Title: Lecture - Statistical Physics, PHYS 602

Speakers: Emilie Huffman

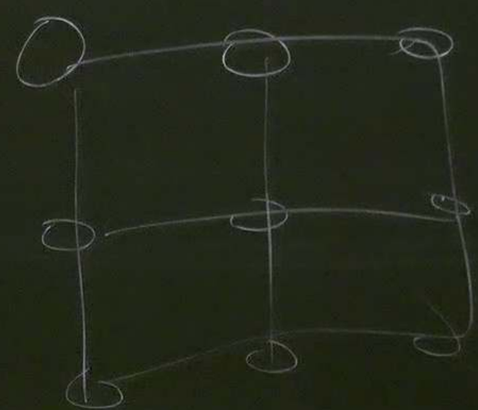
Collection/Series: Statistical Physics (Core), PHYS 602, October 7 - November 6, 2024

Subject: Condensed Matter, Other

Date: November 05, 2024 - 10:45 AM

URL: https://pirsa.org/24110002

System of Interacting Spins



$$H = \sum_{ij} J_{ij}\,\sigma_i \sigma_j$$

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

Access $p(\sigma)$

Sample $p(\sigma)$

Calculate quantities $p(\sigma)$

$\sigma_i \sigma_j$

$H(\sigma)$

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

Access $p(\sigma)$

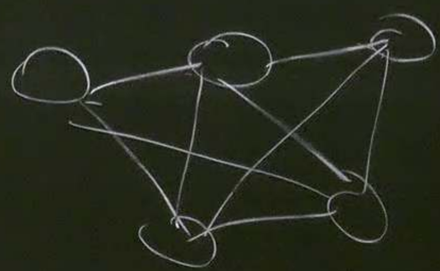
Sample $p(\sigma)$

Calculate quantities $p(\sigma)$

Generative models

Energy-Based Models (EBM)

Hopfield network & Boltzmann Machine



$$E = -\sum_{ij} W_{ij} x_i x_j \,, \quad x_i \in \{-1, 1\}$$

$$\lambda = \{W\}$$

ine

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

$$\underbrace{\{\vec{x}^\mu\}_\mu^N}$$

encode into the low energy states of the HN

$$\arg\min_\lambda \sum_\mu E_\lambda(x^\mu)$$

Hebbian Learning

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

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

choose i, flip $x_i \to -x_i$

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

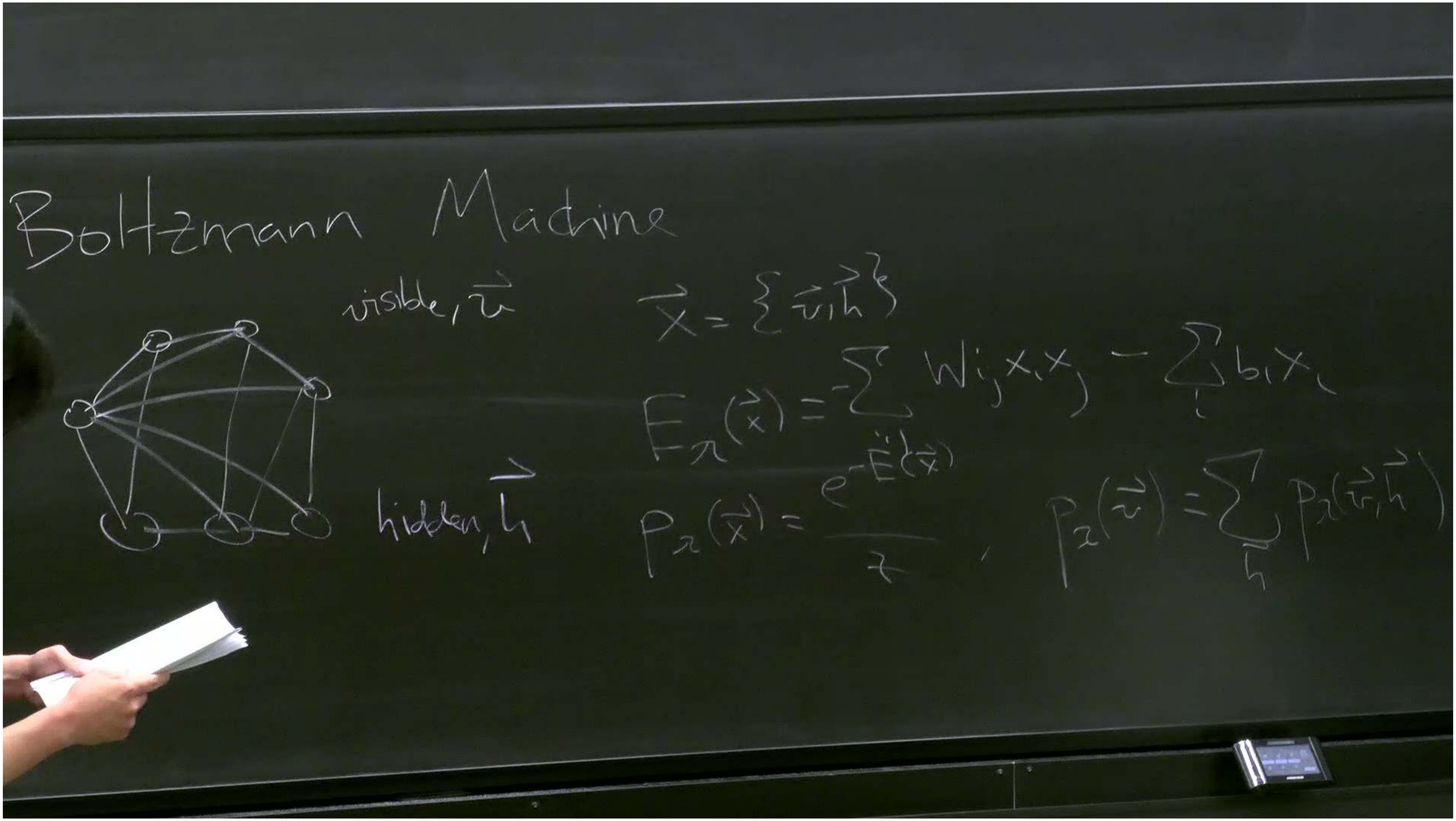
$$\underset{x}{\arg\min}\, E_\lambda(\vec{x})$$

Pr

Problem

Stuck local minima

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


Boltzmann Machine
visible, $\vec{v}$
hidden, $\vec{h}$
$\vec{x} = \{\vec{v}, \vec{h}\}$
$E_\lambda(\vec{x}) = -\sum W_{ij} x_i x_j - \sum_i b_i x_i$
$p_\lambda(\vec{x}) = \frac{e^{-E_\lambda(\vec{x})}}{Z}$, $p_\lambda(\vec{v}) = \sum_{\vec{h}} p_\lambda(\vec{v}, \vec{h})$

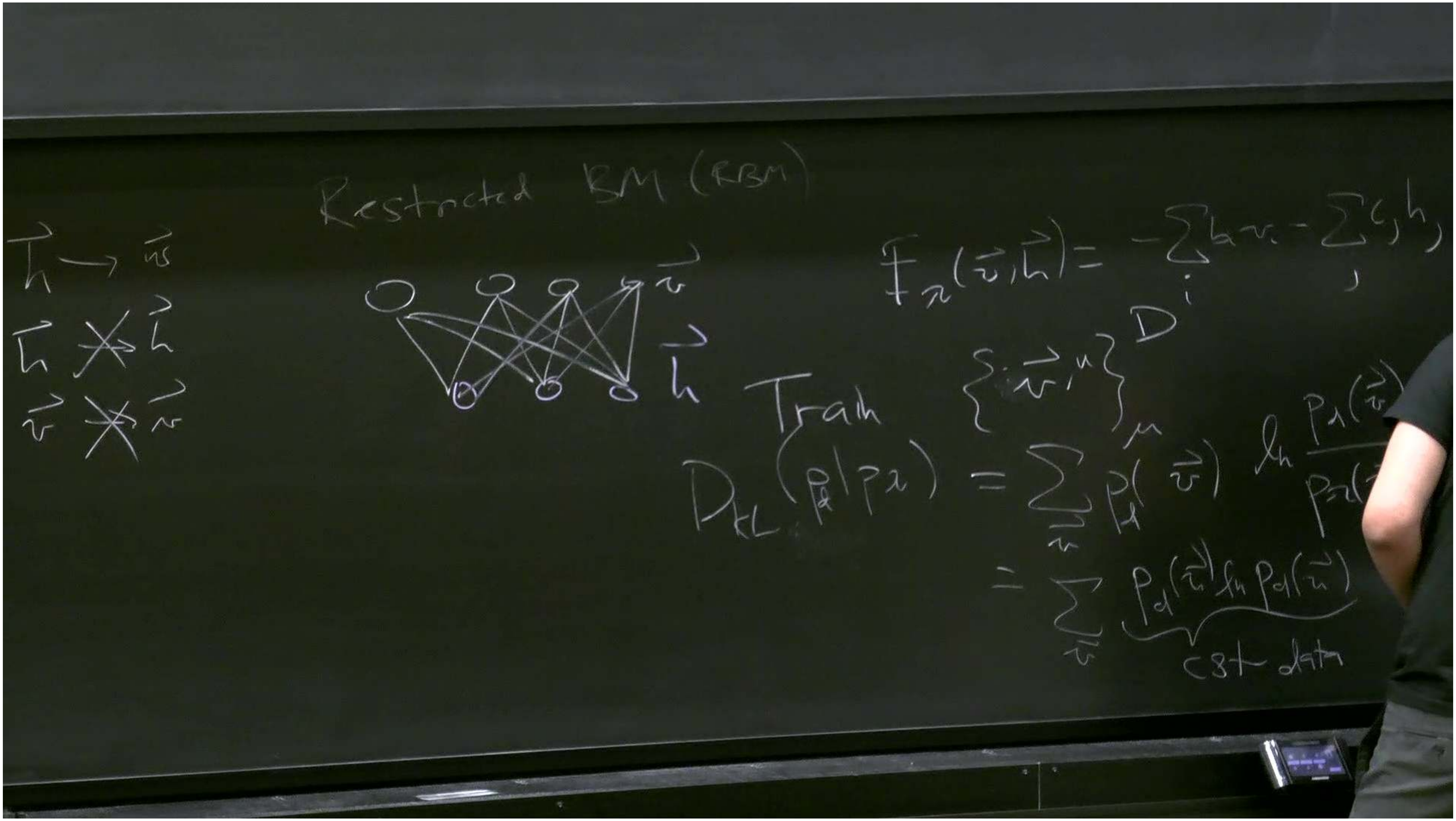
$$- \sum_i b_i x_i$$

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

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

$$\underset{x}{\operatorname{argmax}} \sum_{\mu} p_x(\vec{v}^{\mu})$$

Difficult to train (high interconnectivity $\vec{h}, \vec{v}$; $\vec{h}, \vec{h}$; $\vec{v}, \vec{v}$)



Restricted BM (RBM)
$\vec{h} \to \vec{v}$
$\vec{h} \not\to \vec{h}$
$\vec{v} \not\to \vec{v}$
$\vec{v}$
$\vec{h}$
$F_\lambda(\vec{v},\vec{h}) = -\sum_i b_i v_i - \sum_j c_j h_j$
Train $\{\vec{v}^\mu\}_\mu^D$
$D_{KL}(p_d | p_\lambda) = \sum_{\vec{v}} p_d(\vec{v}) \ln \frac{p_\lambda(\vec{v})}{p_\lambda(\vec{v})}$
$= \sum_{\vec{v}} \underbrace{p_d(\vec{v}) \ln p_d(\vec{v})}_{\text{cst data}}$

$$-\sum_i b_i v_i - \sum_j c_j h_j - \sum_{ij} W_{ij} v_i h_j$$

D

$$\{\ \}_\mu$$

$$p_d(\vec{v}) \ln \frac{p_d(\vec{v})}{p_\lambda(\vec{v})}$$

$$\underbrace{p_d(\vec{v}) \ln p_d(\vec{v})}_{\text{cst data}} - \underbrace{\sum_{\vec{v}} p_d(\vec{v}) \ln p_\lambda(\vec{v})}$$

NLL

$$= -\sum_{\{\vec{v}^\mu\}} \ln p_\lambda(\vec{v}^\mu)$$

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

Training Gradient Descent

$$\mathcal{L}_{\lambda}(\vec{x})$$

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

reduce $\mathcal{L}$

$$\frac{\partial NLL}{\partial \lambda} = \frac{\partial}{\partial}$$

$$\frac{\partial NLL}{\partial x} = \frac{\partial}{\partial x}\Big[\underbrace{\ln \sum_h e^{-E_x(\vec{v},h)}}_{+ve} \; \underbrace{- \ln Z}_{-ve}\Big]$$

tractable

$$-\sum_h p(h|v)\frac{\partial E_x}{\partial x}$$

$$\sum_{\vec{v}} \,,\; 2^N \Rightarrow \text{intractable}$$

MCMC

$$D = \{\sigma\} \xrightarrow{\text{MC step}} D' = \{\sigma'\} \longrightarrow \cdots \xrightarrow{\rho} \text{St}$$

$$D_0 = \{\vec{v}^{\,\mu}\}_\mu \xrightarrow[\text{based model}]{\text{MC step}} D' = \{\vec{v}^{\,\prime\mu}\}_\mu$$

$\{D_0, D'\}_\mu \Rightarrow$ RBM dist = data dist

Contrastive-Divergence

Contrastive Divergence

Sample RBM

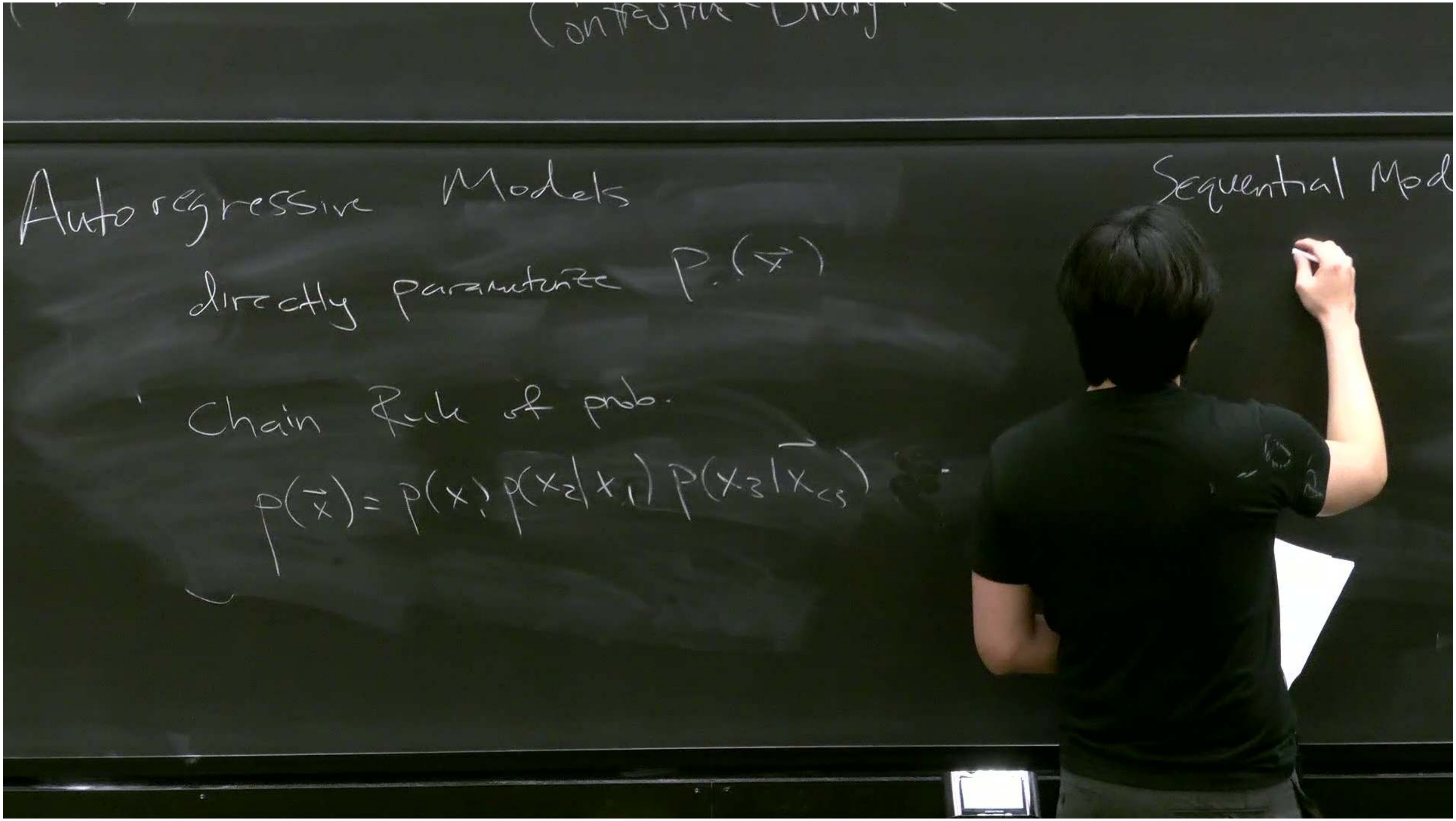
MCMC

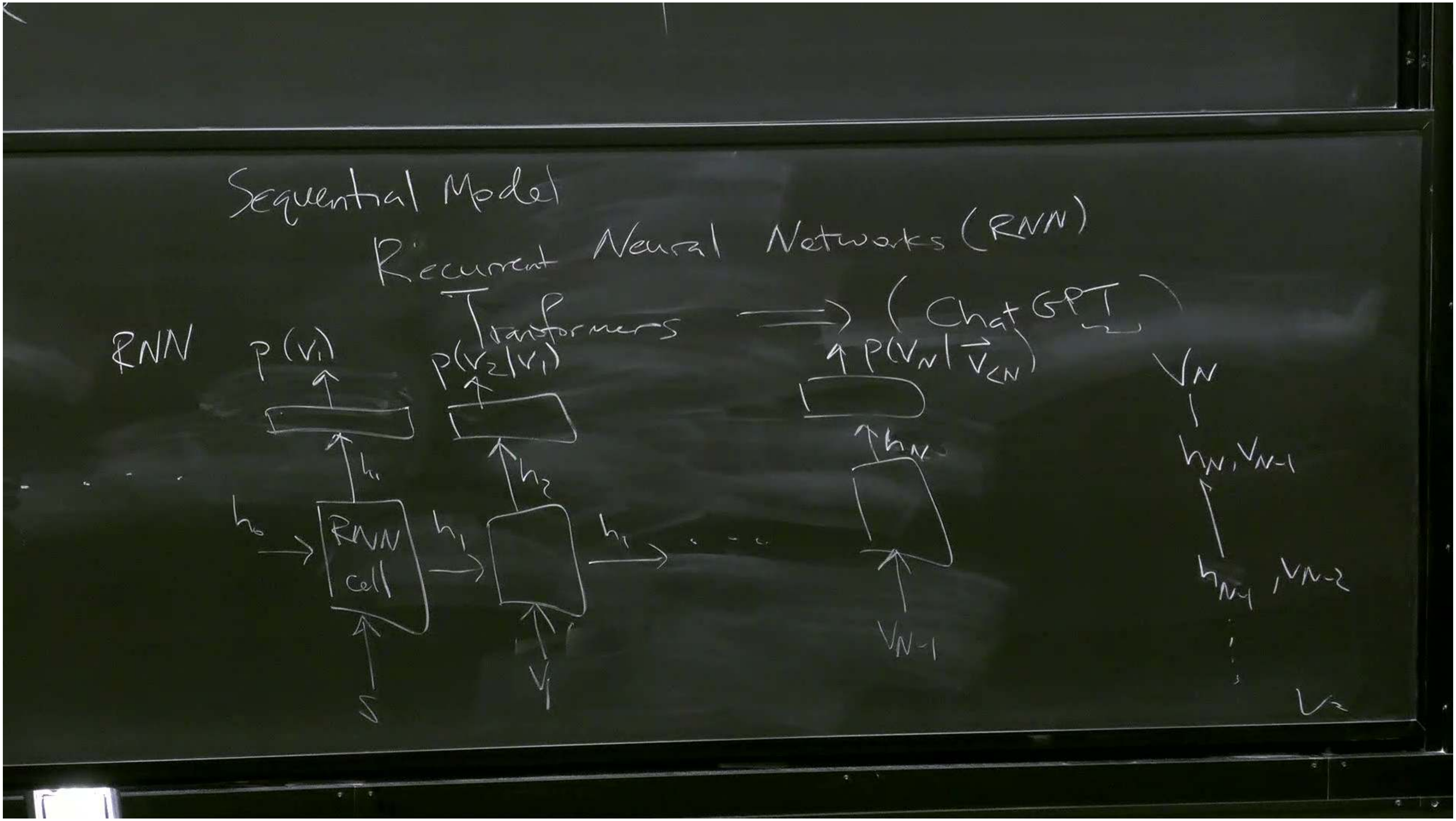
$$\vec{v}^{(0)} \longrightarrow \vec{h}^{(0)} \longrightarrow \vec{v}^{(1)} \longrightarrow \vec{h}^{(1)} \rightarrow \cdots \rightarrow \vec{v}^{(k)}$$

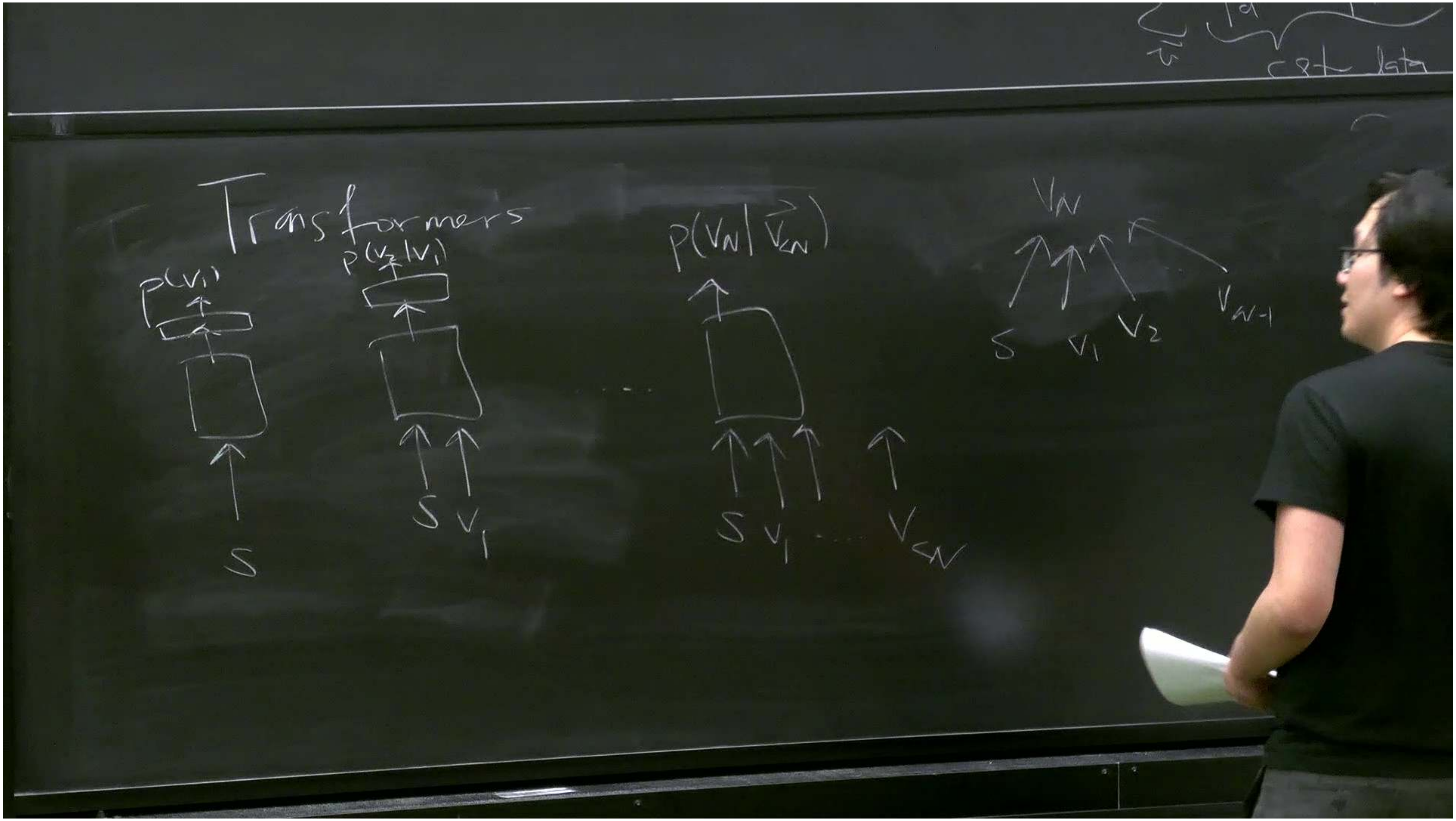
K-block Gibbs Sampling

- Many equil. steps

- Autocorrelation


Autoregressive Models
directly parametrize $p(\vec{x})$
Chain Rule of prob.
$p(\vec{x}) = p(x_1)\, p(x_2|x_1)\, p(x_3|\vec{x}_{<3})$
Sequential Mod


Sequential Model
Recurrent Neural Networks (RNN)
Transformers → (ChatGPT)
RNN
$p(v_1)$
$p(v_2|v_1)$
$p(v_N|\vec{v}_{<N})$
h_1
h_2
h_N
h_0
RNN cell
h_1
s
v_1
v_{N-1}
v_N
h_N, v_{N-1}
h_{N-1}, v_{N-2}
v_2


Transformers
p(v1)
S
p(v2|v1)
S v1
p(vN | v<N)
S v1 ..... v<N
vN
S v1 v2 vN-1

$$\underbrace{p_d(\vec{x}) \ln p_d(\vec{x})}_{\text{const data}} - \sum_{x} p_d(\vec{x}) \ln p \ldots$$

$N-1$

Train

No partition $f^n \Leftarrow p_\lambda(\vec{x})$ normalized

Gradient Descent NLL

Sampling

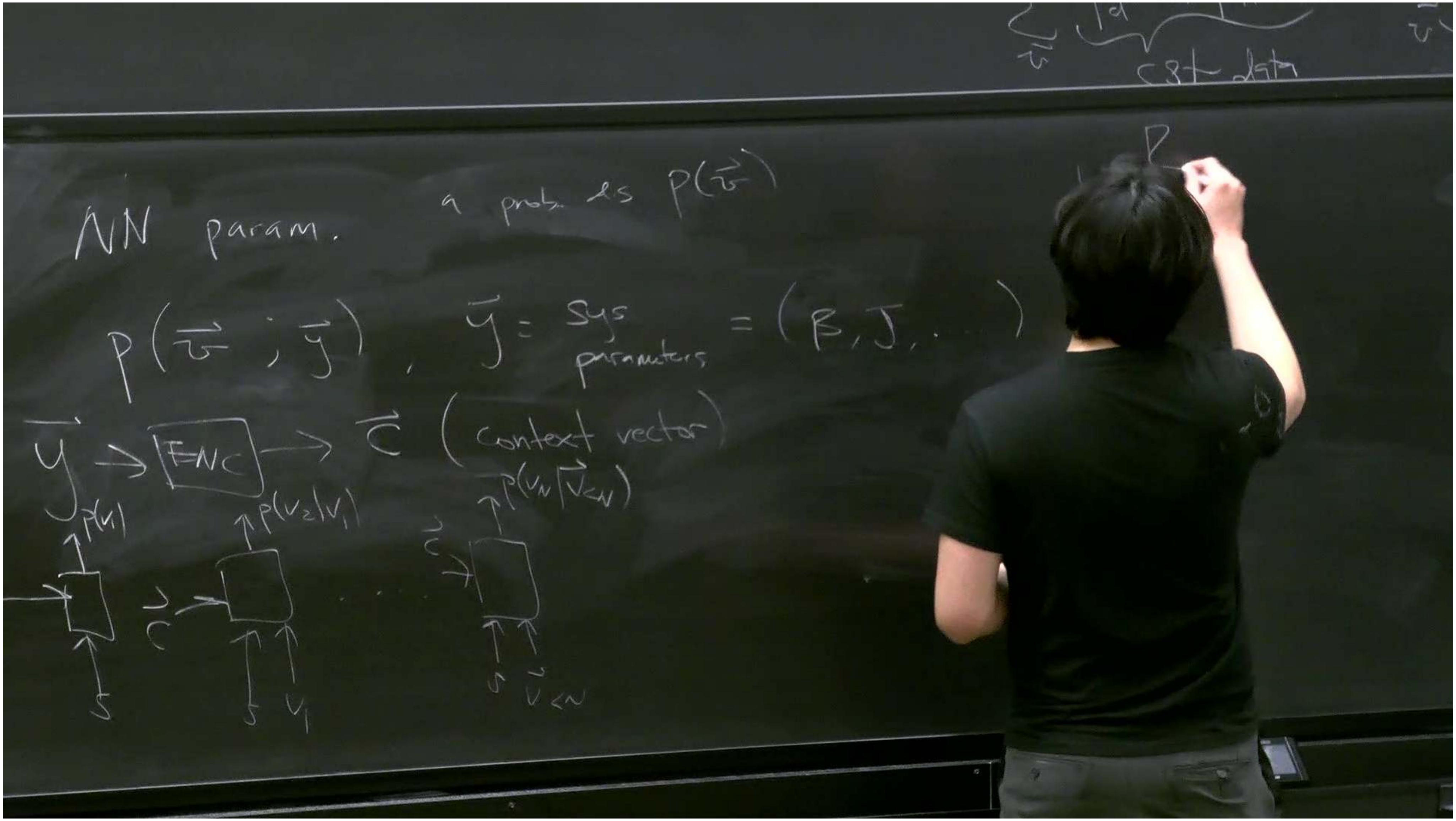
$p_\lambda(v_1)$ $p_\lambda(v_2|v_1)$

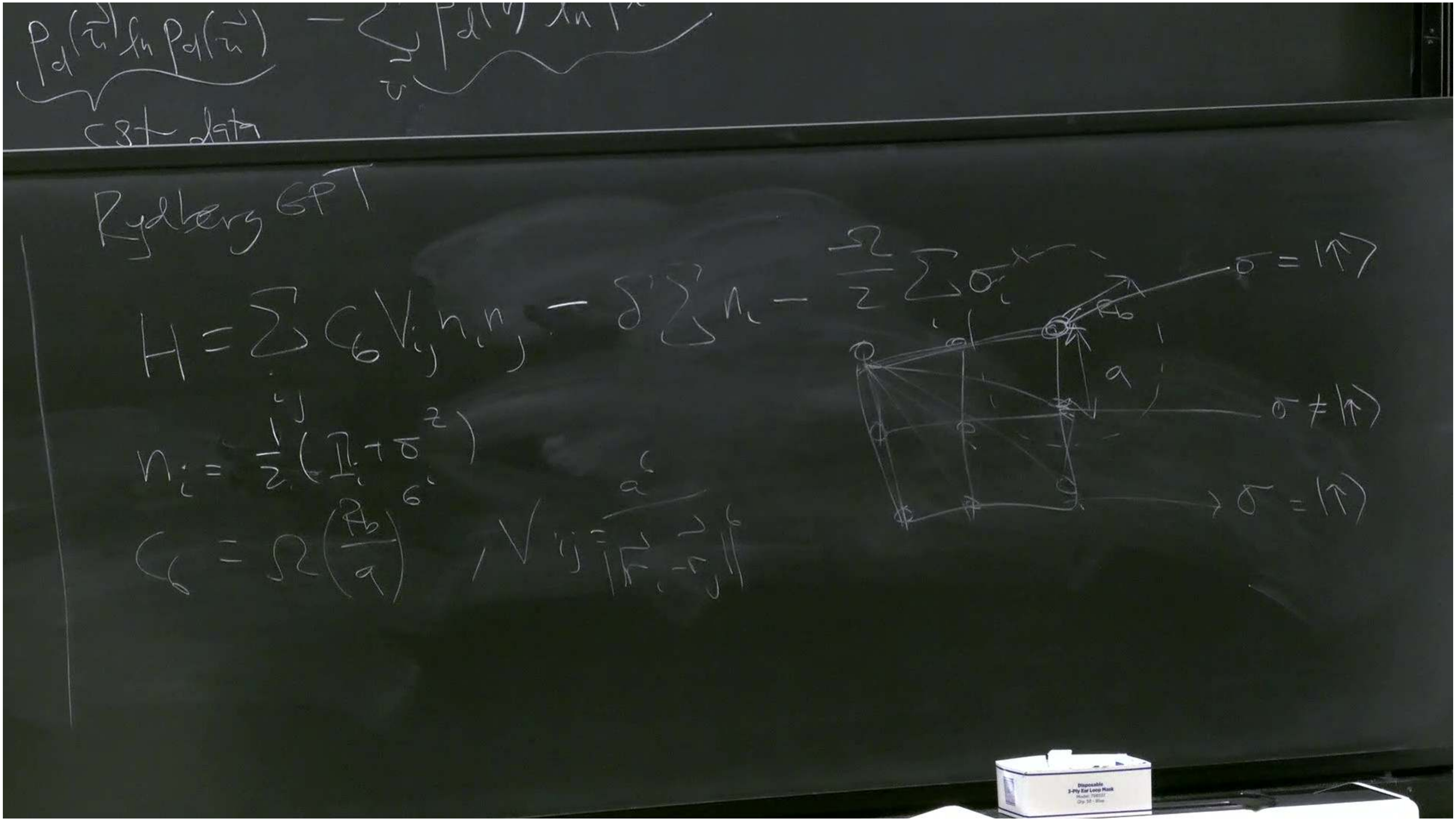
sample

s s, v_1

iid sample

independent & identically distributed


NN param.
a prob. dis p(v)
p(v ; y)
y = sys parameters = (β, J, ...)
ENC
c (context vector)
p(v1)
p(v2|v1)
p(vN|v<N)


$\rho_d(\vec{n}) \ln \rho_d(\vec{n}) - \sum_i$
est data
Rydberg GPT
$H = \sum_{ij} C_6 V_{ij} n_i n_j - \delta \sum_i n_i - \frac{\Omega}{2} \sum_i \sigma_i^x$
$n_i = \frac{1}{2}(\mathbb{I}_i + \sigma_i^z)$
$C_6 = \Omega \left(\frac{R_b}{a}\right)^6, \quad V_{ij} = \frac{a^6}{|\vec{r}_i - \vec{r}_j|^6}$
R_b
a
$\sigma = |\uparrow\rangle$
$\sigma \neq |\uparrow\rangle$
$\sigma = |\uparrow\rangle$

NLL

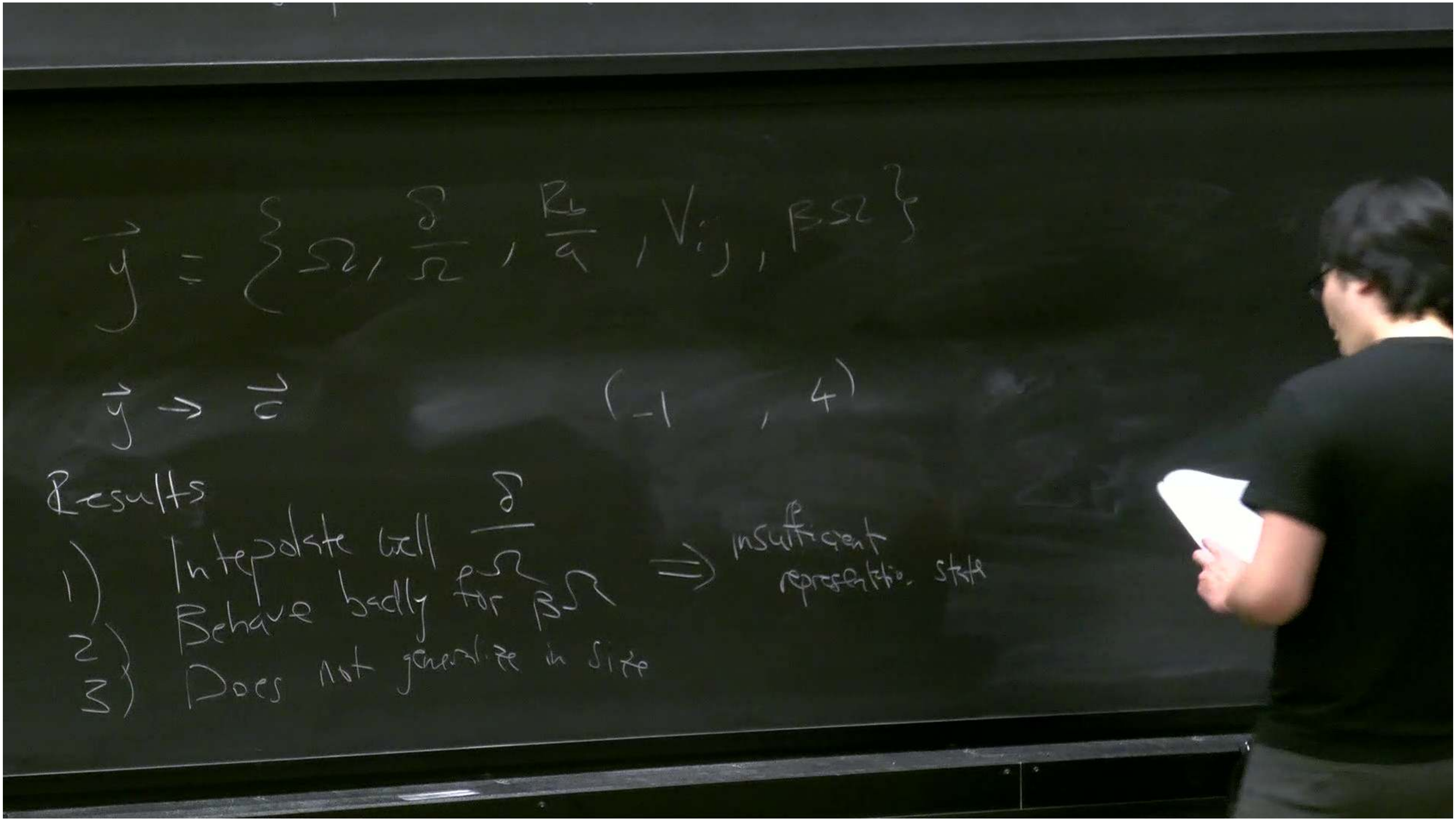
$$\ldots a_i v_i - \sum_j c_j h_j - \sum_{ij} W_{ij} v_i h_j \qquad = -\sum_{\{\vec{v}^\mu\}} \ln p_\lambda(\vec{v}^\mu)$$

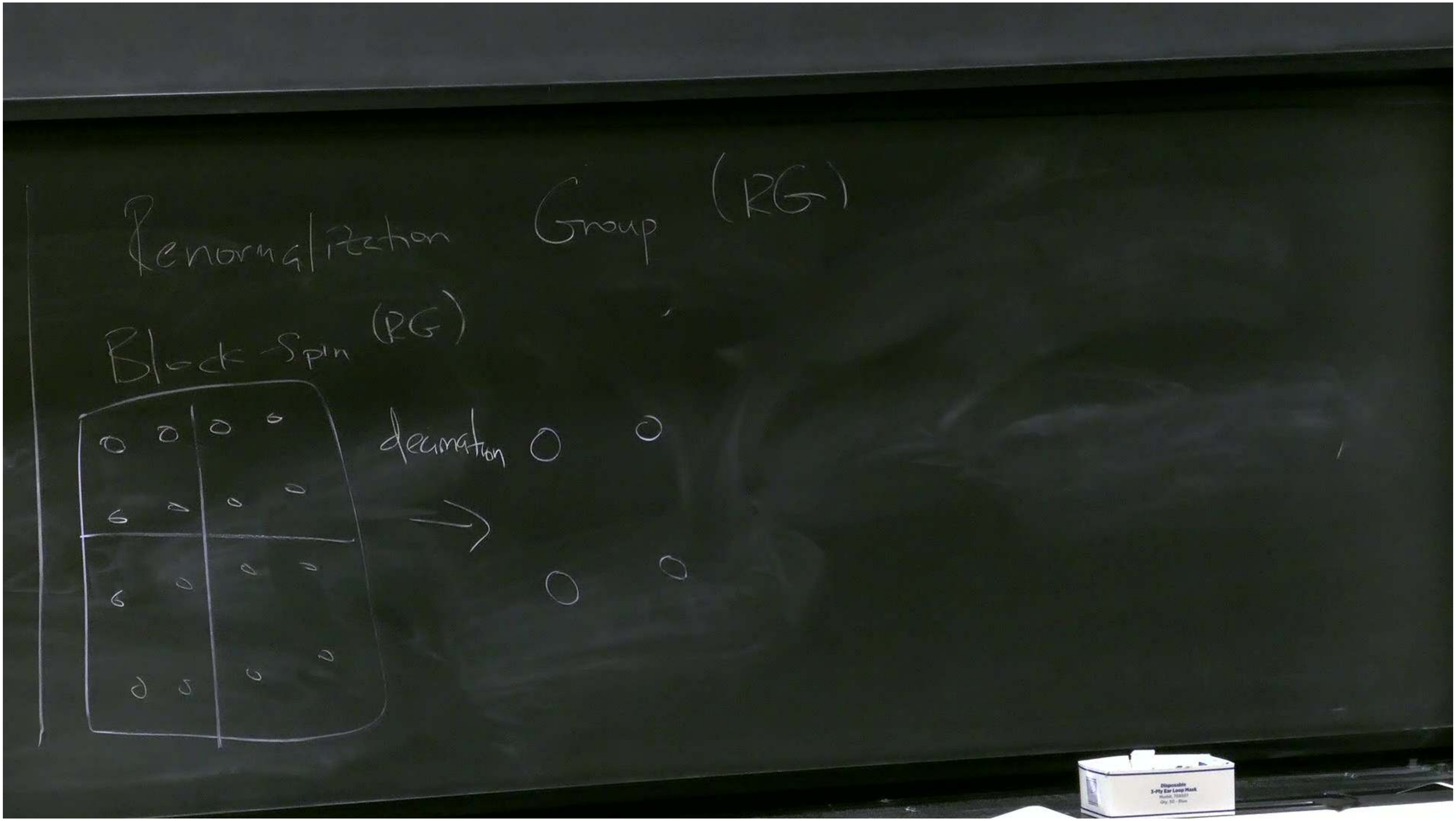
$$D \qquad \Rightarrow \arg\max_\lambda \sum_\mu p_\lambda(\vec{v}^\mu)$$

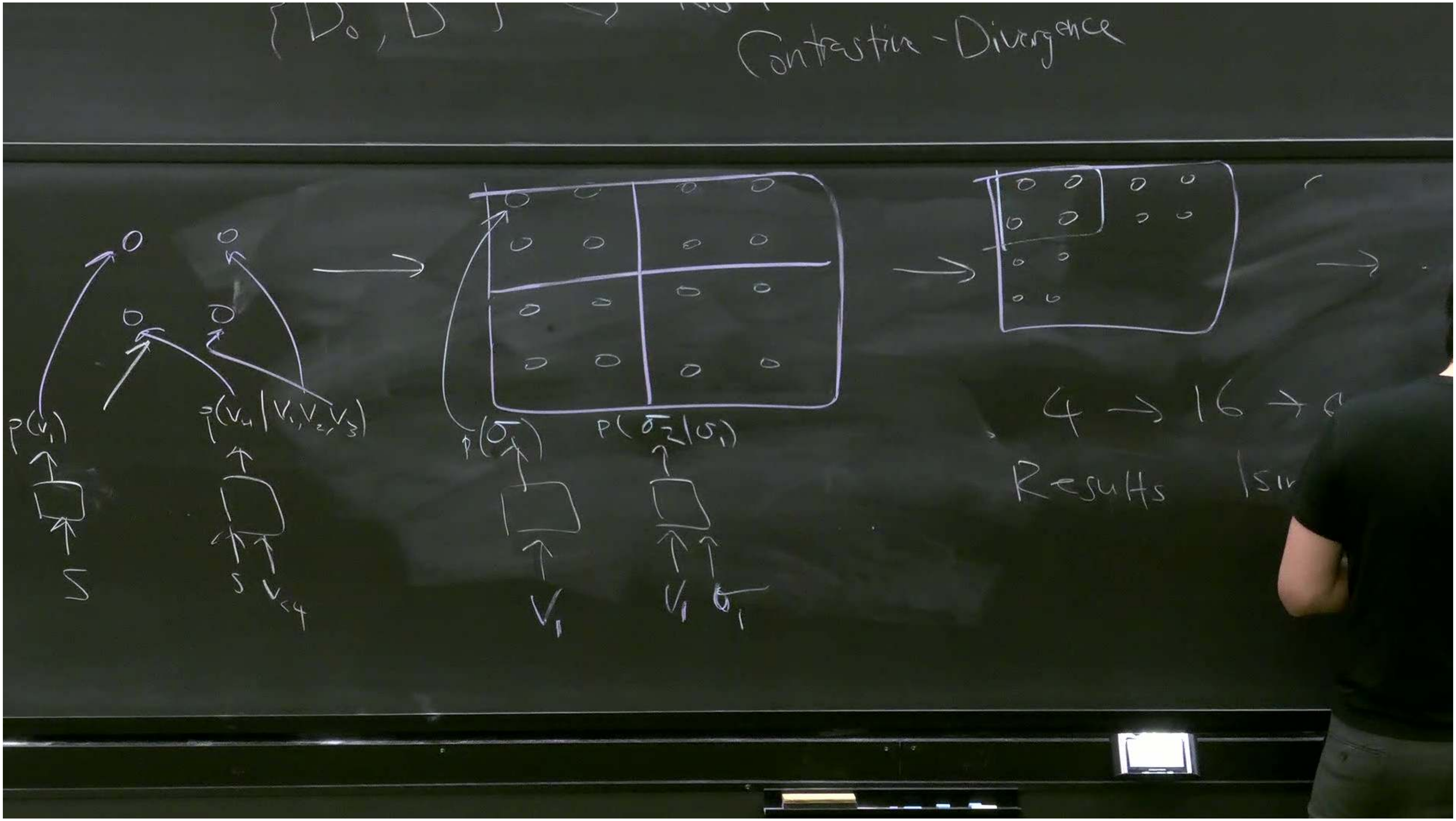
$$\}_\mu$$

$$p_d(\vec{v}) \ln \frac{p_d(\vec{v})}{p_\lambda(\vec{v})}$$

$$\underbrace{p_d(\vec{v}) \ln p_d(\vec{v})}_{\text{cst data}} \underbrace{- \sum_{\vec{v}} p_d(\vec{v}) \ln p_\lambda(\vec{v})}$$


y = { Ω, δ/Ω, Rb/a, Vij, βΩ }
y → c (-1, 4)
Results
1) Interpolate well δ/Ω
2) Behave badly for βΩ ⇒ insufficient representation state
3) Does not generalize in size


Renormalization Group (RG)
Block-Spin (RG)
decimation


Contrastive - Divergence
p(v1)
p(v4 | v1, v2, v3)
S
S v<4
p(σ1)
p(σ2 | σ1)
v1
v1 σ1
4 → 16 →
Results

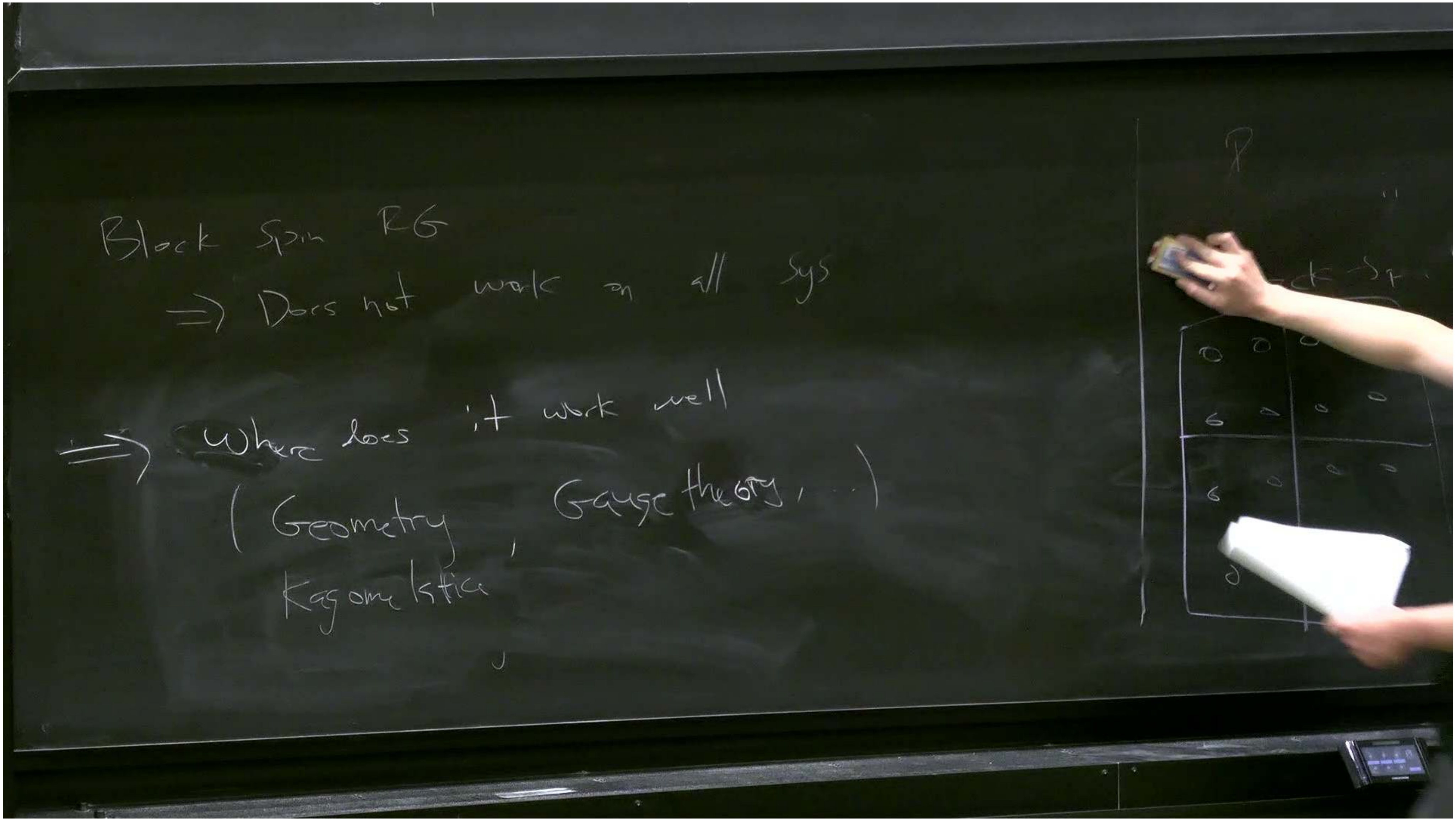
→ 16 → 64 ---- →

ults Ising Model | Rydberg syst model sample

Performed well in size scaling ⟹ trained 8x8 → 16x16

Transfer Learning 100 epochs train 1 epoch → 1 epoch

8x8 → 16x16 → 32x32


Block Spin RG
⇒ Does not work on all sys
⇒ Where does it work well
(Geometry, Gauge theory, ...)
Kagome lattice