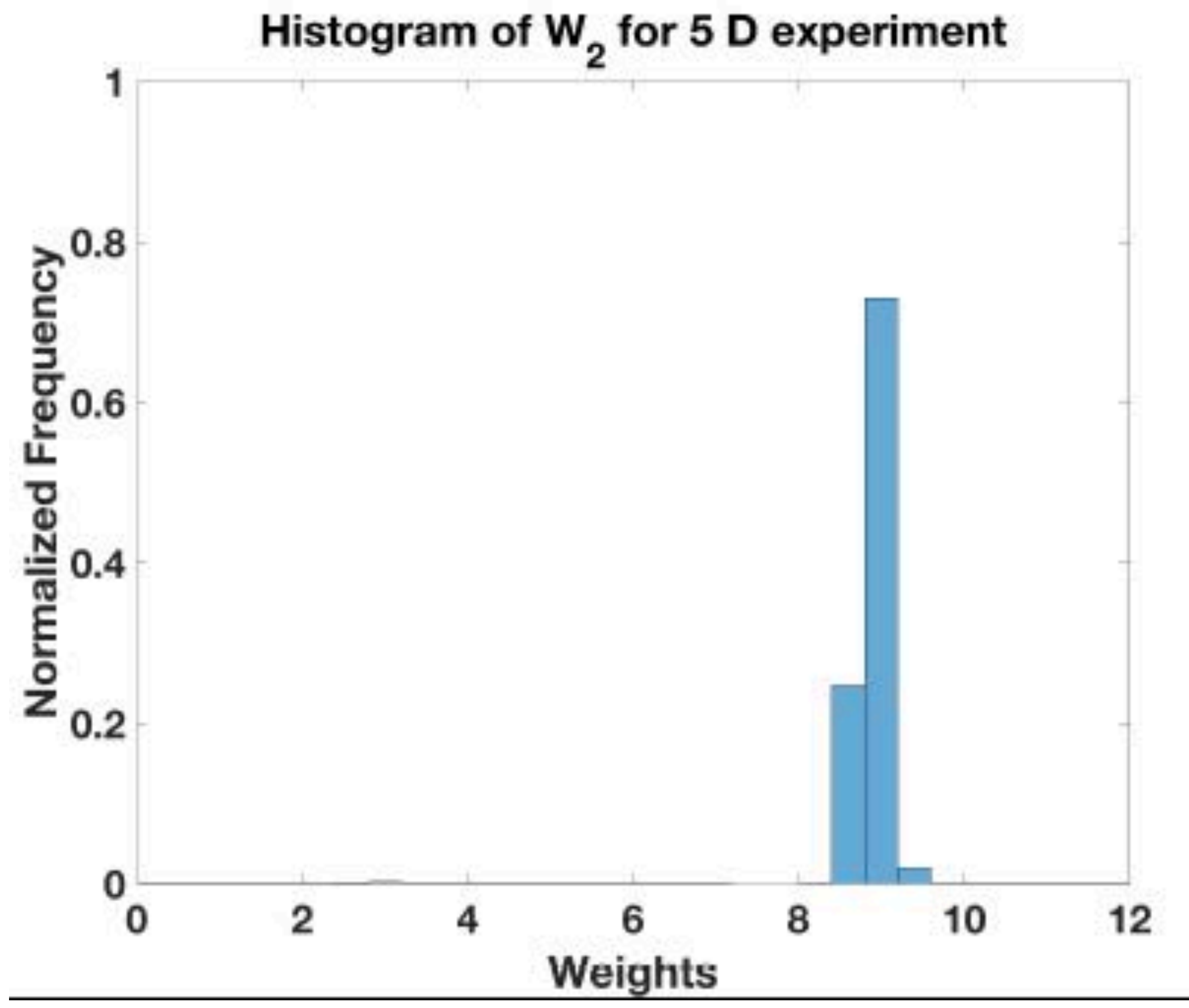
SGDL and SGD observation: summary

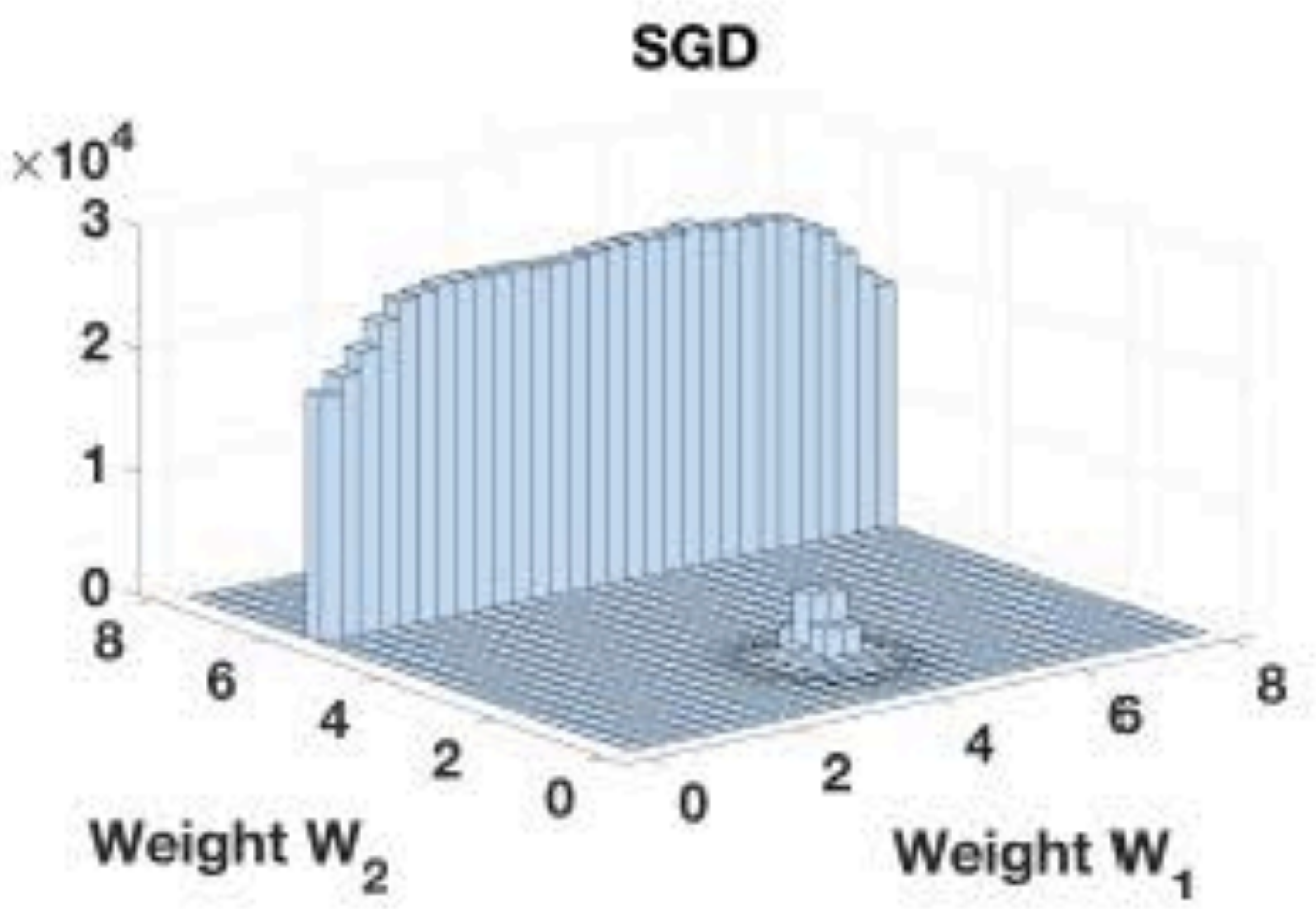
- There are may zero minimizer with overparametrized deep networks because of Bezout theorem

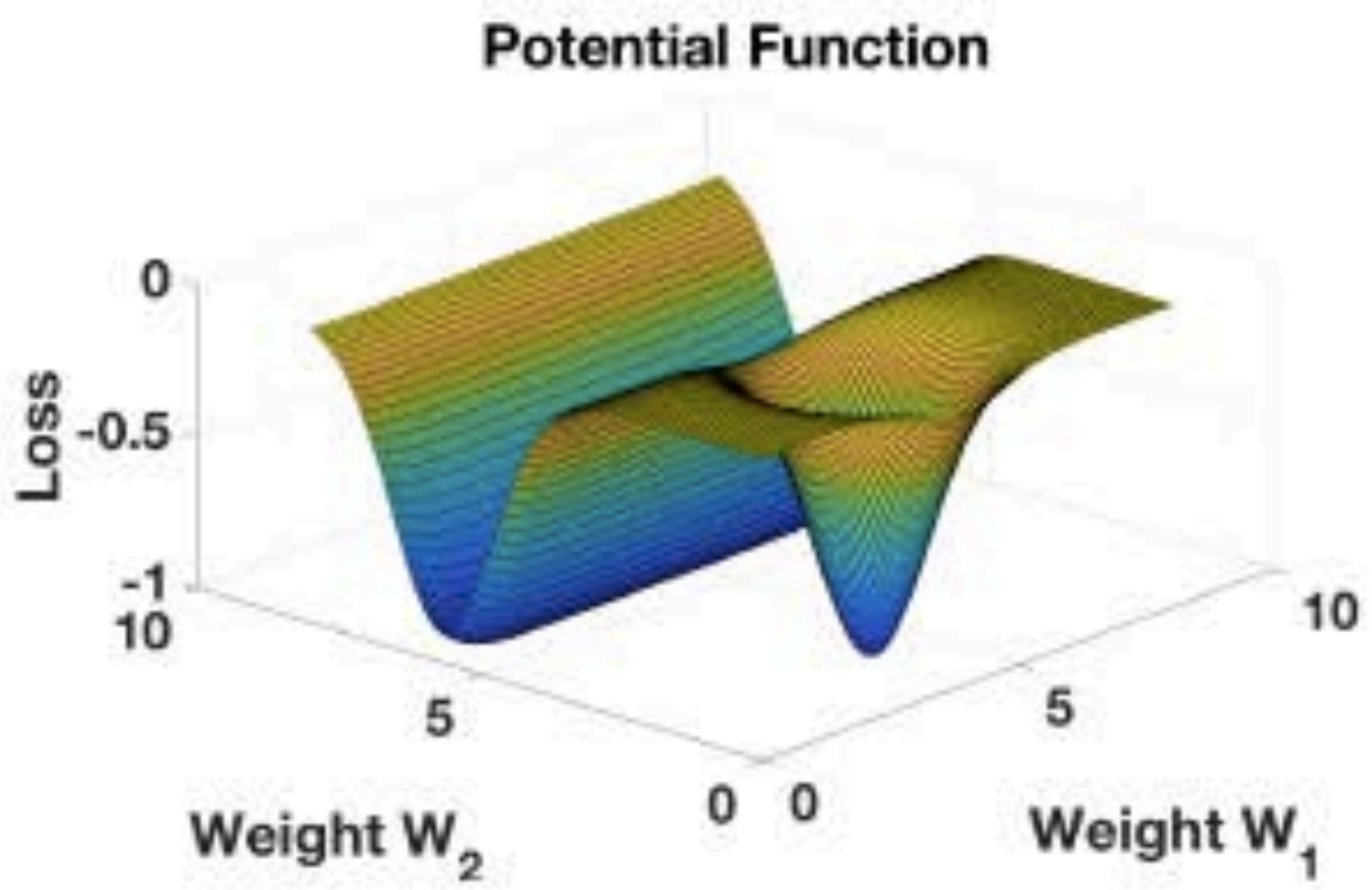
- SGDL finds with very high probability large volume, flat **zero-minimizers**; empirically SGD behaves in a similar way

- Flat minimizers correspond to degenerate zero-minimizers and thus to global minimizers;



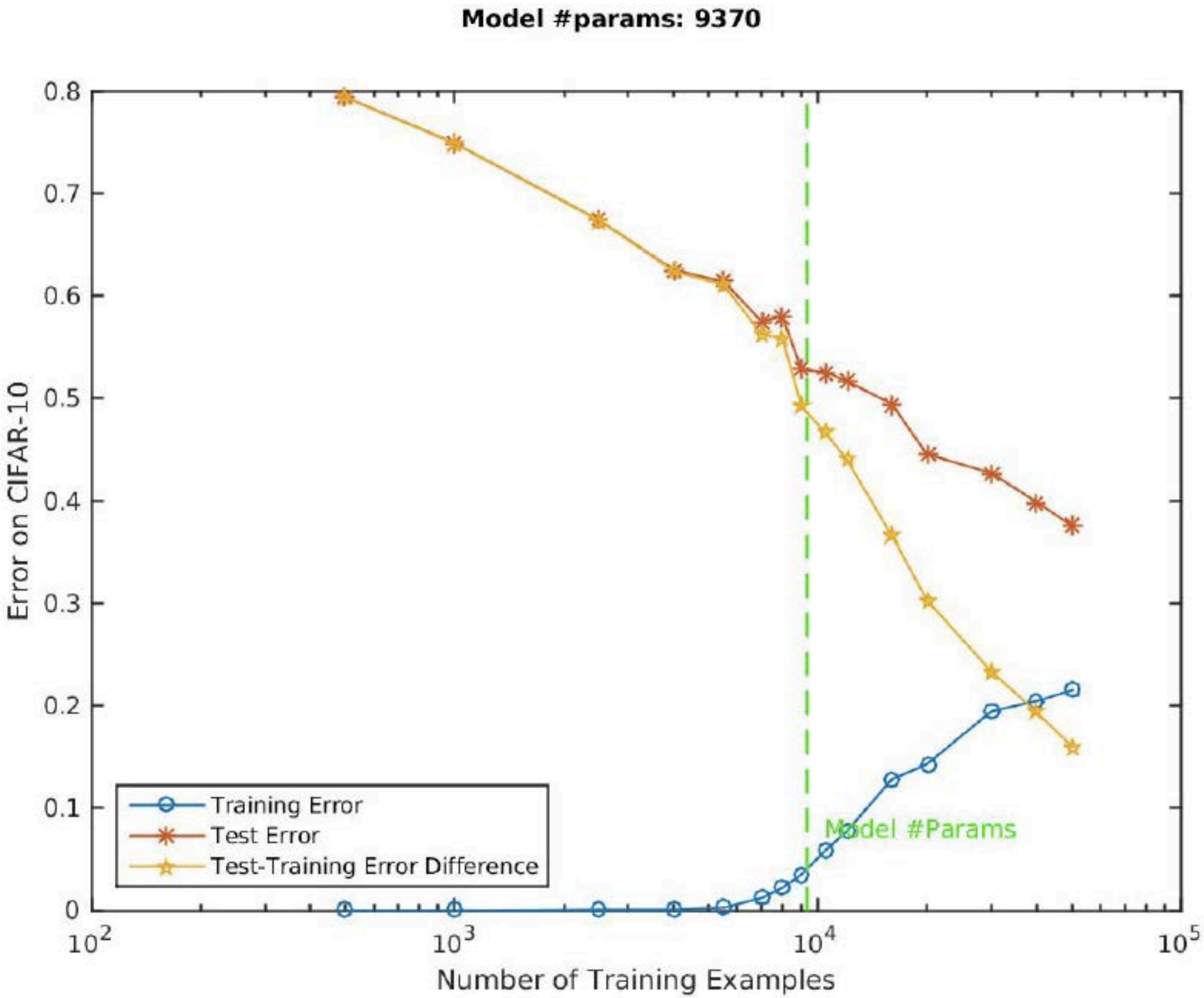






Poggio, Rakhlin, Golovitc, Zhang, Liao, 2017

Theory III: How can underconstrained solutions generalize?

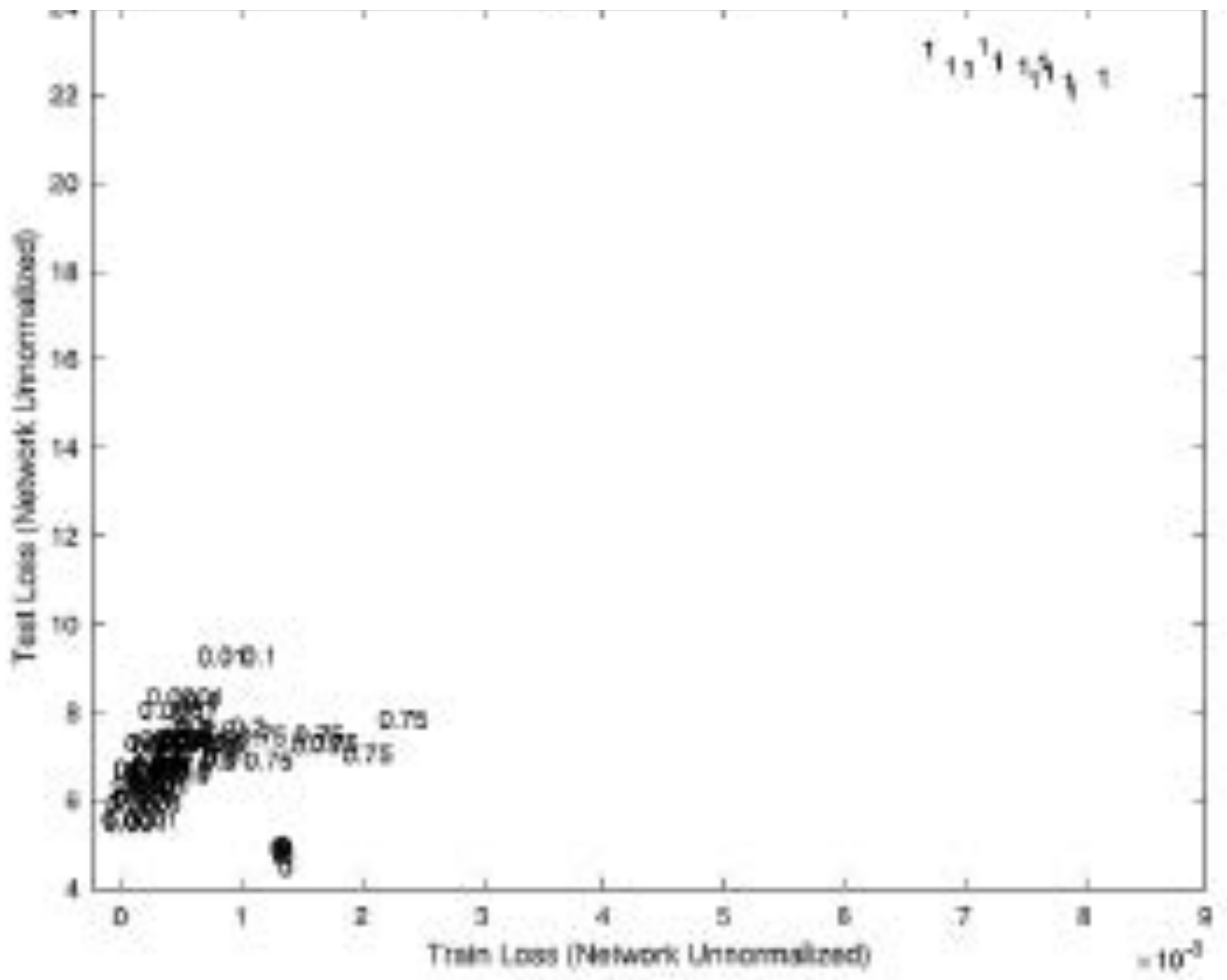


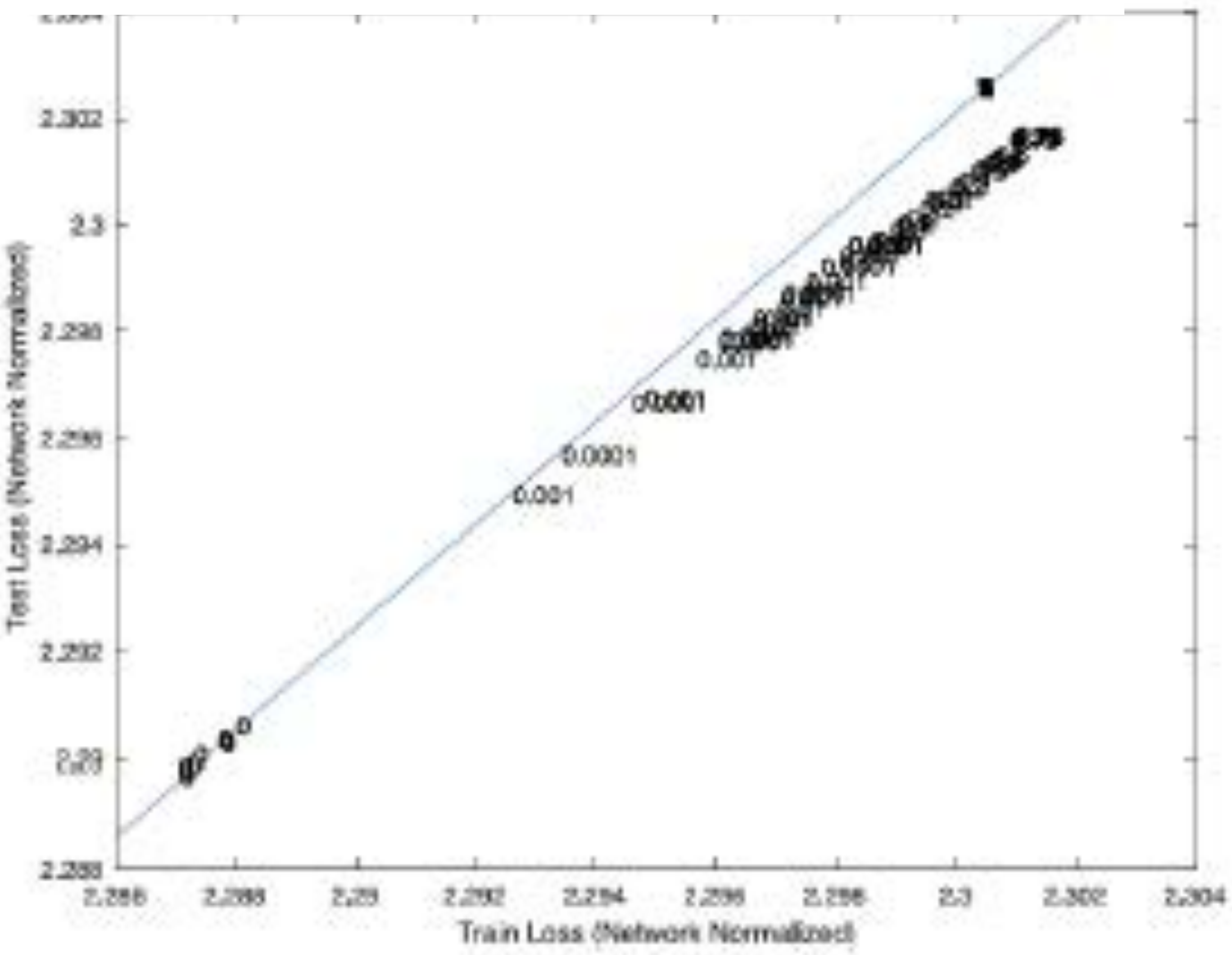
CENTER FOR
Brains
Minds+
Machines

Classical Generalization Bounds

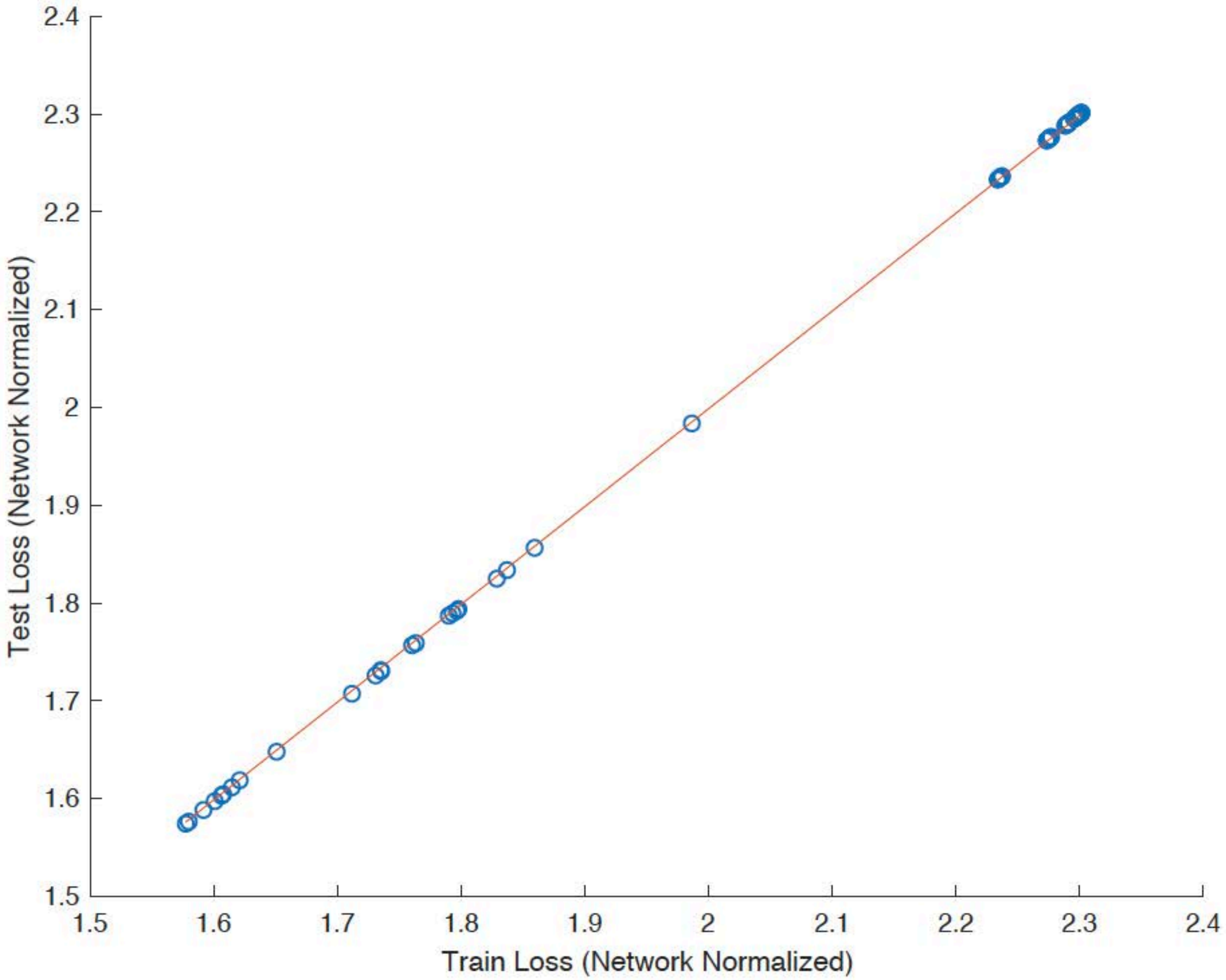
$$\textit{With probability} \geq (1-\delta)\ \forall f$$

$$|\mathbf{E}(\ell) - \mathbf{E}_S(\ell)| \leq \mathbb{C}_{N,\delta},$$





Very good generalization!



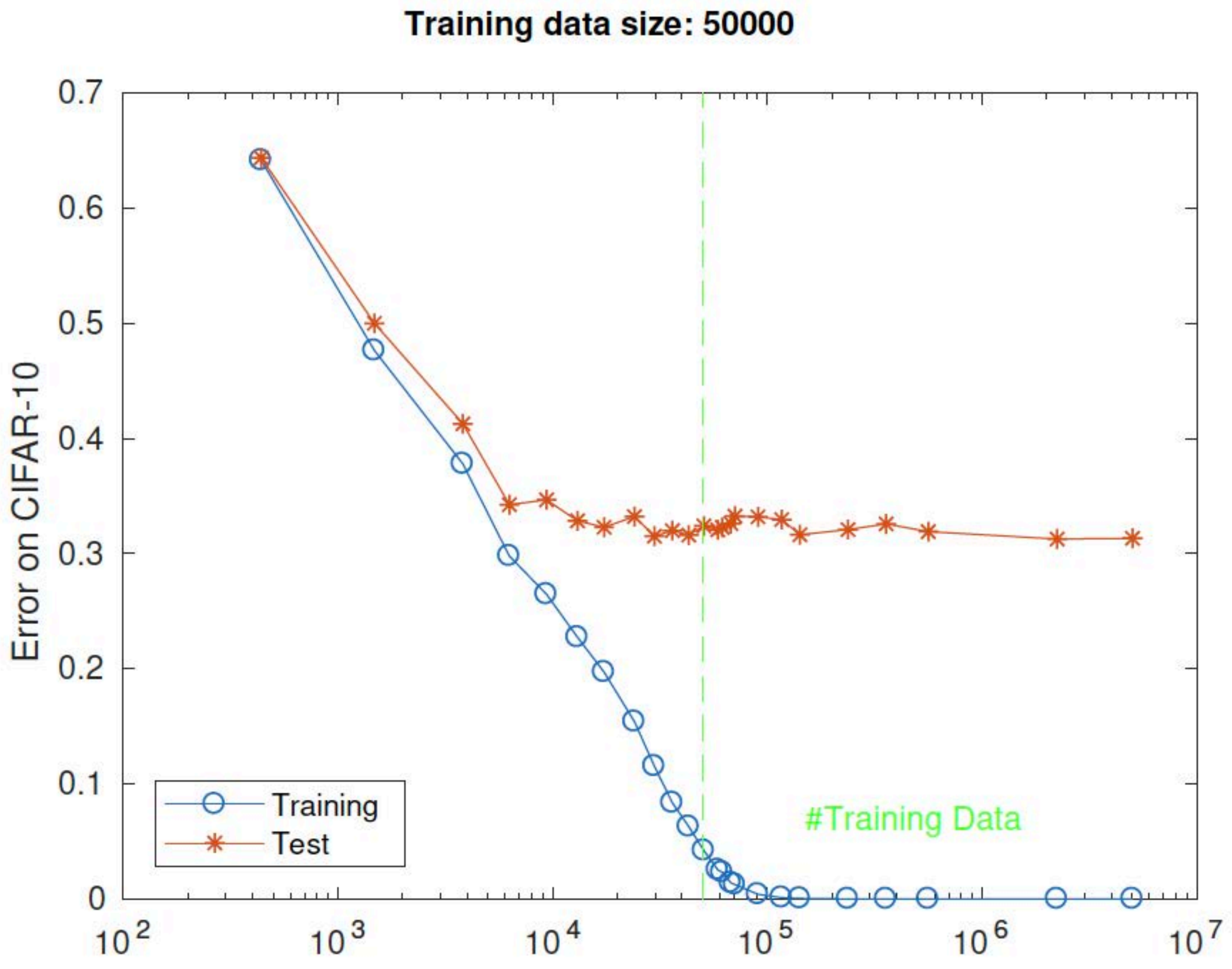
MNIST with different initializations

Three Theory Questions: Summary of Answers

- *Approximation theorems:* for compositional functions deep but not shallow networks avoid the *curse of dimensionality*.

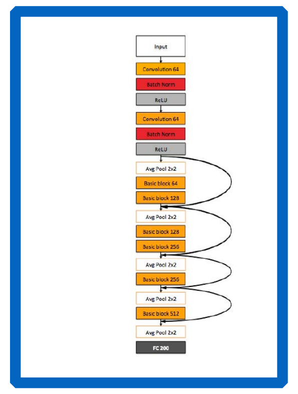
- *Optimization remarks:* SGD finds with high probability global minima which are degenerate.

- *Generalization:* The gradient dynamics of deep networks near global minima converges to <u>minimum norm solution for each layer of weights.</u>



Musings on Near Future Breakthroughs

- new architectures/class of applications from basic DCN block (example GAN + RL/DL + …)



- new semisupervised training framework, avoiding labels: implicit labeling…predicting next “frame”…

- new basic supervised block/circuit $\boxed{?}$

- new learning algorithm (Shim) instead of SGD …

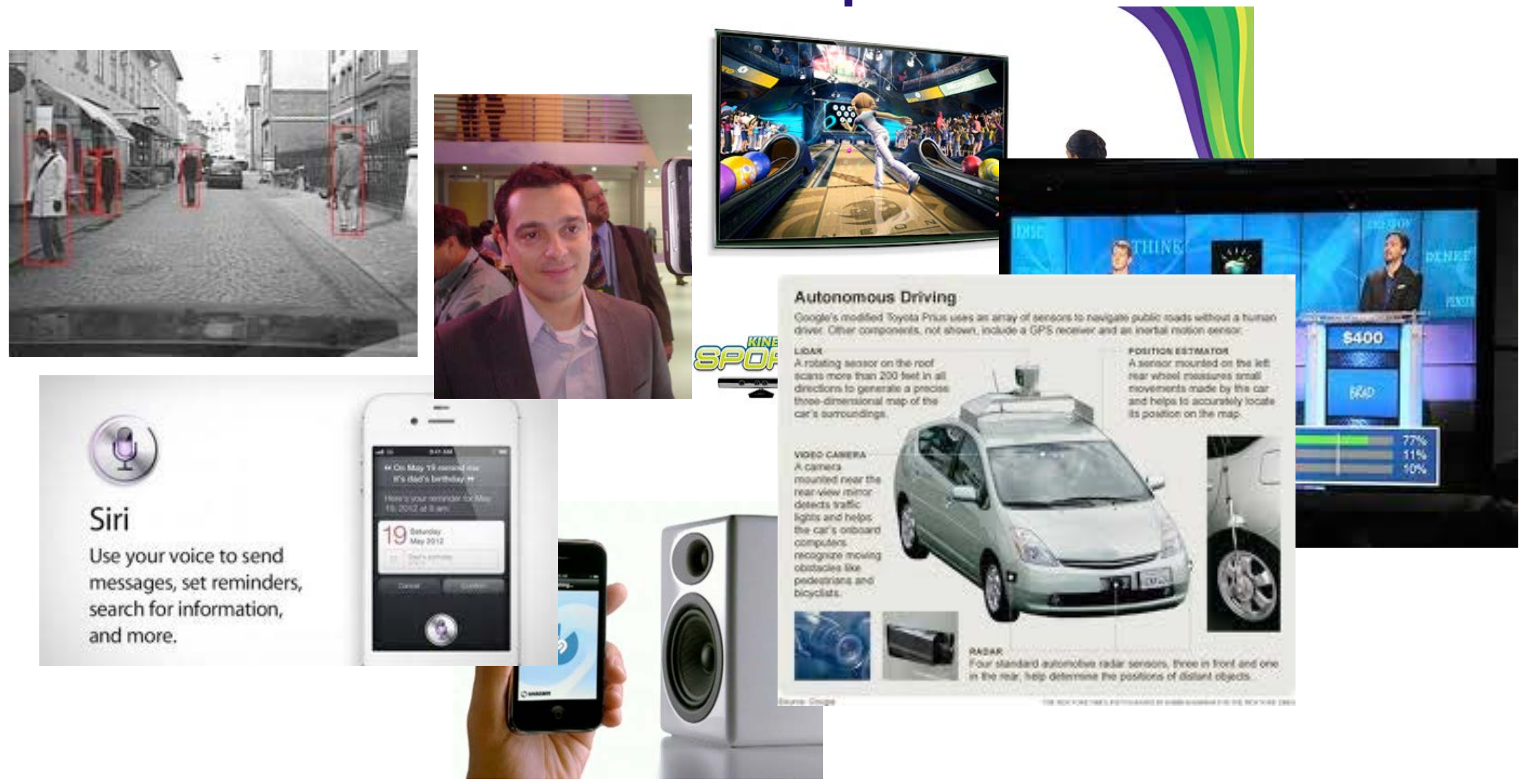
General musings

The evolution of computer science

- there were programmers
- there are now labelers
- there may be schools for bots…

Today's science, tomorrow's engineering: learn like children learn

The first phase (and successes) of ML:
supervised learning, big data: $n \to \infty$



from programmers…
…to labelers…
…to computers that learn like children…

The next phase of ML: implicitly supervised learning,
learning like children do, small data: $n \to 1$