Title: Machine Learning Lecture - 230309

Speakers: Lauren Hayward

Collection: Machine Learning for Many-Body Physics (2022/2023)

Date: March 09, 2023 - 9:00 AM

URL: https://pirsa.org/23030031

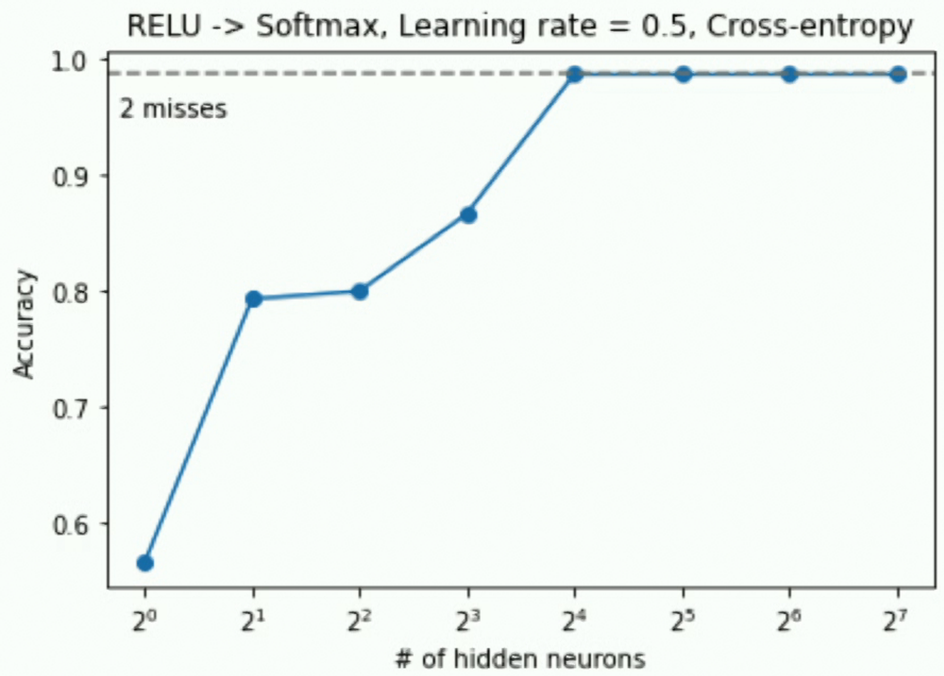
Machine Learning for Many-Body Physics

Lecture 6

PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Accuracy vs. the number of neurons in the hidden layer

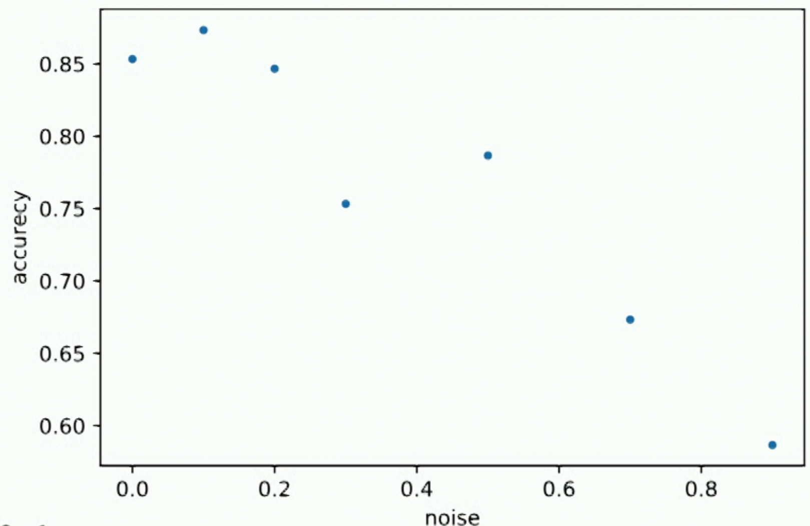


PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Results from 2022 students

Accuracy vs. the magnitude of noise in the data



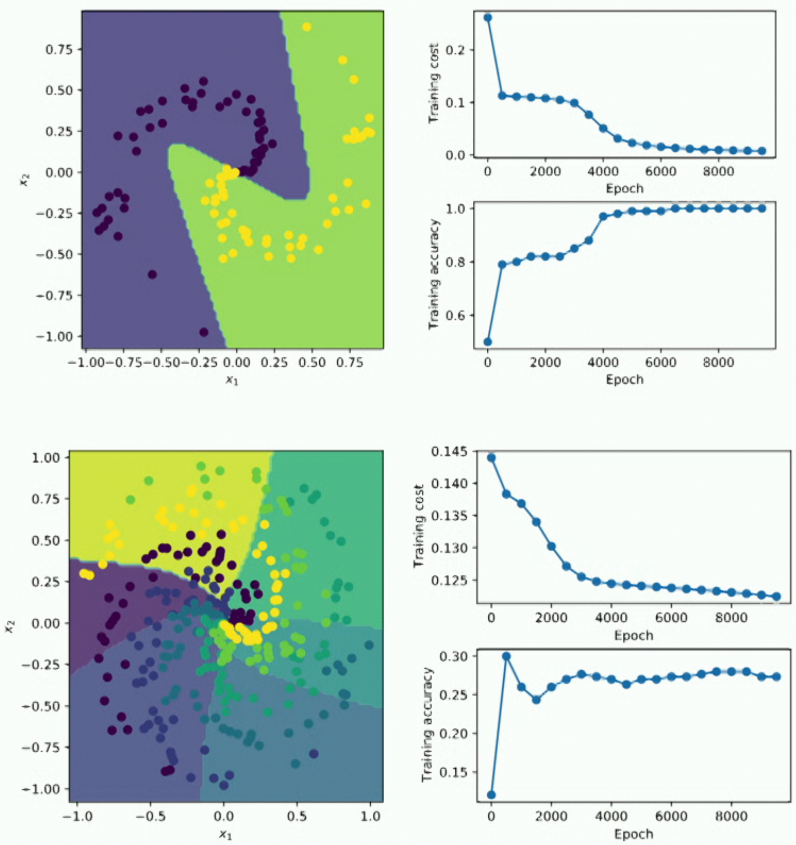
- learning_rate = 0.1
- cost_func = torch.nn.MSELoss()
- activation_function = sigmoid
- number_of_layer and neurons: 1, 4
- N_epochs = 100000

PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Results from 2022 students

Varying the number of different labels (branches)

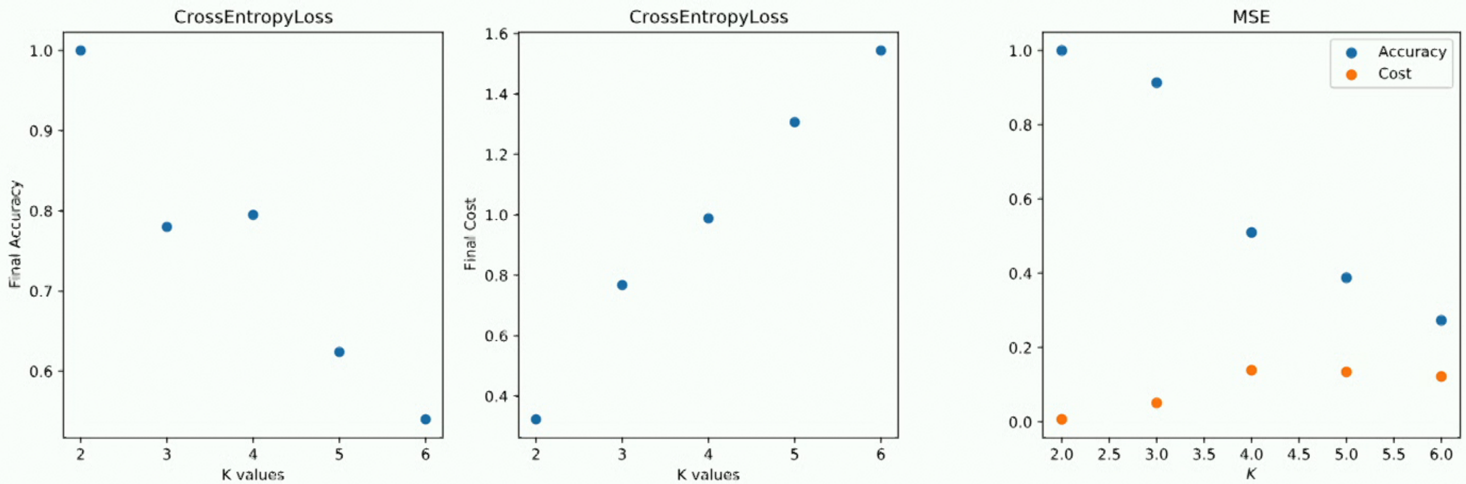


PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Results from 2022 students

Accuracy vs. the number of different labels (branches)

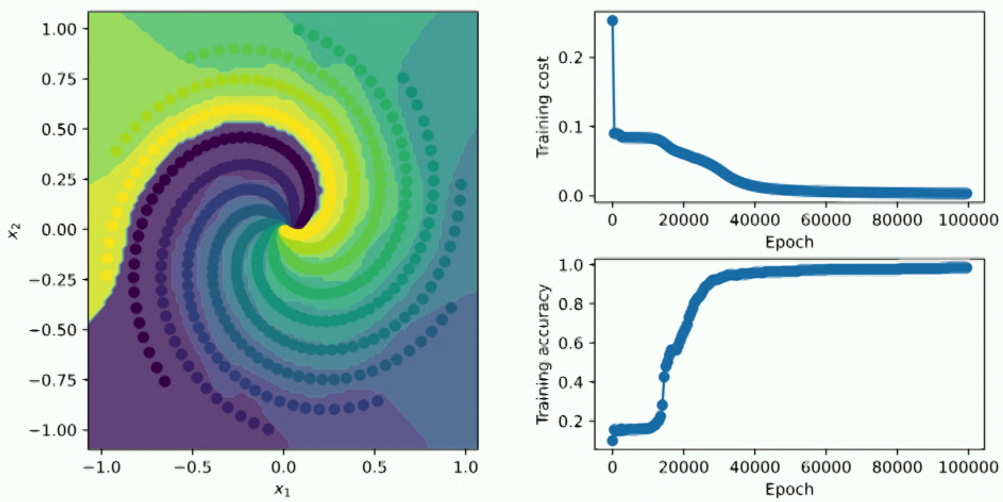



```
- learning rate = 1
- 4 neurons in the hidden layer
- first layer uses sigmoid
- outer layer uses softmax
```

PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Group 3

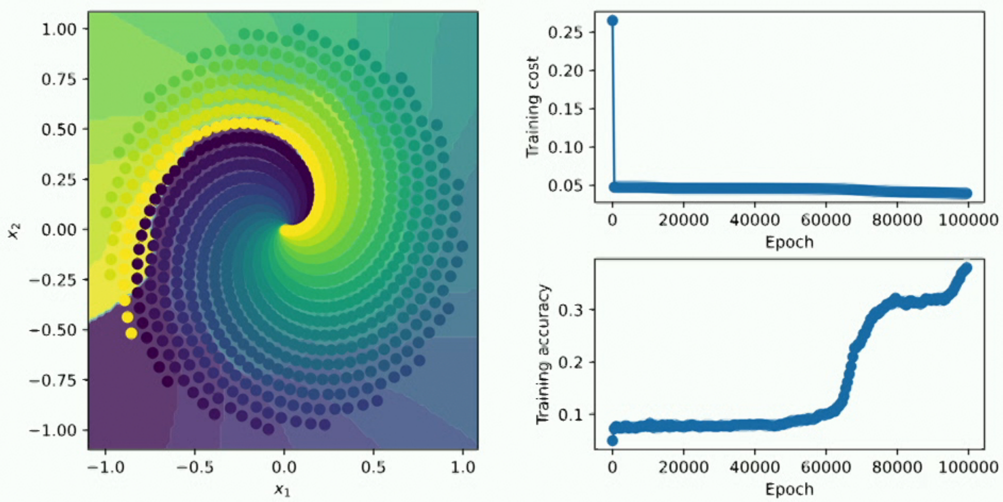


- 4 layers with 2, 20, 40, and k neurons
- Adam optimization wasn't working!
- Sigmoid activation function
- MSELoss

PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Group 3

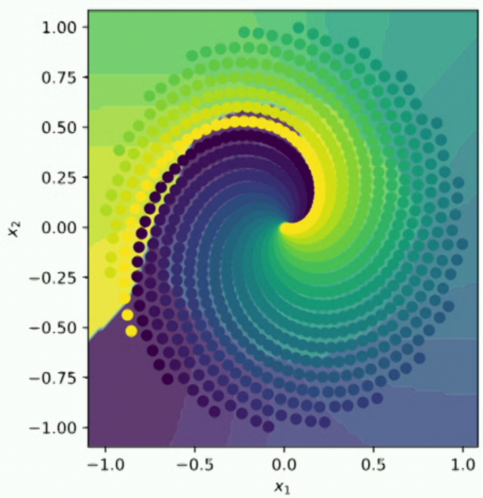


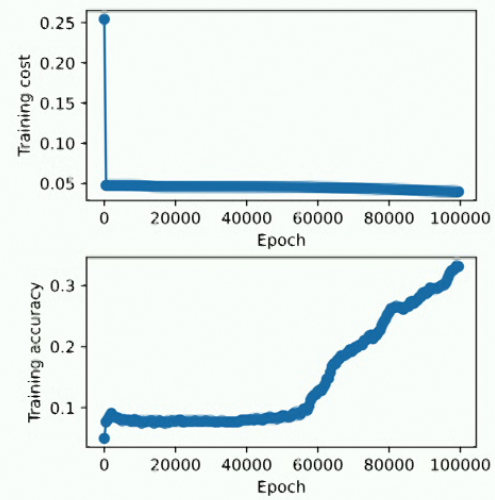
- 4 layers with 2, 20, 40, and k neurons
- Adam optimization wasn't working!
- Sigmoid activation function
- MSELoss

PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Group 3



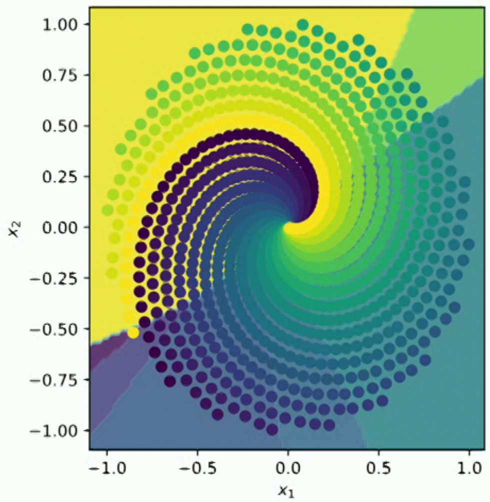


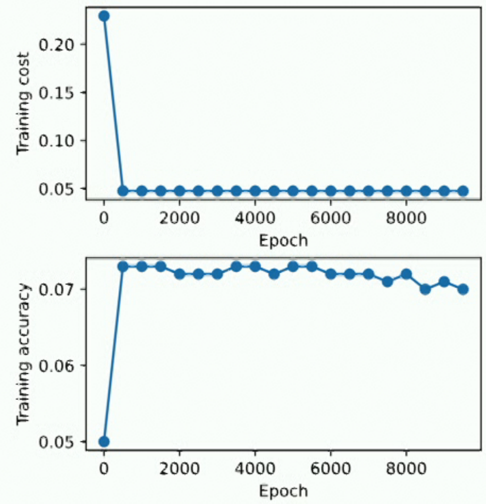
- 4 layers with 2, 20, 40, and k neurons
- Adam optimization wasn't working!
- Sigmoid activation function
- MSELoss

PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Group 3



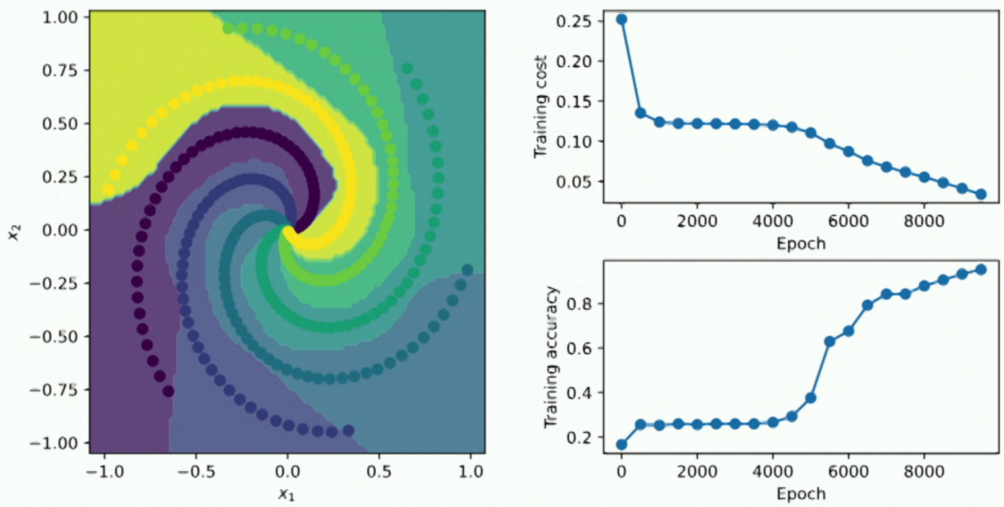


- 4 layers with 2, 20, 40, and k neurons
- Adam optimization wasn't working!
- Sigmoid activation function
- MSELoss

PERIMETER
SCHOLARS
INTERNATIONAL

Tutorial 2, Problem 2

Group 3

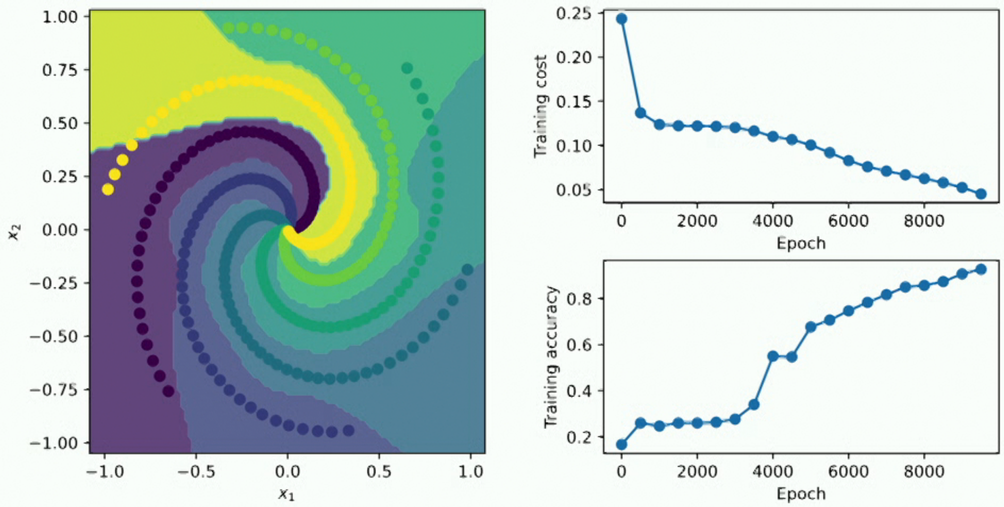


- 4 layers with 2, 20, 40, and k neurons
- Adam optimization wasn't working!
- Sigmoid activation function
- MSELoss

PERIMETER SCHOLARS INTERNATIONAL

Tutorial 2, Problem 2

Group 3



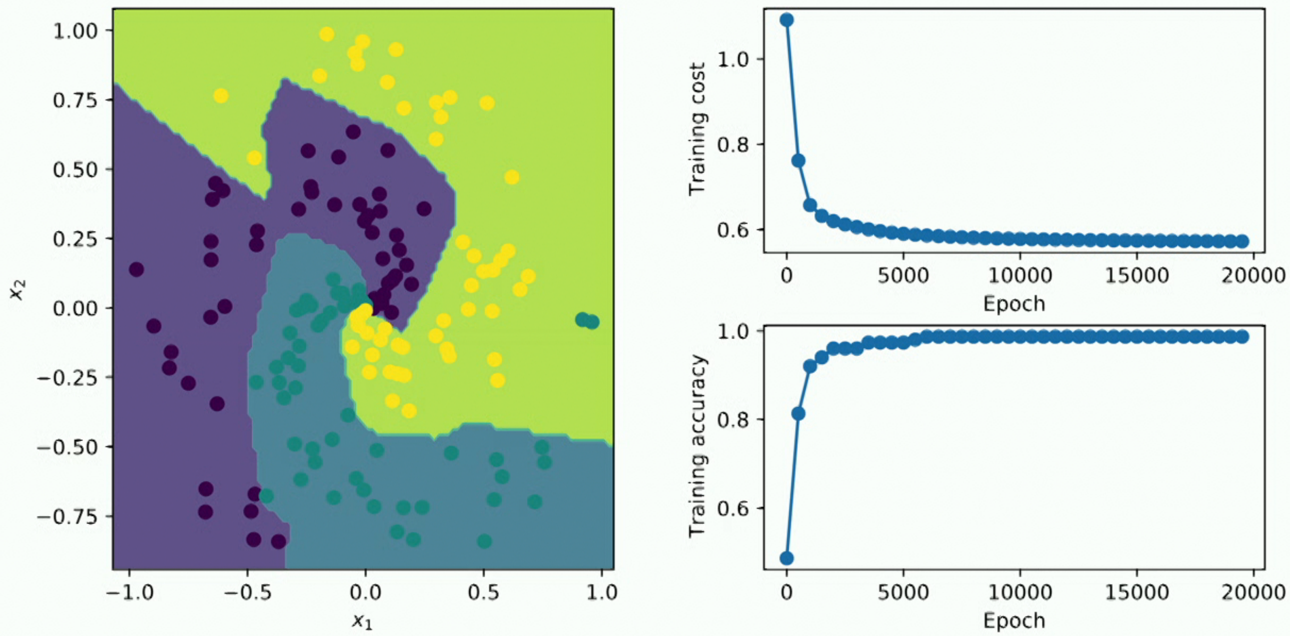
- 4 layers with 2, 20, 40, and k neurons
- Adam optimization wasn't working!
- Sigmoid activation function
- MSELoss

PERIMETER SCHOLARS INTERNATIONAL

Overfitting

Results from 2022 students

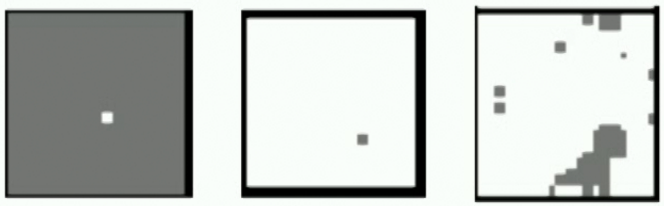
```
mag_noise   = 0.5
hidden_size = 100
```



PERIMETER SCHOLARS INTERNATIONAL

Ising model classification

Ferromagnetic phase:



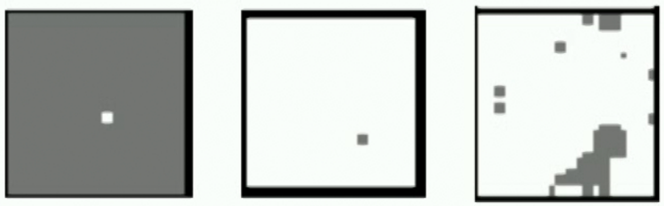
Paramagnetic phase:

Carrasquilla and Melko, arXiv:1605.01735

PERIMETER SCHOLARS INTERNATIONAL

Ising model classification

Ferromagnetic phase:



?

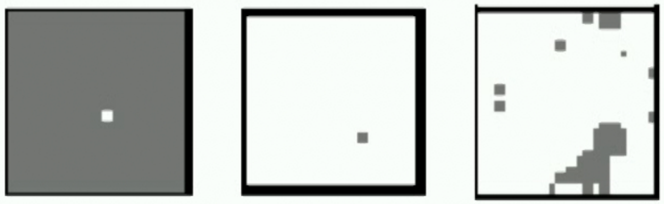
Paramagnetic phase:

Carrasquilla and Melko, arXiv:1605.01735

PERIMETER SCHOLARS INTERNATIONAL

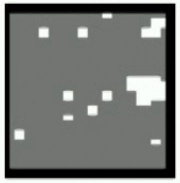
Ising model classification

Ferromagnetic phase:



Paramagnetic phase:

?

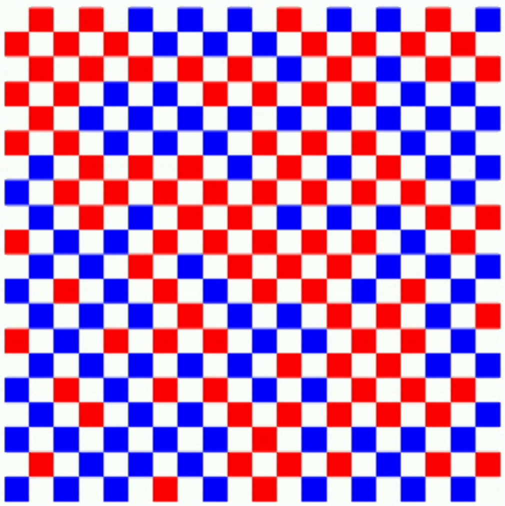


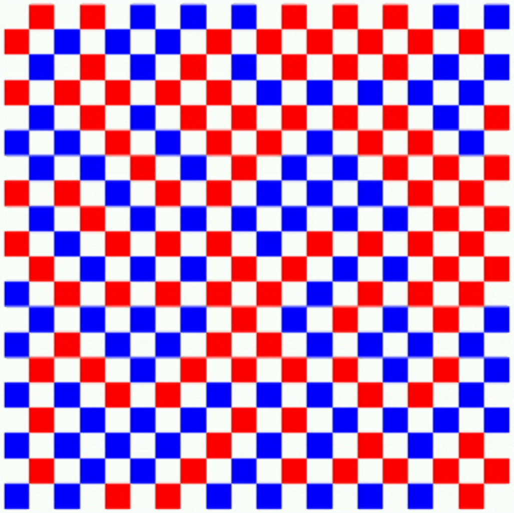
?

Carrasquilla and Melko, arXiv:1605.01735

PERIMETER SCHOLARS INTERNATIONAL

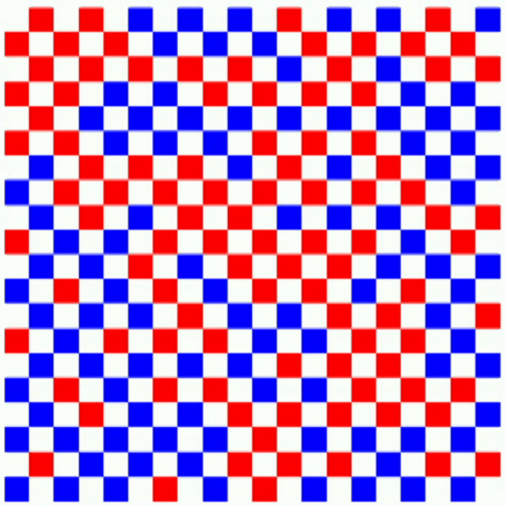
Ising gauge theory classification



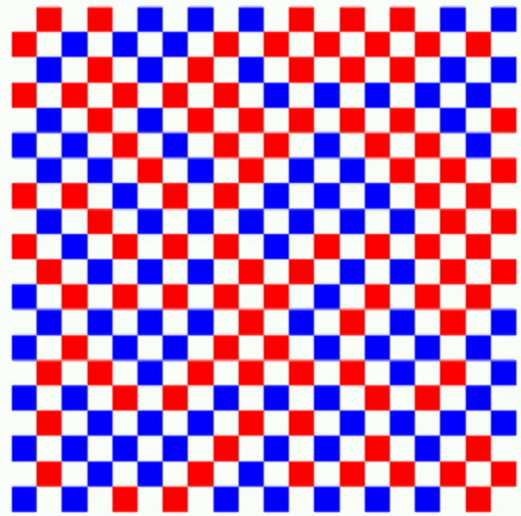


PERIMETER SCHOLARS INTERNATIONAL

Ising gauge theory classification



Random state



Ordered state

Outline for today:

- Using supervised neural networks (NNs) to learn about phases of matter
  - ↳ Ising $\mathbb{Z}_2$ gauge theory
- Recap of hyperparams. in NNs
- Monte Carlo sampling

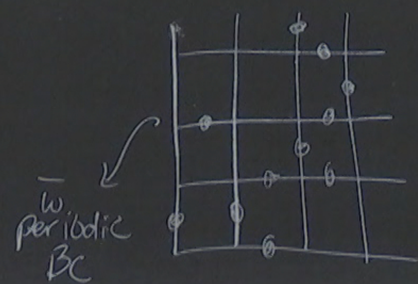
Distinguishing phases of the classical Ising $\mathbb{Z}_2$ gauge theory

$$H = -J \sum_{\substack{\text{plaquettes} \\ p}} \left(\prod_{i \in p} s_i \right)$$

with $s_i = +1$ or -1

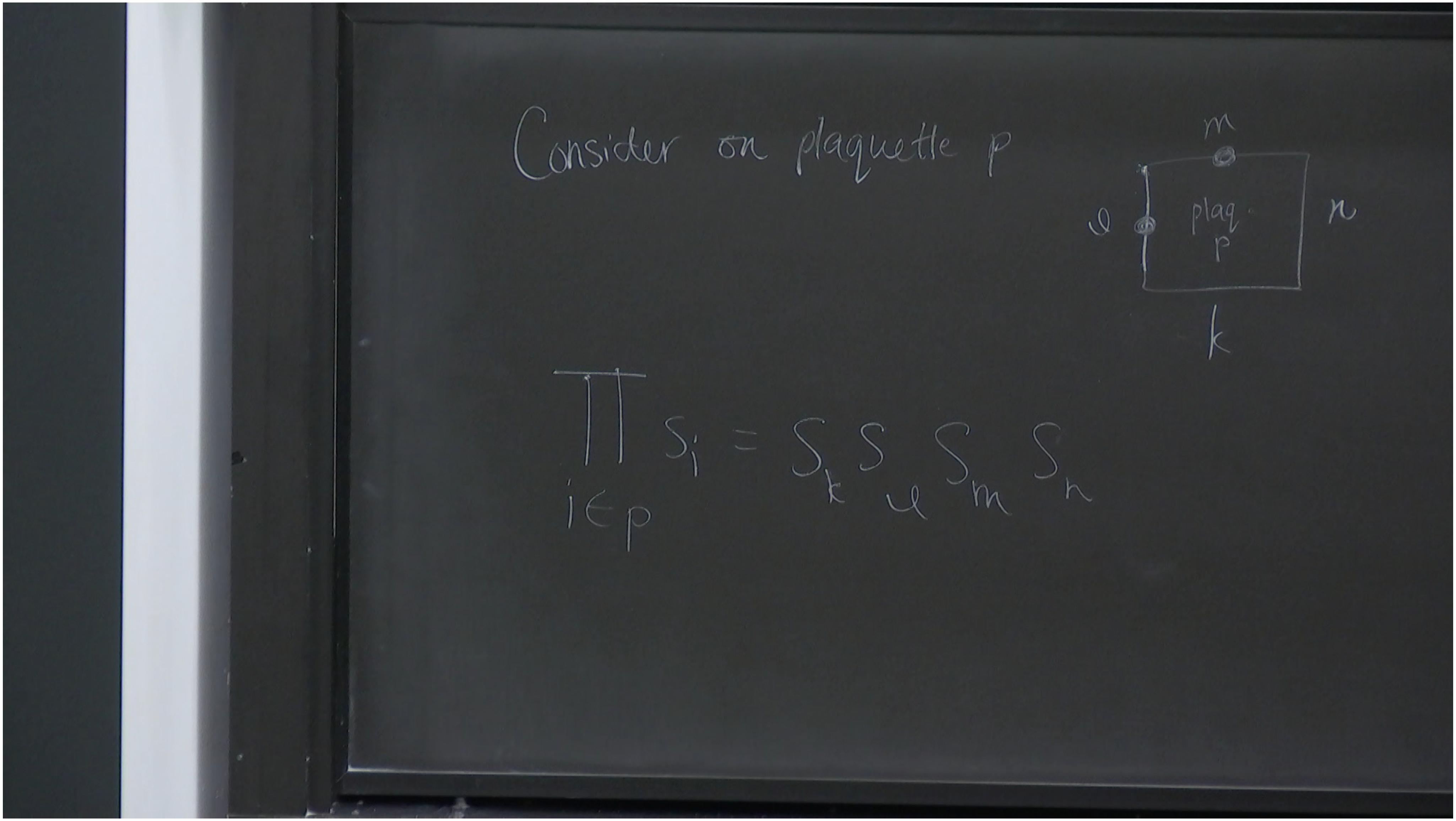
The d.o.f. are on the bonds of the lattice

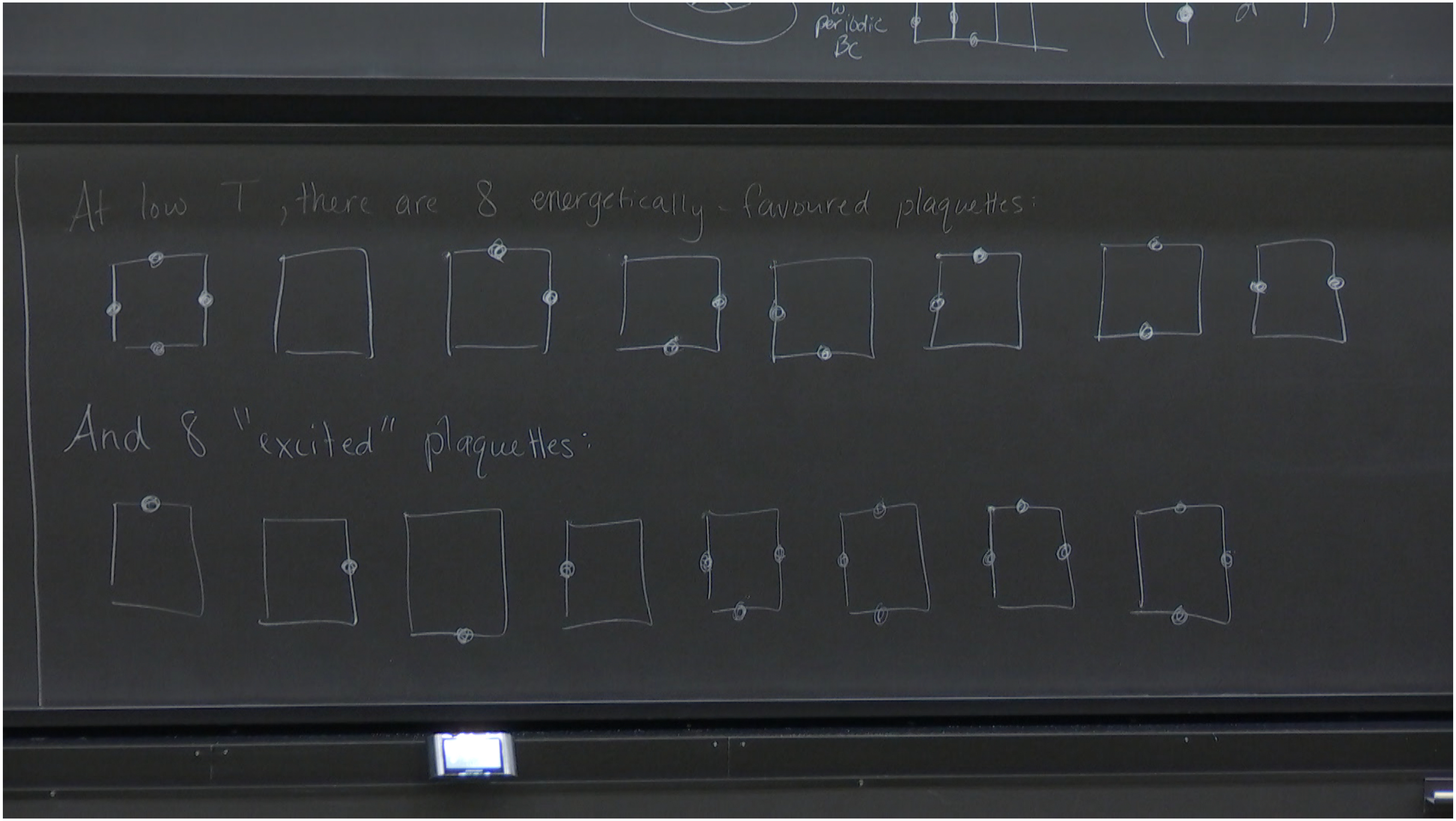
On a torus:



$s_i = +1$ or -1

(• or |)

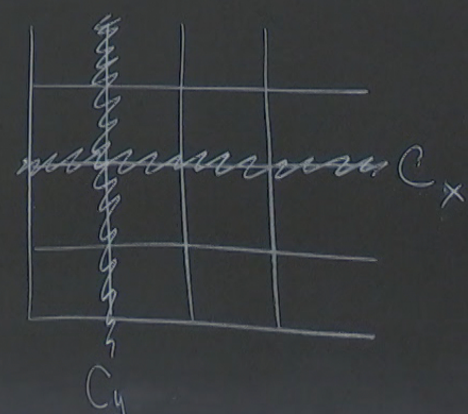

Consider on plaquette p
m
l
plaq. p
n
k
$\prod_{i \in p} S_i = S_k S_l S_m S_n$


periodic BC
At low T, there are 8 energetically-favoured plaquettes:
And 8 "excited" plaquettes:

There is topological order exactly at $T=0$ in the thermodynamic limit and no order for $T>0$

We can use the topological Wilson loop W_c as a measure of this order

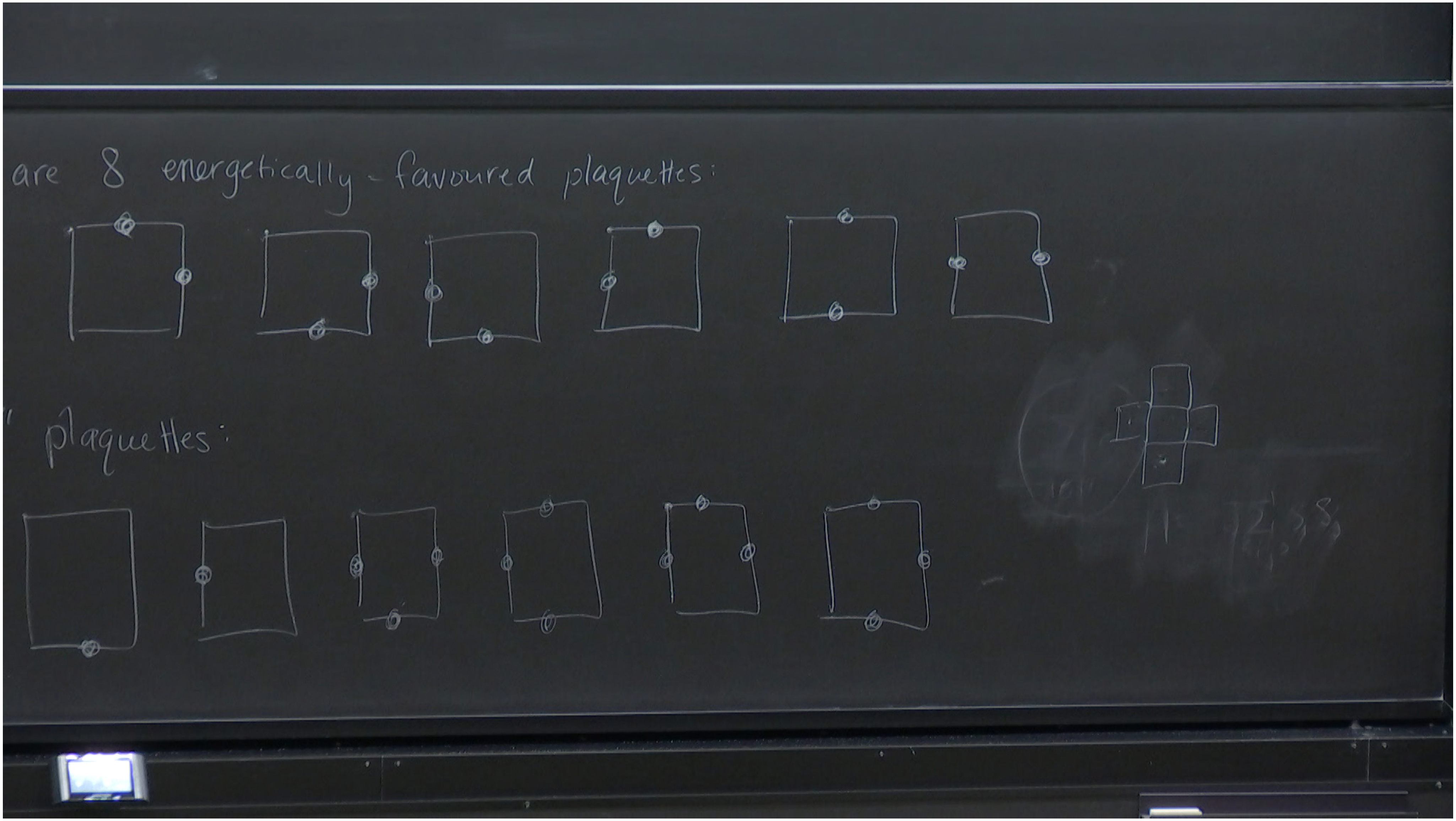
$$W_c = \prod_{i \in C} S_i = \pm 1$$

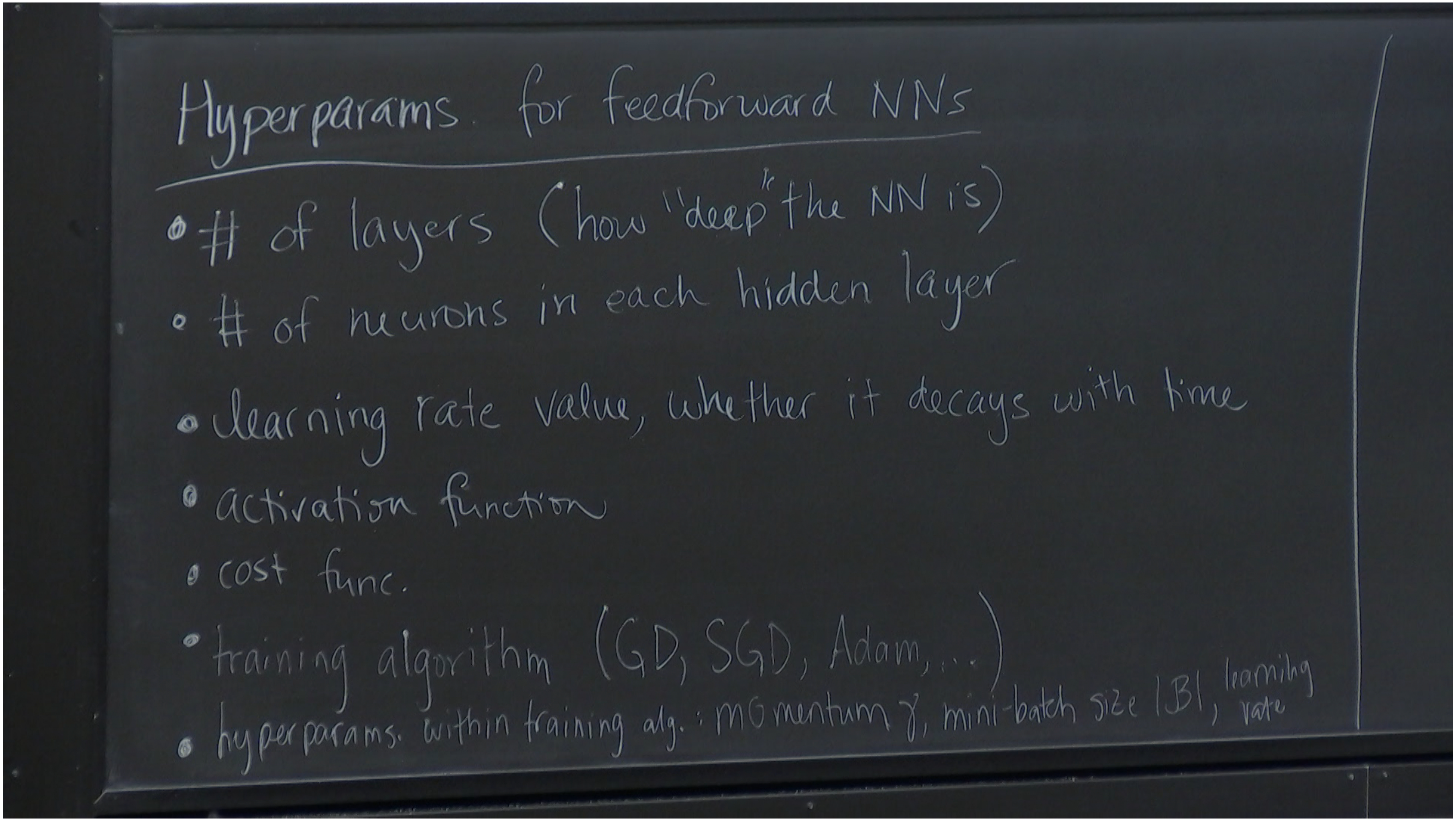


In the topologically-ordered groundstate, W_{C_x} and W_{C_y} are constant for all choices of C_x and C_y

Four possibilities:

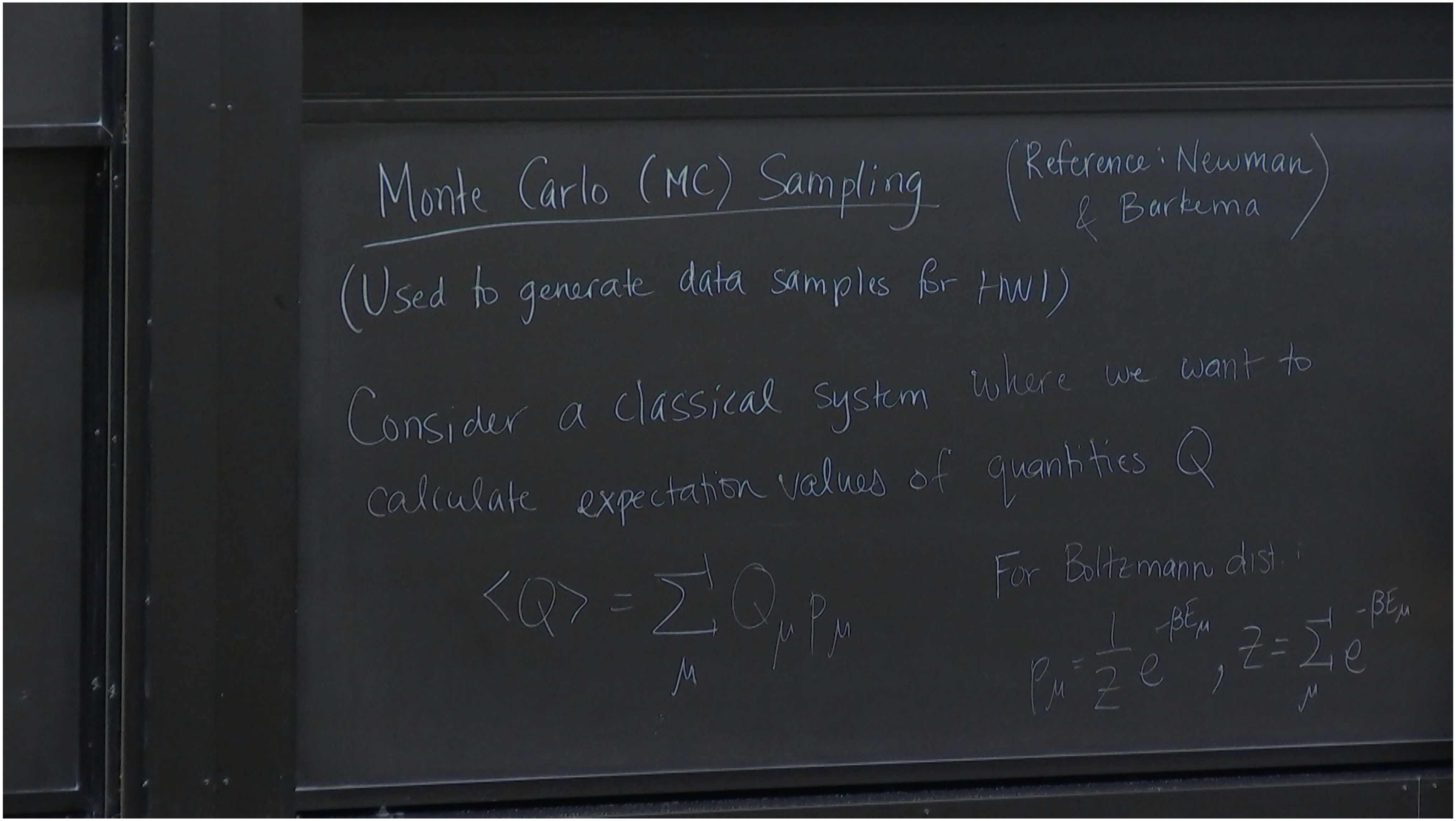
W_{C_x}	W_{C_y}
+1	+1
+1	−1
−1	+1
−1	−1


are 8 energetically-favoured plaquettes:
plaquettes:


Hyperparams for feedforward NNs
• # of layers (how "deep" the NN is)
• # of neurons in each hidden layer
• learning rate value, whether it decays with time
• activation function
• cost func.
• training algorithm (GD, SGD, Adam, ...)
• hyperparams. within training alg.: momentum γ, mini-batch size |B|, learning rate

- how much regularization is added to the cost (param. λ) and what kind (L1, L2, ...)
- masking prob. if using dropout
- how we partition into training, validation, testing
- whether or not we add noise to our data, what kind and how much
- how the weights and biases are initialized

⋮


Monte Carlo (MC) Sampling
(Reference: Newman & Barkema)
(Used to generate data samples for HW1)
Consider a classical system where we want to calculate expectation values of quantities Q
$\langle Q \rangle = \sum_\mu Q_\mu p_\mu$
For Boltzmann dist.:
$p_\mu = \frac{1}{Z} e^{-\beta E_\mu}, \quad Z = \sum_\mu e^{-\beta E_\mu}$

- E_μ is the energy of config μ
- $\beta = \frac{1}{k_B T}$
- The sums $\sum_\mu$ are over huge # of terms

(2^N terms for classical Ising). Analytically intractable

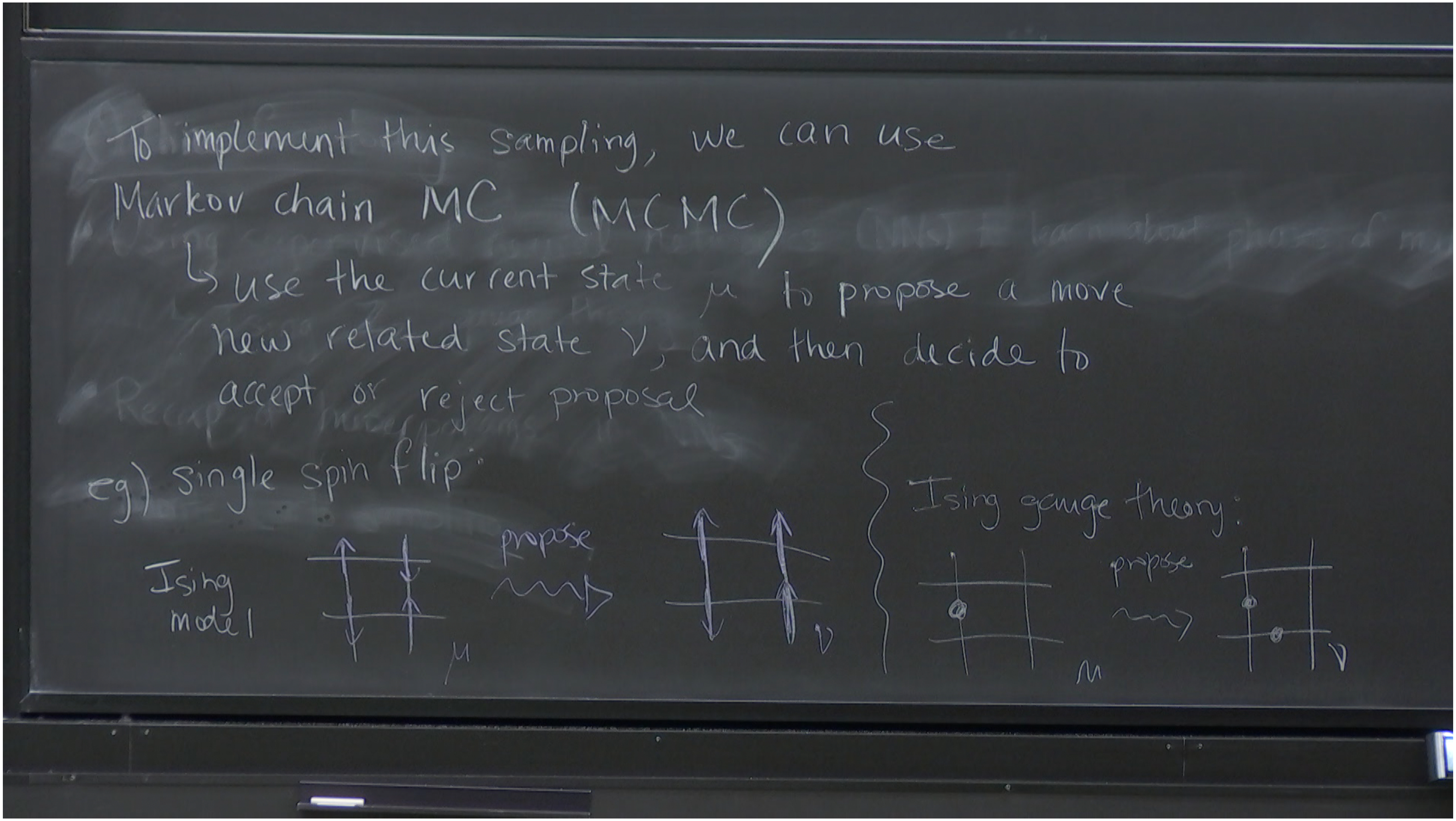
Idea of MC: estimate $\langle Q \rangle$ from M

States $\{\mu_1, \mu_2, \ldots, \mu_M\}$, $M \ll 2^N$

In particular, if each state μ_i is selected from the set of all states according to the desired dist. p_{μ_i}, then

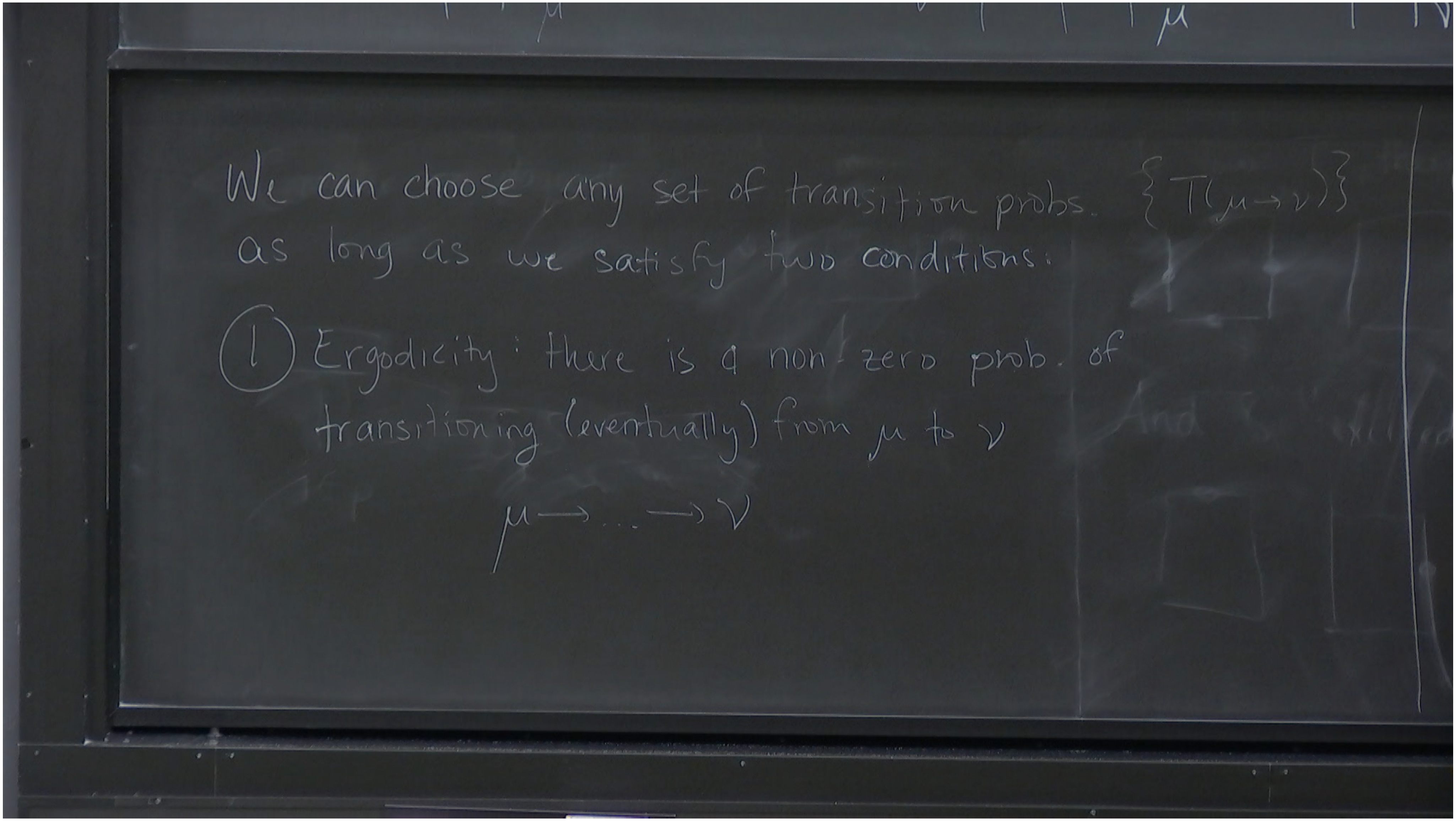
$$\langle Q \rangle \simeq Q_M = \frac{1}{M} \sum_{i=1}^{M} Q_{\mu_i}$$

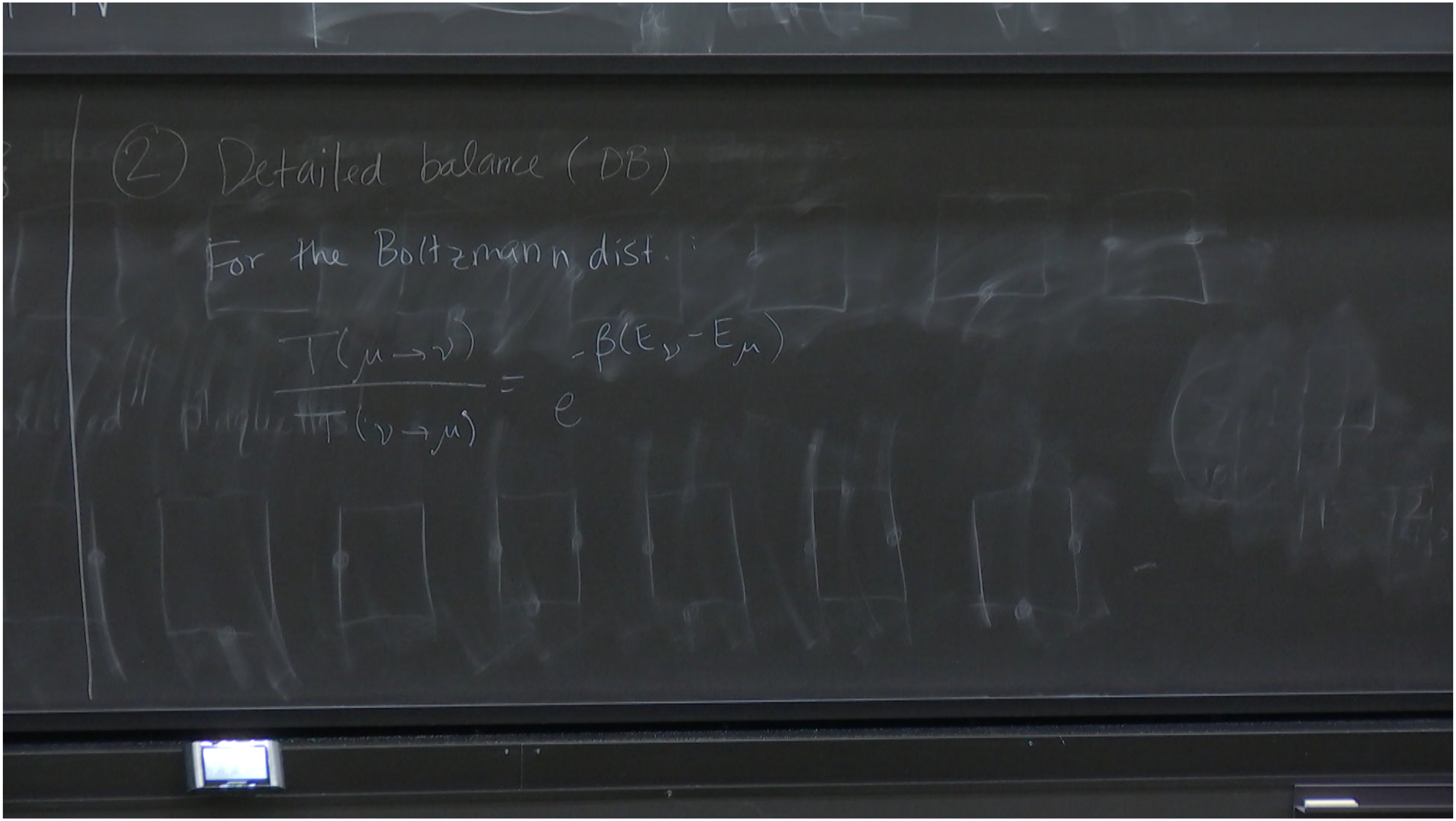
(makes use of "importance sampling")


To implement this sampling, we can use
Markov chain MC (MCMC)
↳ use the current state μ to propose a move
new related state ν, and then decide to
accept or reject proposal
eg) single spin flip:
Ising model
propose
μ
ν
Ising gauge theory:
propose
μ
ν

In priciple, any proposal is allowed

The prob. of moving from μ to ν is the transition prob. $T(\mu \to \nu)$, with $\sum_{\nu} T(\mu \to \nu) = 1$


We can choose any set of transition probs. {T(μ→ν)}
as long as we satisfy two conditions:
1 Ergodicity: there is a non-zero prob. of
transitioning (eventually) from μ to ν
μ→ … →ν


② Detailed balance (DB)
For the Boltzmann dist. :
$\frac{T(\mu \to \nu)}{T(\nu \to \mu)} = e^{-\beta(E_\nu - E_\mu)}$