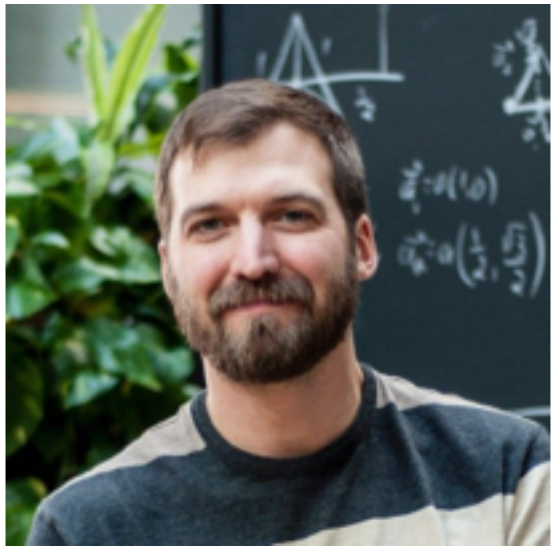
A NEURAL NETWORK PERSPECTIVE ON THE ISING GAUGE THEORY AND THE TORIC CODE

Juan Carrasquilla
D-Wave systems

COLLABORATORS



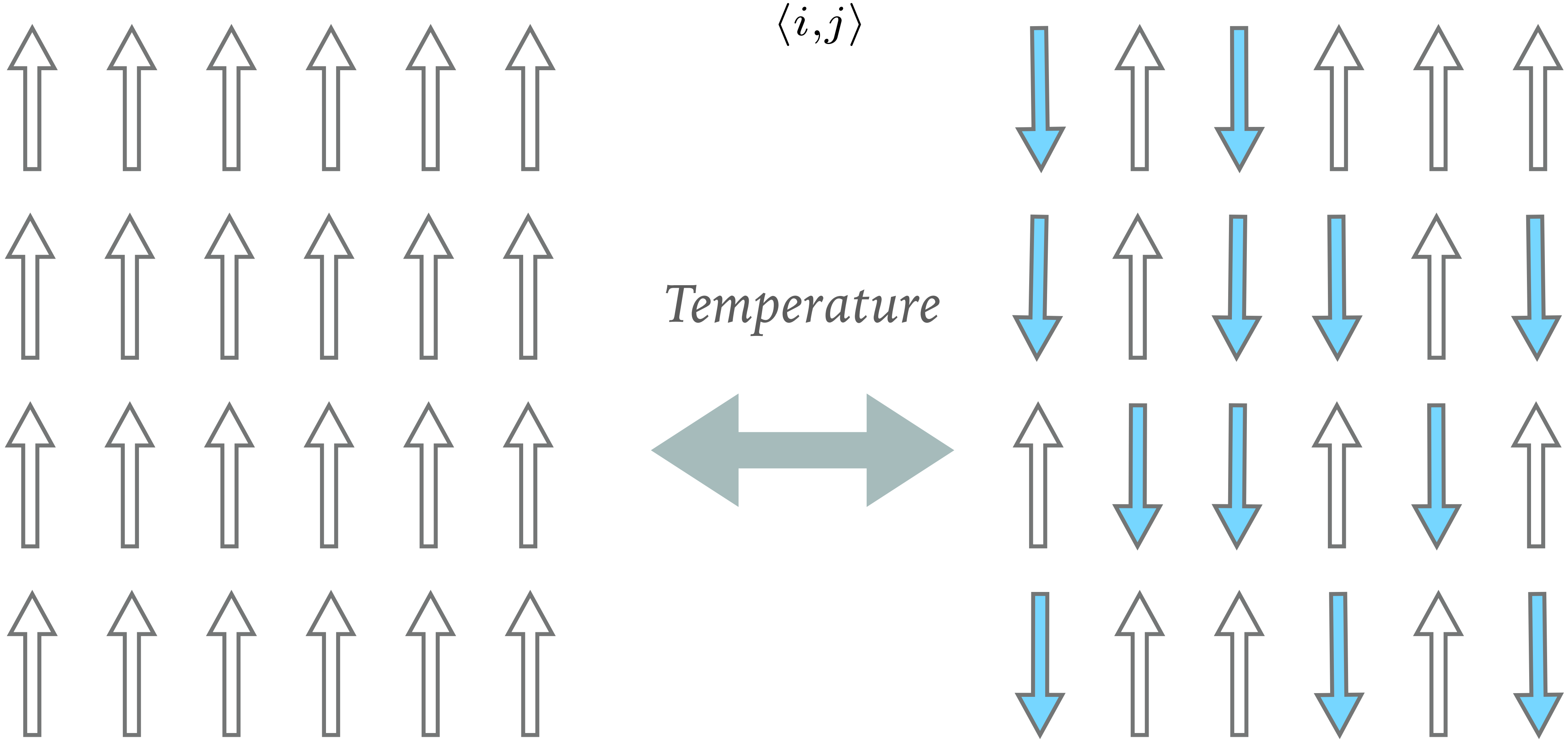
Roger Melko (U. Waterloo and Perimeter)

SUPERVISED LEARNING PERSPECTIVE OF PHASES OF MATTER

PHASES, PHASE TRANSITIONS, AND THE ORDER PARAMETER

Ising ferromagnet in two dimensions

$$E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

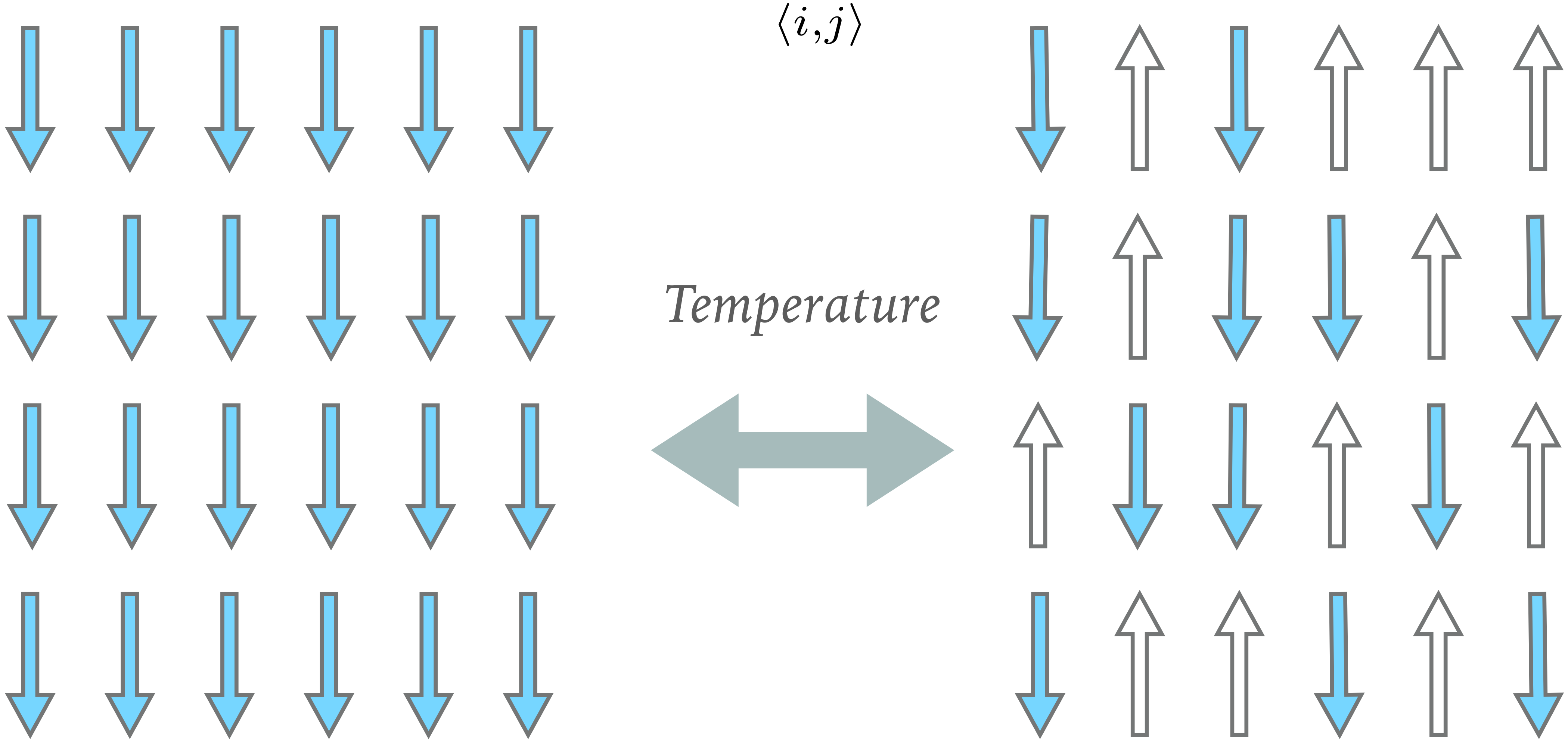


Ferromagnet *Lars Onsager Phys. Rev. 65, 117* *Paramagnet*

PHASES, PHASE TRANSITIONS, AND THE ORDER PARAMETER

Ising ferromagnet in two dimensions

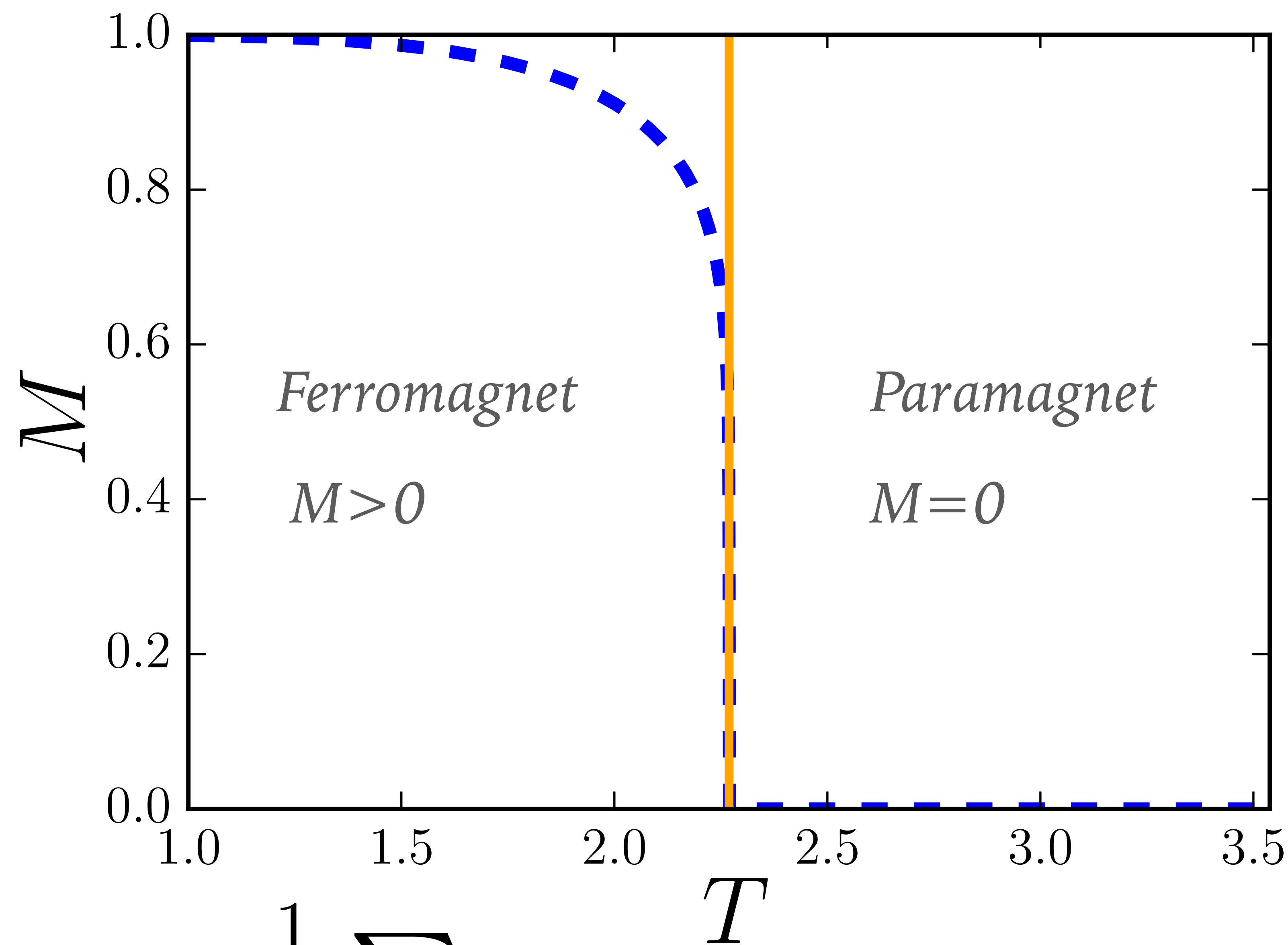
$$E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$



Ferromagnet *Paramagnet*

PHASES, PHASE TRANSITIONS, AND THE ORDER PARAMETER

Ferromagnetic transition: order parameter



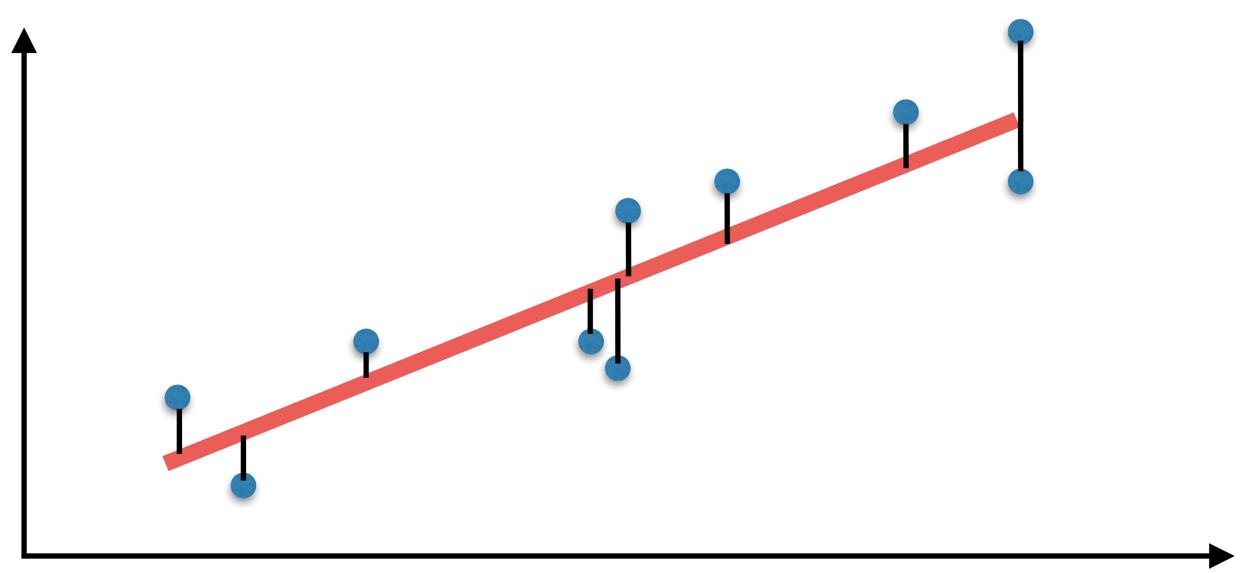
It is a measure of the degree of order in the system

$$M = \frac{1}{N}\sum_i \langle \sigma_i \rangle, \qquad \sigma_i = \pm 1$$

Lars Onsager Phys. Rev. 65, 117 (1944)

WHAT IS SUPERVISED LEARNING?

- Supervised learning is the machine learning task of inferring a function from labeled training data.
- In supervised learning, each example is a pair consisting of an input object (typically a high-d vector x) and a desired output value y (also called the supervisory signal)
- Simplest SL algorithm: linear regression



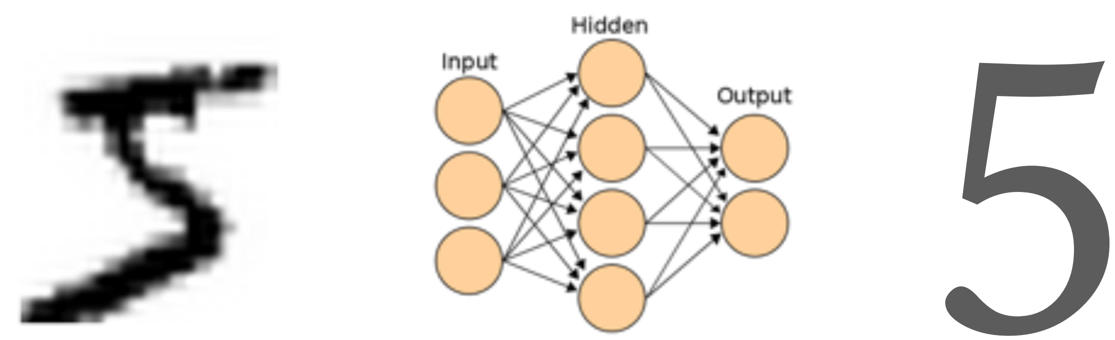
For a linear fit

$$f(a, b) = a + bx$$

M (x,y) pairs

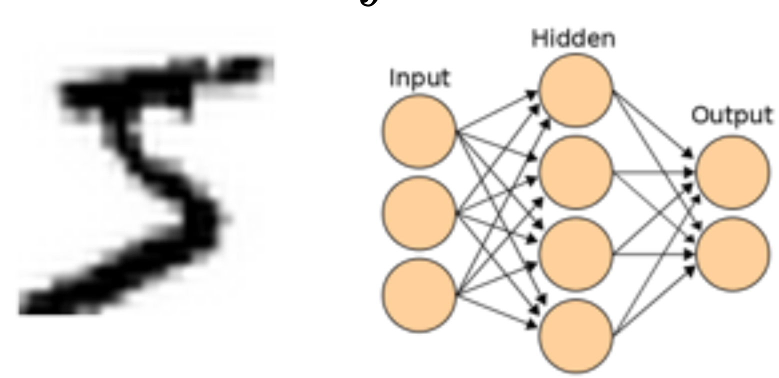
FLUCTUATIONS HANDWRITTEN DIGITS (MNIST)

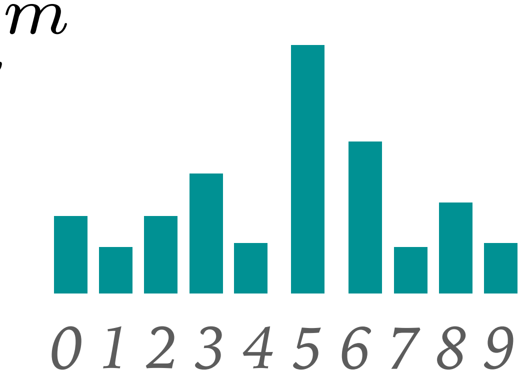
Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.



ML community has developed powerful supervised learning algorithms

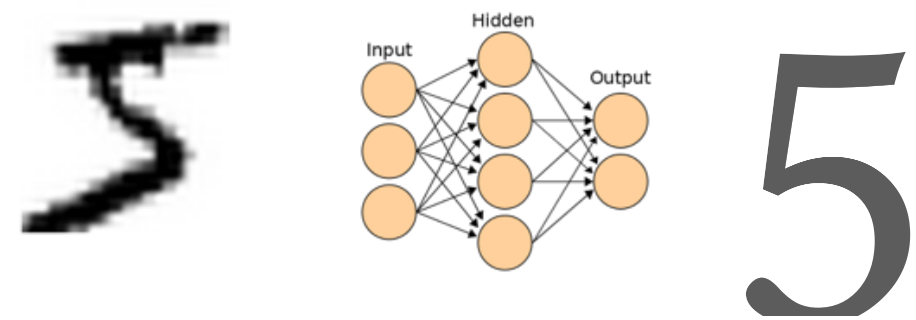
$$f : \mathbf{R}^n \rightarrow \mathbf{R}^m$$





used by ATMs and post offices

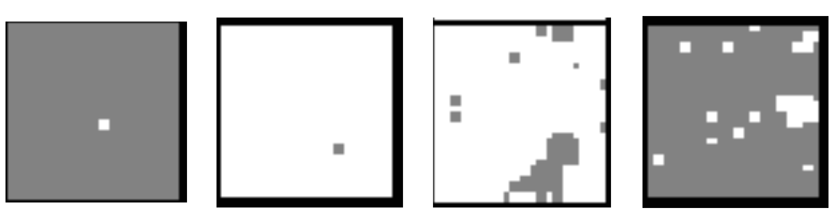
INSPIRATION: FLUCTUATIONS HANDWRITTEN DIGITS (MNIST)

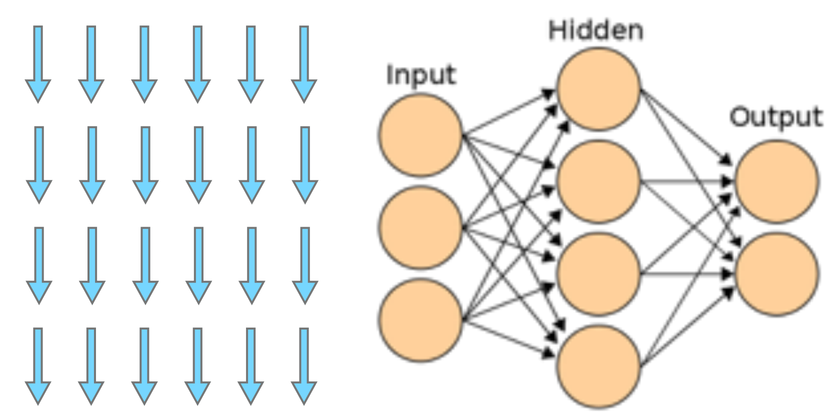


ML community has developed powerful supervised learning

 $= 5 + \textit{fluctuations}$

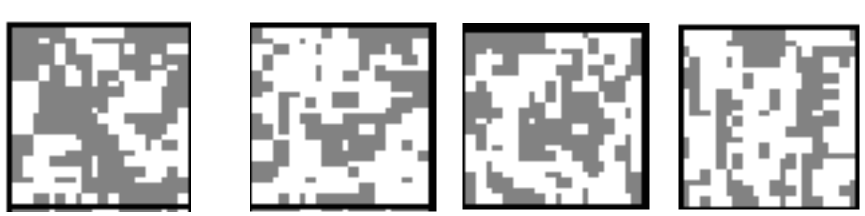
FM phase





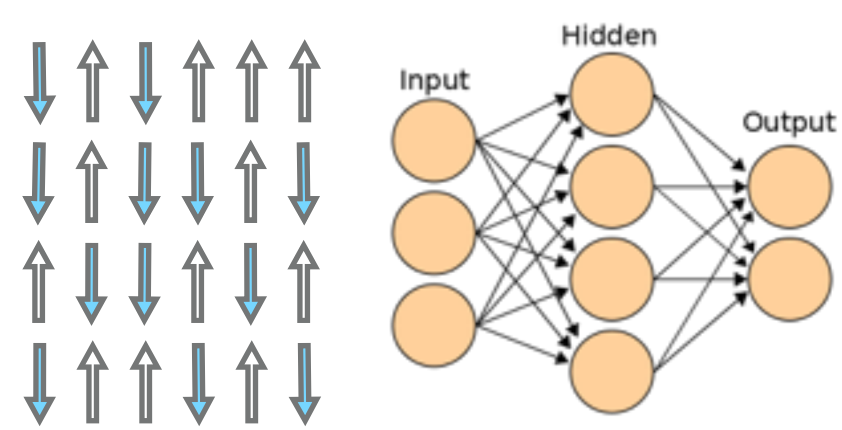
FM (0)

High T phase



gray=spin up

white=spin down

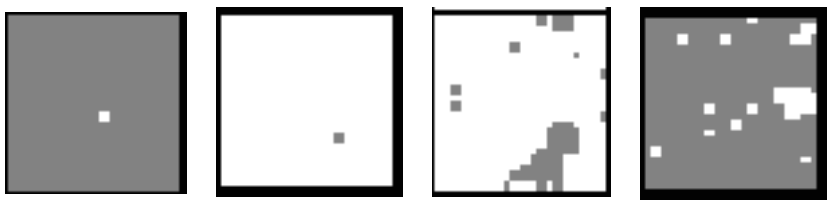


PM (1)

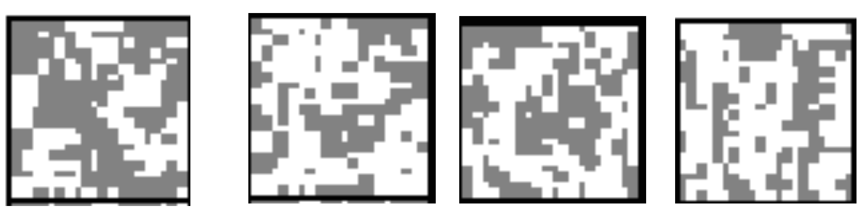
1605.01735

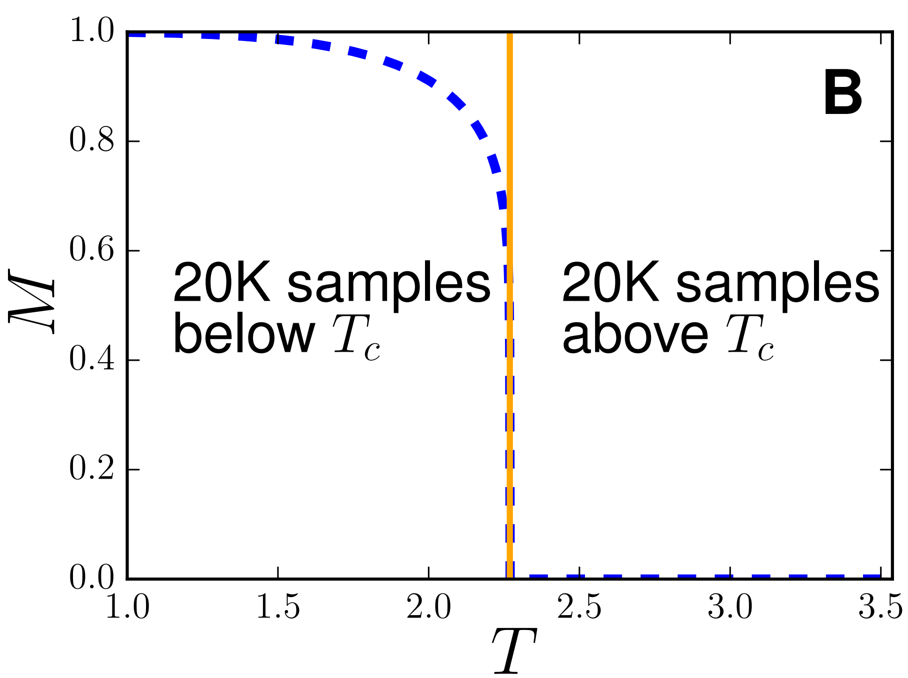
COLLECTING THE TRAINING/TESTING DATA: MC SAMPLING ISING MODEL AND LABELS

2D Ising model in the ordered phase



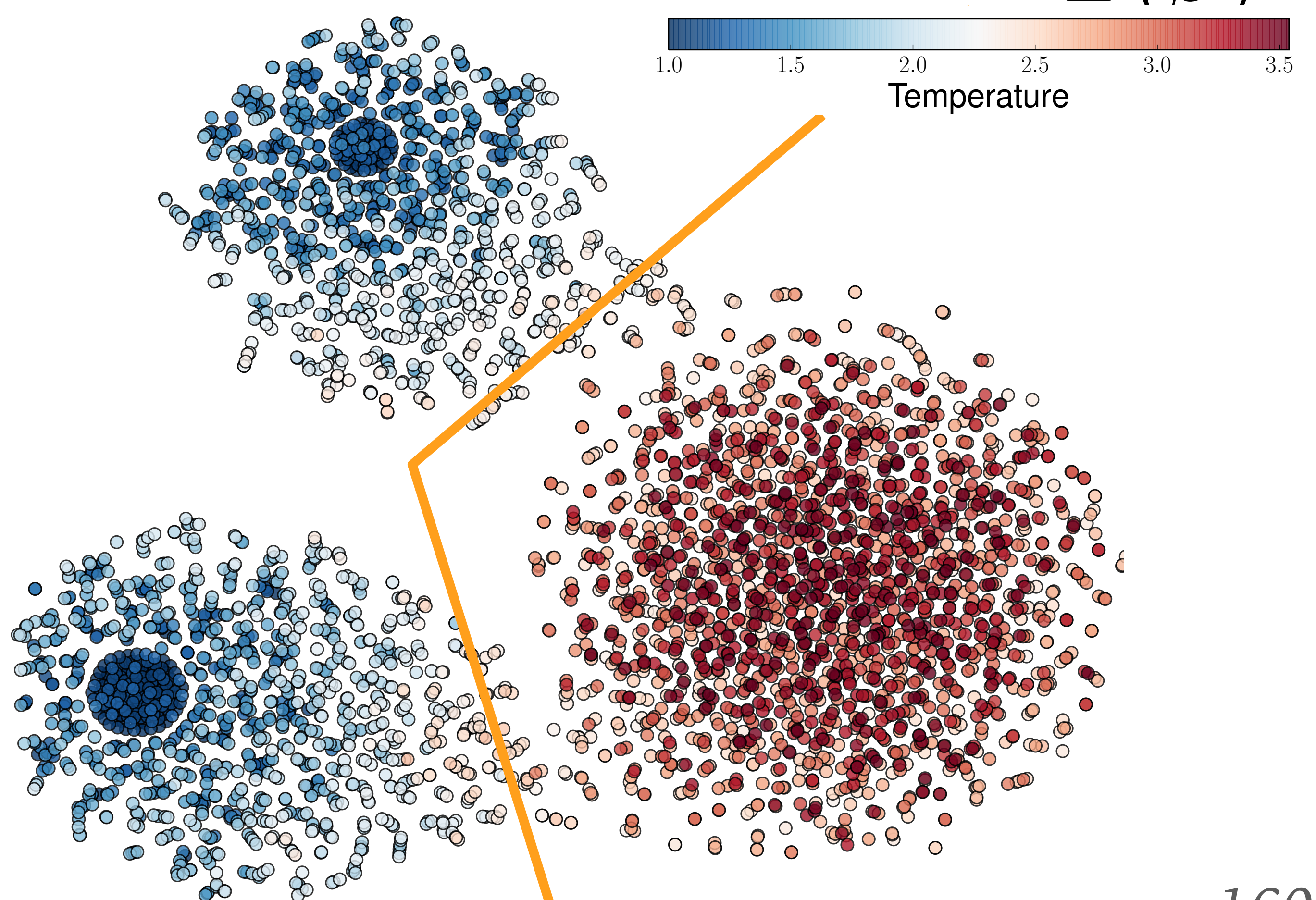
2D Ising model in the disordered phase





Training/testing data is drawn from the Boltzmann distribution

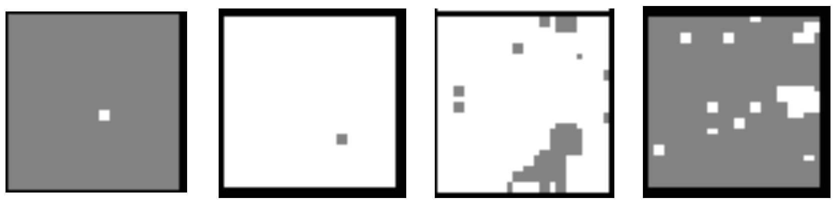
$$p(\sigma_1, \sigma_2, ..., \sigma_N) = \frac{e^{-\beta E(\sigma_1, \sigma_2, ..., \sigma_N)}}{Z(\beta)}$$



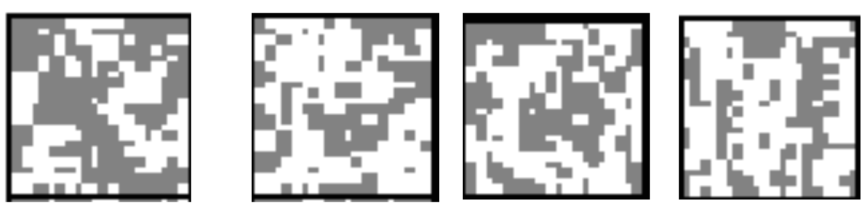
1605.01735

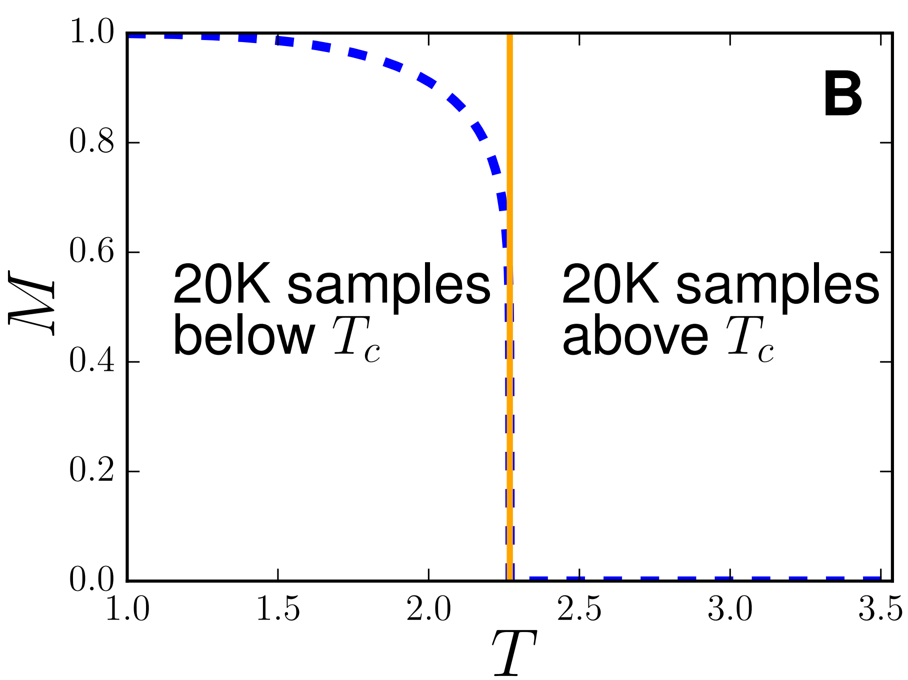
COLLECTING THE TRAINING/TESTING DATA: MC SAMPLING ISING MODEL AND LABELS

2D Ising model in the ordered phase



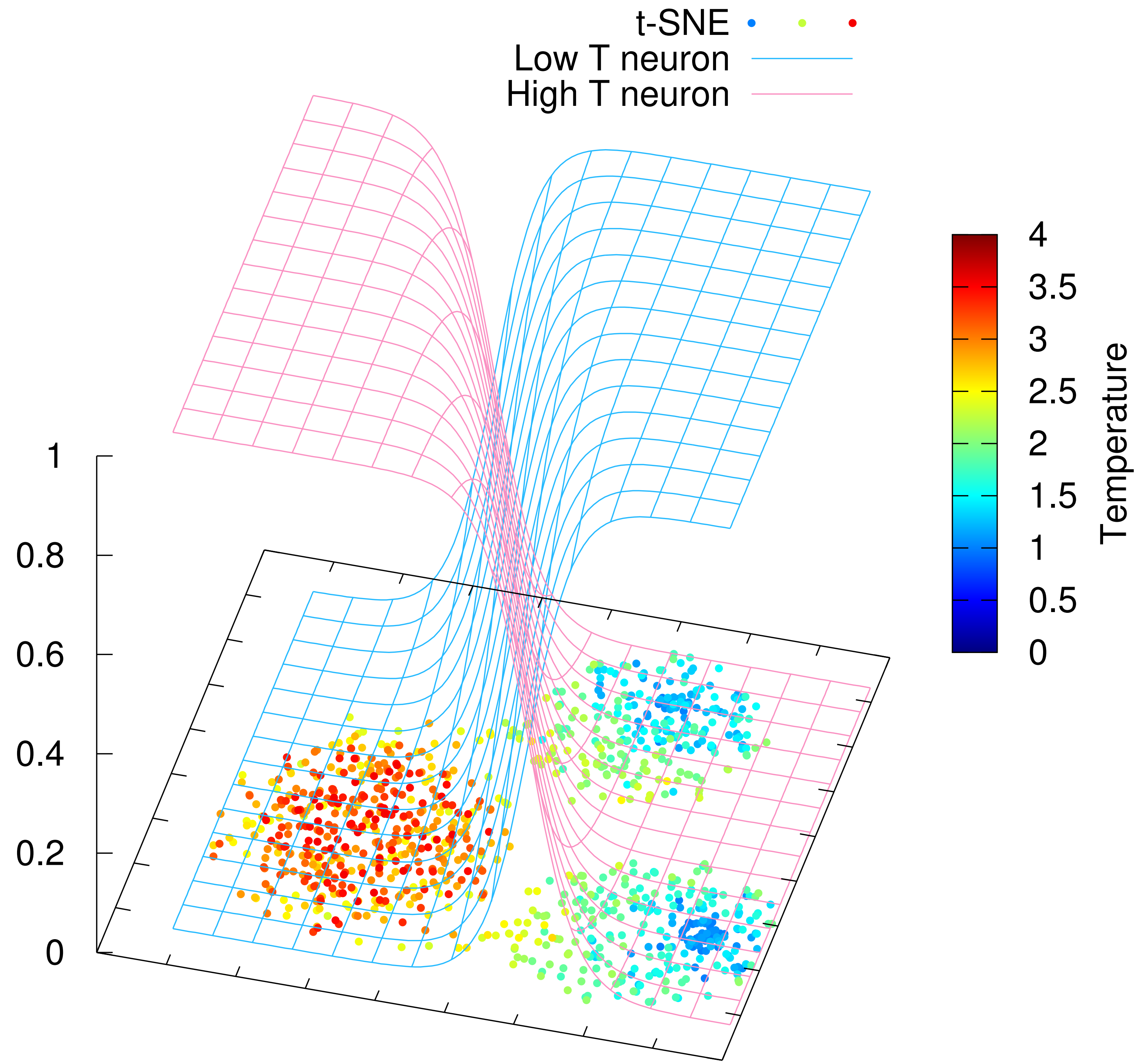
2D Ising model in the disordered phase





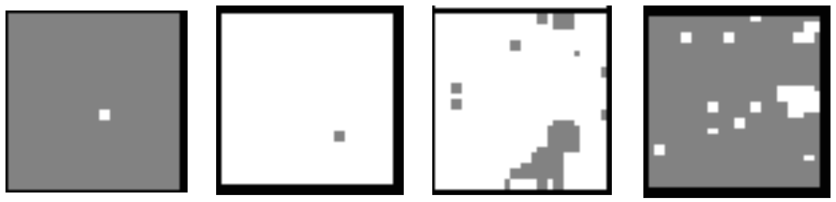
Successful training amounts to finding functions

$F_{\text{High }T}(\sigma_1, ..., \sigma_N)$ $\quad$ $F_{\text{Low }T}(\sigma_1, ..., \sigma_N)$

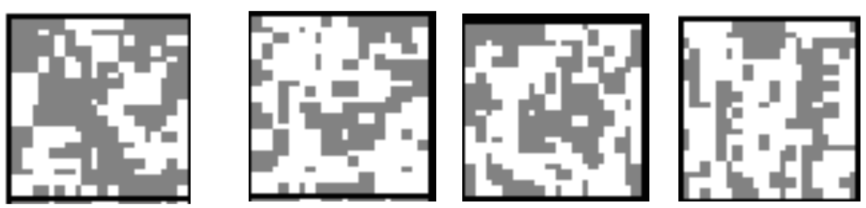


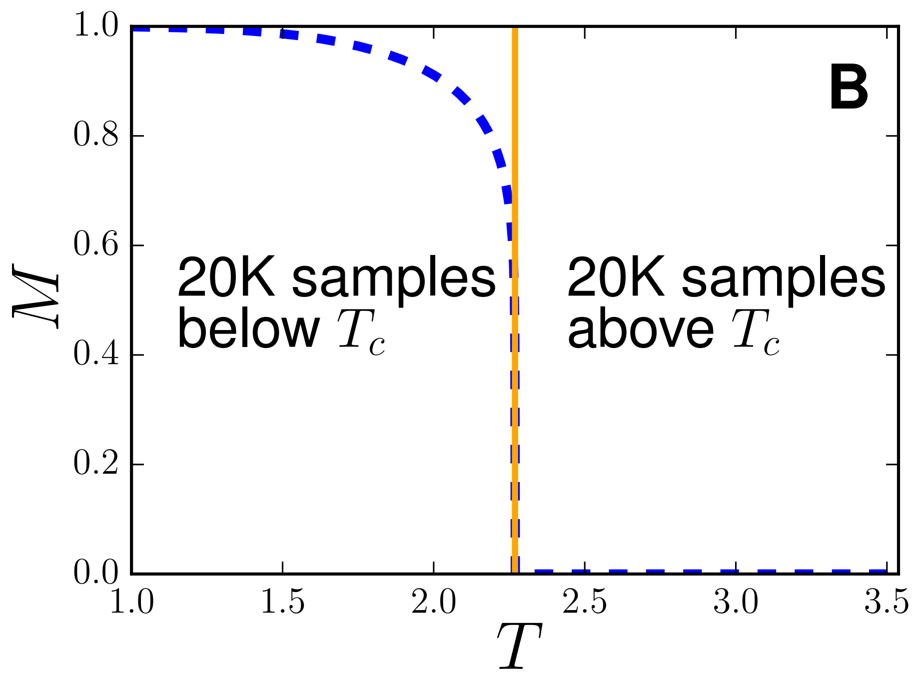
RESULTS: SQUARE LATTICE ISING MODEL (TEST SETS)

2D Ising model in the ordered phase

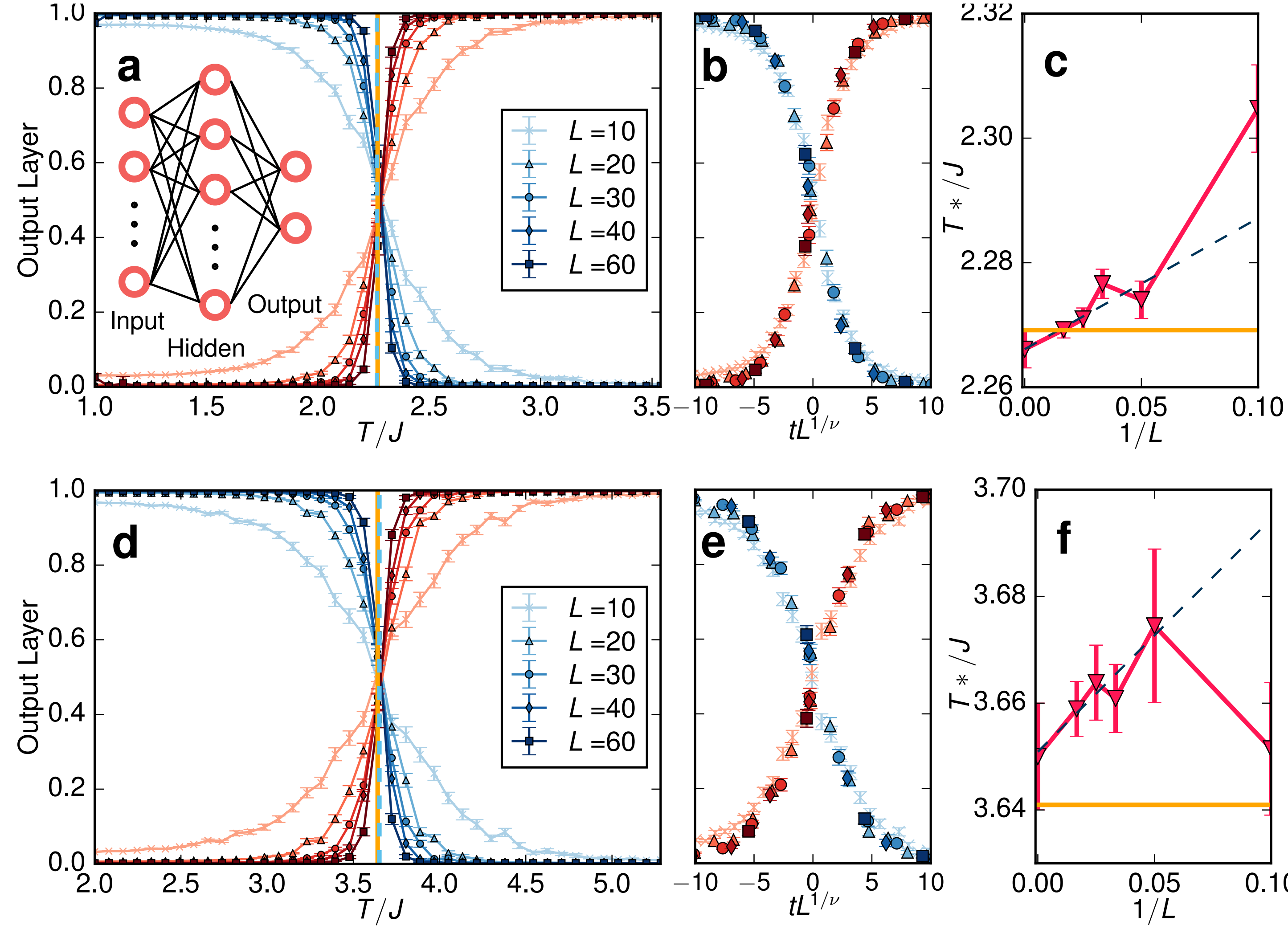


2D Ising model in the disordered phase





$T_c/J = 2/\ln\left(1+\sqrt{2}\right)$



Critical

1605.01735, Nature Physics

Tc of Triangular within <1%

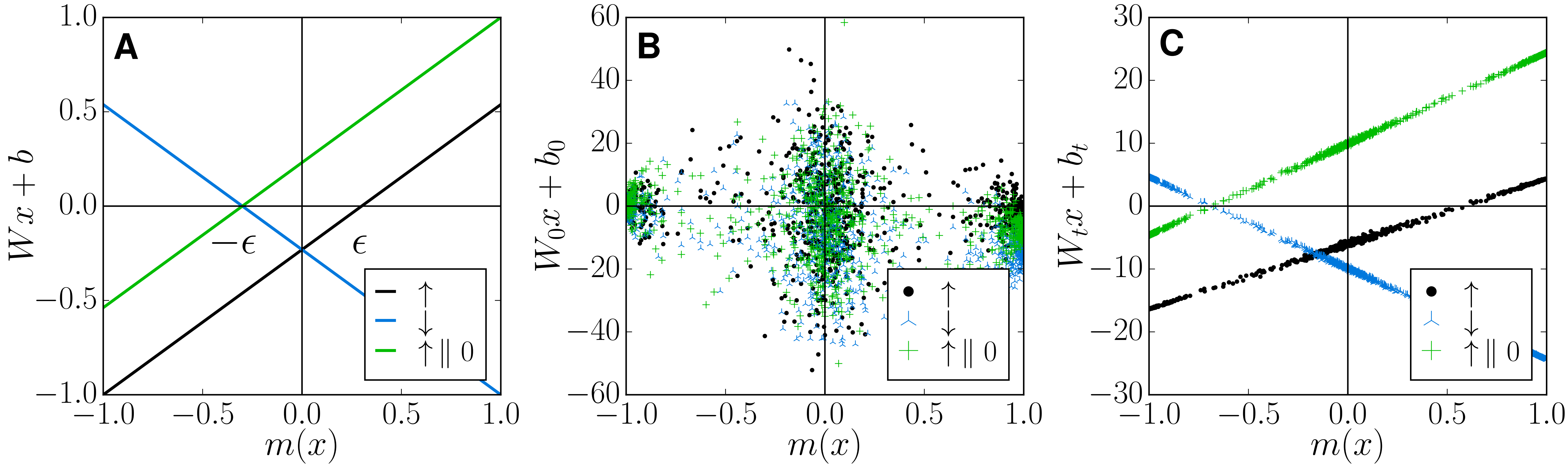
NN knows about criticality

$\nu \approx 1$

Check Ye-Hua Liu's talk for an unsupervised solution

ANALYTICAL UNDERSTANDING

Investigating the argument of the hidden layer during the training



$$W = \frac{1}{N(1+\epsilon)} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ -1 & -1 & \cdots & -1 \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \text{ and } b = \frac{\epsilon}{(1+\epsilon)} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \qquad Wx + b = \frac{1}{(1+\epsilon)} \begin{pmatrix} m(x) - \epsilon \\ -m(x) - \epsilon \\ m(x) + \epsilon \end{pmatrix},$$

$$x = [\sigma_1 \sigma_2, \ldots, \sigma_N]^{\mathrm{T}} \qquad m(x) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i$$

1605.01735

CAN WE DEAL WITH DISORDERED AND TOPOLOGICAL PHASES NOT DESCRIBED BY ORDER PARAMETERS?

PHASES OF MATTER WITHOUT AN ORDER PARAMETER AT T=0

- Topological phases of matter. Examples: Fractional quantum hall effect, quantum spin liquids, Ising gauge theory. Potential applications in topological quantum computing. Interestingly, these phases defy the Landau symmetry breaking classification.

- Coulomb phases = Highly correlated "spin liquids" described by electrodynamics. Examples: Common water ice and spin ice materials ($Ho_2Ti_2O_7$ and $Dy_2Ti_2O_7$)

ST Bramwell, MJP Gingras Science 294 (5546), 1495-1501

PHASES OF MATTER WITHOUT AN ORDER PARAMETER AT T=0,∞

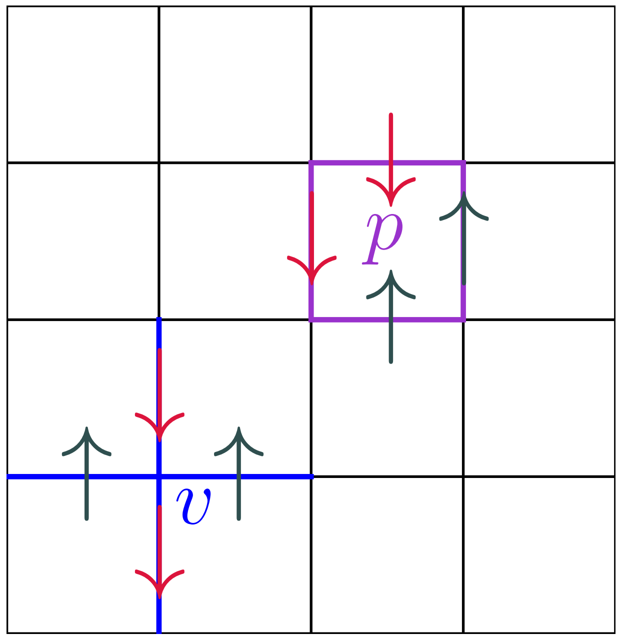
Ising square ice

$$H = J \sum_v Q_v^2$$

$$Q_v = \sum_{i \in v} \sigma_i^z$$

Wegner's Ising gauge theory:

$$H = -J \sum_p \prod_{i \in p} \sigma_i^z$$



F.J. Wegner, J. Math. Phys. 12 (1971) 2259

(Kogut Rev. Mod. Phys. 51, 659 (1979))

Degenerate classical Coulomb phase

The ground state is a degenerate manifold with algebraically decaying spin–spin correlations.

The ground state is again a degenerate manifold with exponentially decaying spin–spin correlations.

The grandmother of all lattice models for topological quantum computation

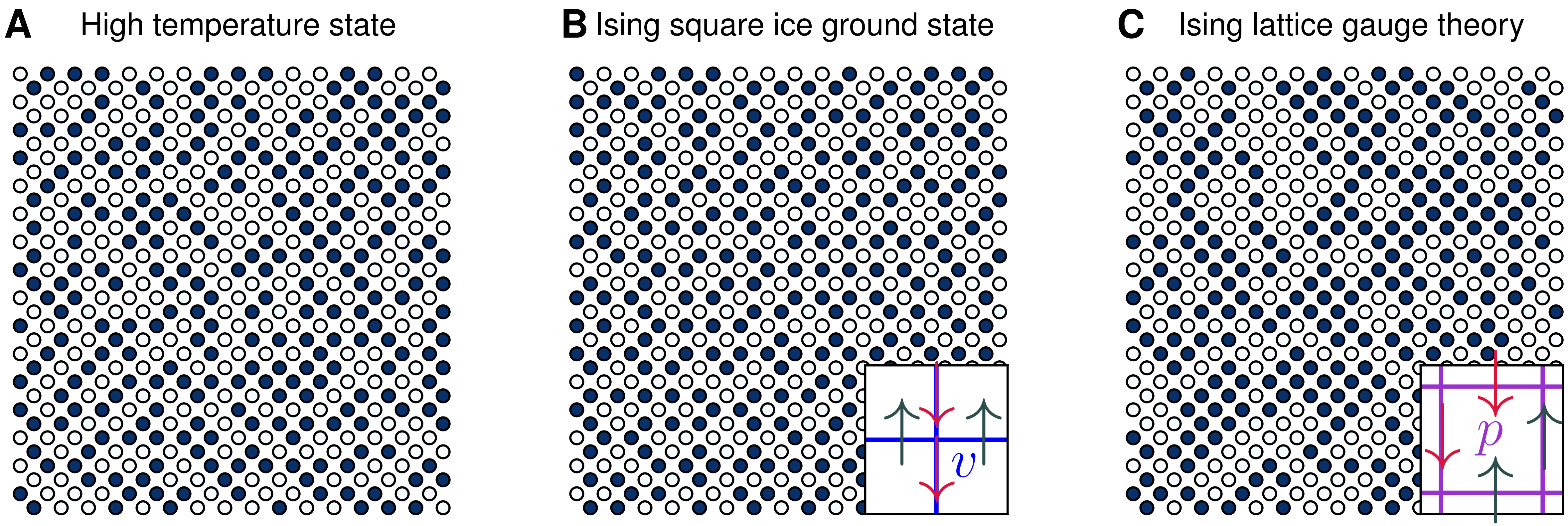
Ground state is a classical disordered topologically ordered phase

Castelnovo and Chamon Phys. Rev. B 76, 174416 (2007)

T=0

high temperature phase

PHASES OF MATTER WITHOUT AN ORDER PARAMETER AT T=0,∞



Ising square ice

$$H = J\sum_v Q_v^2$$

$$Q_v = \sum_{i\in v}\sigma_i^z$$

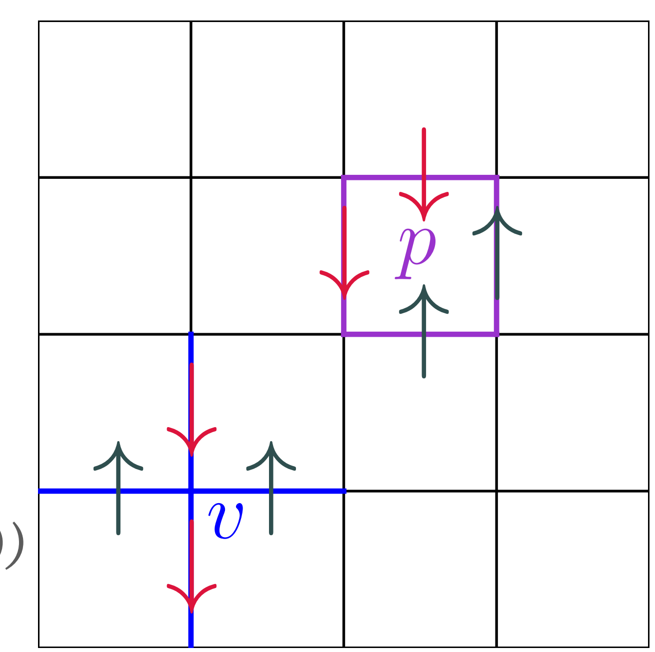
Loop update + spin flip MC

Ising gauge theory

$$H = -J\sum_p \prod_{i\in p}\sigma_i^z$$

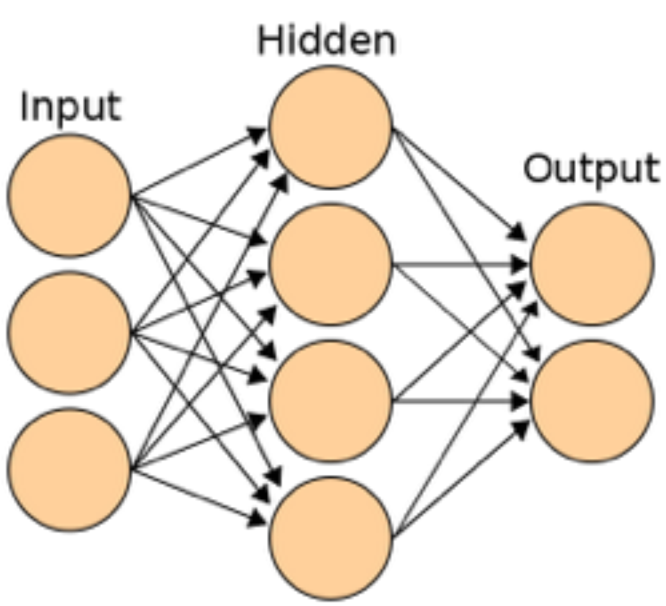
(Kogut Rev. Mod. Phys. 51, 659 (1979))

Gauge update + spin flip MC



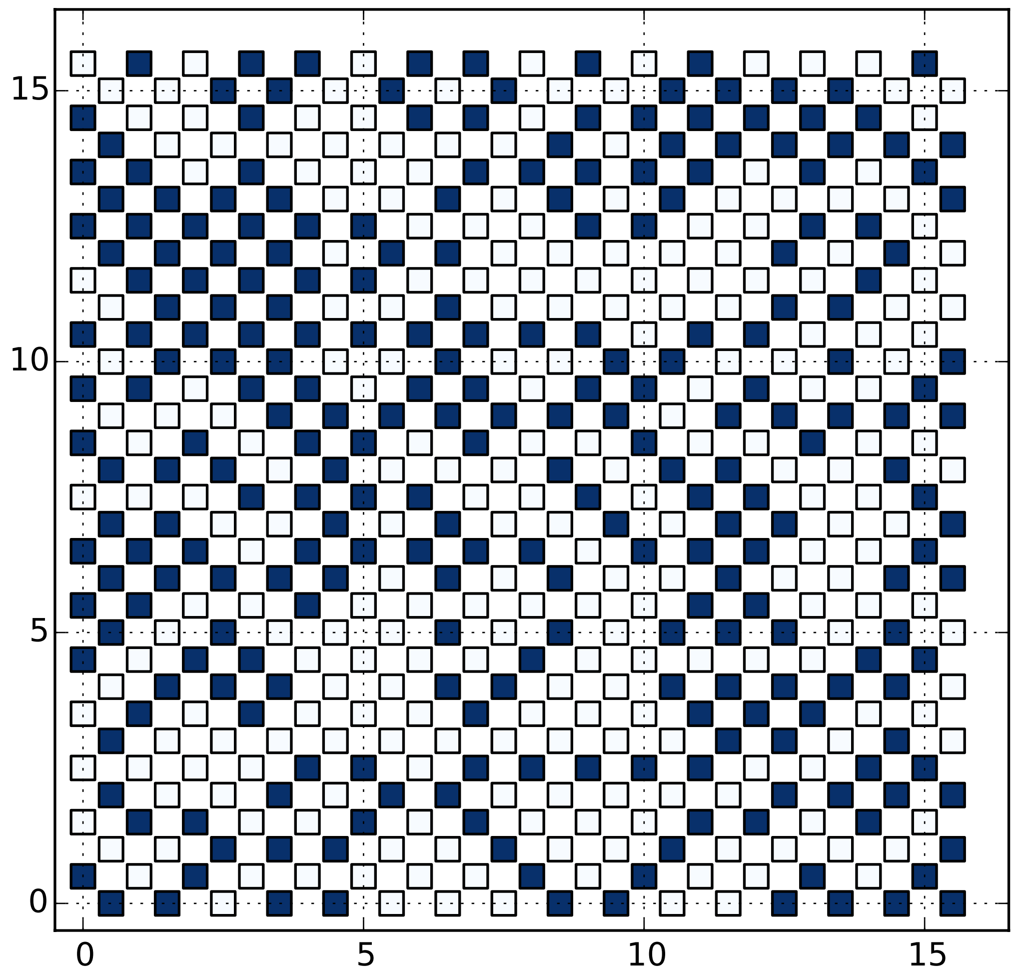
PHASES OF MATTER WITHOUT AN ORDER PARAMETER AT T=0

- Neural nets capture the subtle differences between low- and high-temperature states successfully!
- Square ice: 99% accuracy
- Ising gauge theory: 50% (guessing) with a fully-connected neural net. Training fails. How to overcome this issue?

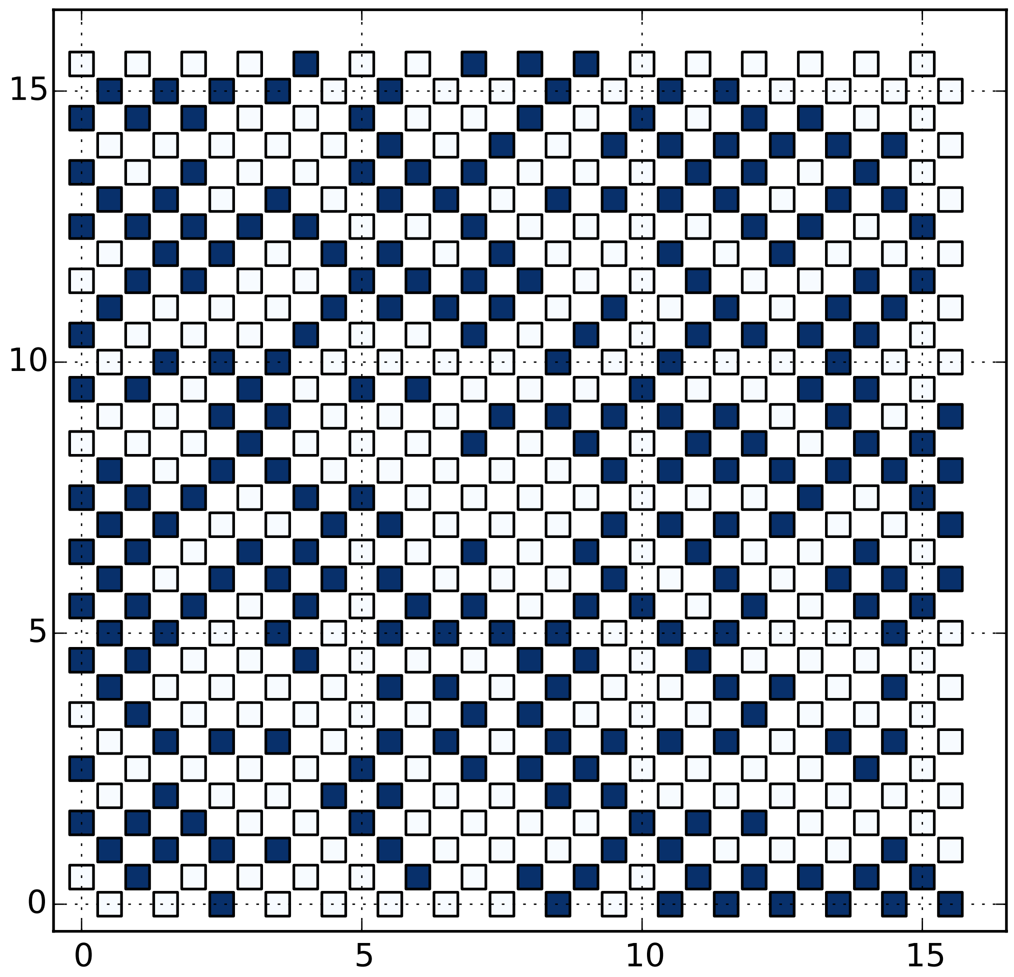


Difficult to train for the Ising gauge theory

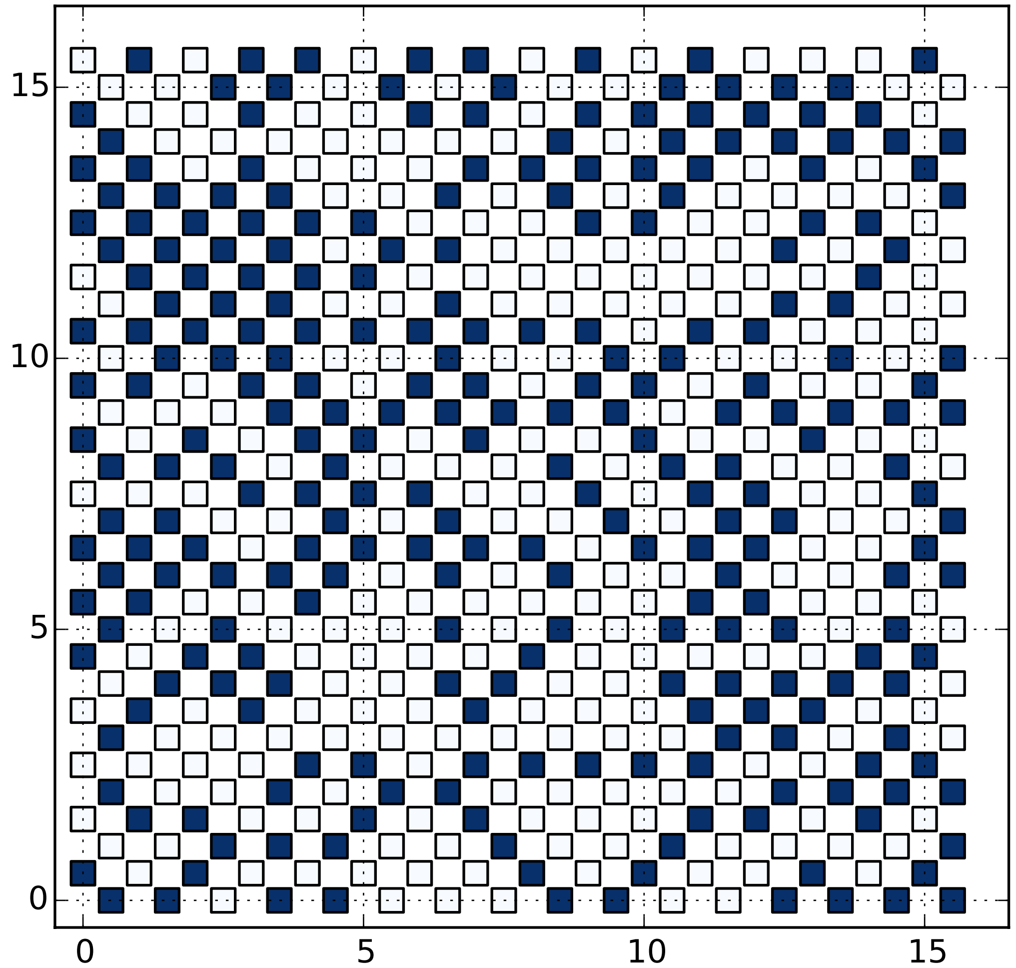
For two configurations



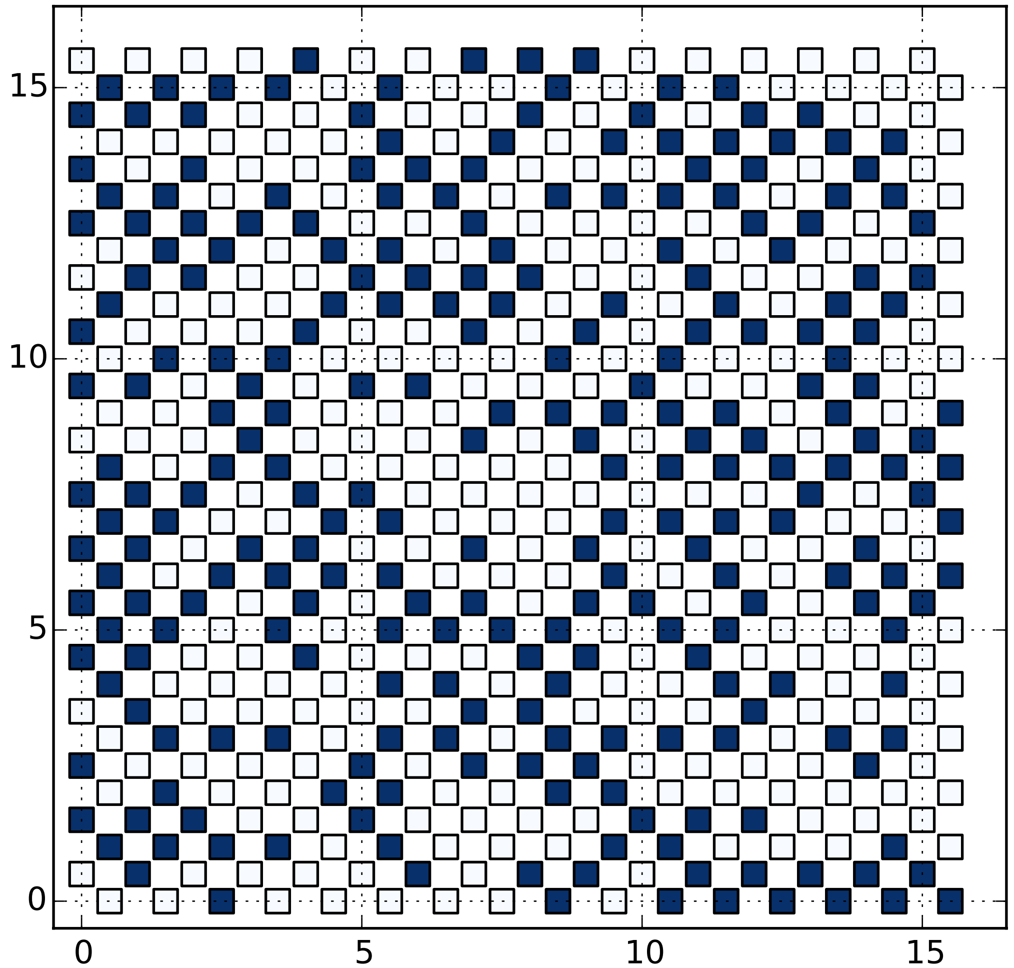
?



For two configurations



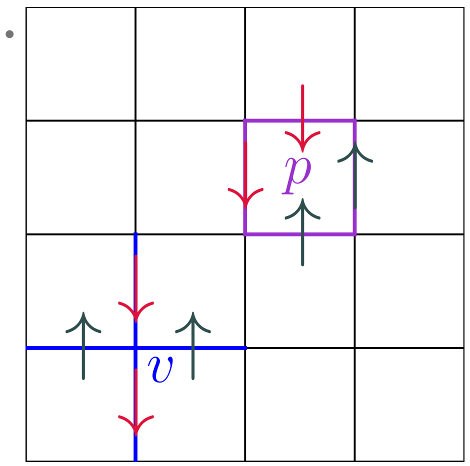
?



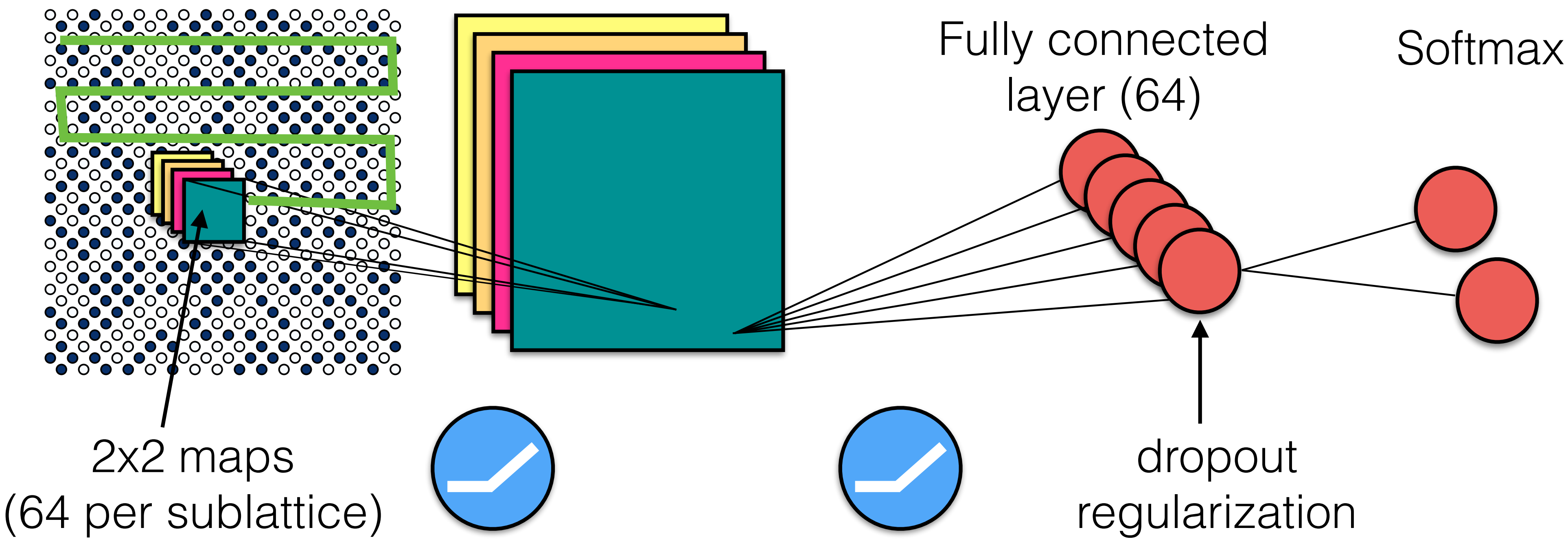
Ground state

ISING GAUGE THEORY

F.J. Wegner, J. Math. Phys. 12 (1971) 2259



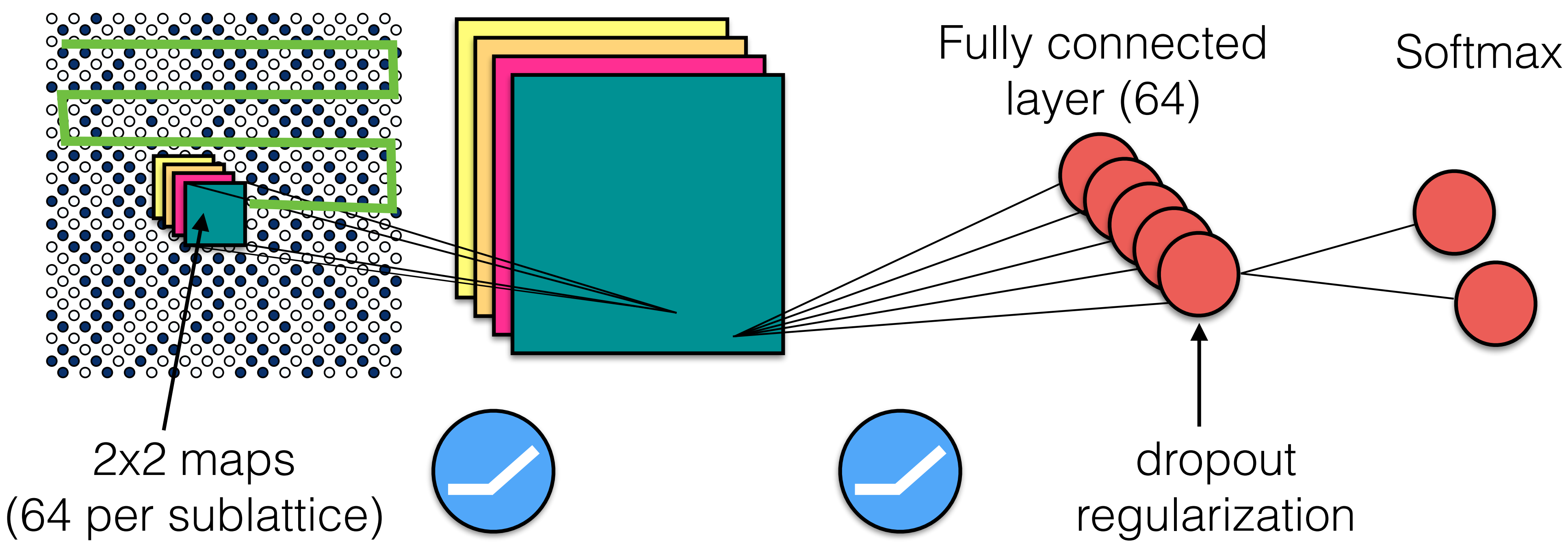
$$H = -J \sum_p \prod_{i\in p} \sigma_i^z$$

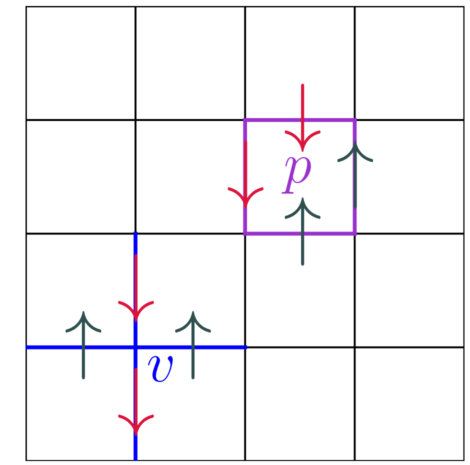


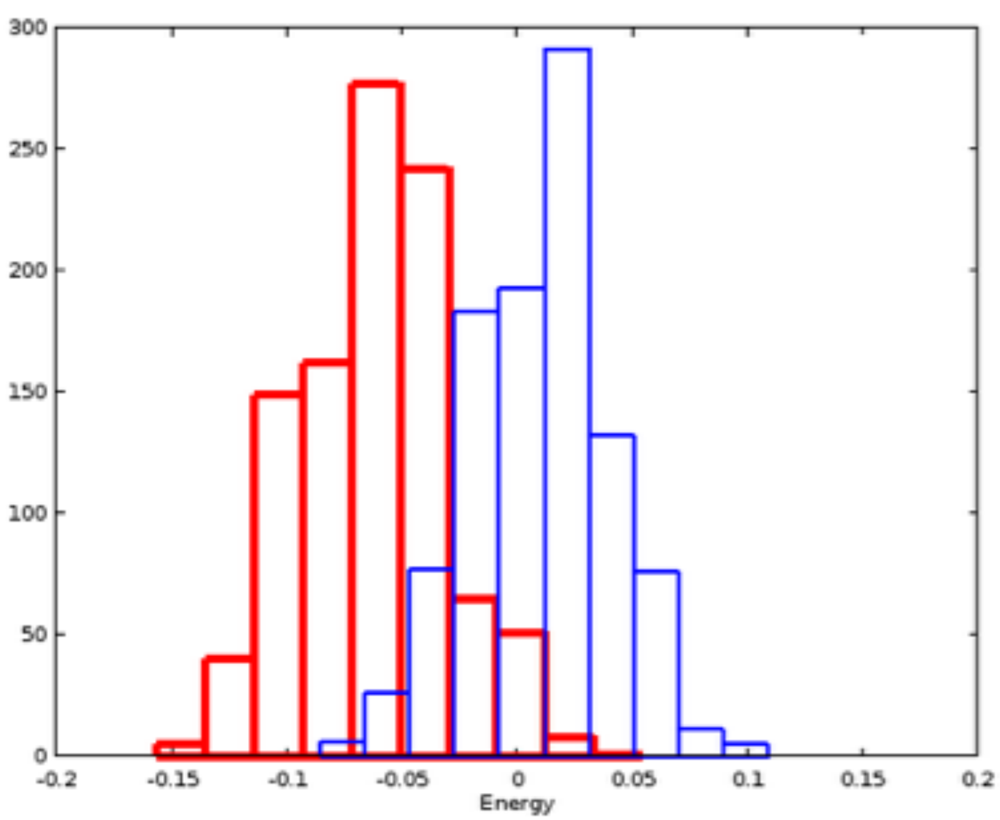
99% accuracy
easy to train

The picture we draw for what the CNN is using to distinguish the phases is that of the detection of satisfied local constraints. Not optimal since it is not generic. Check Frank's talk for a general solution

ISING GAUGE THEORY: GENERATING ADVERSARIAL EXAMPLES







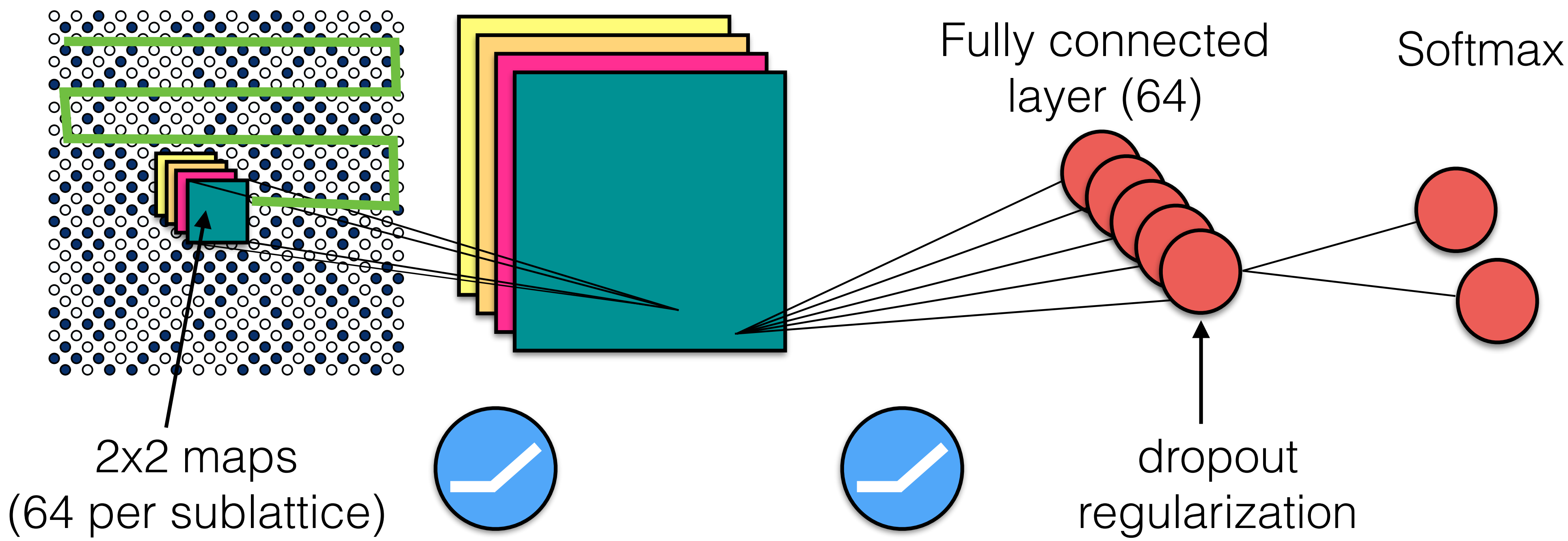
Inquire the NN: What do you think a ground state configuration should look like?

Maximize the readout layers with Metropolis MC

It barely of knows that GS configurations should have low energy. The resulting model is lazy due to the training setup that we used

ISING GAUGE THEORY *F.J. Wegner, J. Math. Phys. 12 (1971) 2259*

$$H = -J \sum_p \prod_{i \in p} \sigma_i^z$$



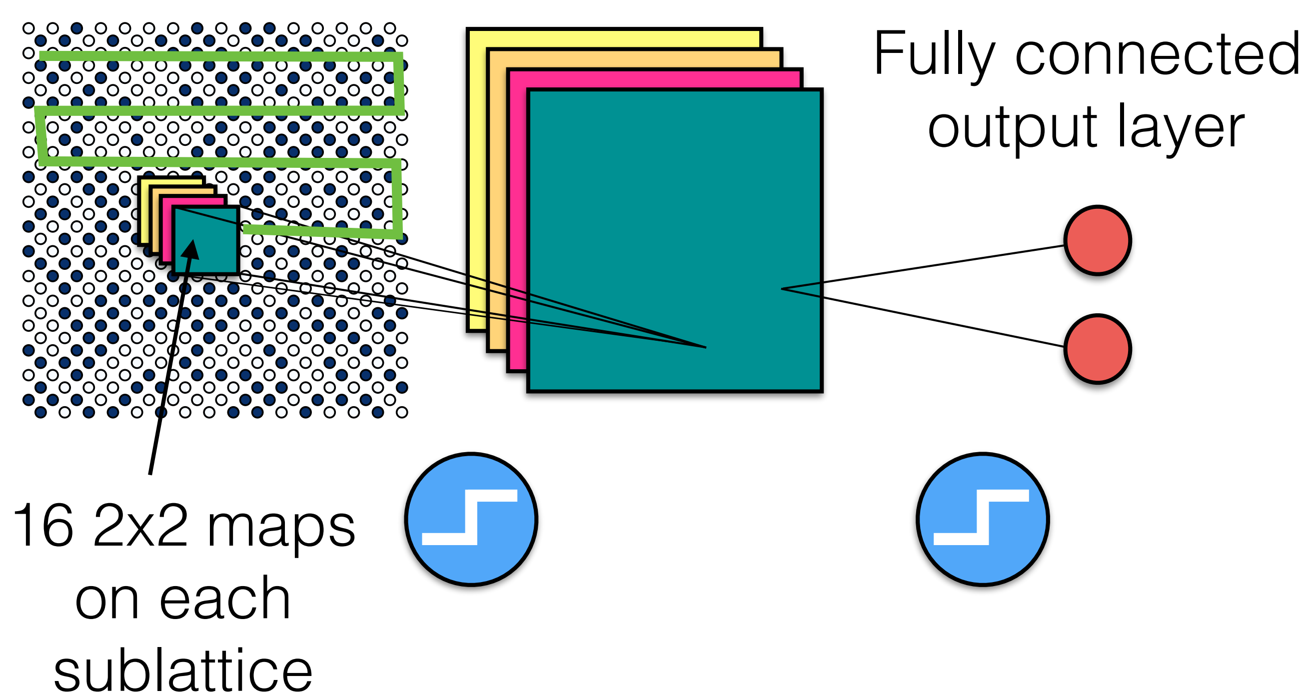
Can we do better? What is the the ideal neuron's output?

high T: [0,1] *Let's construct an analytical model that does that*

Low T: [1,0]

ANALYTICAL UNDERSTANDING: WHAT DOES THE CNN USE TO MAKE PREDICTIONS?

- The convolutional neural net relies on the detection of satisfied local constraints to make accurate predictions of whether a state is drawn at low or infinite temperature.
- Based on this observation we derived the weights of a streamlined convolutional network *analytically* designed to work pretty well on our test sets.

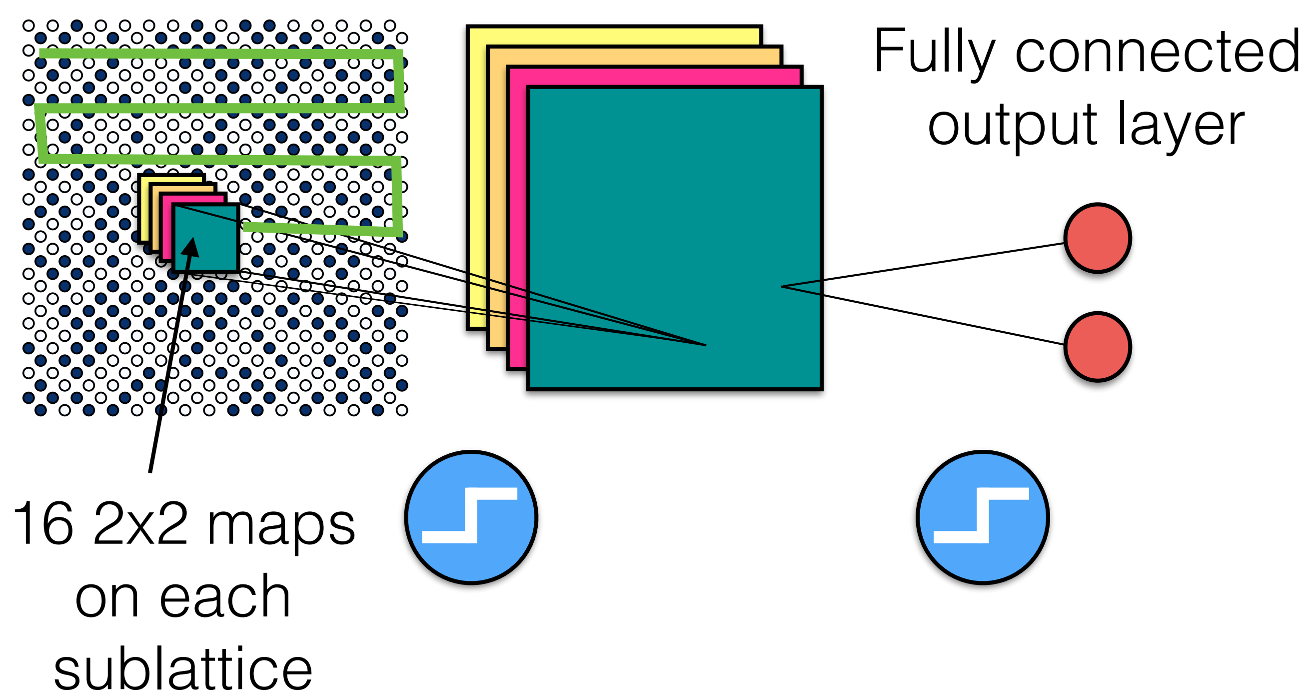


100% accuracy on test sets

1605.01735, Nature Physics

ANALYTICAL UNDERSTANDING: WHAT DOES THE CNN USE TO MAKE PREDICTIONS?

- CNN checks every plaquette whether constrains are satisfied or not (basically checks the energy of each plaquette)
- Fully connected layer counts defects (or computes the total energy) and the last perceptron decides if the configuration is ground state or not.



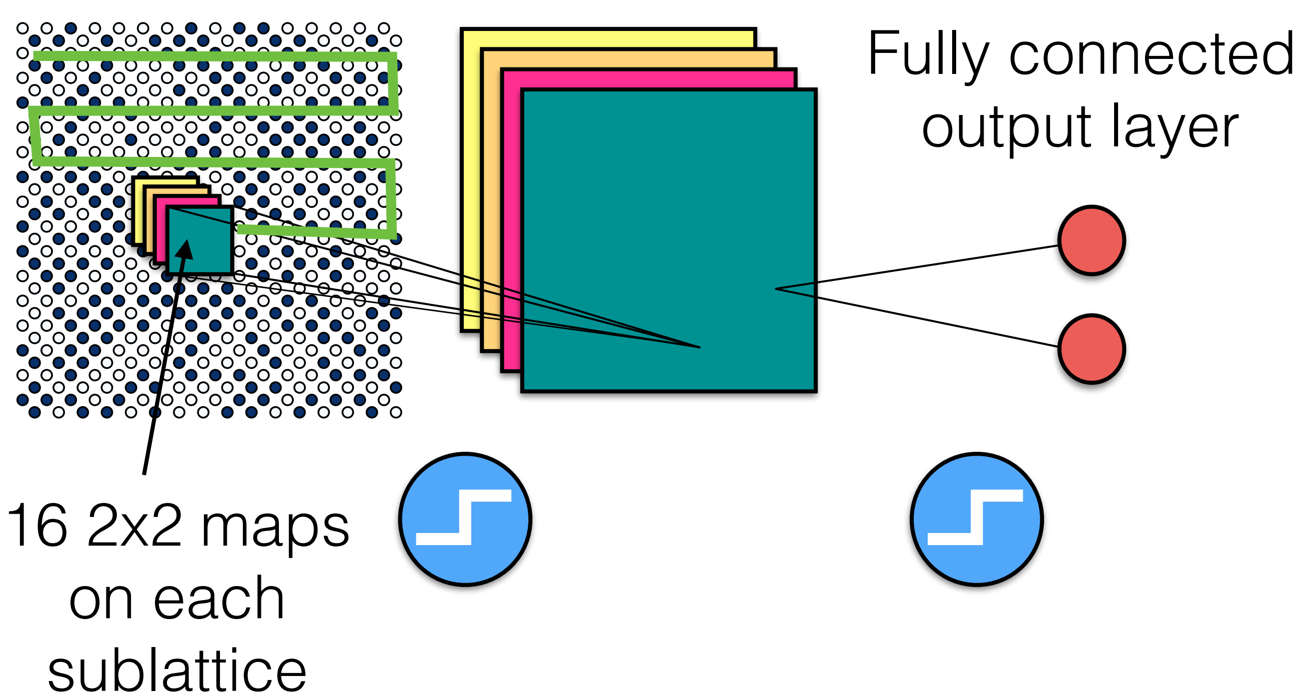
100% accuracy on test sets

1605.01735, Nature Physics

ANALYTICAL UNDERSTANDING: WHAT DOES THE CNN USE TO MAKE PREDICTIONS?

- Adequate output neutron behavior.
- It turns out that the cold neutron behaves as the Boltzmann weight of the model at zero temperature

$$O_{\text{cold}}(\sigma_1, ..., \sigma_N) \propto \lim_{\beta \to \infty} \exp \beta J \sum_p \prod_{i \in p} \sigma_i^z$$



100% accuracy on test sets

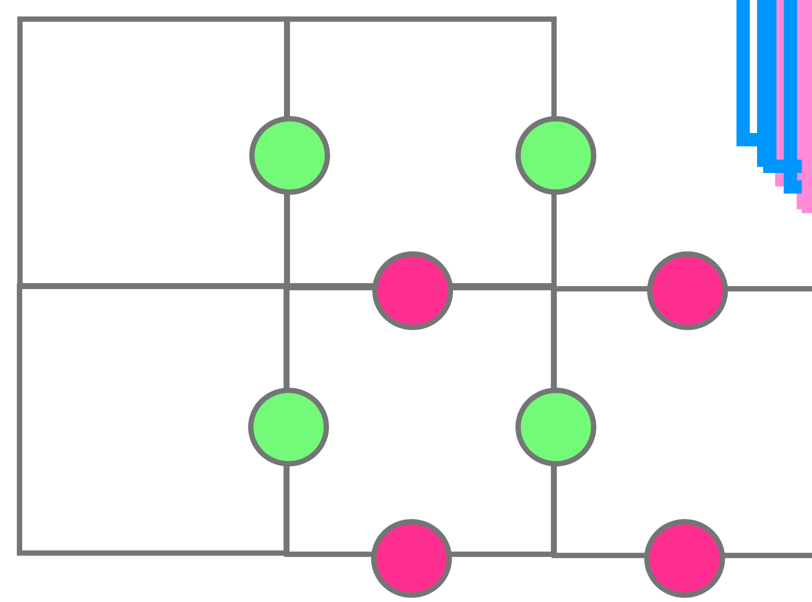
1605.01735, Nature Physics

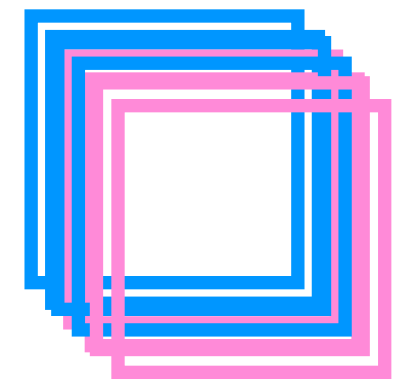
IMPLEMENTING LOCAL CONSTRAINTS WITH A CNN

Convolutional layer

f	s=A	s=B	f	s=A	s=B
1	$\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$	9	$\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\ 0 & 0\end{pmatrix}$
2	$\begin{pmatrix}-1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\ 0 & 0\end{pmatrix}$	10	$\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\ 0 & 0\end{pmatrix}$
3	$\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\ 0 & 0\end{pmatrix}$	11	$\begin{pmatrix}1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$
4	$\begin{pmatrix}1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\ 0 & 0\end{pmatrix}$	12	$\begin{pmatrix}-1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$
5	$\begin{pmatrix}-1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$	13	$\begin{pmatrix}-1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\ 0 & 0\end{pmatrix}$
6	$\begin{pmatrix}-1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\ 0 & 0\end{pmatrix}$	14	$\begin{pmatrix}-1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\ 0 & 0\end{pmatrix}$
7	$\begin{pmatrix}1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\ 0 & 0\end{pmatrix}$	15	$\begin{pmatrix}-1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\ 0 & 0\end{pmatrix}$
8	$\begin{pmatrix}-1 & 0\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\ 0 & 0\end{pmatrix}$	16	$\begin{pmatrix}-1 & 0\\ -1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\ 0 & 0\end{pmatrix}$

$$b_c = -(2+\epsilon)\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}$$





Different sublattices

different channel

fully-connected layer

$$W_o = \begin{pmatrix} \overbrace{1 \quad \dots \quad 1}^{8L^2 \text{ terms}} & \overbrace{-L^2 \dots -L^2}^{8L^2 \text{ terms}} \\ -1 \quad \dots \quad -1 & L^2 \dots L^2 \end{pmatrix}, \quad \text{and } b_o = \begin{pmatrix}0\\ 0\end{pmatrix}$$

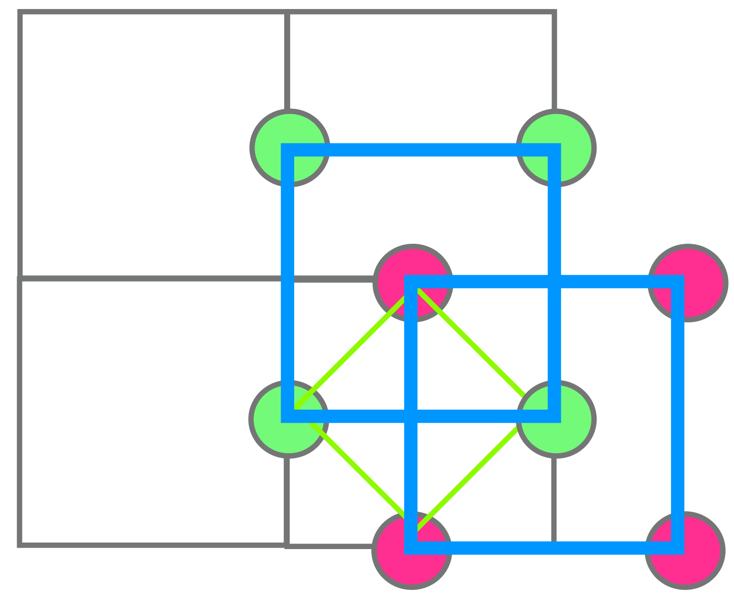
Basically the Conv. layer encodes the Hamiltonian

HOW TO IMPLEMENT THIS LOCAL CONSTRAINT WITH A CONVNET

Convolutional layer

f	s=A	s=B	f	s=A	s=B
1	$\begin{pmatrix}1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix}$	9	$\begin{pmatrix}1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\0 & 0\end{pmatrix}$
2	$\begin{pmatrix}-1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\0 & 0\end{pmatrix}$	10	$\begin{pmatrix}1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\0 & 0\end{pmatrix}$
3	$\begin{pmatrix}1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\0 & 0\end{pmatrix}$	11	$\begin{pmatrix}1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix}$
4	$\begin{pmatrix}1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\0 & 0\end{pmatrix}$	12	$\begin{pmatrix}-1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix}$
5	$\begin{pmatrix}-1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix}$	13	$\begin{pmatrix}-1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\0 & 0\end{pmatrix}$
6	$\begin{pmatrix}-1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\0 & 0\end{pmatrix}$	14	$\begin{pmatrix}-1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\0 & 0\end{pmatrix}$
7	$\begin{pmatrix}1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & -1\\0 & 0\end{pmatrix}$	15	$\begin{pmatrix}-1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\0 & 0\end{pmatrix}$
8	$\begin{pmatrix}-1 & 0\\1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & 1\\0 & 0\end{pmatrix}$	16	$\begin{pmatrix}-1 & 0\\-1 & 0\end{pmatrix}$	$\begin{pmatrix}-1 & -1\\0 & 0\end{pmatrix}$

$$b_c = -(2+\epsilon)\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}$$



Different sub lattices

different channel

image recognition R G ~~B~~

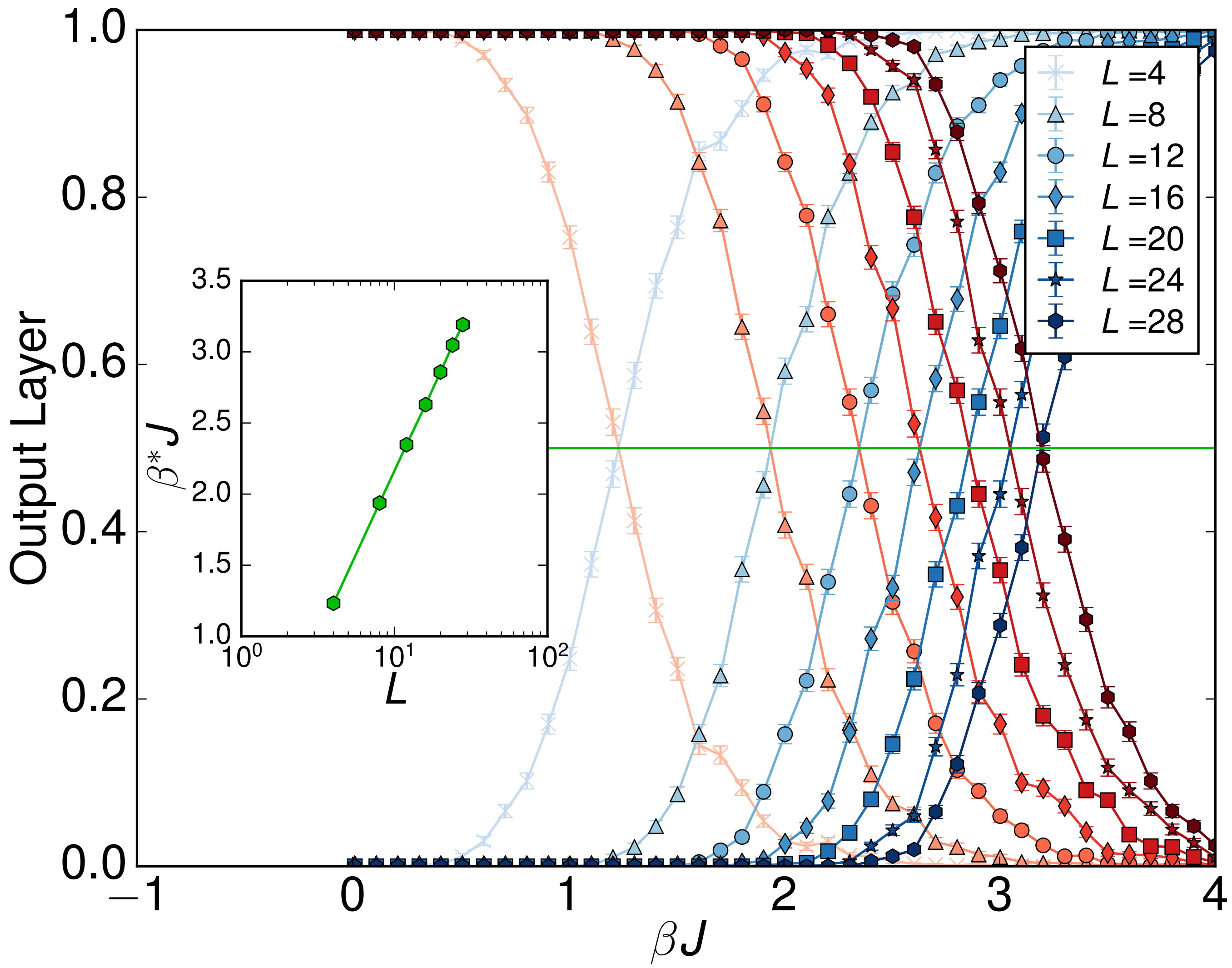
fully-connected layer

$$W_o = \begin{pmatrix} \overbrace{1 \ \dots \ 1}^{8L^2 \text{ terms}} & \overbrace{-L^2 \dots -L^2}^{8L^2 \text{ terms}} \\ -1 \ \dots \ -1 & L^2 \dots L^2 \end{pmatrix}, \quad \text{and } b_o = \begin{pmatrix}0\\0\end{pmatrix}$$

Basically the Conv. layer encodes the Hamiltonian

Fully connected layer counts defects

LOGARITHMIC CROSSOVER OF THE ISING GAUGE THEORY



NN is confused = crossover temperature

slowly cross over to the high-temperature phase. The cross-over temperature T^* happens as the number of thermally excited defects $\sim N\exp(-2J\beta)$ is of the order of one, implying $T^*/J \sim 1/\ln\sqrt{N}$.[23] As the presence of local defects is the mechanism through which the CNN decides

Only one defect on average distorts the Wilson loops. Same crossover observed in the topological entanglement entropy.

Nature Physics 13, 431–434 (2017)

1605.01735

THE GROUND STATE OF THE TORIC CODE

GROUND STATE OF THE TORIC CODE IS THE RK WAVE FUNCTION OF THE ISING GAUGE THEORY

$$|\Psi_{\mathrm{RK}}\rangle = \frac{1}{\sqrt{Z(\beta)}} \sum_{\mathcal{C}} e^{-\beta/2E_{\mathcal{C}}} |\mathcal{C}\rangle$$

classical d -> quantum d

Quantum system admitting an SMF decomposition	Classical system
Hilbert space with basis $\mathcal{B}$ labeled by $\mathcal{S}$	Configuration space $\mathcal{S}$
Ground state wavefunction	Boltzmann distribution
Quantum phase transitions	Classical phase transitions

…. and more

Castelnovo et al. Annals of Physics, Vol. 318, Issue 2, August 2005, Pages 316-344

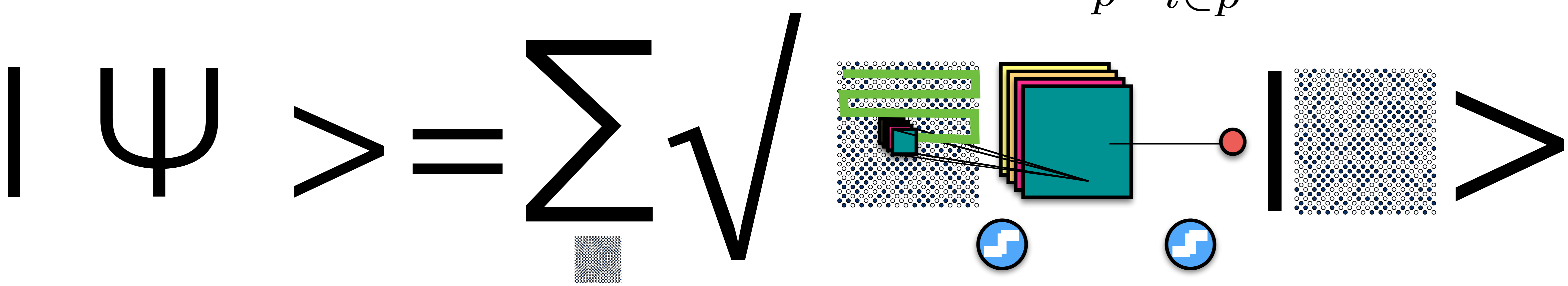
GROUND STATE OF THE TORIC CODE IS THE RK WAVE FUNCTION OF THE ISING GAUGE THEORY

$$H = -J_p \sum_p \prod_{i \in p} \sigma_i^z - J_v \sum_v \prod_{i \in v} \sigma_i^x$$

$$|\Psi_{\mathrm{TC}}\rangle \propto \lim_{\beta \to \infty} \sum_{\sigma_1, \ldots, \sigma_N} e^{\frac{\beta}{2} J \sum_p \prod_{i \in p} \sigma_i^z} |\sigma_1, \ldots, \sigma_N\rangle$$

PEPS : *F. Verstraete, M. M. Wolf, D. Perez-Garcia, J. I. Cirac Phys. Rev. Lett. 96, 220601 (2006).*

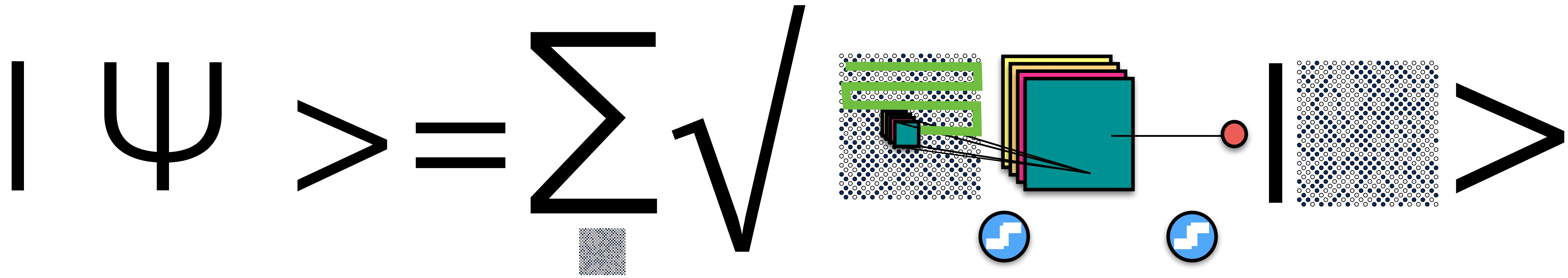
$$O_{\mathrm{cold}}(\sigma_1, \ldots, \sigma_N) \propto \lim_{\beta \to \infty} \exp \beta J \sum_p \prod_{i \in p} \sigma_i^z$$



1605.01735 *Nature Physics 13, 431–434 (2017)*

GROUND STATE OF THE TORIC CODE IS THE RK WAVE FUNCTION OF THE ISING GAUGE THEORY

$$H = -J_p \sum_p \prod_{i \in p} \sigma_i^z - J_v \sum_v \prod_{i \in v} \sigma_i^x$$



Neural net represents a ground state of the toric code: equal weight superposition of closed string states Nature Physics 13, 431–434 (2017)

Optimize using VMC Sorella, Imada and Carleo 's style

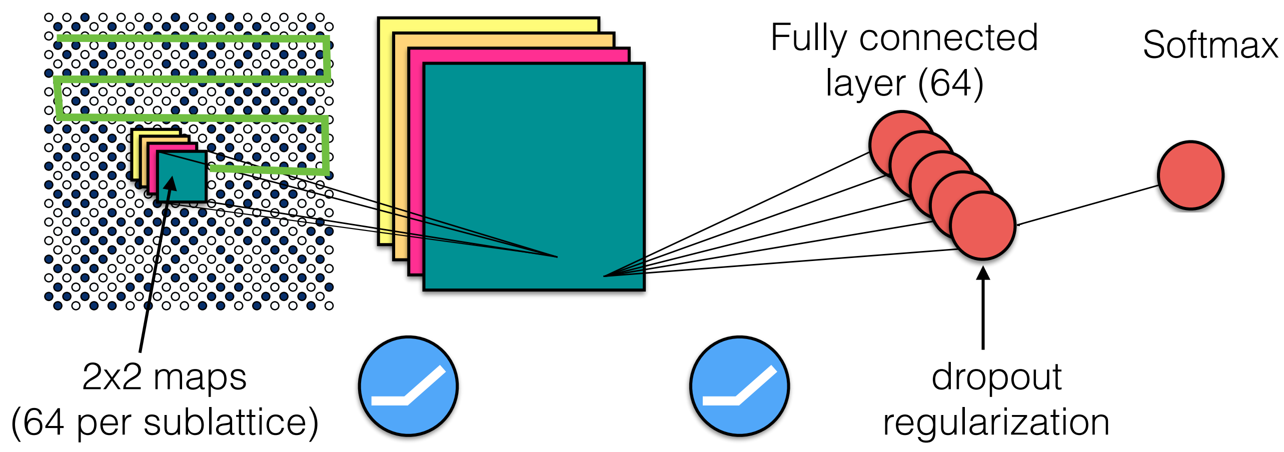
Closely related to Dong-Ling Deng et al arXiv:1609.09060, Phys. Rev. X 7, 021021 (2017)

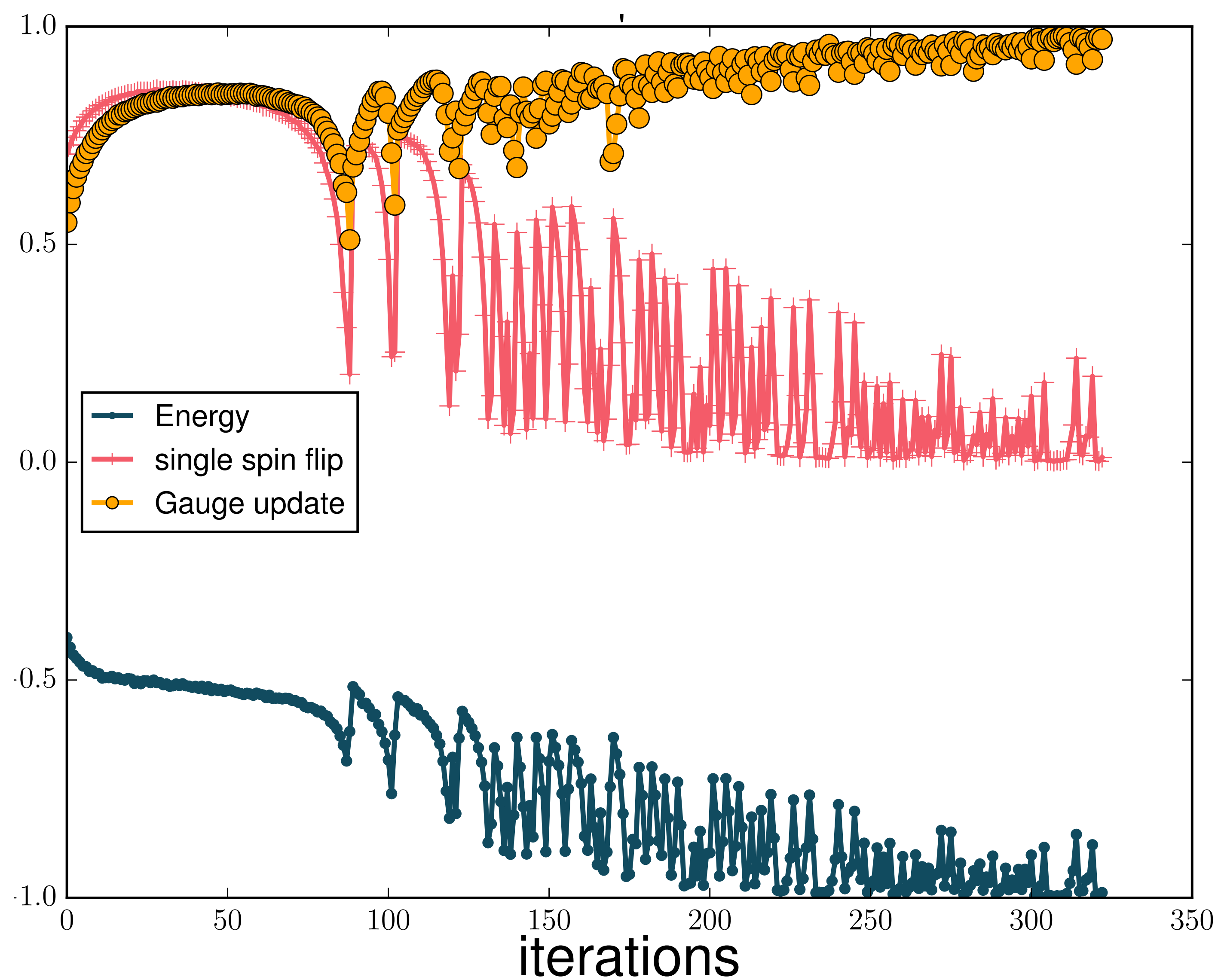
Jing Chen, Song Cheng, Haidong Xie, Lei Wang, Tao Xiang arXiv:1701.04831 RBMs

Xun Gao and Lu-Ming Duan, 1701.05039

Artificial neural network methods in quantum mechanics. Computer Physics Communications 104 (1997) 1-14

TOWARD DEEP LEARNING GROUND STATES (WORK IN PROGRESS)





3-layer CNN

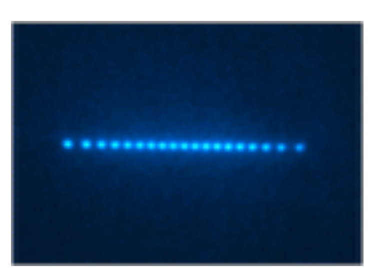
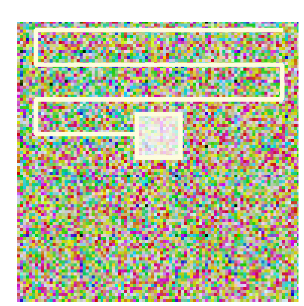
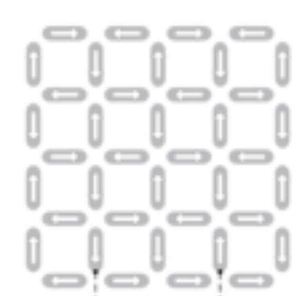
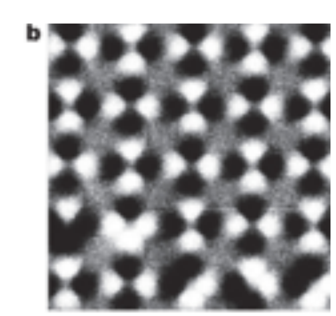
Tensorflow

CONCLUSION

- ➤ We encode and discriminate phases and phase transitions, both conventional and topological, using neural network technology.
- ➤ We have a solid understanding of what the neural nets do in those cases through controlled analytical models.

OUTLOOK

- ➤ We expect a rapid adoption of ML techniques as a tool in condensed matter physics.
- ➤ Variational interpretation of CNNs and their optimization for ground state.

QUANTUM MACHINE LEARNING POSITIONS AT D-WAVE

There are several positions at D-Wave at the intersection between

Machine learning, reinforcement learning, quantum physics, and quantum computing.

Friendly research atmosphere.

Great cities (Vancouver, Toronto)

email me at: jcarrasquilla@dwavesys.com

https://www.dwavesys.com/careers

D:WAVE
The Quantum Computing Company™