

**Title:** Lecture - Statistical Physics, PHYS 602

**Speakers:** Emilie Huffman

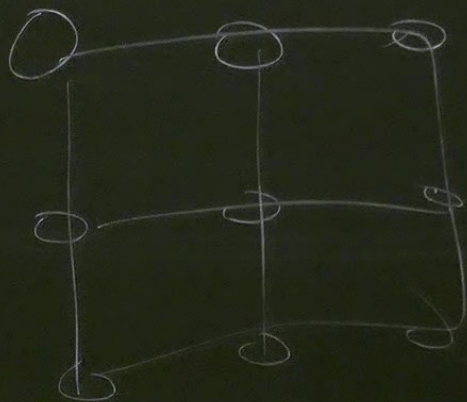
**Collection/Series:** Statistical Physics (Core), PHYS 602, October 7 - November 6, 2024

**Subject:** Condensed Matter, Other

**Date:** November 05, 2024 - 10:45 AM

**URL:** <https://pirsa.org/24110002>

# System of Interacting Spins



$$H = \sum_{\langle ij \rangle} J_{ij} \sigma_i \sigma_j$$

$$P(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z}$$

Access  $p(\sigma)$

Sample  $p(\sigma)$

Calculate quantities  $p(\sigma)$

$\sigma_i \sigma_j$

$H(\sigma)$

$$\beta = \frac{1}{T}$$

Access  $p(\sigma)$

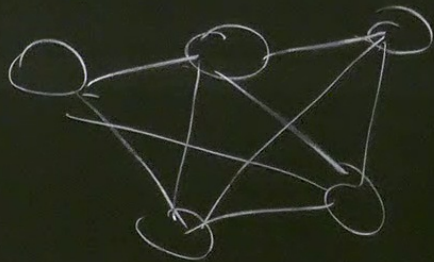
Sample  $p(\sigma)$

Calculate quantities  $p(\sigma)$

Generative model,

# Energy-Based Models (EBM)

Hopfield network & Boltzmann Machine



$$E = -\sum w_{ij} x_i x_j, \quad x_i \in \{-1, 1\}$$

$$\mathcal{Z} = \{w\}$$

$$\left\{ \vec{x}^{\mu} \right\}_{\mu}^N$$

ine  
 encode into the low energy states of the HW

$$x_i \in \{-1, 1\}$$

$$\text{argmin}_{\lambda} \sum_{\mu} E_{\lambda}(\vec{x}^{\mu})$$

Hebbian Learning

$$W = \frac{1}{N} \sum_{\mu} \vec{x}^{\mu} (\vec{x}^{\mu})^T$$

$\vec{x}(t_0) = \text{random start}$

choose  $i$ , flip  $x_i \rightarrow -x_i$

$E(\vec{x}(t_1)) < E(\vec{x}(t_0)) \rightarrow \text{accept change}$

$\underset{x}{\operatorname{argmin}} E_{\vec{x}}(x)$

Pr

Problem

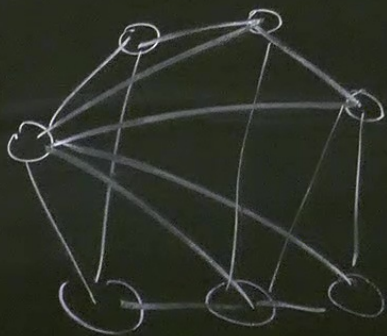
Stuck local minima

Restricted capacity (Linear amount patterns  
in # units ( $N$ ))

change



# Boltzmann Machine



visible,  $\vec{v}$

$$\vec{x} = \{ \vec{v}, \vec{h} \}$$

$$E_{\lambda}(\vec{x}) = - \sum_{i,j} W_{ij} x_i x_j - \sum_i b_i x_i$$

hidden,  $\vec{h}$

$$P_{\lambda}(\vec{x}) = \frac{e^{-E_{\lambda}(\vec{x})}}{Z}$$

$$P_{\lambda}(\vec{v}) = \sum_{\vec{h}} P_{\lambda}(\vec{v}, \vec{h})$$

$$\sum_i \text{bit}_i$$

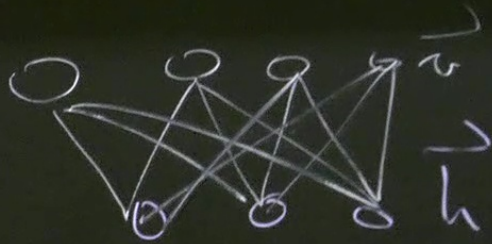
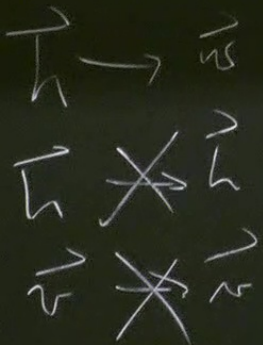
$$p(\vec{v}) = \sum_{\vec{h}} p_x(\vec{v}, \vec{h})$$

Train:  $\left\{ \vec{v}^\mu \right\}_\mu$   $2^D$

$$\text{argmax}_\lambda \sum_\mu p_x(\vec{v}^\mu)$$

Difficult to train (high interconnectivity  $\begin{pmatrix} \vec{v} \\ \vec{h} \\ \vec{h} \\ \vec{v} \end{pmatrix}$ )

# Restricted BM (RBM)



$$F_{\pi}(\vec{v}, \vec{h}) = -\sum_i b_i v_i - \sum_j c_j h_j - \sum_{i,j} w_{ij} v_i h_j$$

Train  $\{ \vec{v}_i, \vec{h}_i \}_{i=1}^D$

$$P_{KL}(P_1 | P_2) = \sum_{\vec{v}, \vec{h}} \frac{P_1(\vec{v}, \vec{h})}{P_2(\vec{v}, \vec{h})} \ln \frac{P_1(\vec{v}, \vec{h})}{P_2(\vec{v}, \vec{h})}$$

$$= \sum_{\vec{v}, \vec{h}} \underbrace{P_1(\vec{v}, \vec{h})}_{\text{CS+ data}} \ln \frac{P_1(\vec{v}, \vec{h})}{P_2(\vec{v}, \vec{h})}$$

NLL

$$\underbrace{-\sum_i b_i v_i - \sum_j c_j h_j - \sum_k W_{ij} v_i h_j}_{D} = -\sum_{\vec{v}^m} \ln P_2(\vec{v}^m)$$

$$\Rightarrow \arg \max_{\lambda} \sum_M P_2(\vec{v}^m)$$

$\underbrace{P_d(\vec{v}) \ln P_d(\vec{v})}_{\text{CST data}} - \sum_{\vec{v}} P_d(\vec{v}) \ln P_2(\vec{v})$

Training Gradient Descent

$$L_{\lambda}(\vec{x})$$

$$\lambda' = \lambda - \eta \frac{\partial L}{\partial \lambda}$$

reduce  $L$

$$\frac{\partial \text{MSE}}{\partial \lambda} = \frac{\partial}{\partial \lambda}$$

$\frac{\partial}{\partial \lambda}$   
cost data

$$\frac{\partial \text{MLL}}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left[ \underbrace{\ln \sum_h e^{-E_2(\vec{v}, h)}}_{+ve} - \underbrace{\ln Z}_{-ve} \right]$$

tractable  $\frac{\partial E_2}{\partial \alpha}$   
 $-\sum_h p(h|v) \frac{\partial E_2}{\partial \alpha}$

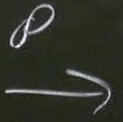
$\sum_{\vec{v}} Z^N \Rightarrow \text{intractable}$

# MCMC

$$D = \{ \sigma \}$$

MC step

$$\rightarrow D' = \{ \sigma' \}$$



sta

$$D_0 = \{ \vec{x}^m \}$$

MC step  
based model

$$D' = \{ \vec{x}'^m \}$$

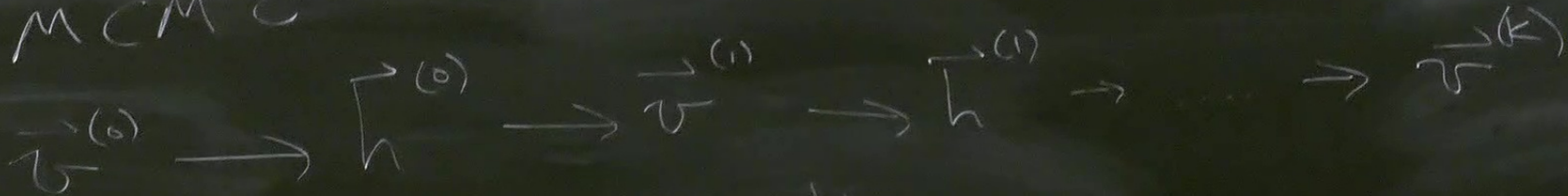
$$\{ D_0, D' \}^m \Rightarrow$$

RBM list = data dist  
Contrastive Divergence

(contrastive Diving)

Sample RBM

MCMC



K-block Gibbs Sampling

- Many equil. steps
- Auto correlation



(contrastive - Dividing)

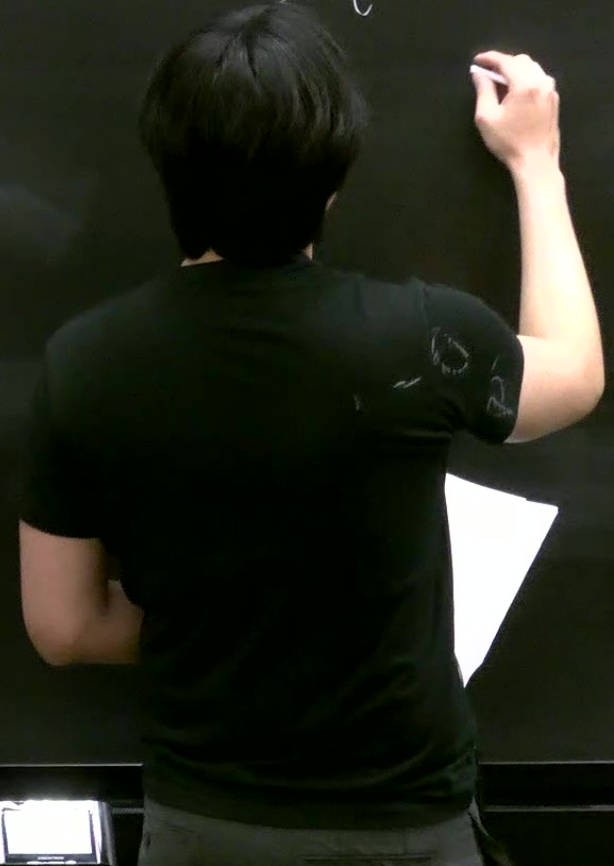
# Autoregressive Models

directly parameterize  $P(\vec{x})$

Chain Rule of prob.

$$P(\vec{x}) = P(x_1) P(x_2|x_1) P(x_3|x_{1:2})$$

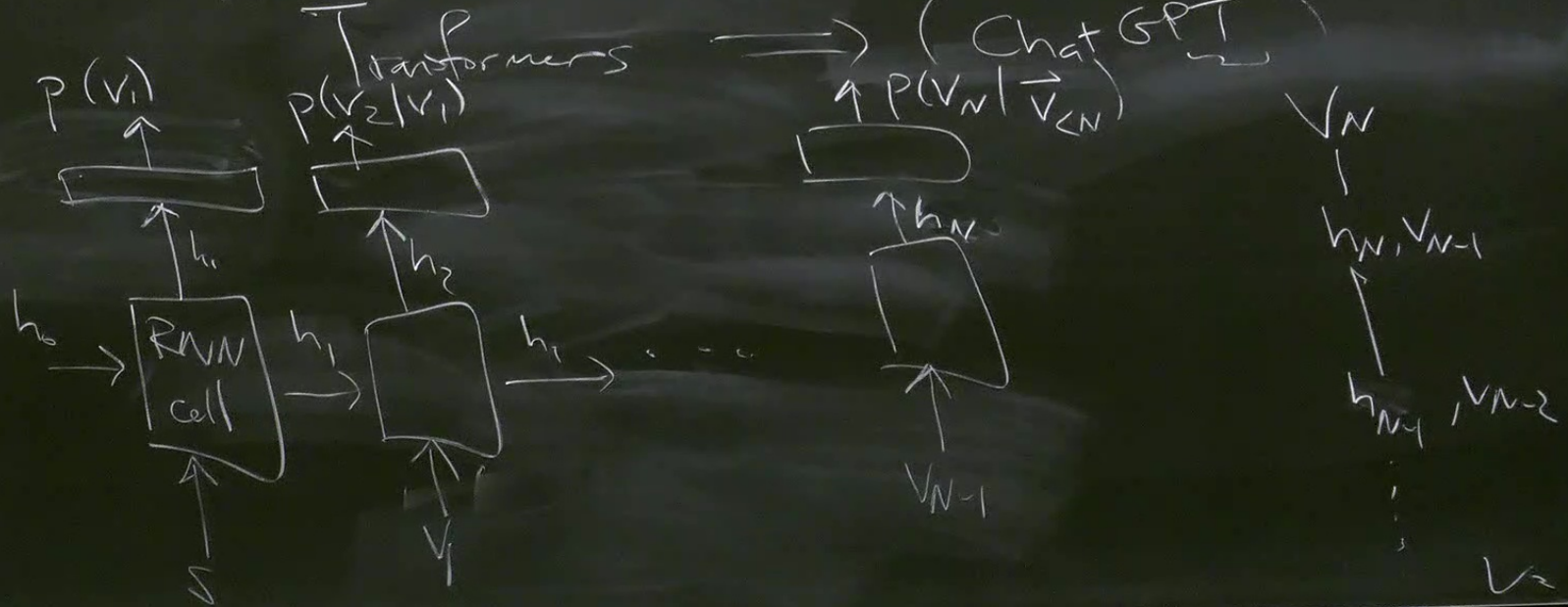
# Sequential Mod



# Sequential Model

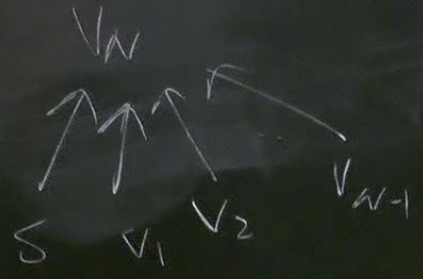
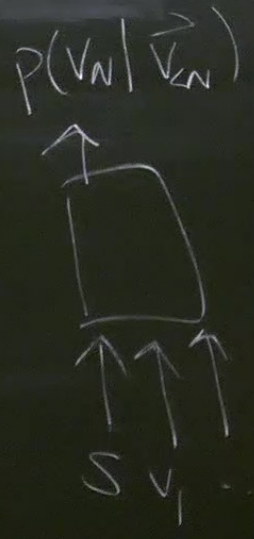
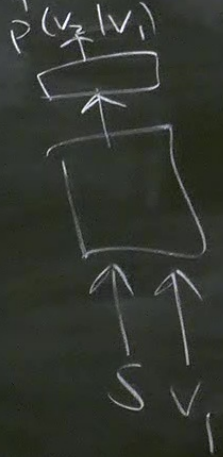
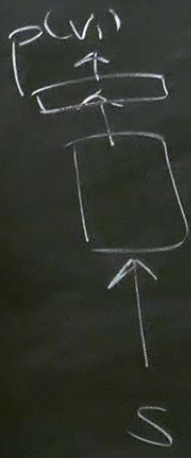
## Recurrent Neural Networks (RNN)

RNN



set data

# Transformers



$$d(\vec{z}) \ln p_d(\vec{z}) - \sum_{\vec{z}} p_d(\vec{z}) \ln p_d(\vec{z})$$

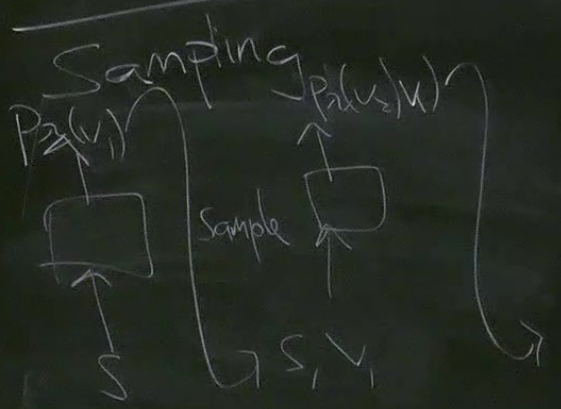
cost data

Train

No partition  $f^n \leftarrow P_2(x)$  normalized

Gradient Descent NLL

w-1



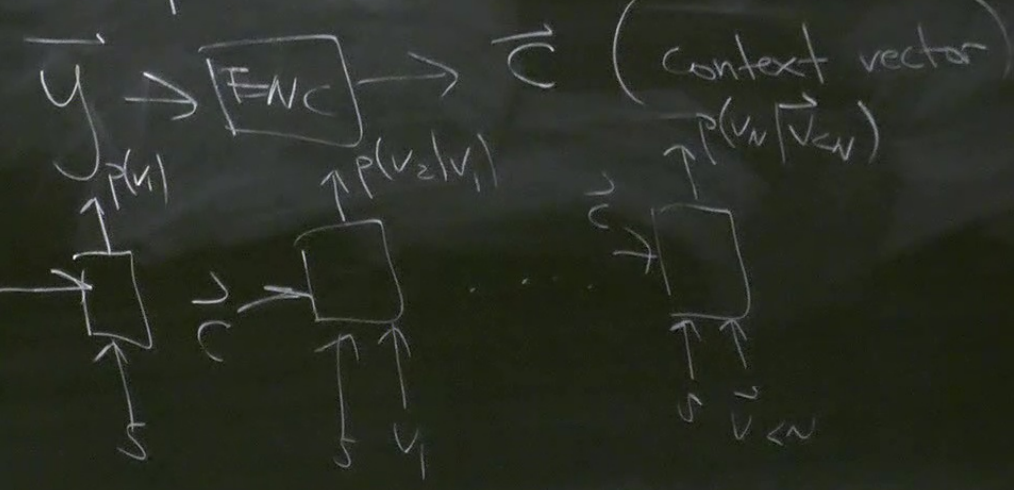
iid sample  
independent & identically distributed

13  
 $\int_{13}$   
 CST data

NN param.

a prob. is  $P(\vec{v})$

$P(\vec{v}; \vec{y})$ ,  $\vec{y} = \text{sys parameters} = (\beta, J, \dots)$



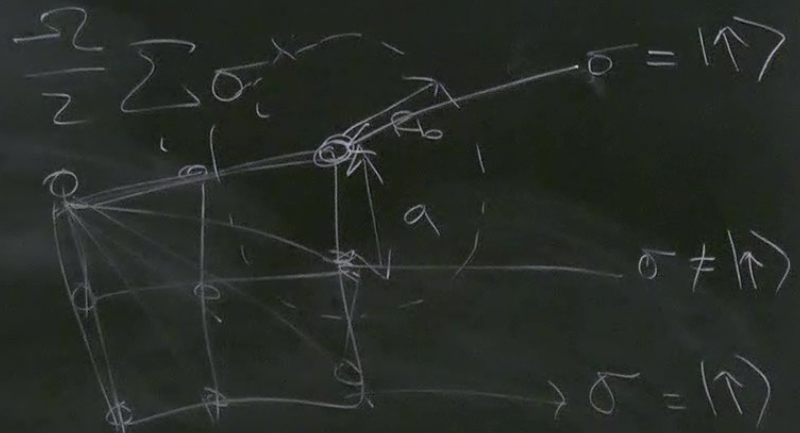
$$\underbrace{P_d(\vec{r}) \ln P_d(\vec{r})}_{\text{CS+ data}} - \underbrace{\sum_{\vec{r}} P_d(\vec{r}) \ln P_d(\vec{r})}_{\text{CS+ data}}$$

Rydberg GPT

$$H = \sum_{i,j} C_{ij} V_{ij} n_i n_j - \delta \sum_i n_i - \frac{\Omega}{2} \sum_i \sigma_i^x$$

$$n_i = \frac{1}{2} (\mathbb{I} + \sigma_i^z)$$

$$C_{ij} = \Omega \left( \frac{R_{ij}}{a} \right)^6, \quad V_{ij} = \frac{1}{|\vec{r}_i - \vec{r}_j|}$$



NLL

$$D = \left( \sum_i c_i h_i, \sum_i W_{ij} \sigma_i h_i \right) = - \sum_{\vec{v}^m} \ln P_2(\vec{v}^m)$$

$$\Rightarrow \operatorname{argmax}_{\lambda} \sum_{\mu} P_2(\vec{v}^{\mu})$$

$\ln \frac{P_1(\vec{v})}{P_2(\vec{v})}$   
 $\underbrace{P_d(\vec{v}) \ln P_d(\vec{v})}_{\text{CST data}} - \sum_{\vec{v}} P_d(\vec{v}) \ln P_2(\vec{v})$

$$\vec{y} = \left\{ \Omega, \frac{\delta}{\Omega}, \frac{R_b}{a}, V_{ij}, \beta \Omega \right\}$$

$$\vec{y} \rightarrow \vec{d} \quad (-1, 4)$$

### Results

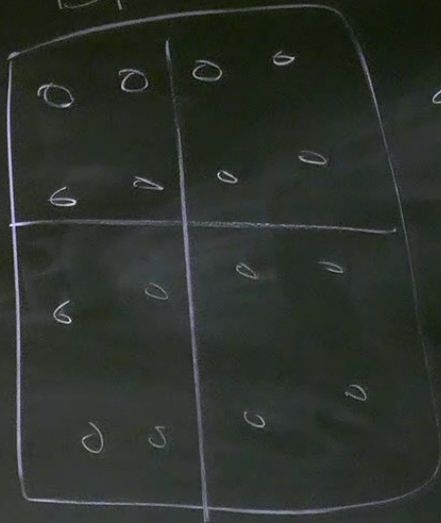
- 1) Intepolste well  $\frac{\delta}{\Omega}$
  - 2) Behave badly for  $\beta \Omega$
  - 3) Does not generalize in size
- $\Rightarrow$  insufficient representatio. state



Renormalization

Group (RG)

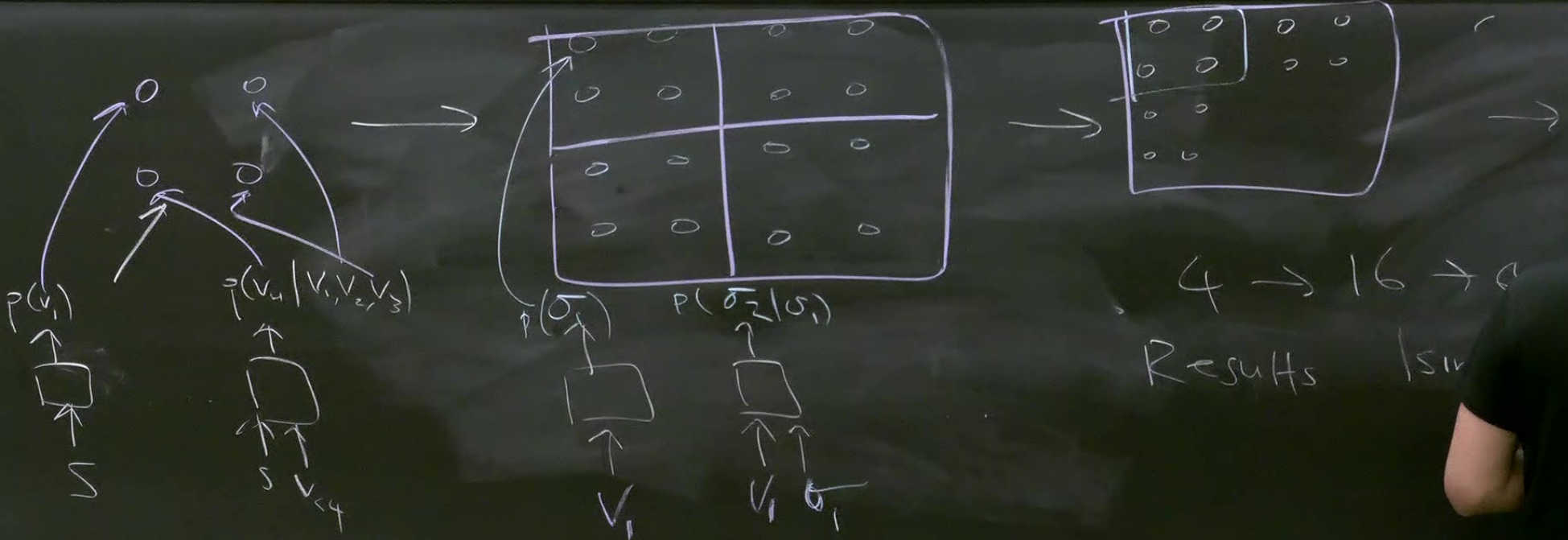
Block-Spin (RG)



decimation



$\{D_0, D\}$  Contrastive Divergence





→ 16 → 64 →

ults using Model Rydberg system  
 Performed well in size scaling ⇒ model trained 8x8 sample 16x16  
 Transfer Learning 100 qubits train 1 epoch 1 patch  
 8x8 → 16x16 → 32x32

Block Spin RG

$\Rightarrow$  Does not work on all sys

$\Rightarrow$  Where does it work well  
( Geometry, Gauge theory, ... )  
Kagome lattice

