Physics inspired Machine Learning

Mykola Maksymenko, Research Lead at SoftServe R&D

Research Lead at SoftServe R&D

- PhD in Theoretical Physics

- Max-Planck Institute for the Physics of Complex Systems (Germany)

- Weizmann Institute of Science (Israel)

- Institute for Condensed Matter Physics of NASU

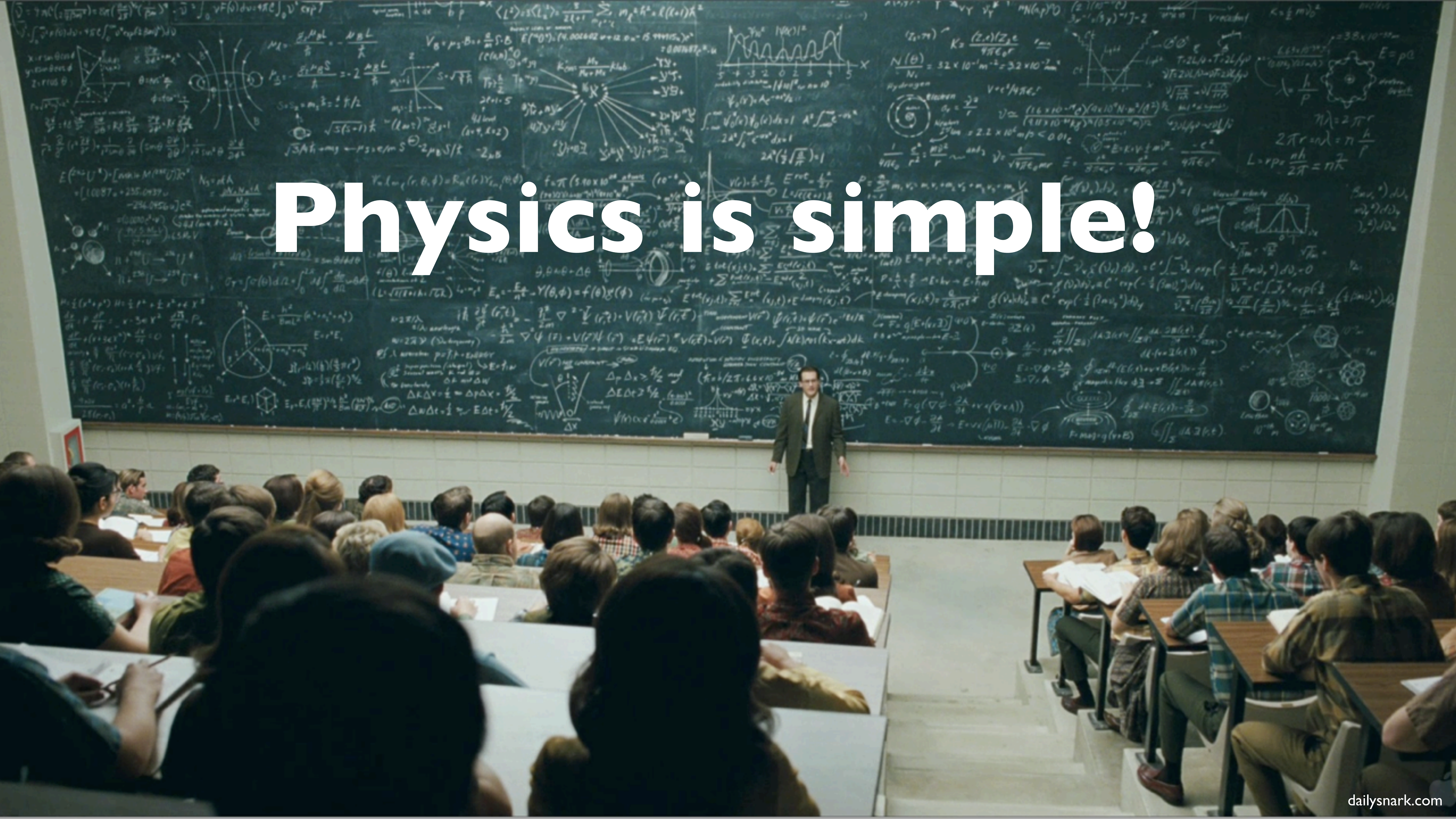
Physics of complex systems

Quantum phases of matter

Complex Networks

Non-equilibrium systems

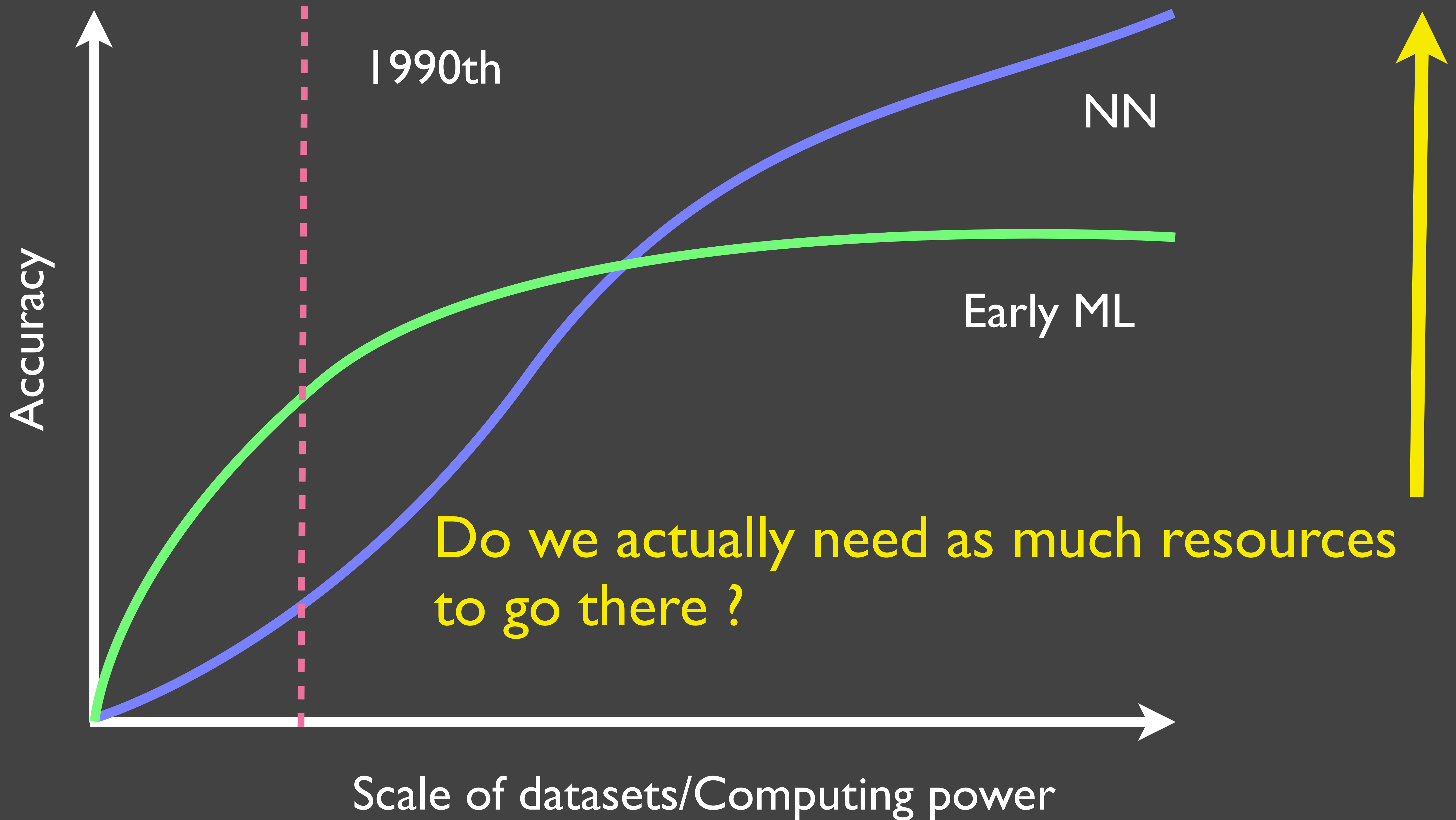
Exotic magnetism



Physics is simple!
dailysnark.com

Building the Deep Learning architectures

A naive picture of a recent progress



Do we actually need as much resources to go there ?

Resources are expensive

Andrej Karpathy
@karpathy

Читати

TitanX runs at ~200W (0.72MJ/h). In ~12h that's ~8MJ. Energy content of wood is 20MJ/kg, so running 1 TitanX overnight burns 1 pound of wood

11:00 - 13 жовт. 2015

The goals:

Optimal models

Easier training

Universality

Outline

Physics of Learning

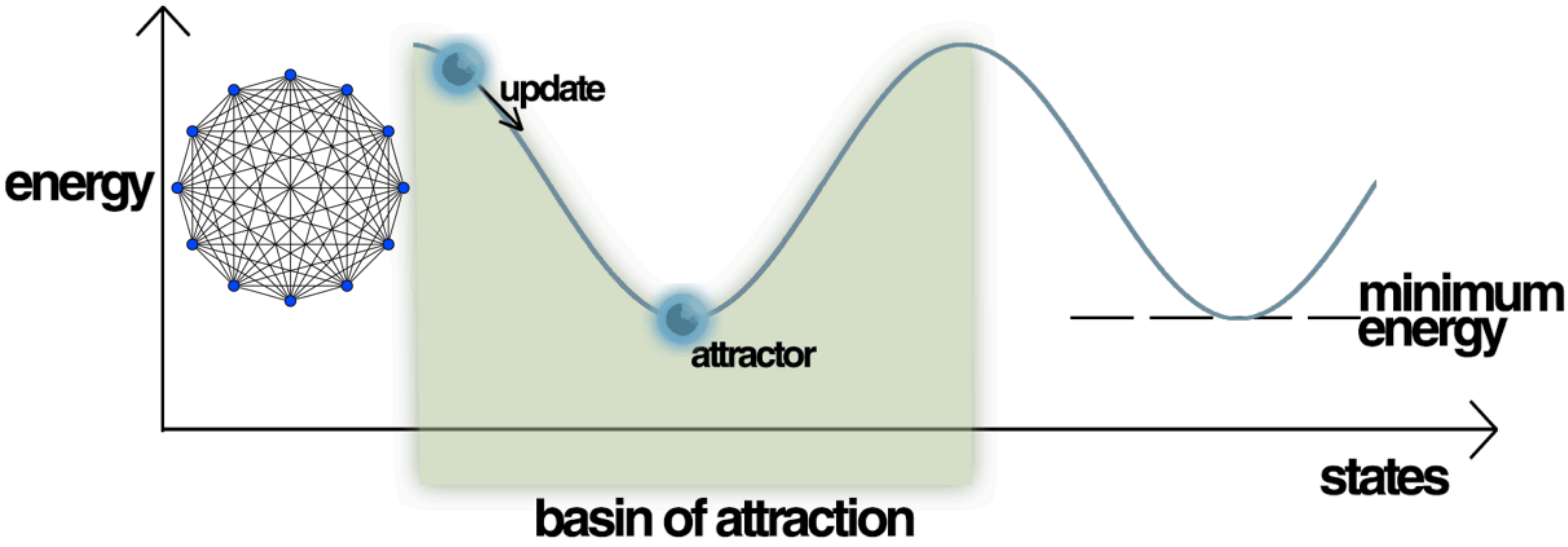
Redundancy of Neural Nets are redundant ?

How physics helps for better learning ?

Neural nets and quantum wave functions

Learning the Tensor networks

Hopfield network
J. Hopfield PNAS 1982



wikipedia.com

Fully recurrent

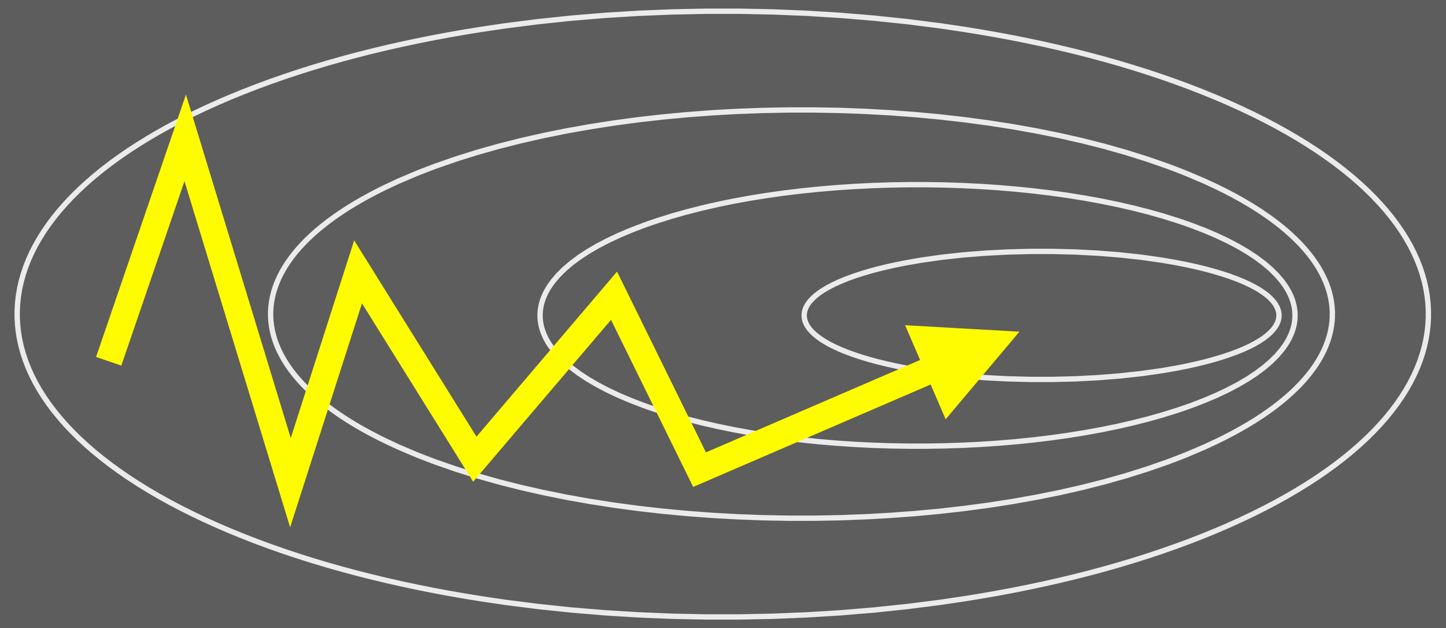
Associative memory as valleys in energy space

Statistical Physics of learning from examples

Seung et. al., PRA 1992

Gradient descent as a Langevin equation

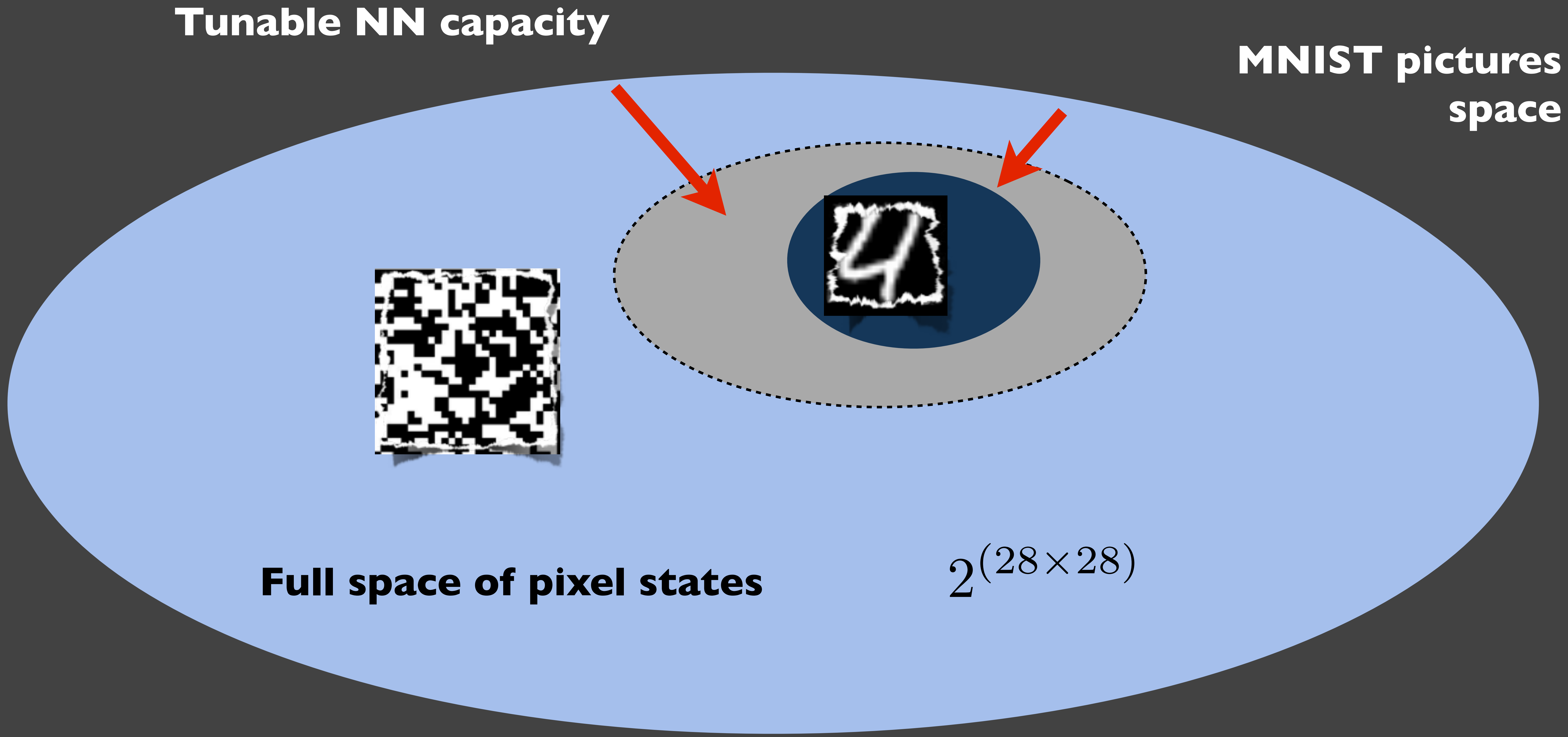
$$\frac{\partial \mathbf{W}}{\partial t} = -\nabla_{\mathbf{W}} E(\mathbf{W}) + \eta(t)$$



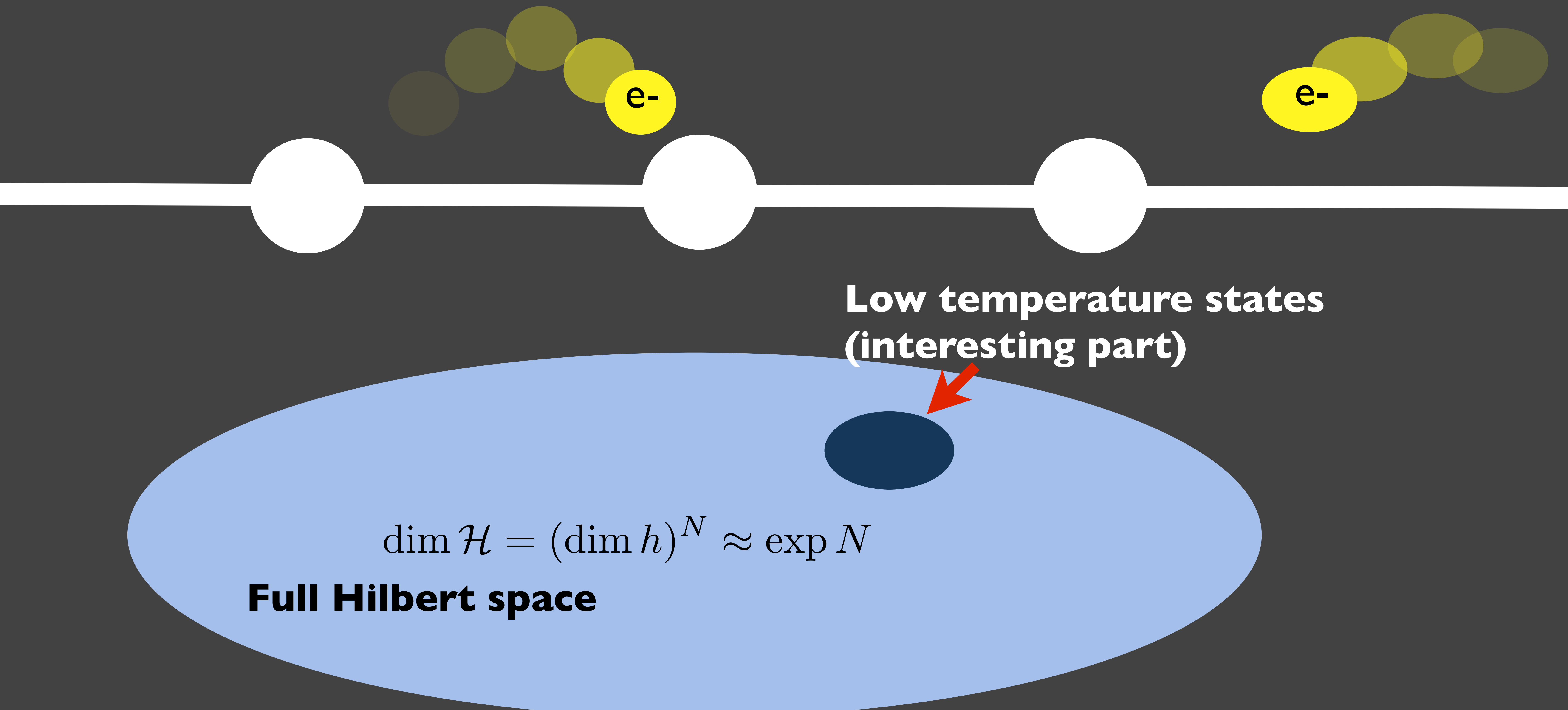
$$P(W) = \frac{e^{-\beta E((W))}}{Z}$$

in the long time limit

Curse of dimensionality in ML



Grows of Quantum State space

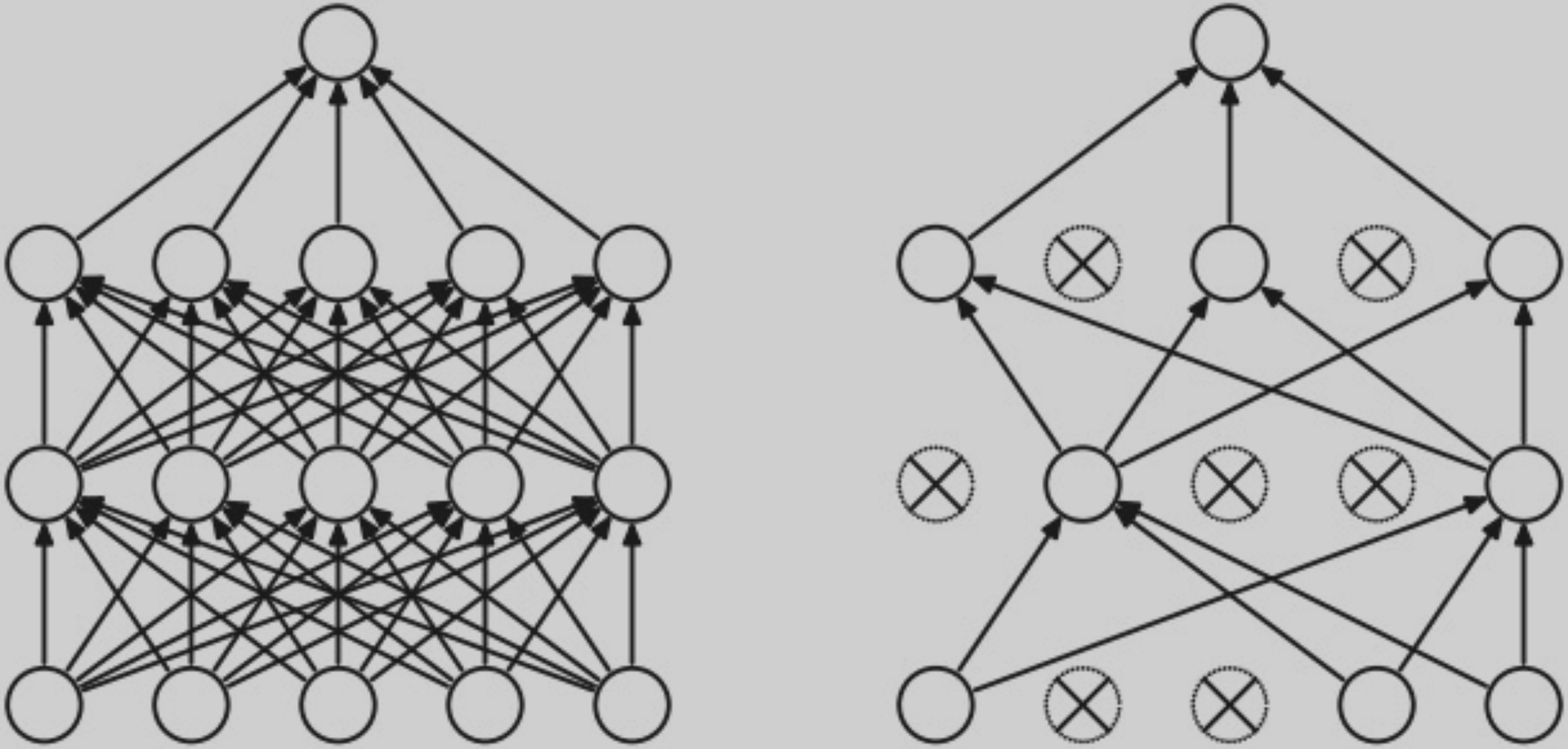



Industry
Fundamental
Research

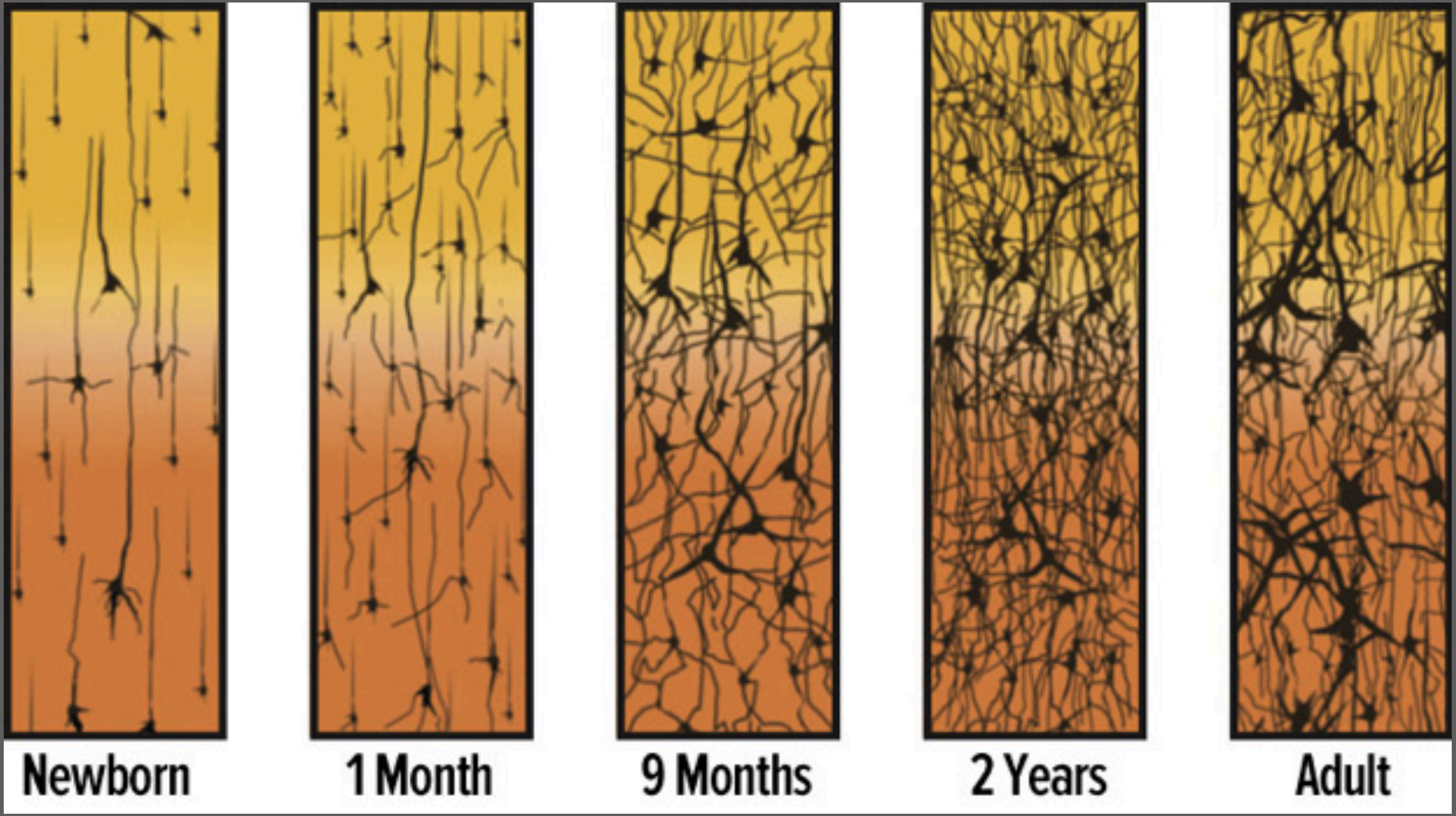
Redundancy of Neural Nets

NN overfit - need Regularized training

Srivastava et.al. JMLR 2014



Compression with prunning



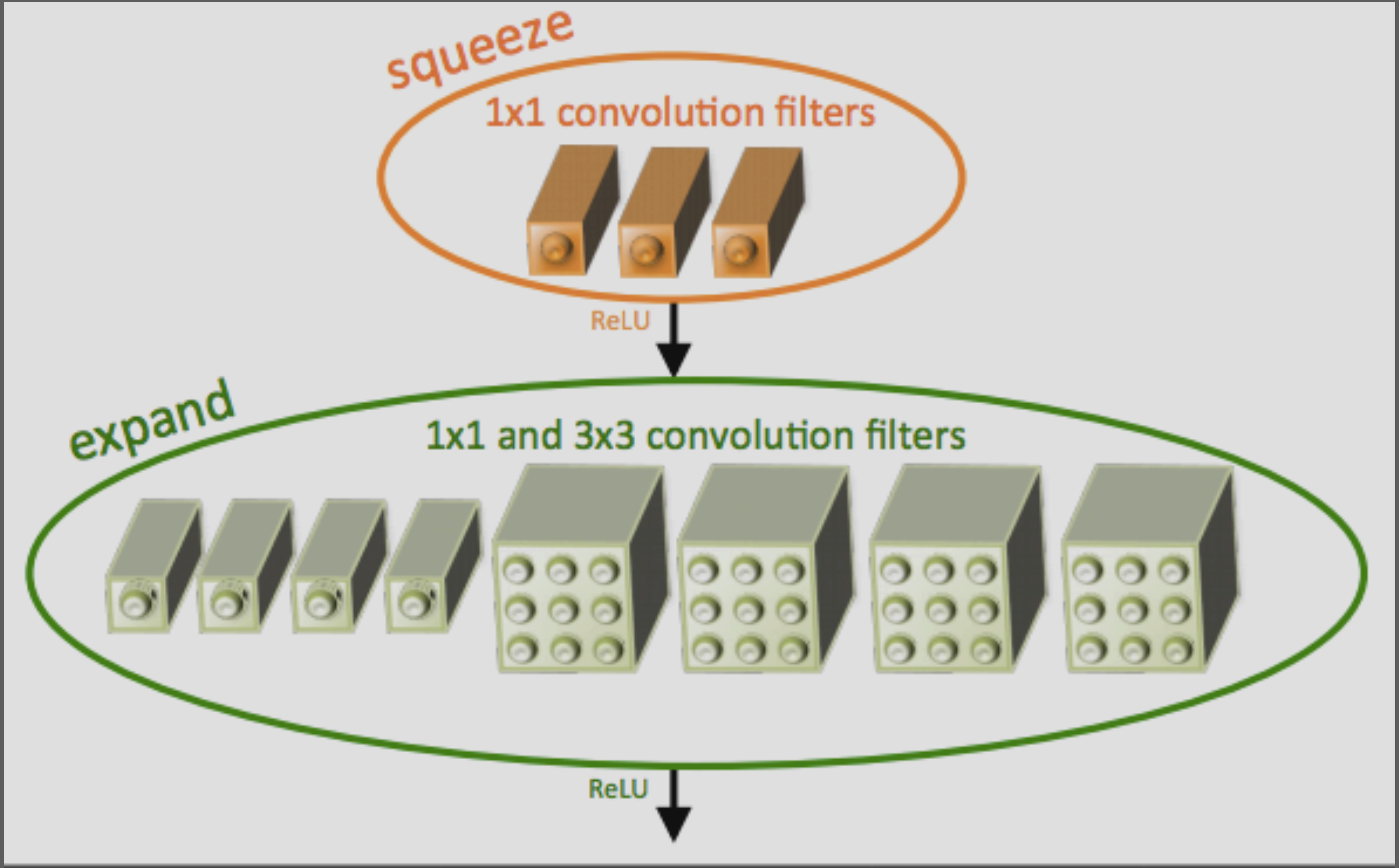
Corel JL, The postnatal development of the human cerebral cortex. 1975

Compression with prunning

Network	Top-1 Error	Top-5 Error	Parameters	Compression Rate
LeNet-300-100 Ref	1.64%	-	267K	
LeNet-300-100 Pruned	1.59%	-	**22K**	**12**×
LeNet-5 Ref	0.80%	-	431K	
LeNet-5 Pruned	0.77%	-	**36K**	**12**×
AlexNet Ref	42.78%	19.73%	61M	
AlexNet Pruned	42.77%	19.67%	**6.7M**	**9**×
VGG16 Ref	31.50%	11.32%	138M	
VGG16 Pruned	31.34%	10.88%	**10.3M**	**13**×

Han et. al. Deep Compression ICLR 2016

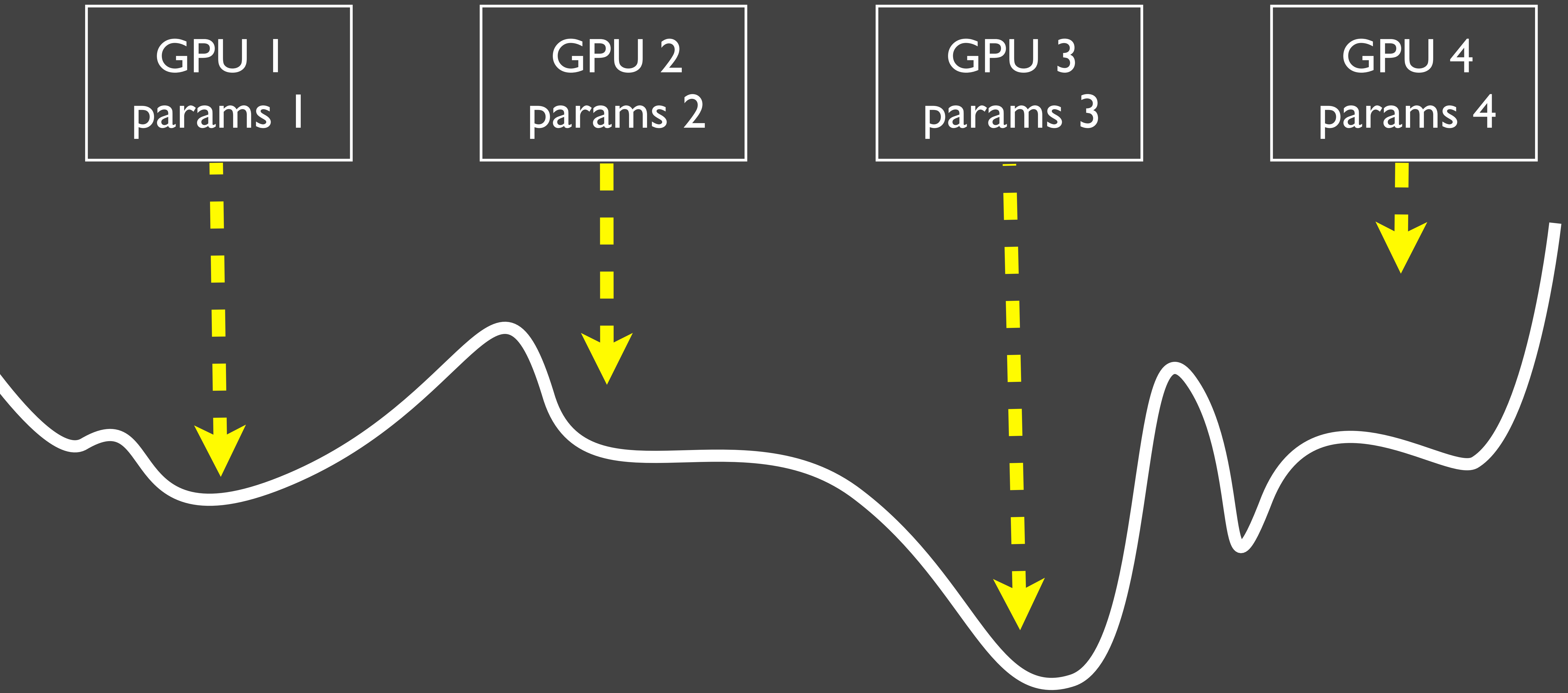
Better architectures pop up regularly



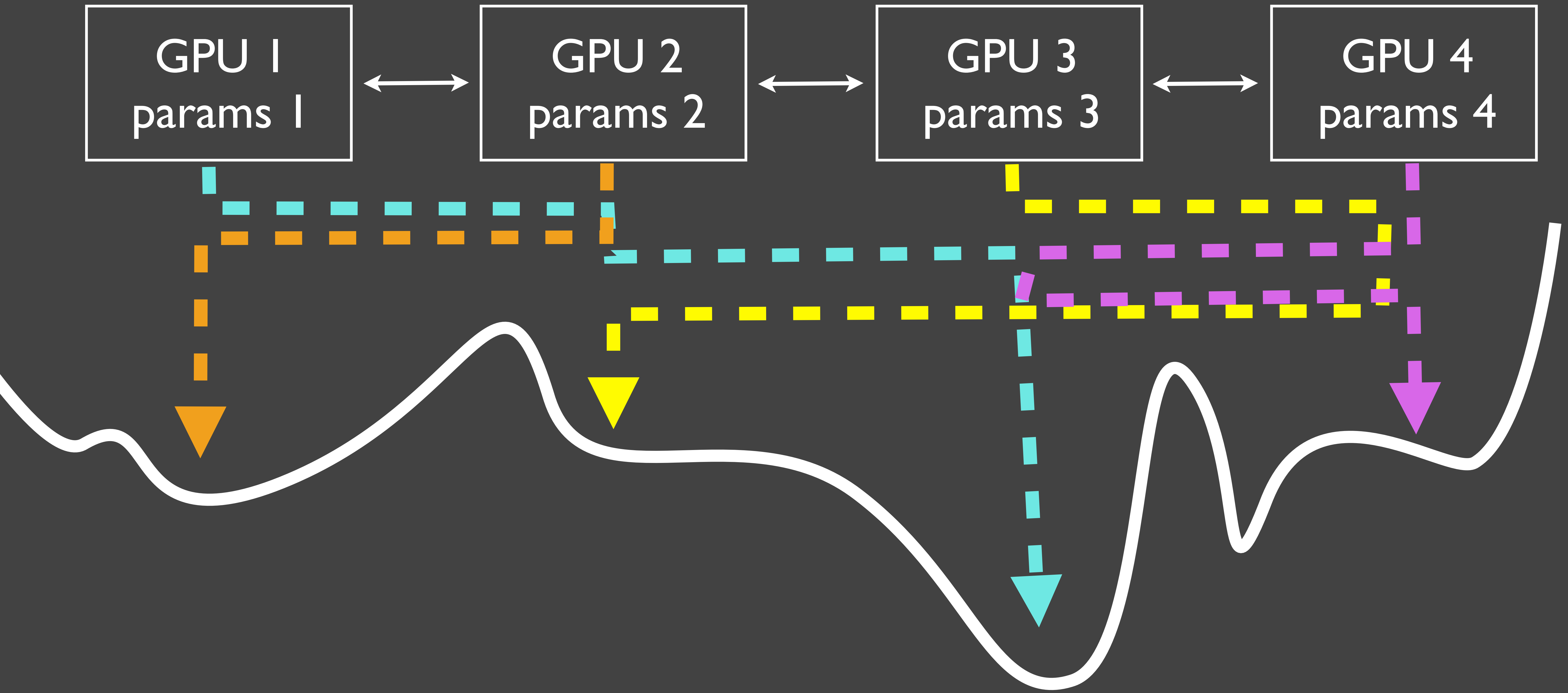
Iandola et.al. arXiv:1602.07360

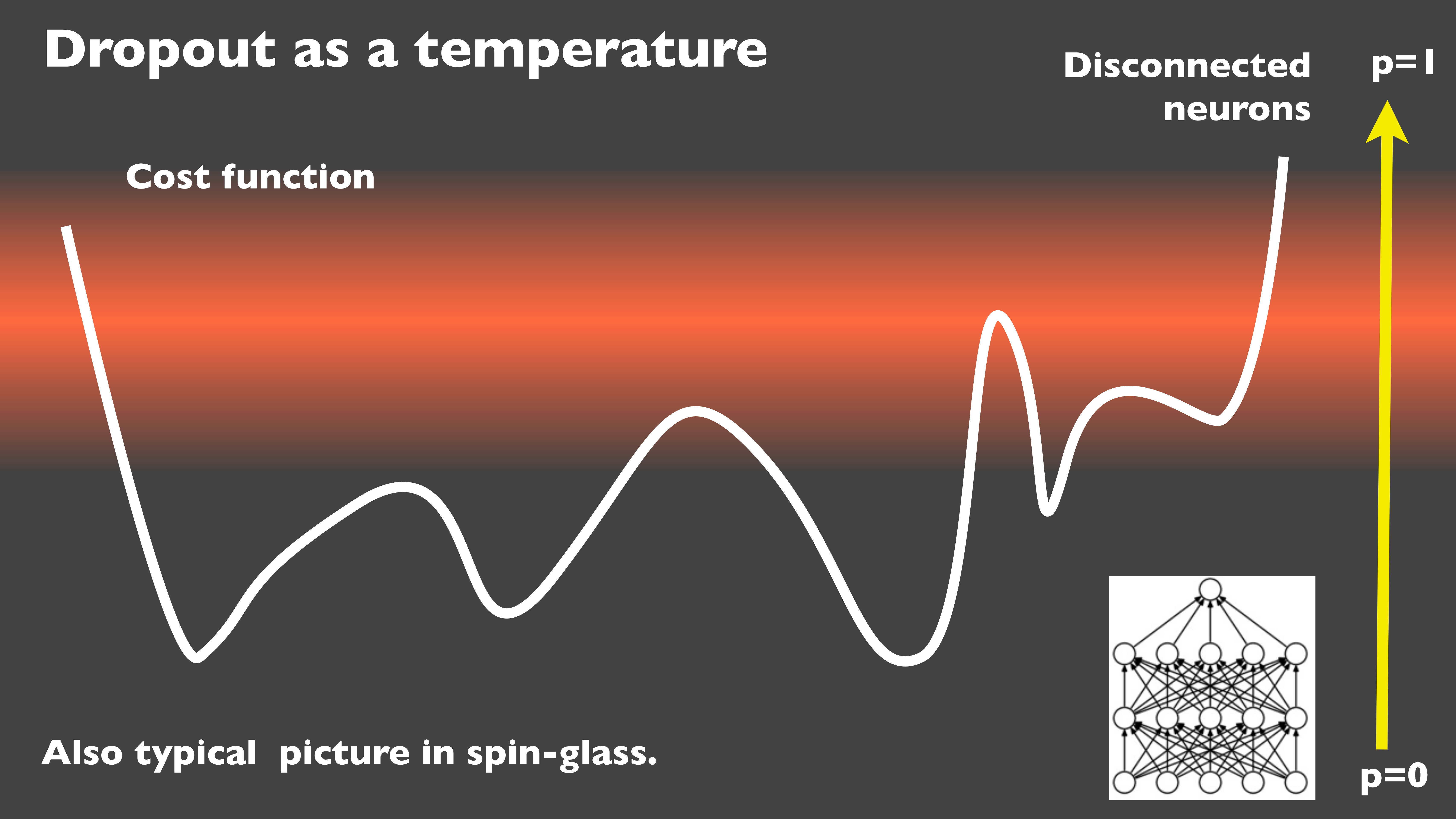
Physics for better learning

Typical search for good parameters



Physicist search for good parameters




Dropout as a temperature
Disconnected
neurons
p=1
Cost function
p=0
Also typical picture in spin-glass.

Dropout as a temperature

Statistical Physics of learning from examples

Seung et. al., PRA 1992

$$\langle\langle E(\mathbf{W})\rangle\rangle = P\epsilon(\mathbf{W})$$

$P = \alpha N$ number of examples scales with respect to network size

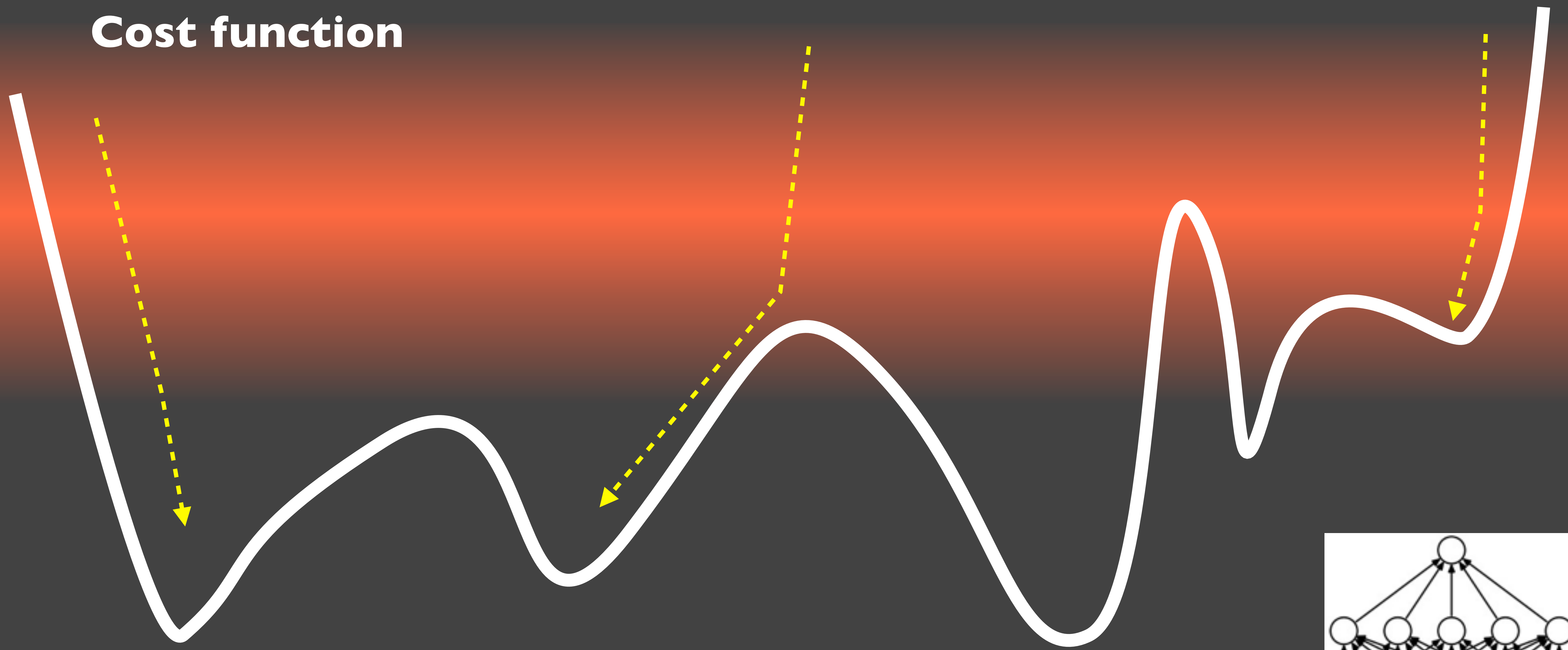
$$P_0(\mathbf{W}) = \frac{\exp\left[-N\beta\alpha\epsilon(\mathbf{W})\right]}{Z}$$ in high-temperature limit

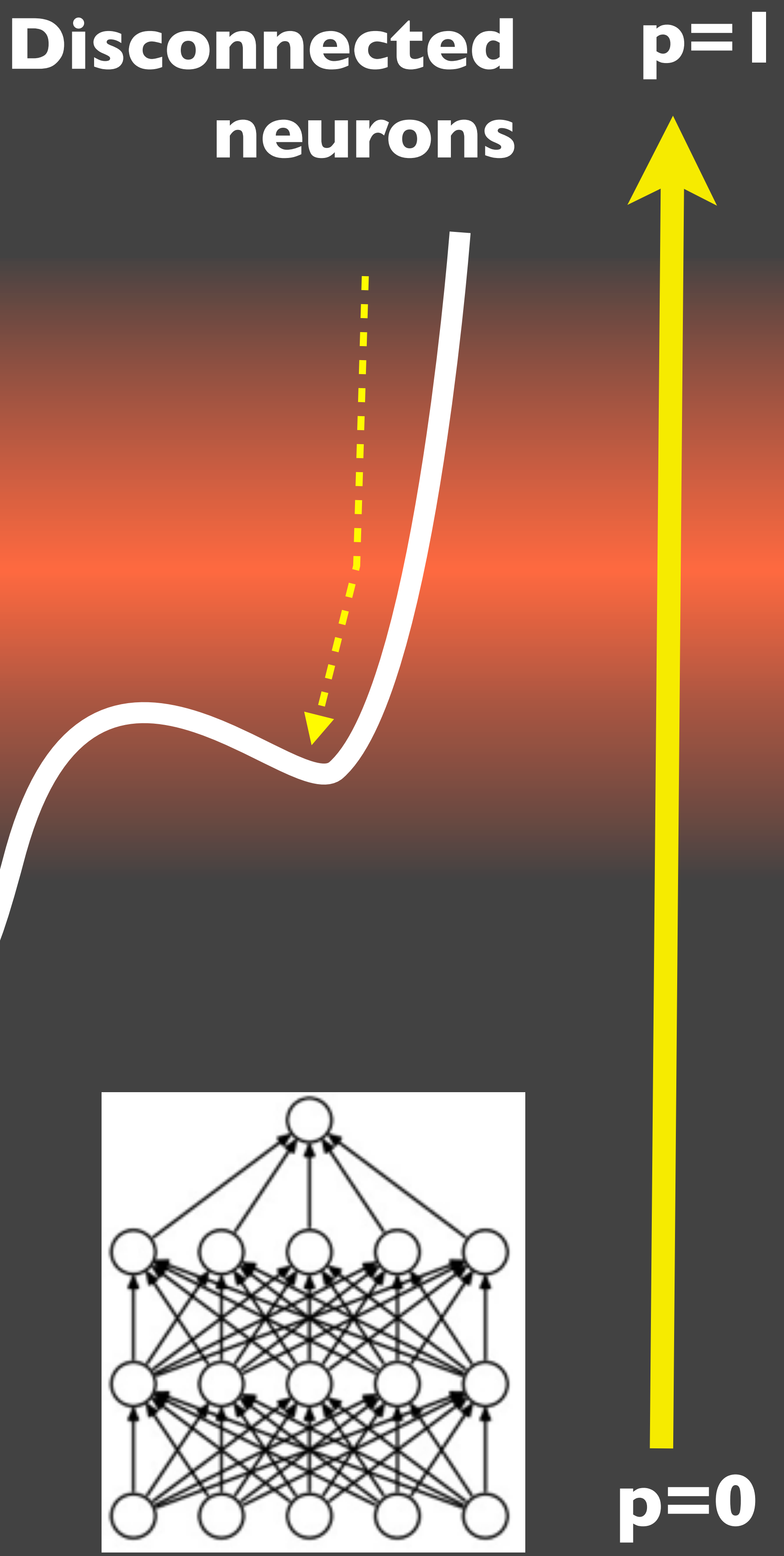
with effective temperature T/α

Annealed Dropout

Rennie et.al. IEEE 2014 (IBM group)

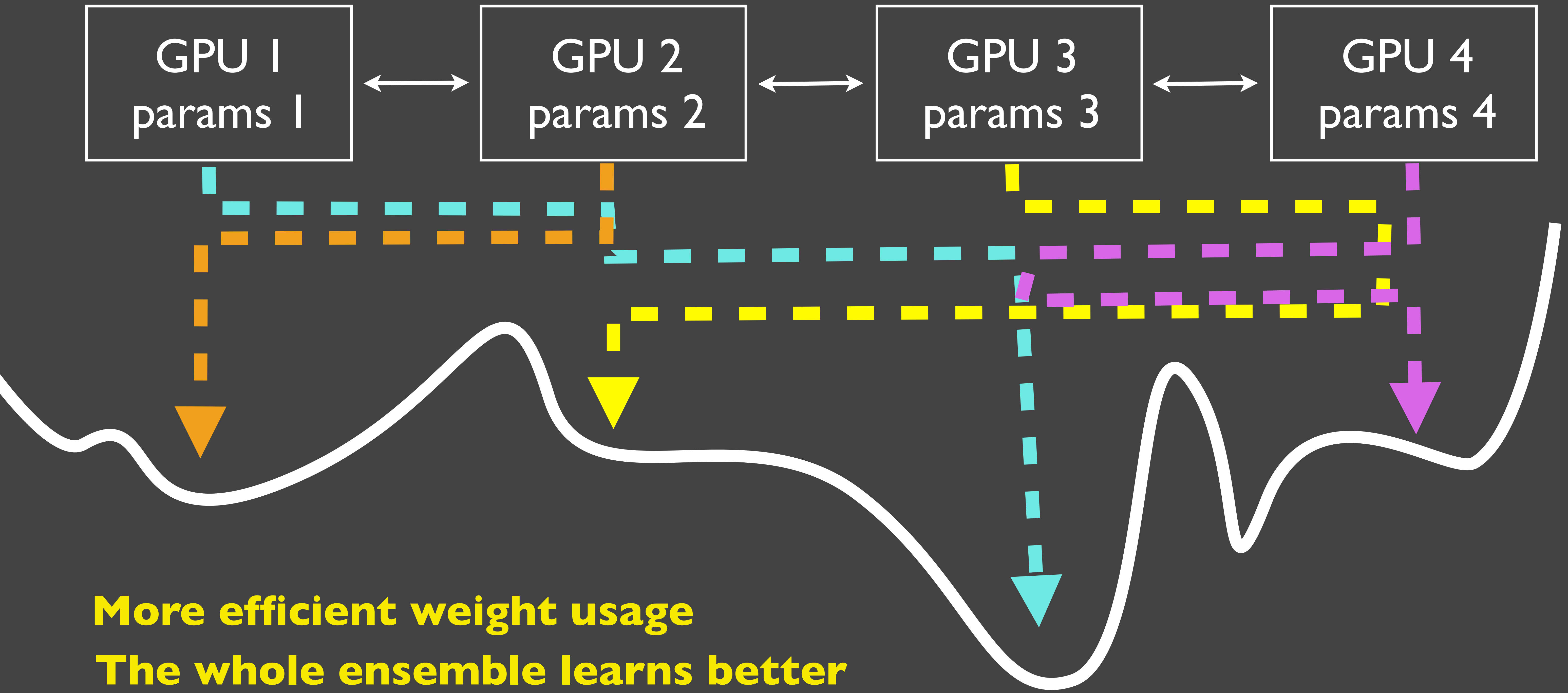
Dropout as a temperature





Annealed Dropout Rennie et.al. IEEE 2014 (IBM group)

Learning with parallel tempering

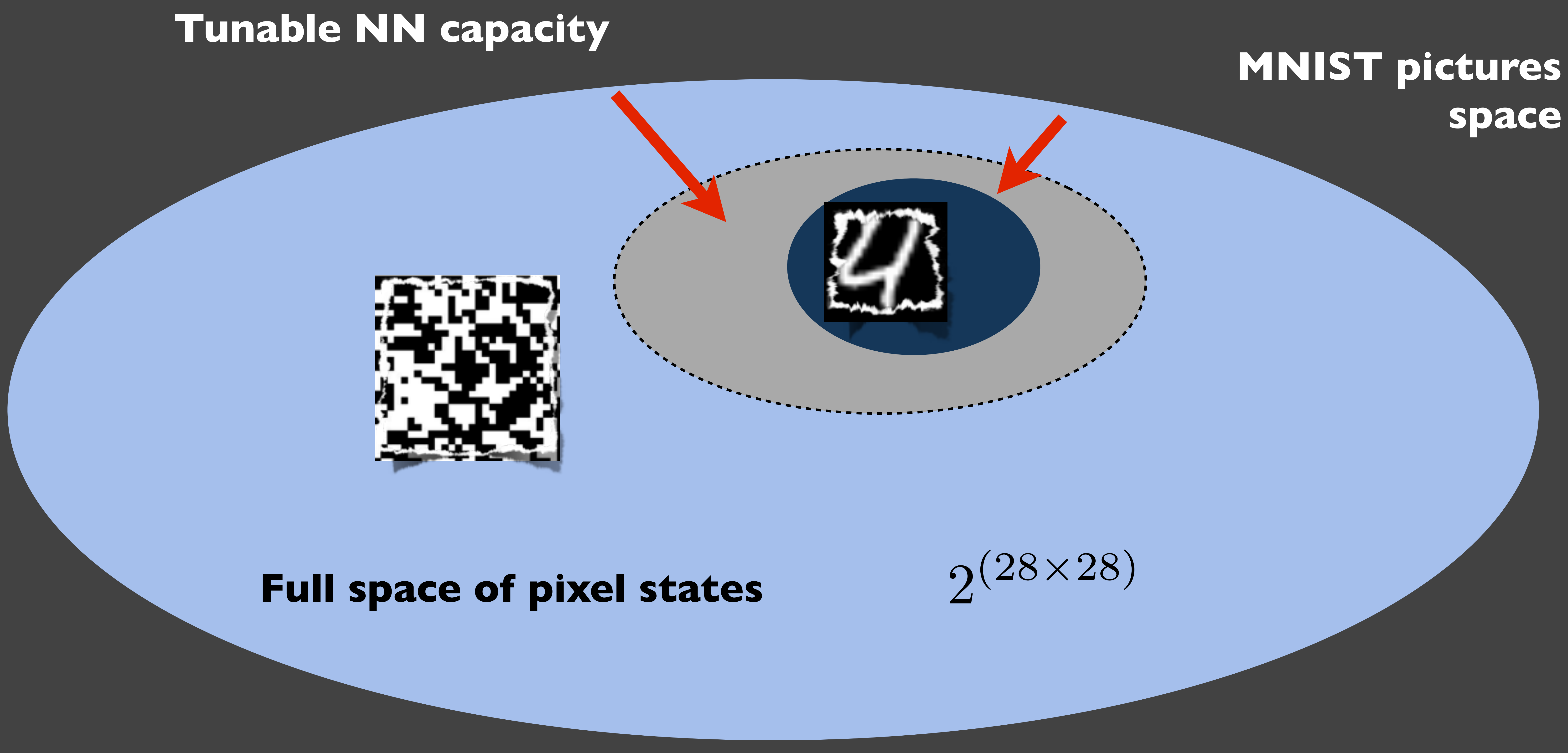


More efficient weight usage

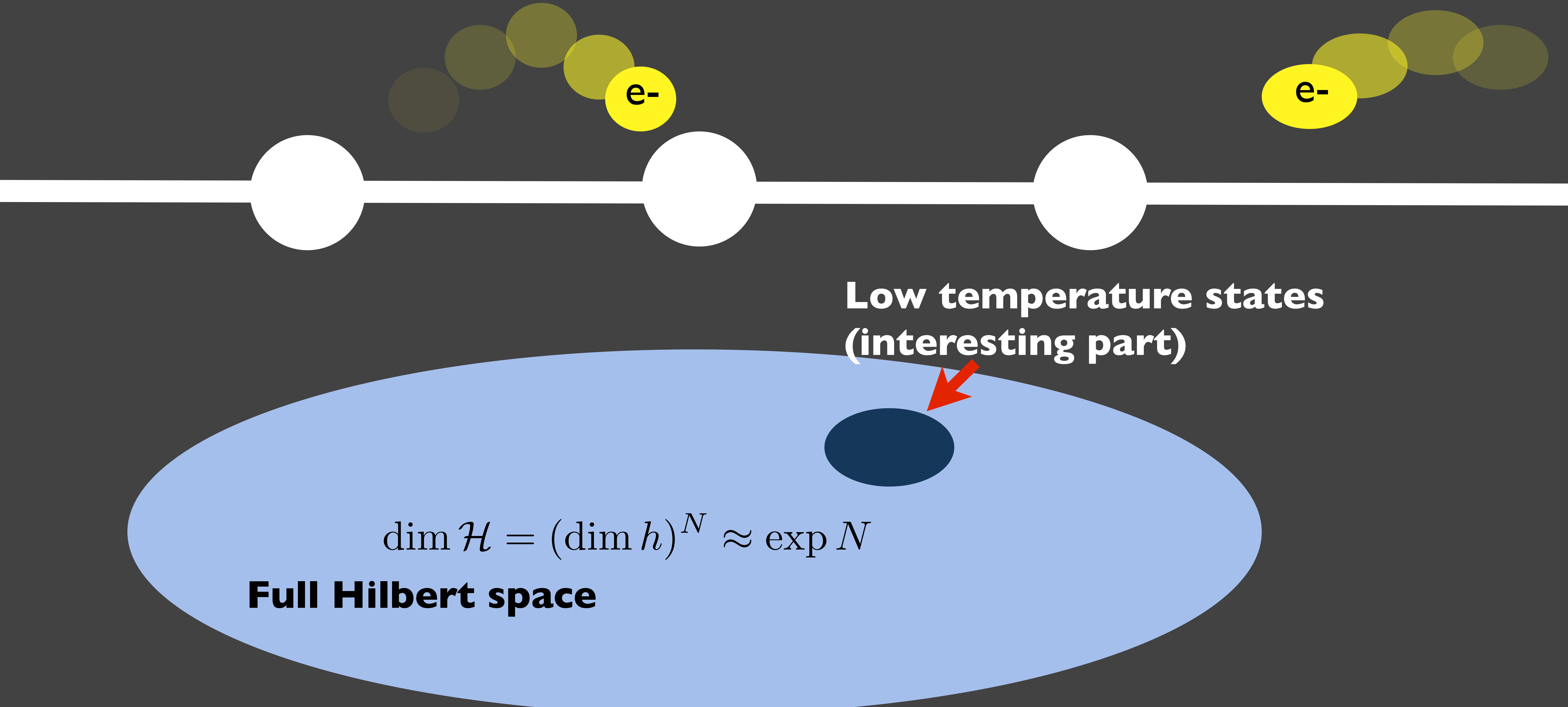
The whole ensemble learns better

Neural Nets
and
Quantum wave functions

Curse of dimensionality in ML



Grows of Quantum State space



What is the quantum wave function ?

What is the quantum wave function ?

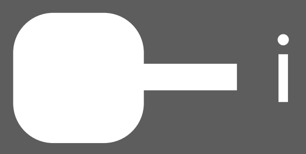
$$|\phi\rangle = \sum_{j_1=\uparrow,\downarrow \ldots j_N=\uparrow,\downarrow} c_{j_1,\ldots,\mathrm{j}_N} |j_1\rangle \ldots |j_N\rangle$$

Large tensor of probabilities for each configuration

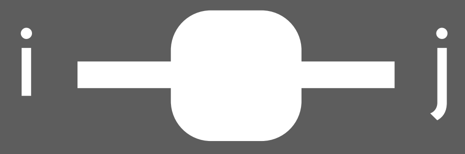
$$|\phi(\uparrow\downarrow\uparrow)\rangle = 0.1$$

Graphical notations for tensors

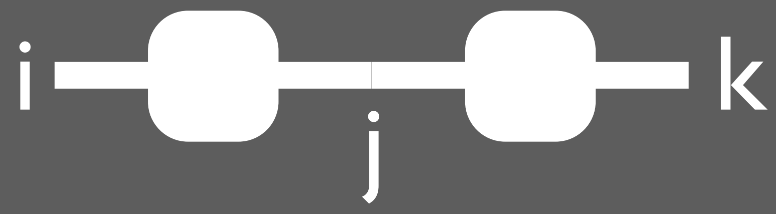
constant



vector v_i



matrix M_{ij}



matrix product $\sum_j M_{ij} M_{jk}$

Representing Quantum states with tensor networks

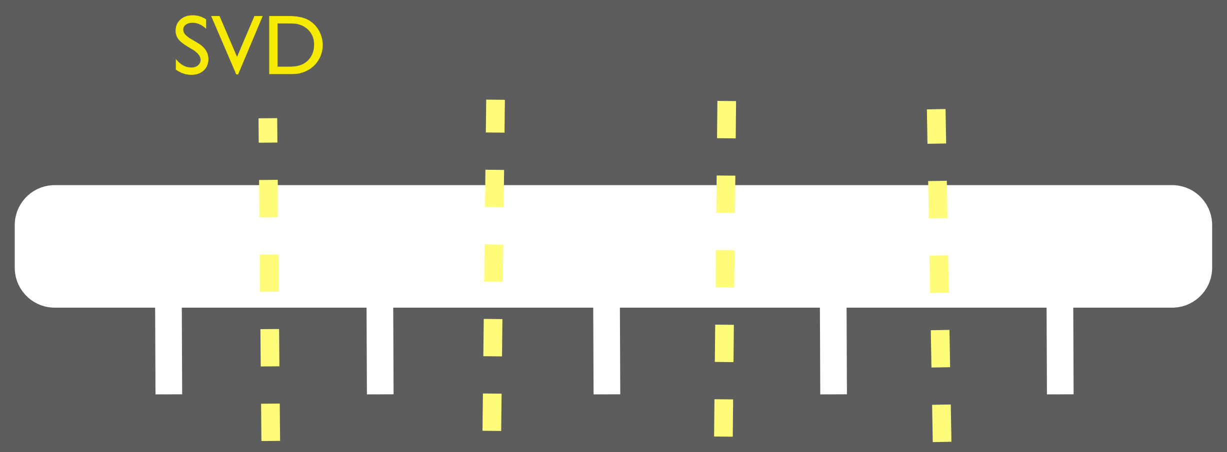
$$c_{j_1 \ldots j_N}$$

Large tensor of probabilities for each configuration

$$|\phi(\uparrow\downarrow\uparrow)\rangle = 0.1$$

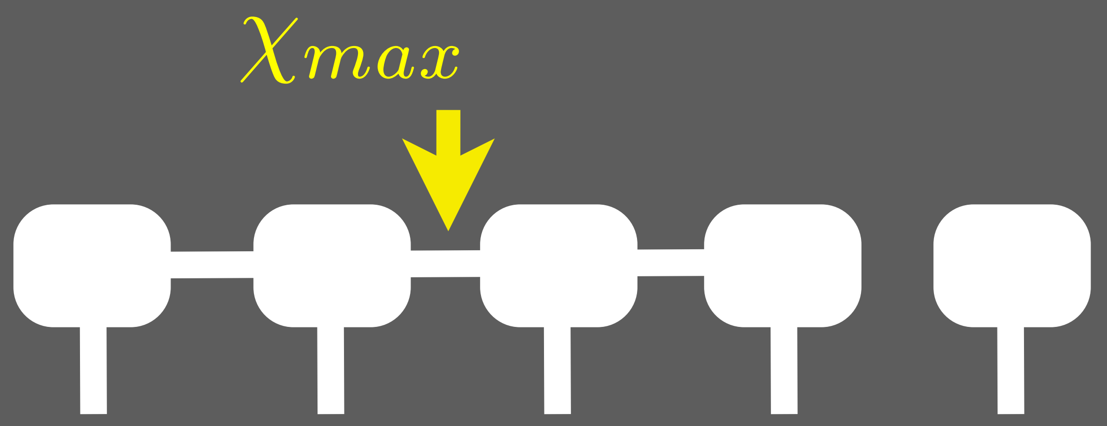
Representing Quantum states with tensor networks

$$c_{j_1 \ldots j_N}$$

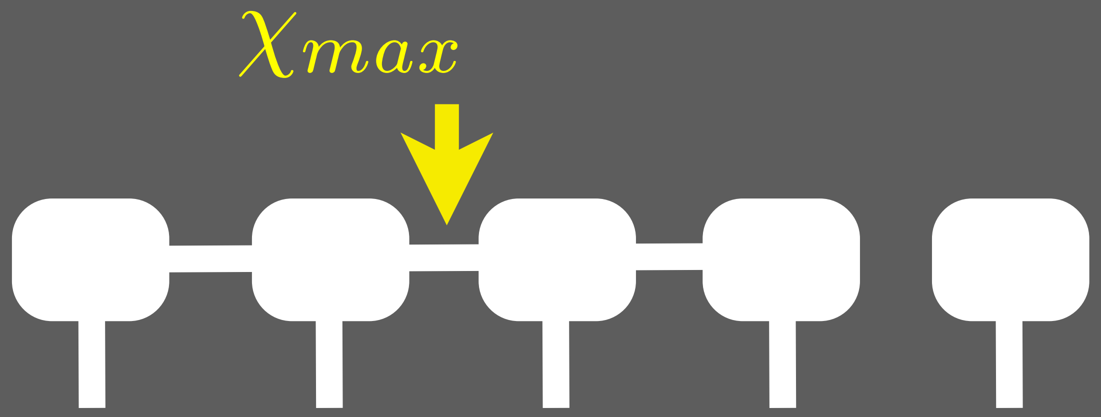


$$\approx$$

$$A^{[1]}_{j_1,\alpha,\beta} A^{[2]}_{j_2,\alpha,\beta} \cdots A^{[N]}_{j_N,\alpha,\beta}$$



Representing Quantum states with tensor networks

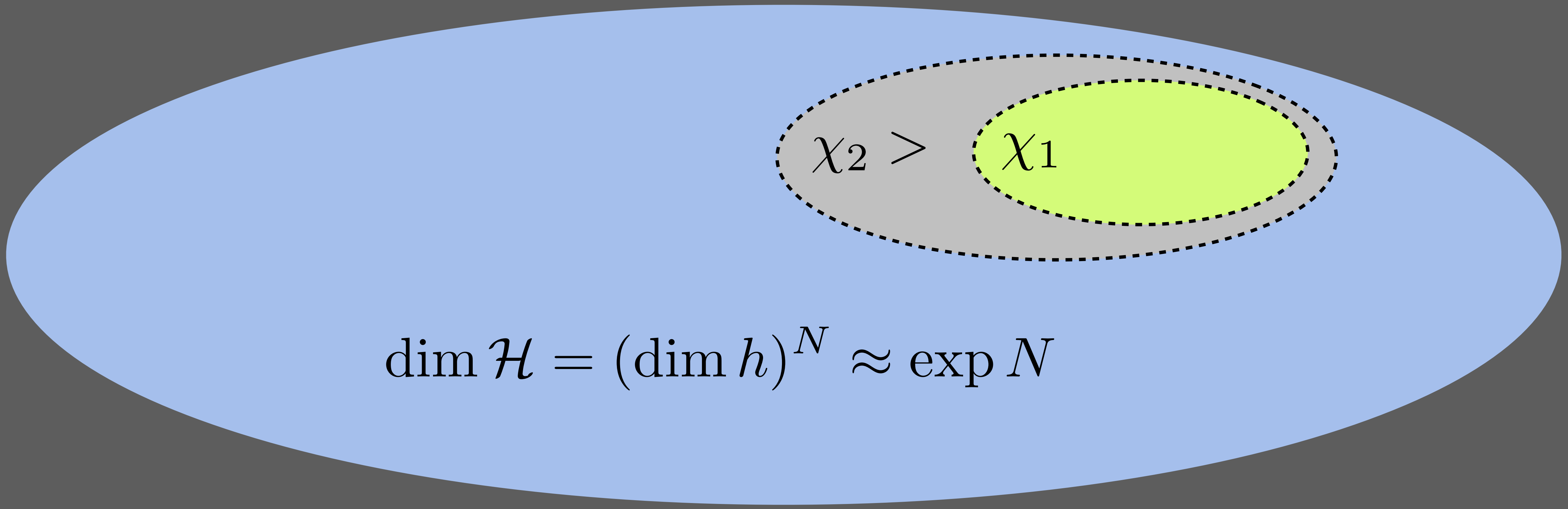


$$\exp(N) \quad \text{vs} \quad Nd\chi_{max}^2$$

Strong reduction of complexity

Representing Quantum states with tensor networks

$$\exp(N) \quad \textbf{vs} \quad N d \chi_{max}^2$$



~~Quantum~~ Machine Learning with Tensor Networks

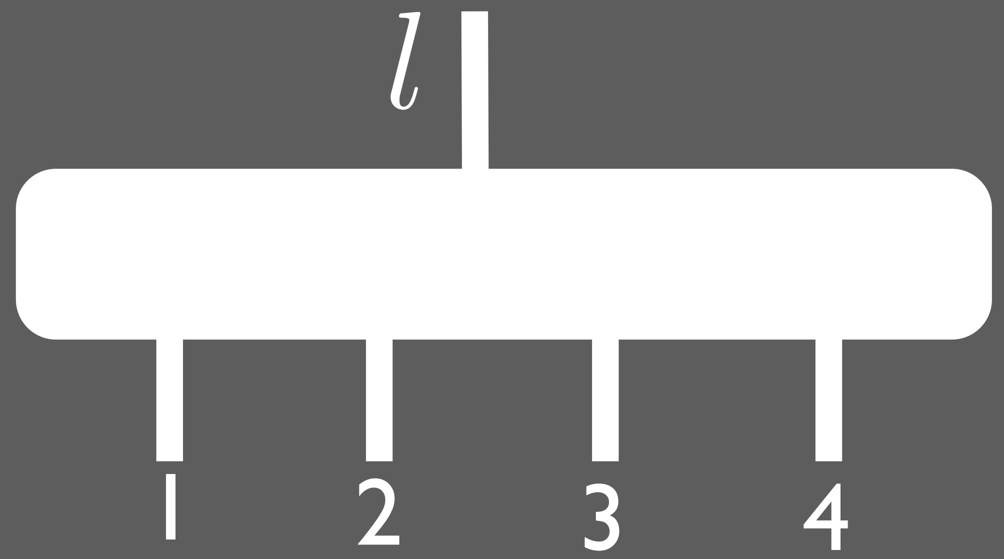
Stoudenmire et. al., NIPS 29, 4799 (2016)
Han et.al., ArXiv:1709.01662 (2017)

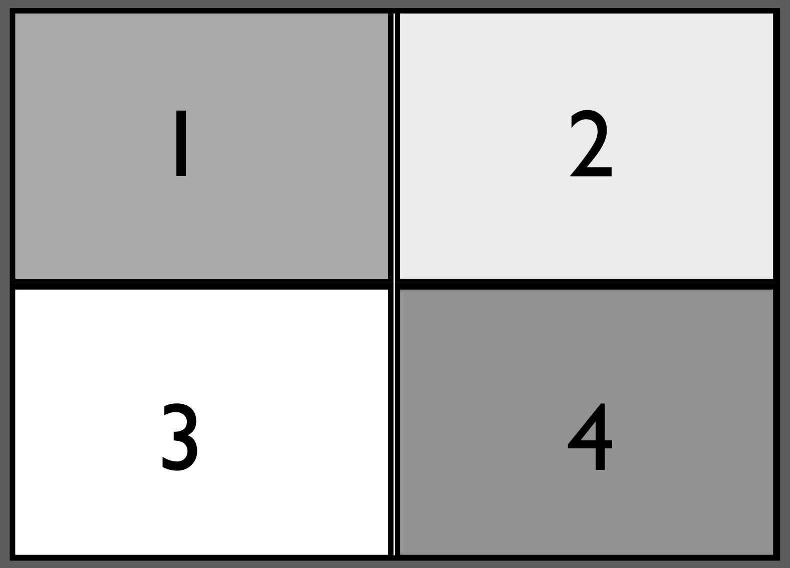
ML with Tensor Networks

$$f^l(\mathbf{x}) = W^l \cdot \Phi(\mathbf{x})$$

ML with Tensor Networks

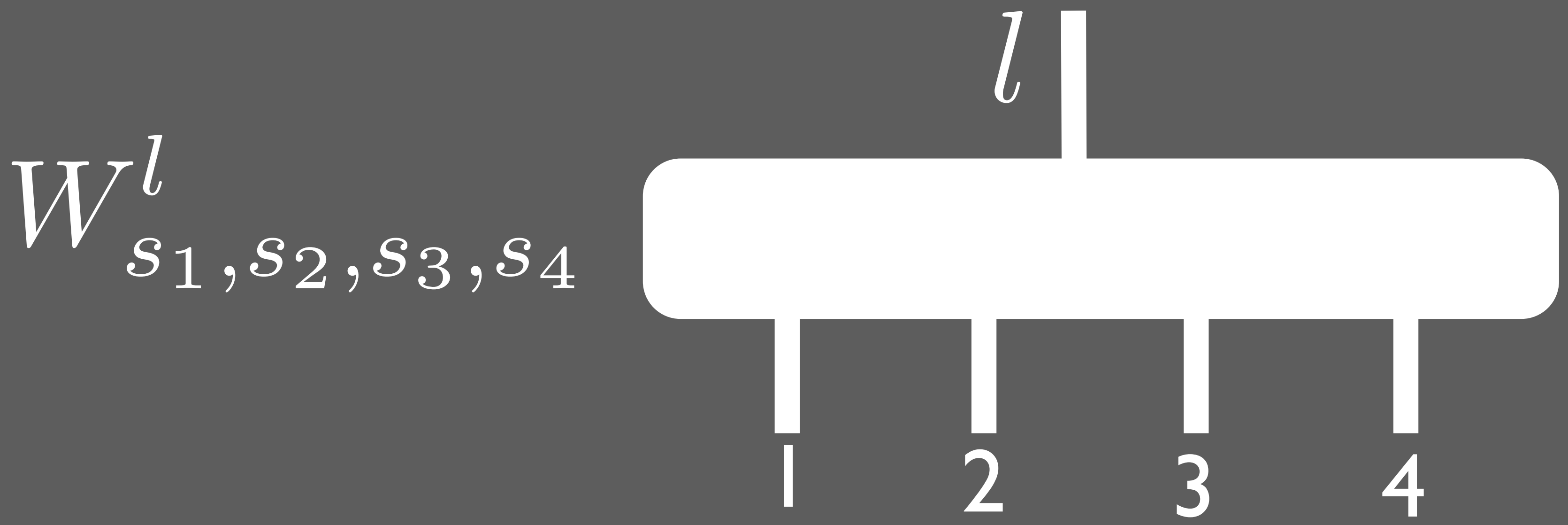
$$f^l(\mathbf{x}) = W^l \cdot \Phi(\mathbf{x})$$

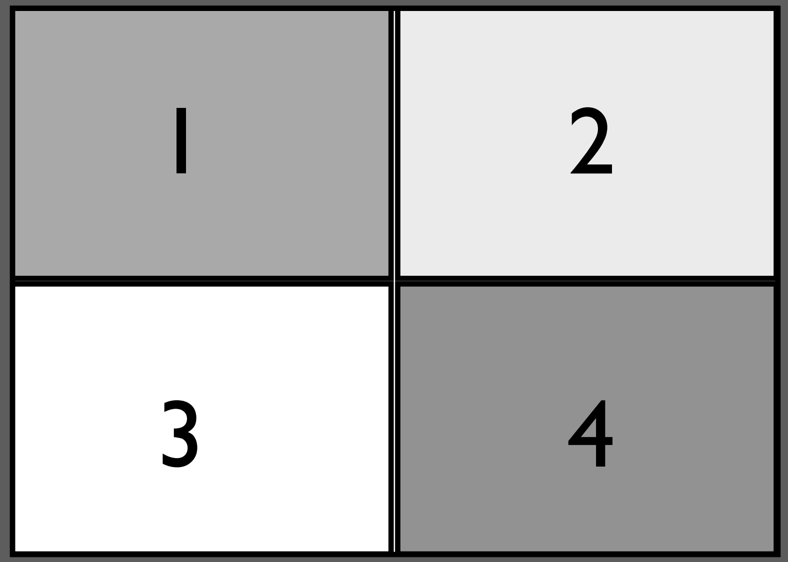




Tensorizing weights and data

$$f^l(\mathbf{x}) = W^l \cdot \Phi(\mathbf{x})$$

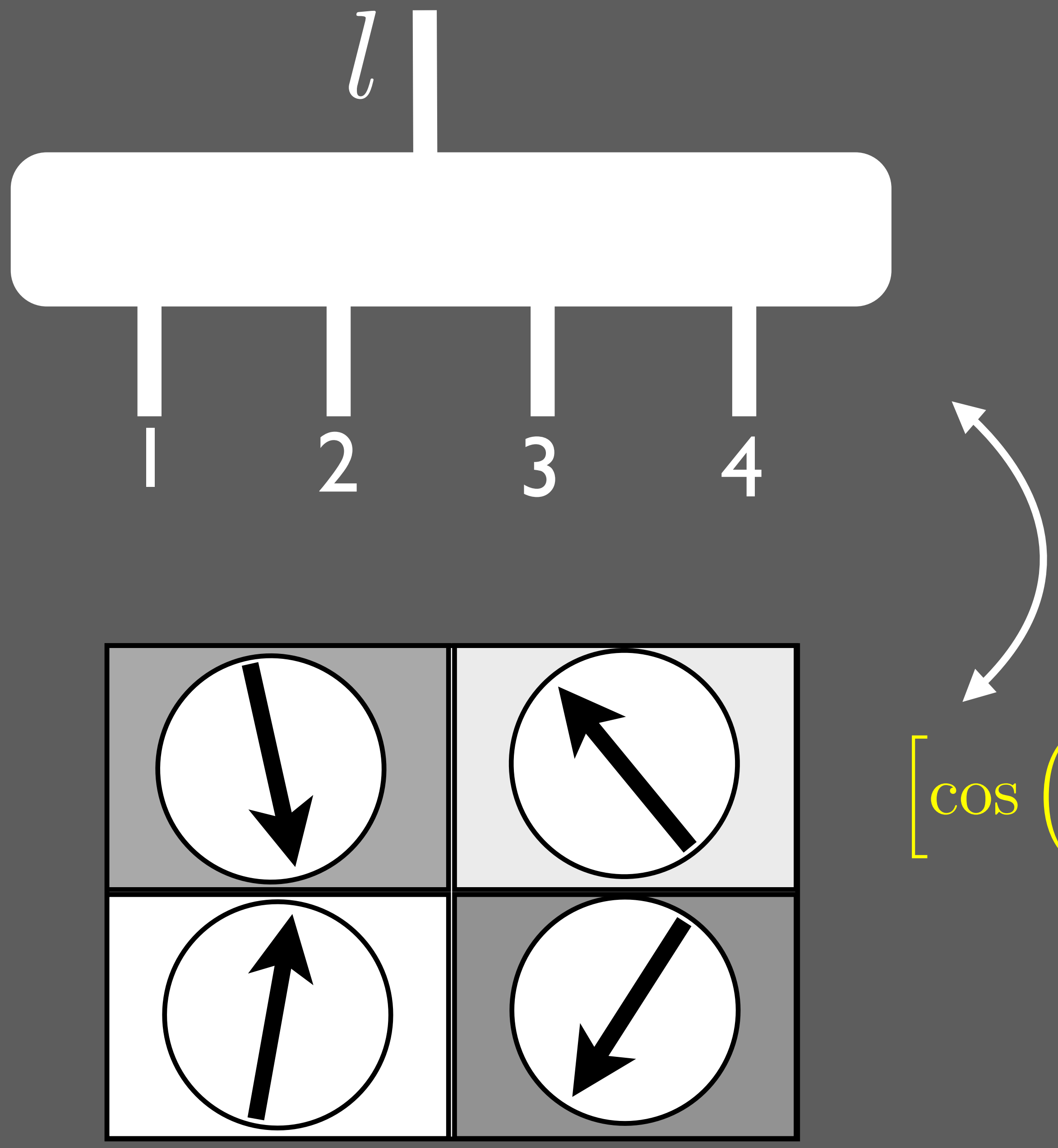




Tensorizing weights and data

$$f^l(\mathbf{x}) = W^l \cdot \Phi(\mathbf{x})$$

$$W^l_{s_1,s_2,s_3,s_4}$$



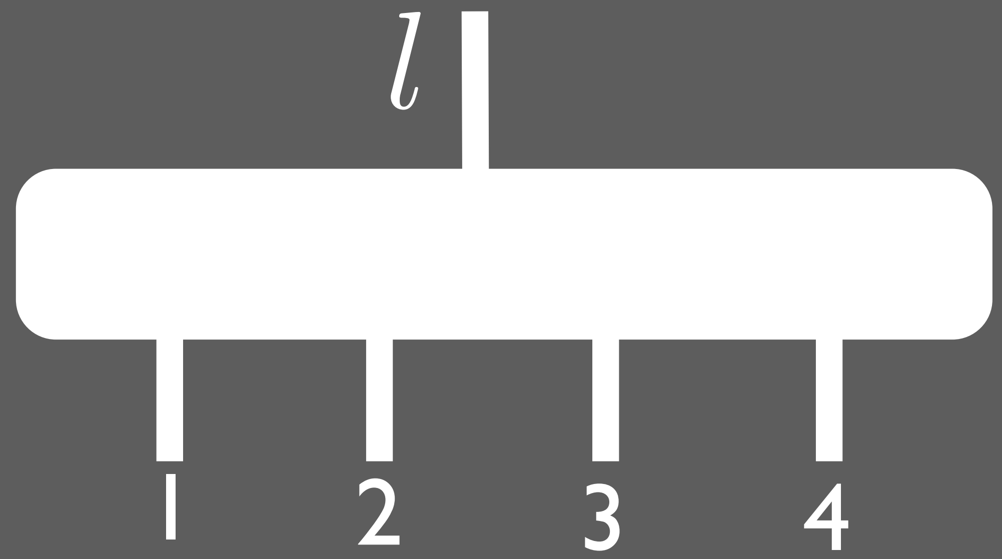
map pixels to vectors

$$\left[\cos\left(\frac{\pi}{2}x_j\right), \sin\left(\frac{\pi}{2}x_j\right)\right]$$

Tensorizing weights and data

$$f^l(\mathbf{x}) = W^l \cdot \Phi(\mathbf{x})$$

$W^l_{s_1,s_2,s_3,s_4}$

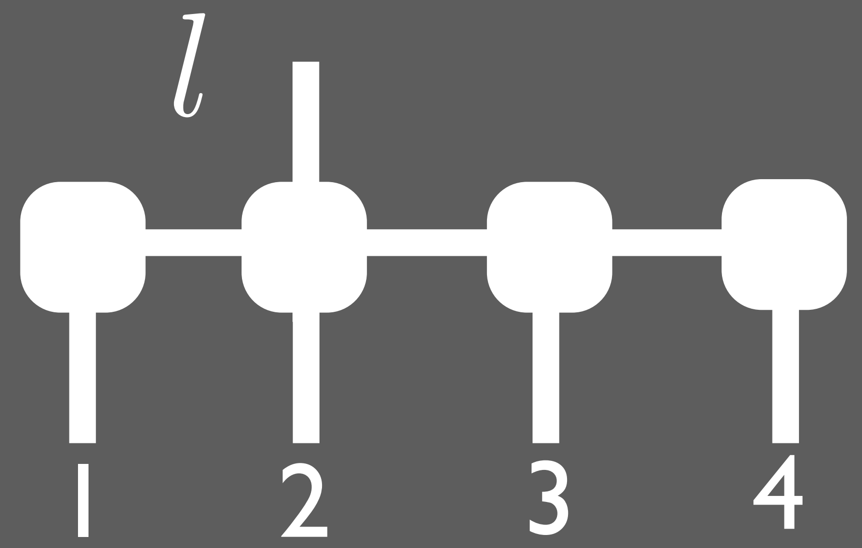


$\phi^{s_1}(x_1) \times \phi^{s_2}(x_2)$
$\times \phi^{s_3}(x_3) \times \phi^{s_4}(x_4)$

Tensorizing weights and data

$$f^l(\mathbf{x}) = W^l \cdot \Phi(\mathbf{x})$$

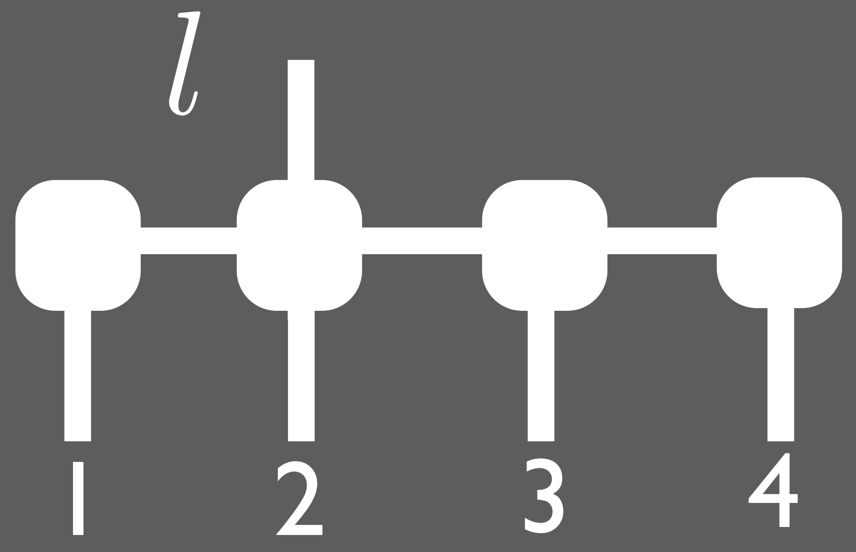
$W^l_{s_1,s_2,s_3,s_4}$



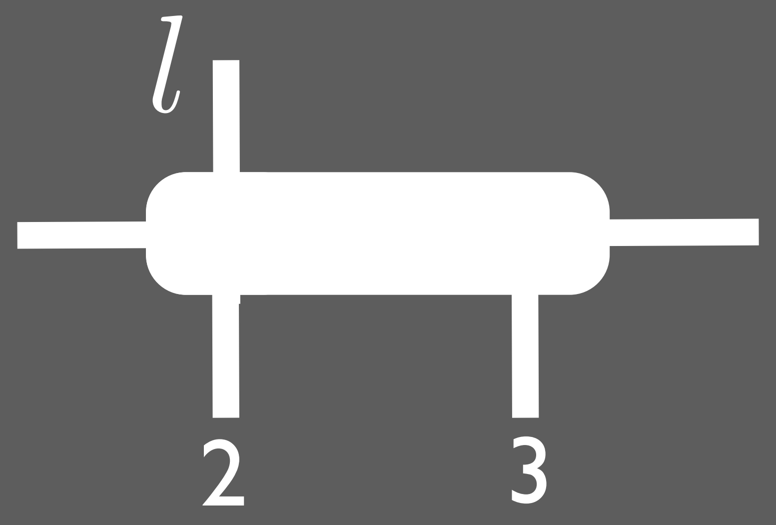
tensorize weights matrix

$$\phi^{s_1}(x_1) \times \phi^{s_2}(x_2) \times \phi^{s_3}(x_3) \times \phi^{s_4}(x_4)$$

Updating the weights

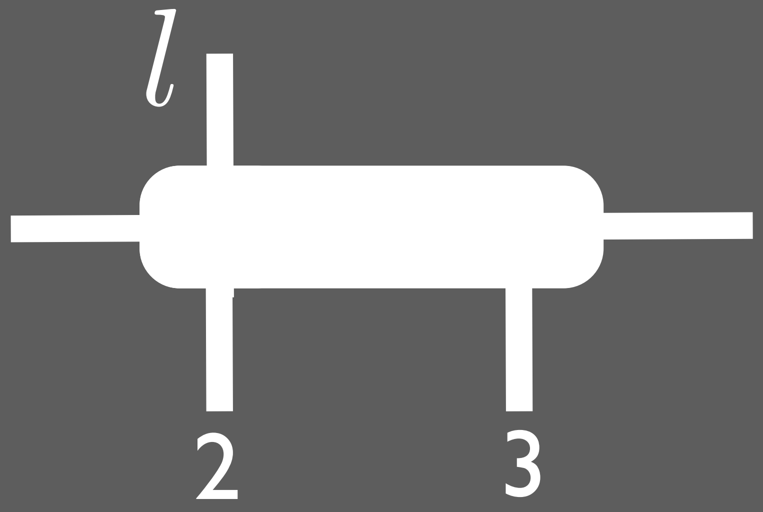


Updating the weights



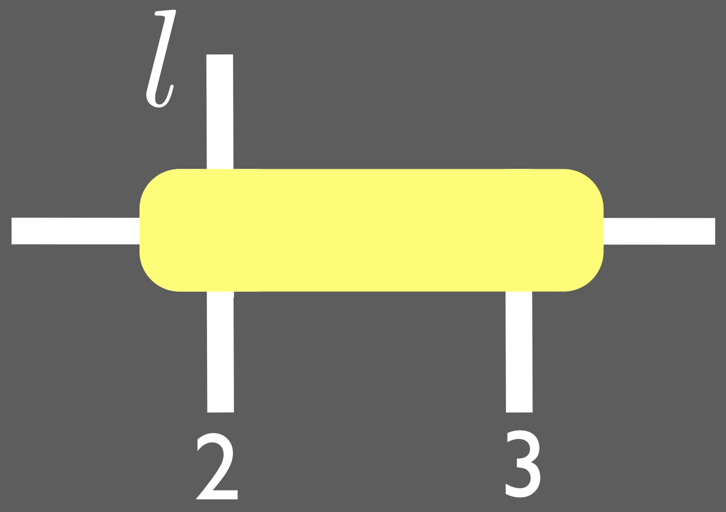
Updating the weights

B^l

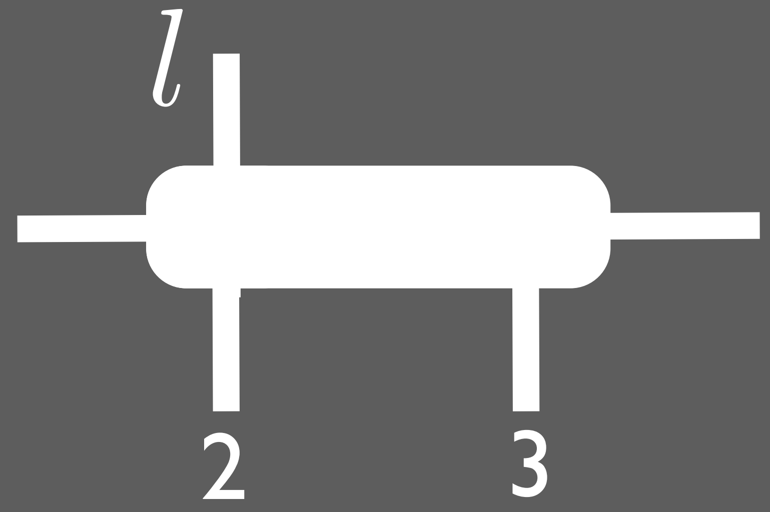


$+$

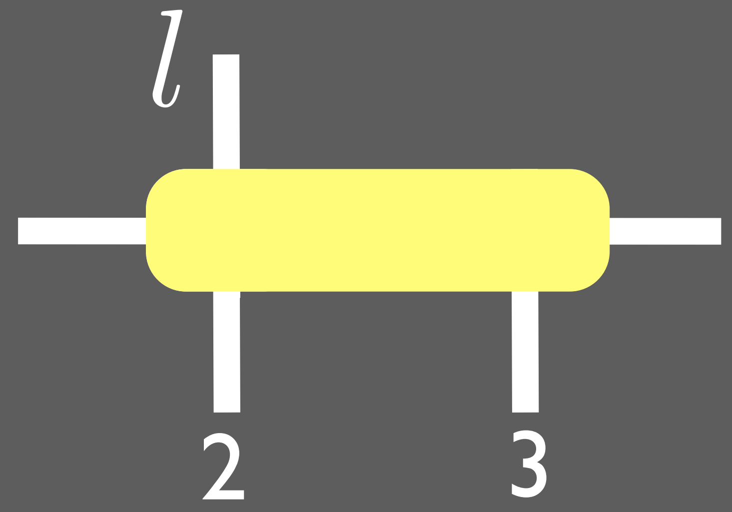
ΔB^l



Updating the weights

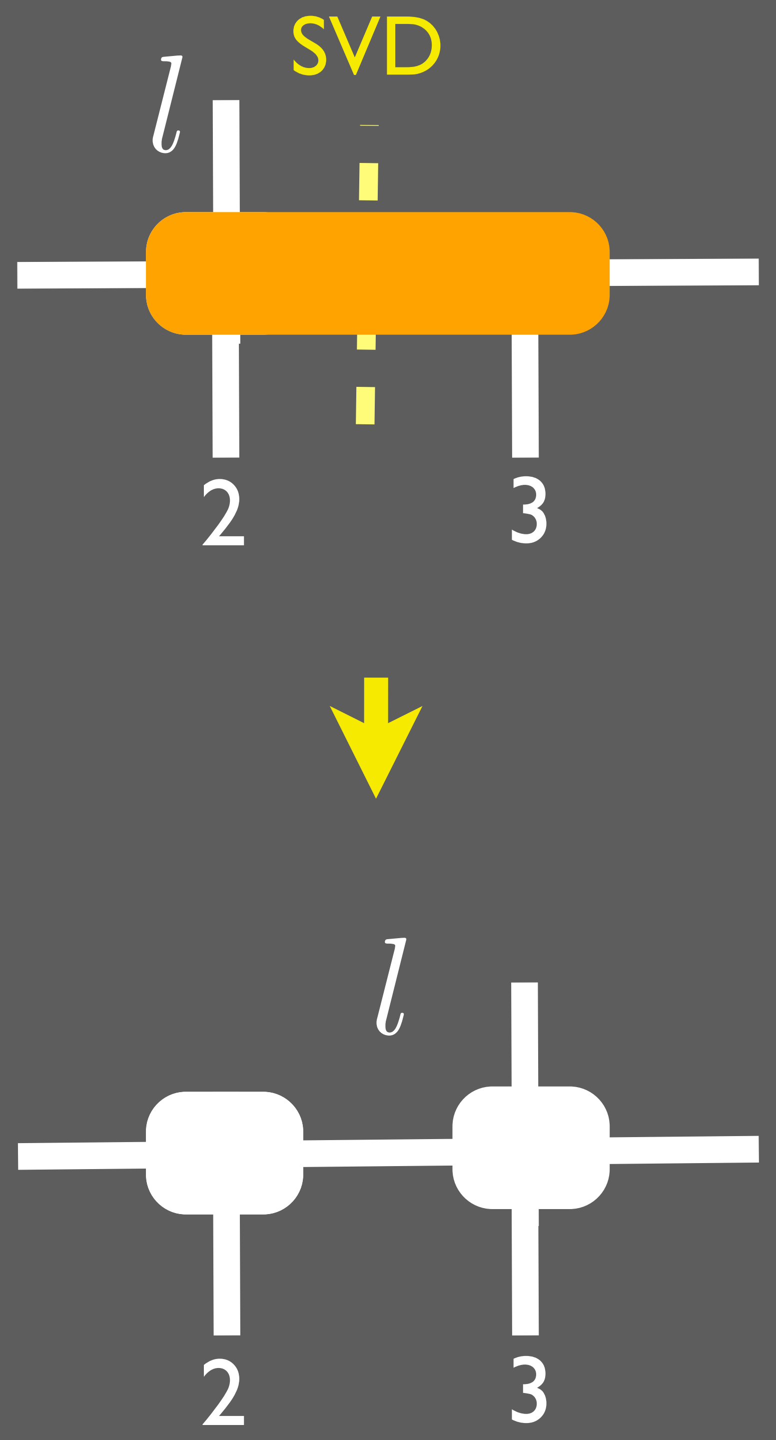
B^l 

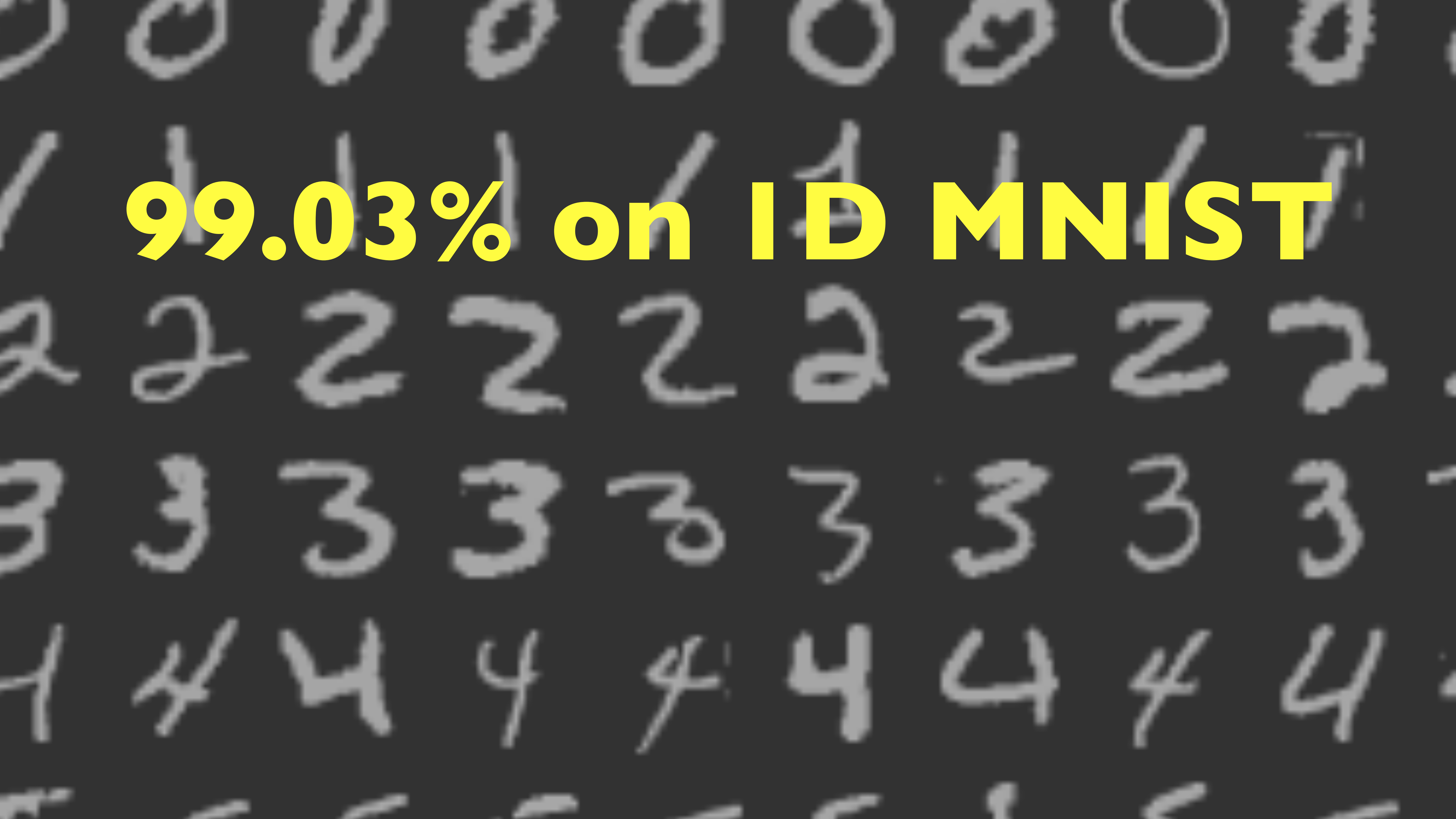
$+$

ΔB^l 

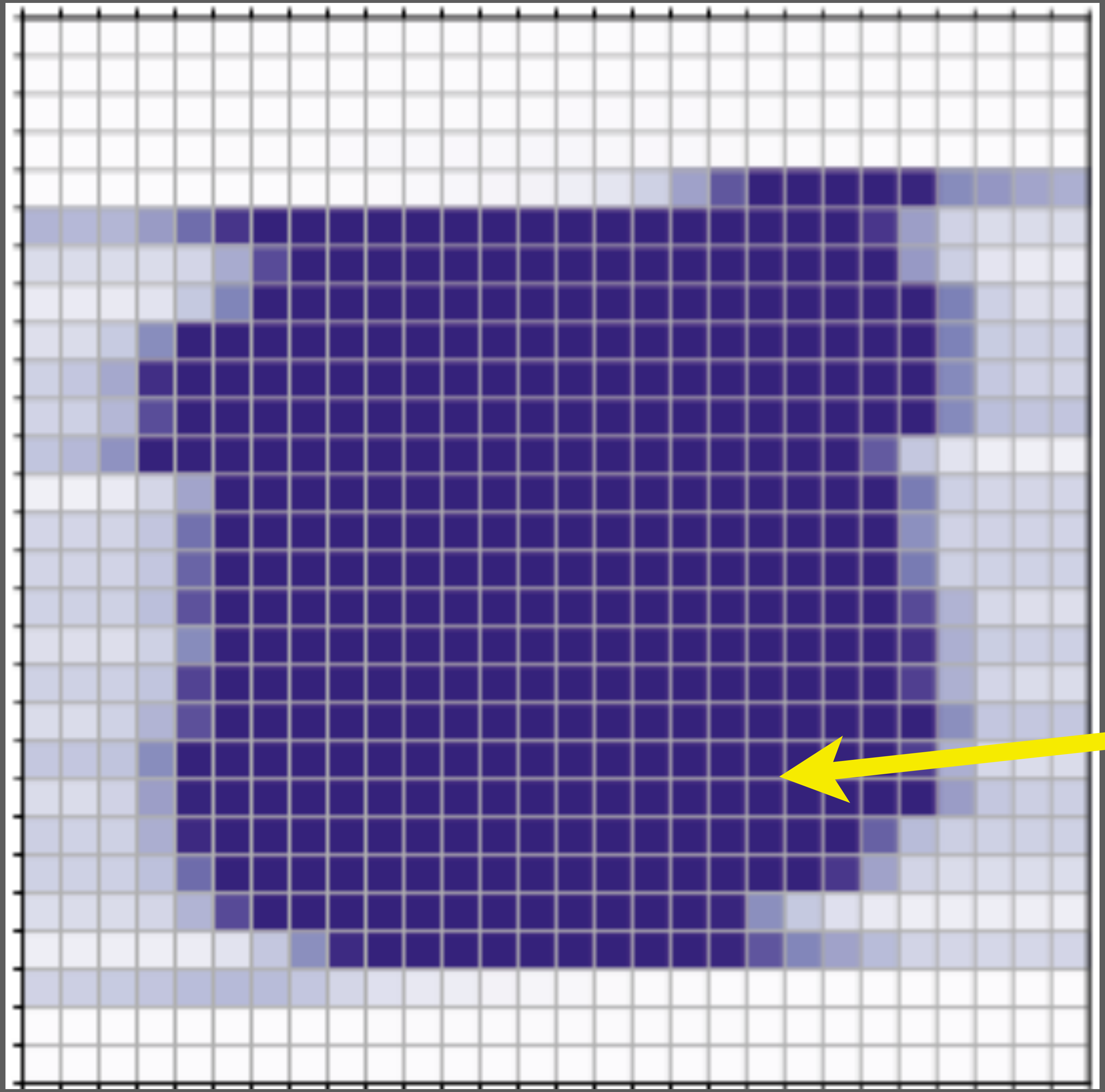
Updating the weights

$\tilde{B}^l$



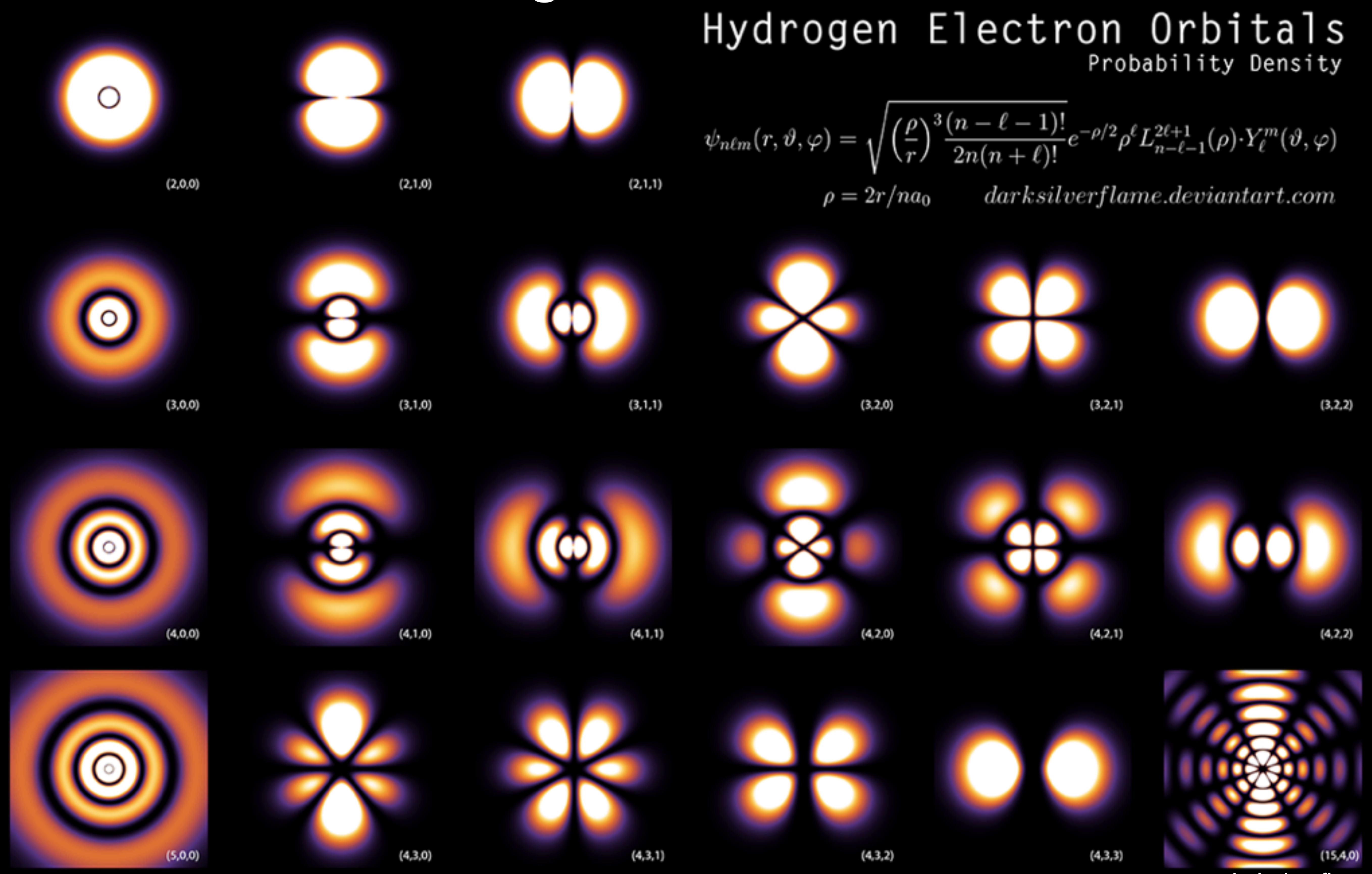

99.03% on ID MNIST

Network adjusts itself according to data complexity



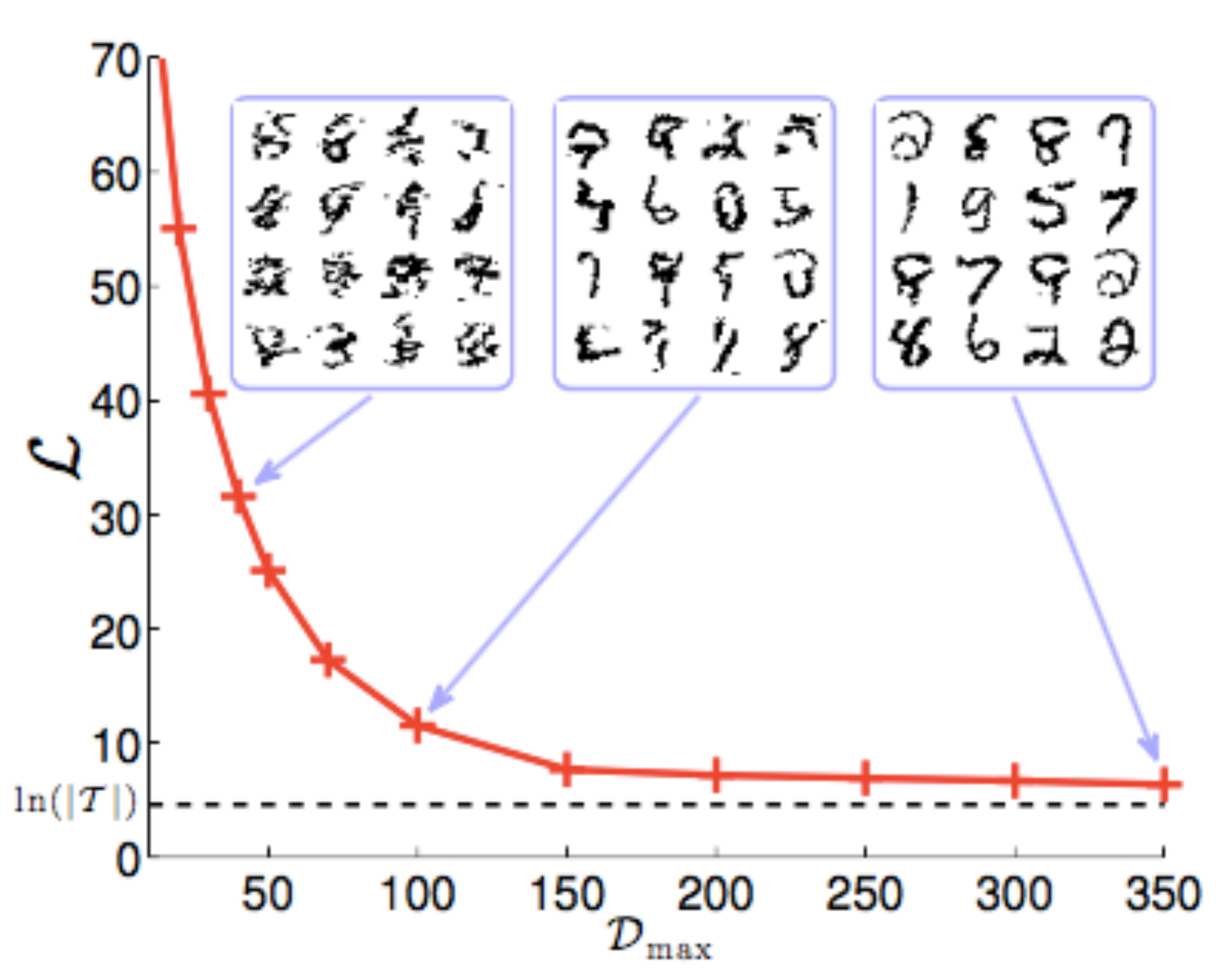
Han et.al., ArXiv:1709.01662 (2017)

Optimal distribution modelling

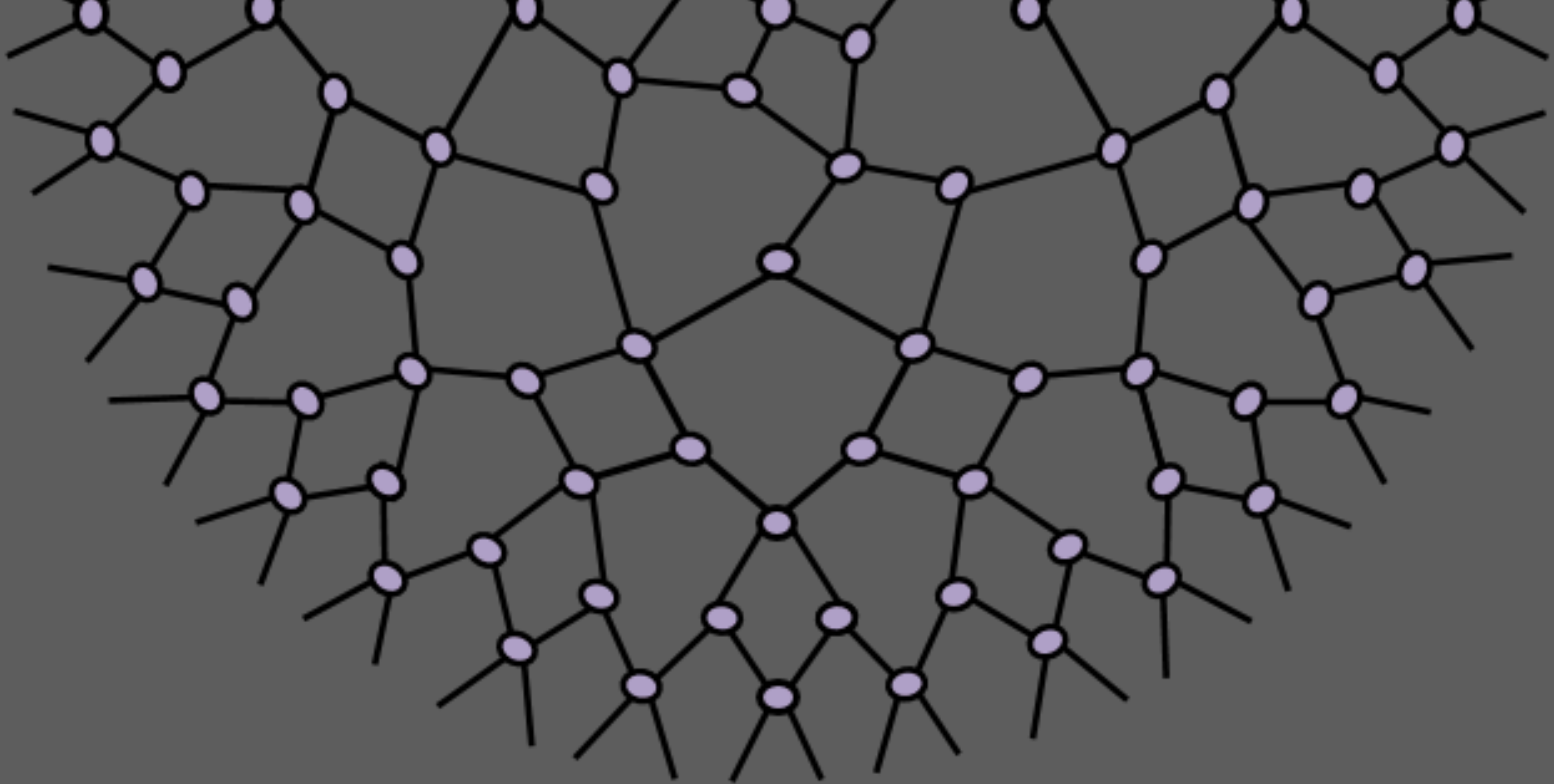


darksilverflame.deviantart.com

Optimal distribution modelling

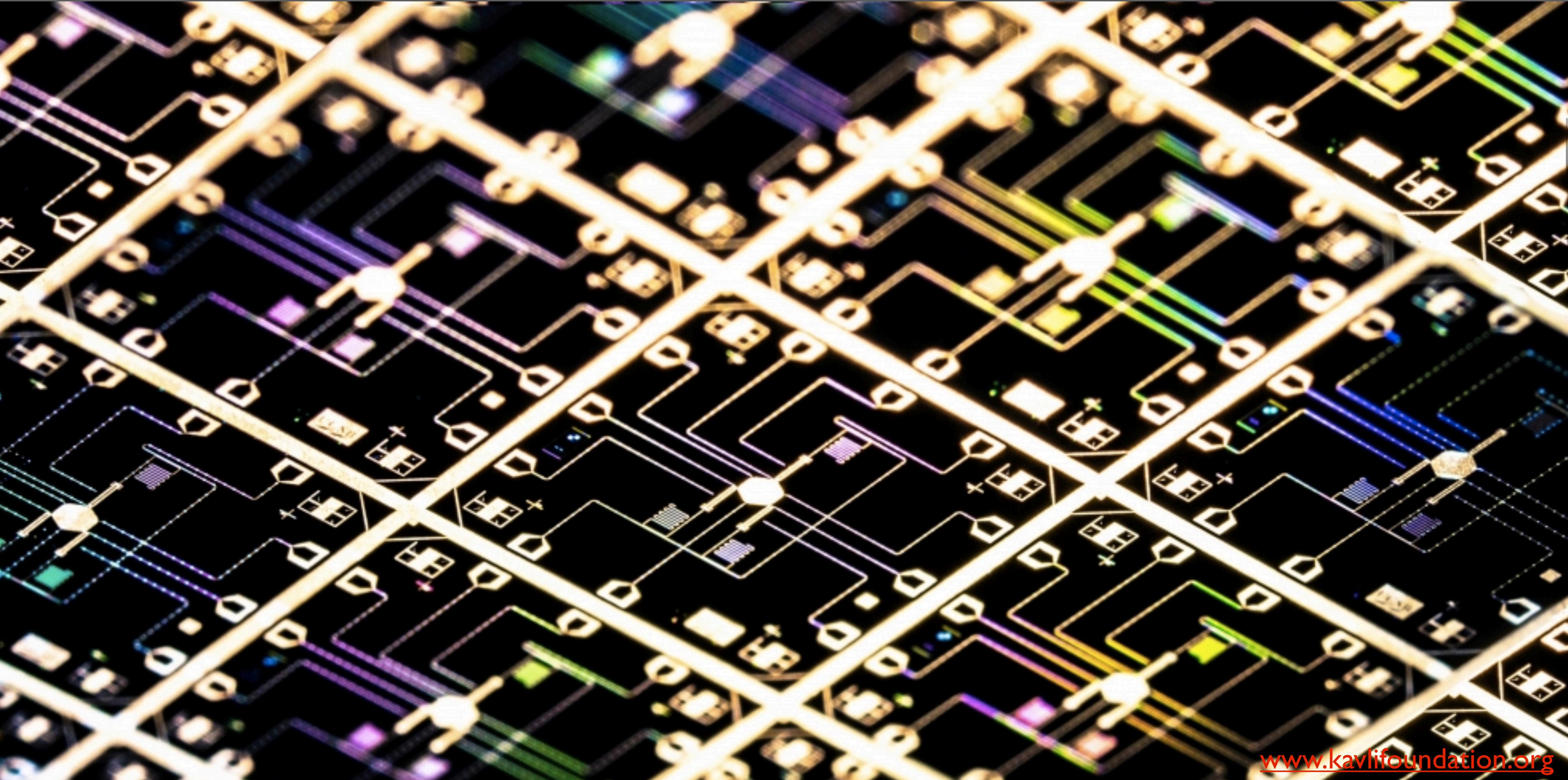


Han et.al., ArXiv:1709.01662 (2017)



A Zoo of open architectures and tricks!

Extrapolation to Quantum Qubit arrays ?



www.kavlifoundation.org

Thank you!