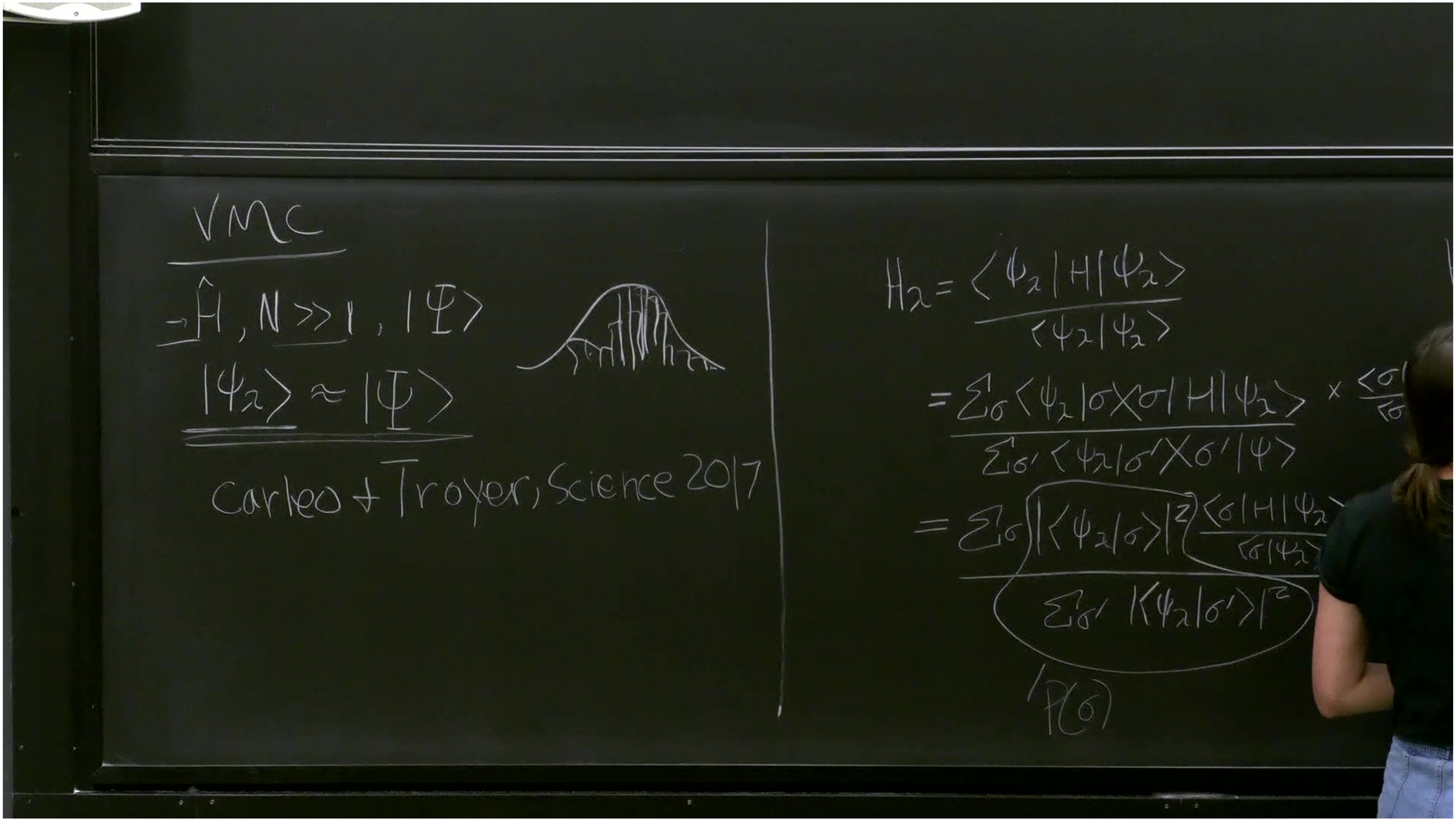
Title: Recurrent Neural Networks (RNNs)

Speakers: Megan Schuyler Moss

Collection: Navigating Quantum and AI Career Trajectories: A Beginnerâ€™s Mini-Course on Computational Methods and their Applications

Date: May 23, 2024 - 11:30 AM

URL: https://pirsa.org/24050086


VMC
$\rightarrow \hat{H},\ N \gg 1,\ |\Psi\rangle$
$|\psi_\lambda\rangle \approx |\Psi\rangle$
Carleo + Troyer, Science 2017
$H_\lambda = \frac{\langle \psi_\lambda | H | \psi_\lambda \rangle}{\langle \psi_\lambda | \psi_\lambda \rangle}$
$= \frac{\sum_\sigma \langle \psi_\lambda | \sigma \rangle \langle \sigma | H | \psi_\lambda \rangle}{\sum_{\sigma'} \langle \psi_\lambda | \sigma' \rangle \langle \sigma' | \psi \rangle} \times$
$= \frac{\sum_\sigma |\langle \psi_\lambda | \sigma \rangle|^2 \frac{\langle \sigma | H | \psi_\lambda \rangle}{\langle \sigma | \psi_\lambda \rangle}}{\sum_{\sigma'} |\langle \psi_\lambda | \sigma' \rangle|^2}$
$P(\sigma)$

$$H_\lambda \cong \frac{1}{N_s} \sum_{\sigma \sim p_\lambda} \frac{\langle\sigma|H|\psi_\lambda\rangle}{\langle\sigma|\psi_\lambda\rangle}$$

$$\boxed{|\langle\sigma|\psi_\lambda\rangle|^2}$$

$\lambda\rangle \times \frac{\langle\sigma|\psi_\lambda\rangle}{\langle\sigma|\psi_\lambda\rangle}$

$\psi\rangle$

$\frac{H|\psi_\lambda\rangle}{\sigma|\psi_\lambda\rangle}$

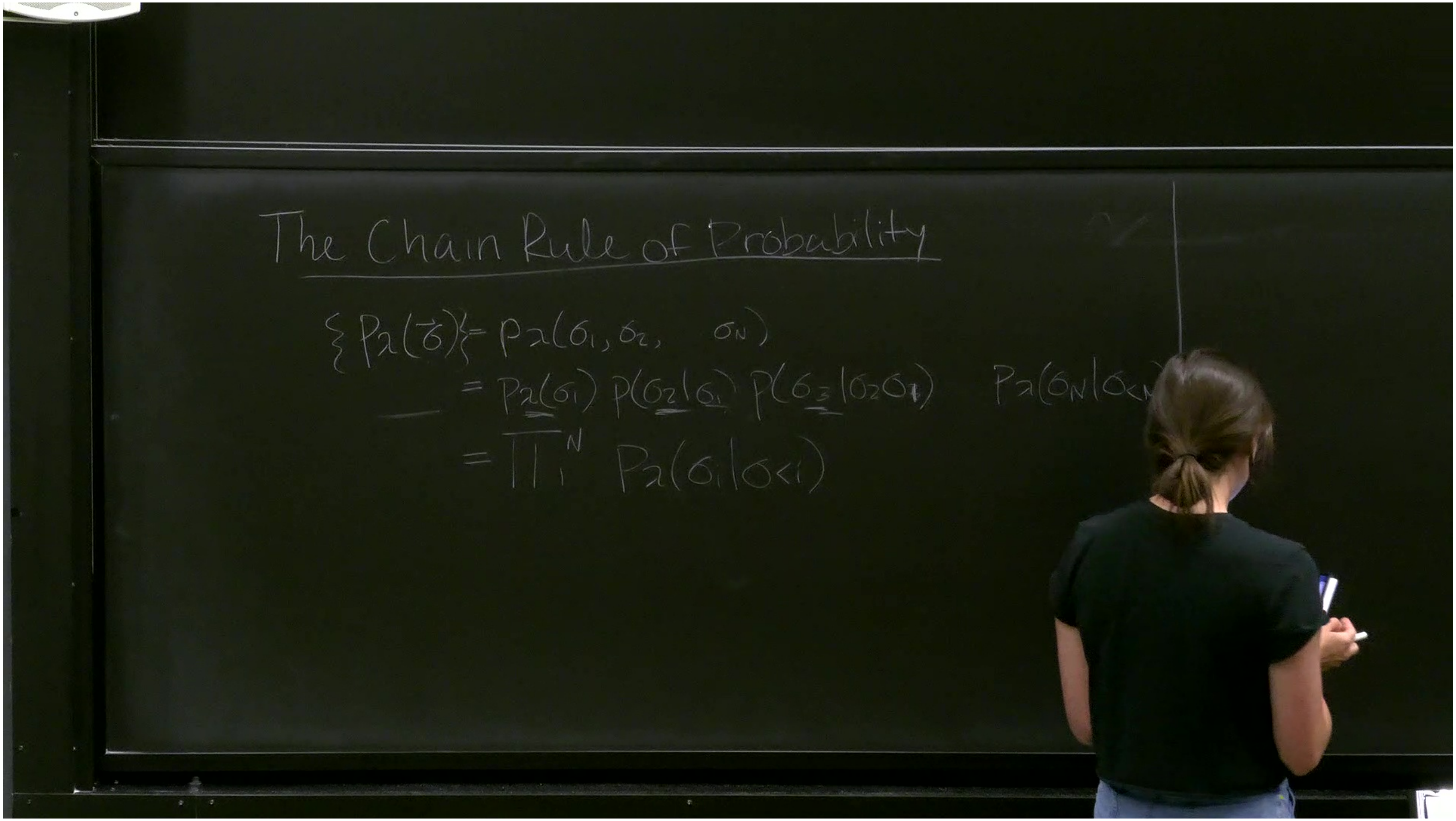
$|^2$

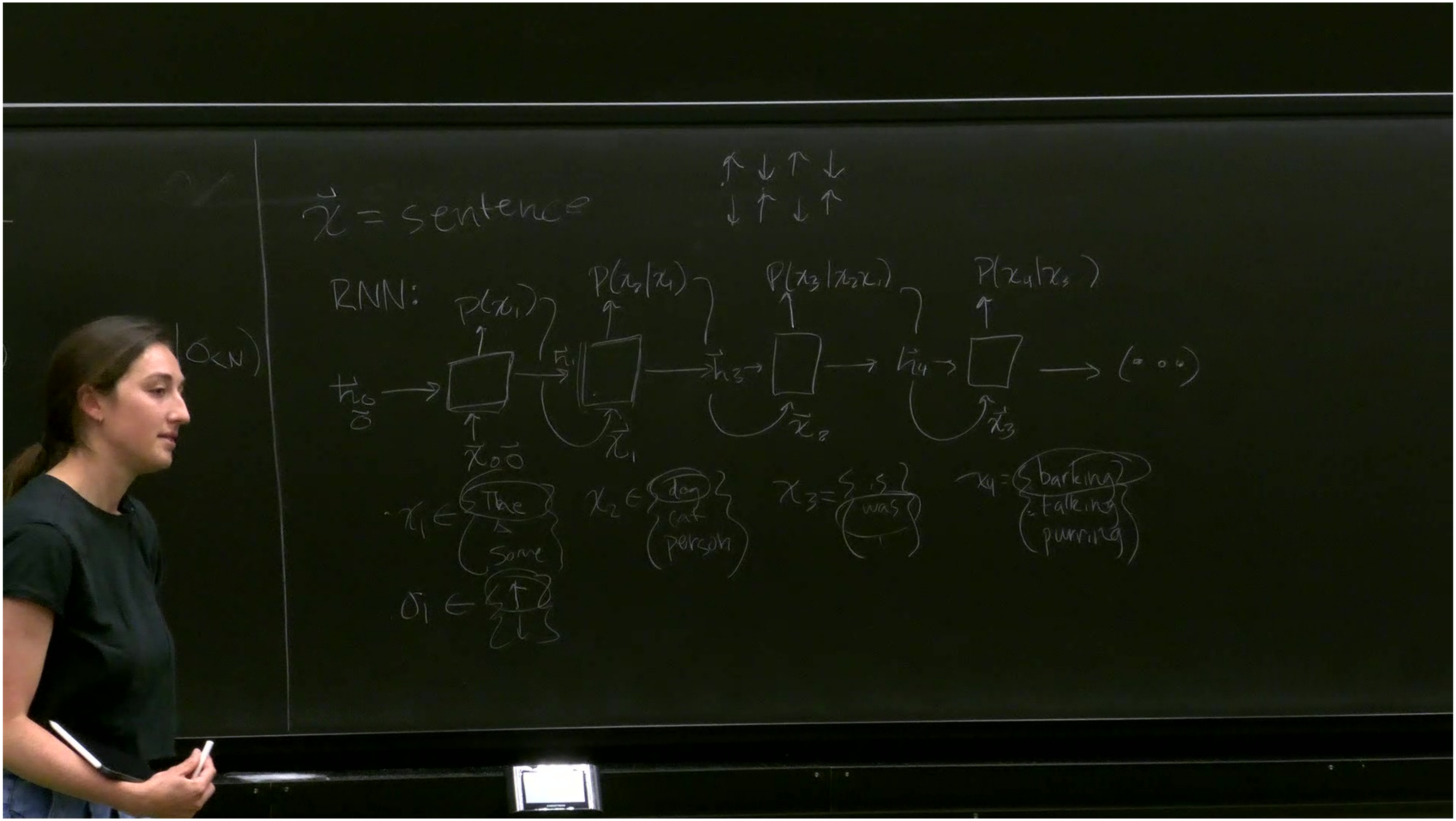
$$\vec{\sigma} = (\sigma_1, \sigma_2, \ \sigma_N)$$

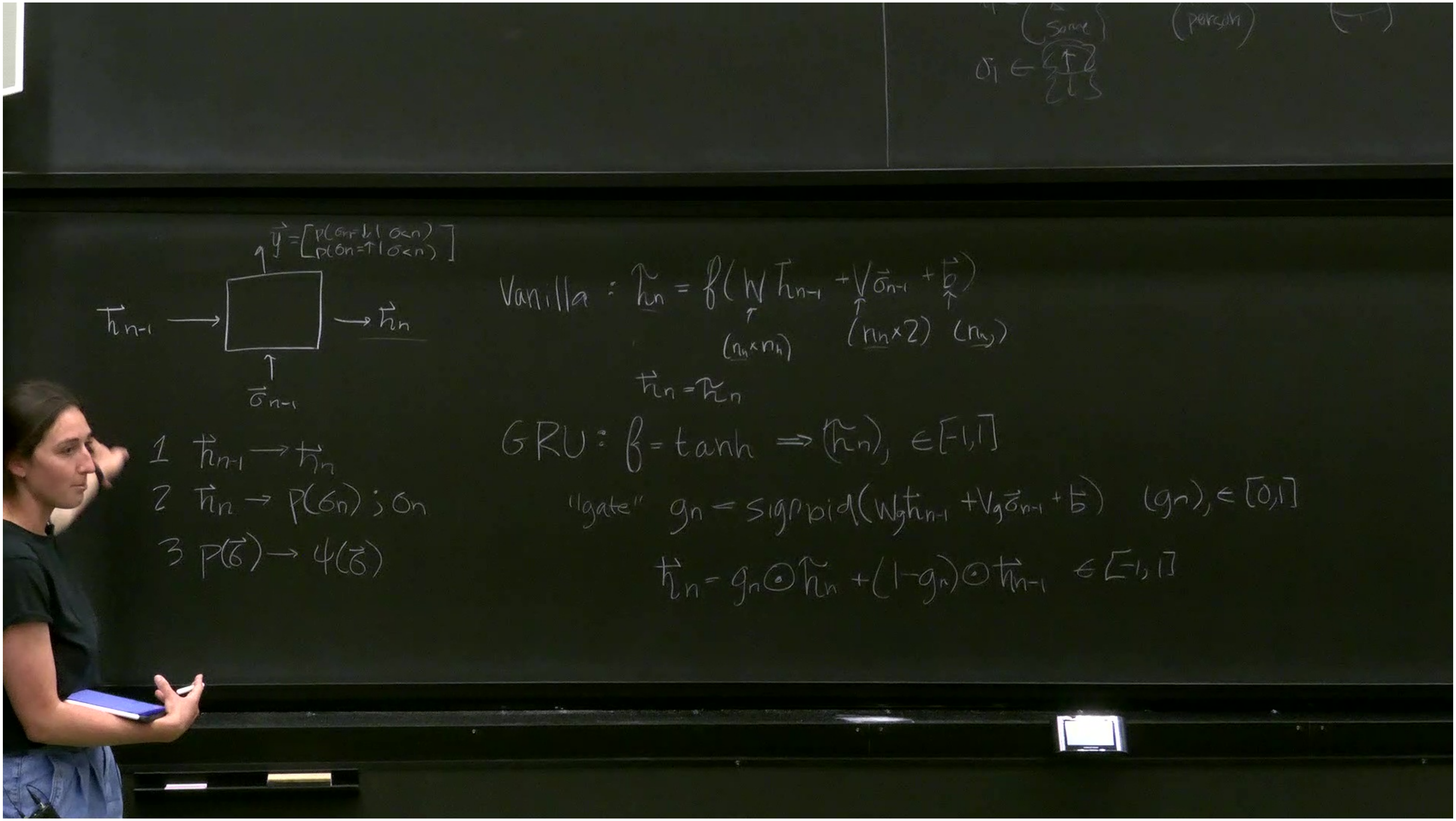
$p_\lambda(\vec{\sigma})$

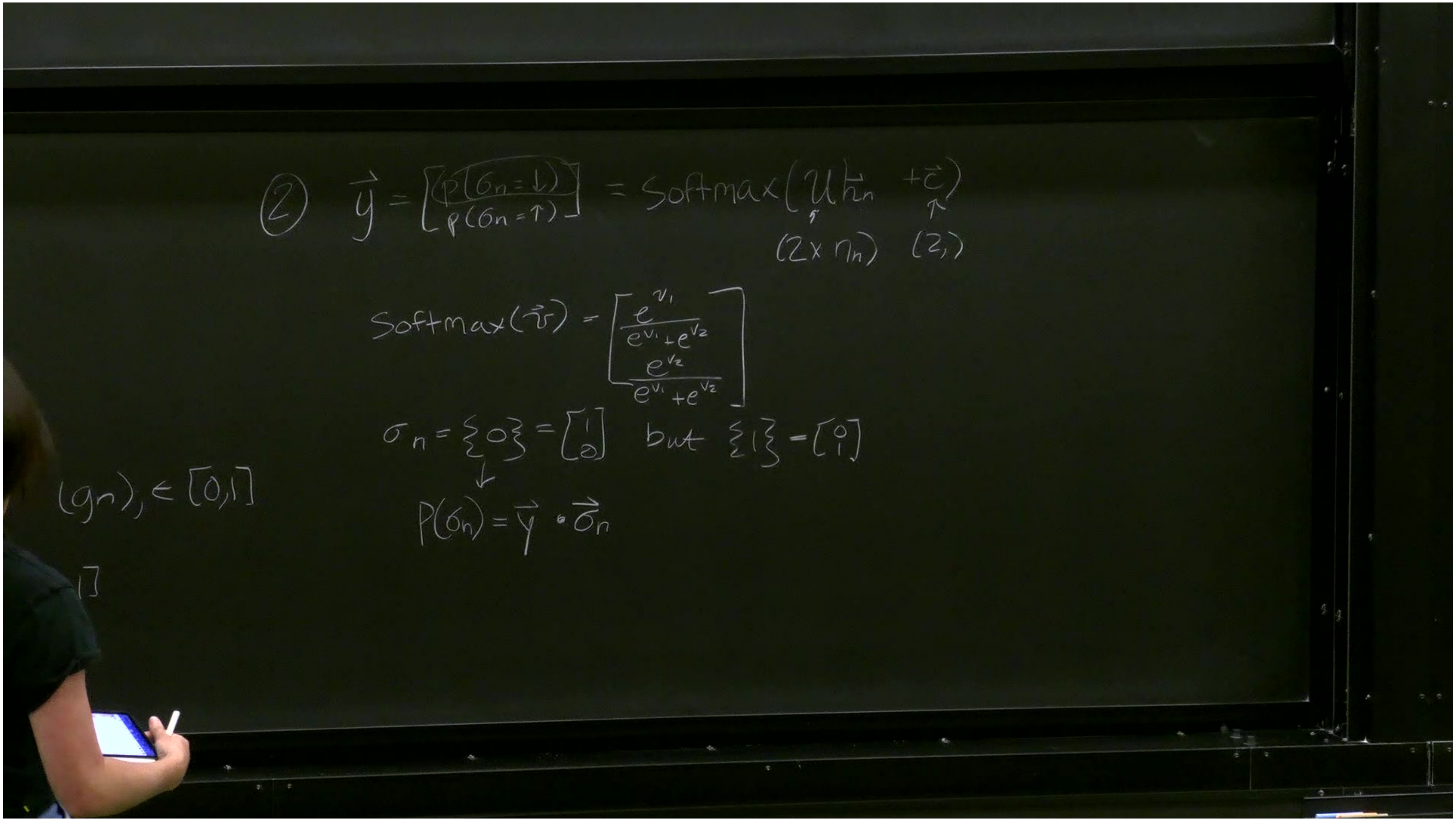
explicit density — directly encodes $p_\lambda(\vec{\sigma})$

intractible	tractable
RBM	RNN
$\tilde{p}_\lambda$ unnorm	p_λ NORMALIZED


The Chain Rule of Probability
$\{p_\lambda(\vec{\sigma})\} = p_\lambda(\sigma_1, \sigma_2, \ldots \sigma_N)$
$= p_\lambda(\sigma_1)\, p(\sigma_2|\sigma_1)\, p(\sigma_3|\sigma_2\sigma_1) \ldots p_\lambda(\sigma_N|\sigma_{<N})$
$= \prod_i^N p_\lambda(\sigma_i|\sigma_{<i})$


x = sentence
RNN:
p(x1)
P(x2|x1)
P(x3|x2x1)
P(x4|x3)
The
Some
dog
cat
person
is
was
barking
talking
purring


Some
person
$\sigma_1 \in \{\uparrow, \downarrow\}$
$\vec{y} = \begin{bmatrix} p(\sigma_n = \downarrow \mid \sigma_{<n}) \\ p(\sigma_n = \uparrow \mid \sigma_{<n}) \end{bmatrix}$
$\vec{h}_{n-1} \longrightarrow \boxed{} \longrightarrow \vec{h}_n$
$\vec{\sigma}_{n-1}$
Vanilla : $\tilde{h}_n = f(W \vec{h}_{n-1} + V \vec{\sigma}_{n-1} + \vec{b})$
$(n_h \times n_h)$ $(n_h \times 2)$ (n_h)
$\vec{h}_n = \tilde{h}_n$
1 $\vec{h}_{n-1} \longrightarrow \vec{h}_n$
2 $\vec{h}_n \rightarrow p(\sigma_n)$; σ_n
3 $p(\vec{\sigma}) \rightarrow \psi(\vec{\sigma})$
GRU : $f = \tanh \Rightarrow (\tilde{h}_n)_i \in [-1,1]$
"gate" $g_n = \mathrm{sigmoid}(W_g \vec{h}_{n-1} + V_g \vec{\sigma}_{n-1} + \vec{b})$ $(g_n)_i \in [0,1]$
$\vec{h}_n = g_n \odot \tilde{h}_n + (1 - g_n) \odot \vec{h}_{n-1} \in [-1,1]$



② $\vec{y} = \begin{bmatrix} p(\sigma_n=\downarrow) \\ p(\sigma_n=\uparrow) \end{bmatrix} = \mathrm{Softmax}(U\vec{h}_n + \vec{c})$
$(2\times n_h)$ $(2,)$
$\mathrm{Softmax}(\vec{v}) = \begin{bmatrix} \frac{e^{v_1}}{e^{v_1}+e^{v_2}} \\ \frac{e^{v_2}}{e^{v_1}+e^{v_2}} \end{bmatrix}$
$\sigma_n = \{0\} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ but $\{1\} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
$P(\sigma_n) = \vec{y} \cdot \vec{\sigma}_n$
$(g_n)_i \in [0,1]$

$$\text{Softmax}(\vec{v}) = \begin{bmatrix} \dfrac{e^{v_1}}{e^{v_1}+e^{v_2}} \\ \dfrac{e^{v_2}}{e^{v_1}+e^{v_2}} \end{bmatrix}$$

$$\sigma_n = \{0\} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{but} \quad \{1\} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$(y_n)_i \in [0,1]$$

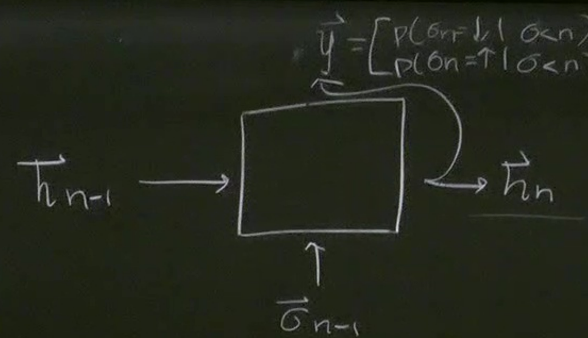
$$\downarrow$$

$$P(\sigma_n) = \vec{y} \cdot \vec{\sigma}_n$$

$$[-1,1]$$

$$3 \quad P(\sigma_n | \sigma_{<n}) \ \forall n \longrightarrow P_\lambda(\vec{\sigma}) = \prod_i^N p_\lambda(\sigma_i | \sigma_{<i})$$

$$\psi_\lambda = \sqrt{P_\lambda(\vec{\sigma})}$$



① $\vec{h}_{n-1} \to \vec{h}_n$

2 $\vec{h}_n \to p(\sigma_n)$; σ_n

3 $p(\vec{\sigma}) \to \psi(\vec{\sigma})$

① Vanilla : $\tilde{h}_n = f(\underset{(n_h \times n_h)}{W} \vec{h}_{n-1} + \underset{(n_h \times 2)}{V} \vec{\sigma}_{n-1} + \underset{(n_h)}{\vec{b}})$

$\vec{h}_n = \tilde{h}_n$

GRU : $f = \tanh \Rightarrow (\tilde{h}_n)_i \in [-1,1]$

"gate" $g_n = \mathrm{sigmoid}(W_g \vec{h}_{n-1} + V_g \vec{\sigma}_{n-1} + b$

$\vec{h}_n = g_n \odot \tilde{h}_n + (1-g_n) \odot \vec{h}_{n-1}$

$$\mathrm{Softmax}(\vec{v}) = \begin{bmatrix} \dfrac{e^{v_1}}{e^{v_1}+e^{v_2}} \\ \dfrac{e^{v_2}}{e^{v_1}+e^{v_2}} \end{bmatrix}$$

$$\sigma_n = \{0\} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{but} \quad \{1\} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\downarrow$$

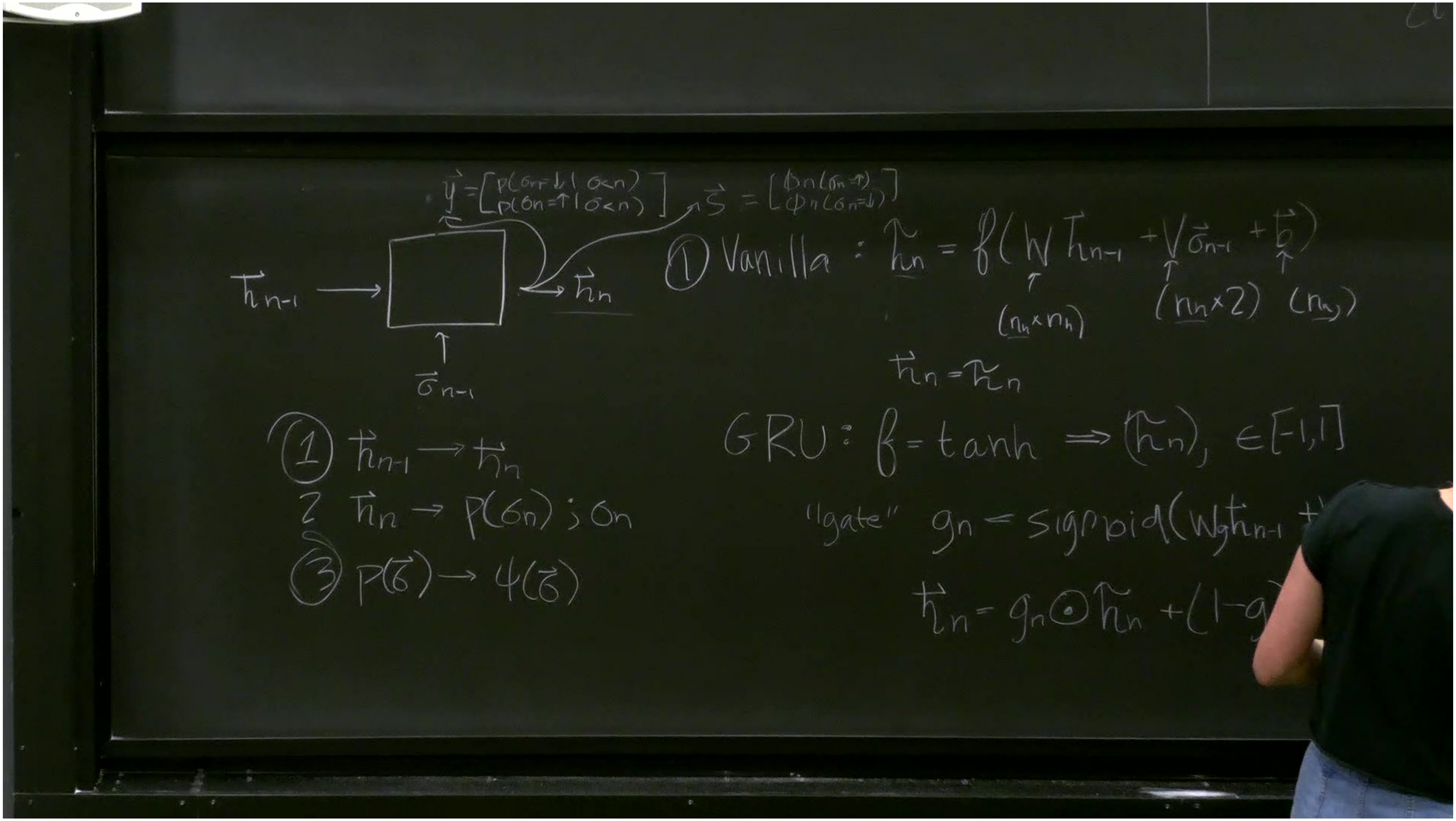
$(y_n)_i \in [0,1]$

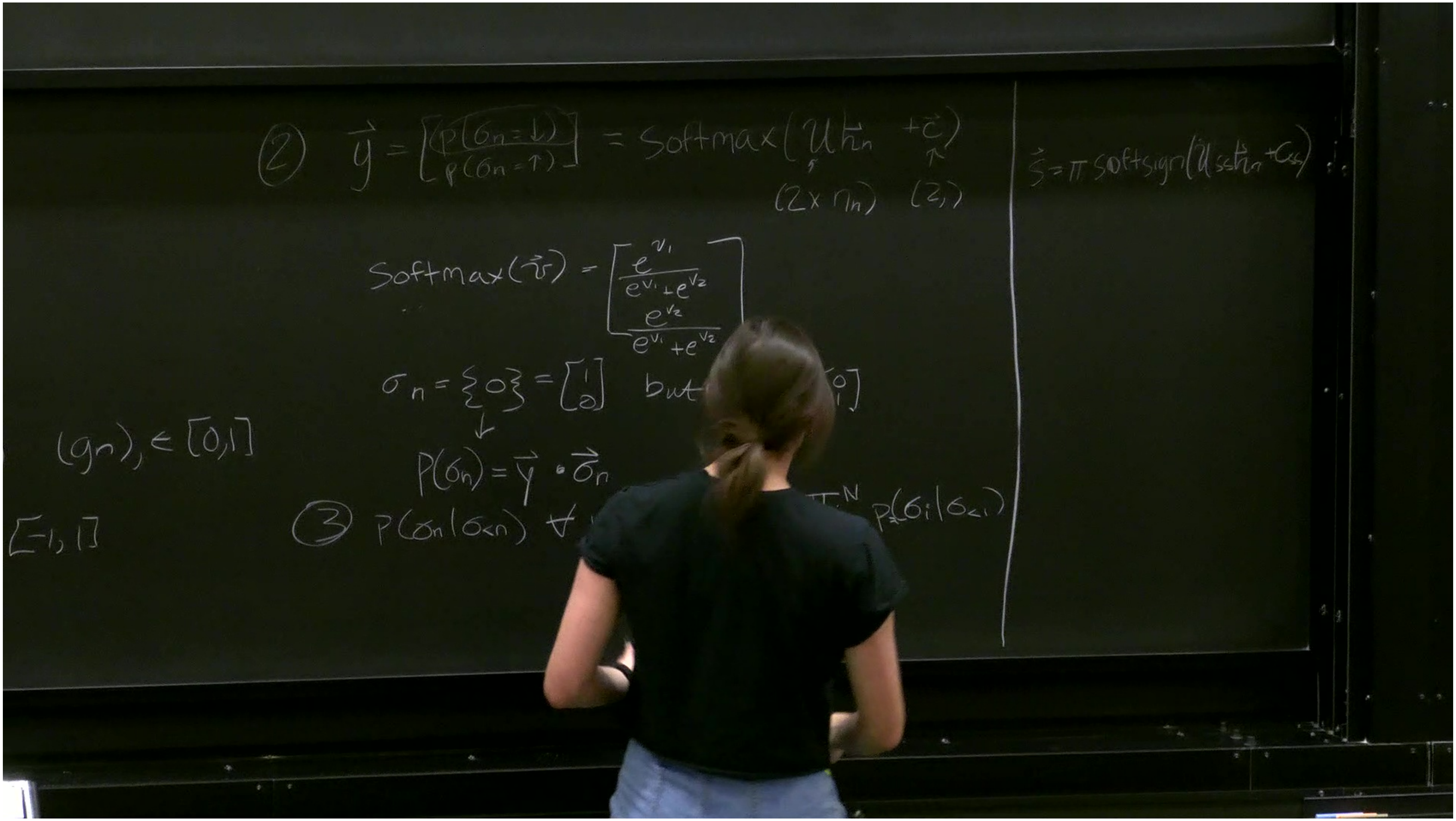
$$P(\sigma_n) = \vec{y} \cdot \vec{\sigma}_n$$

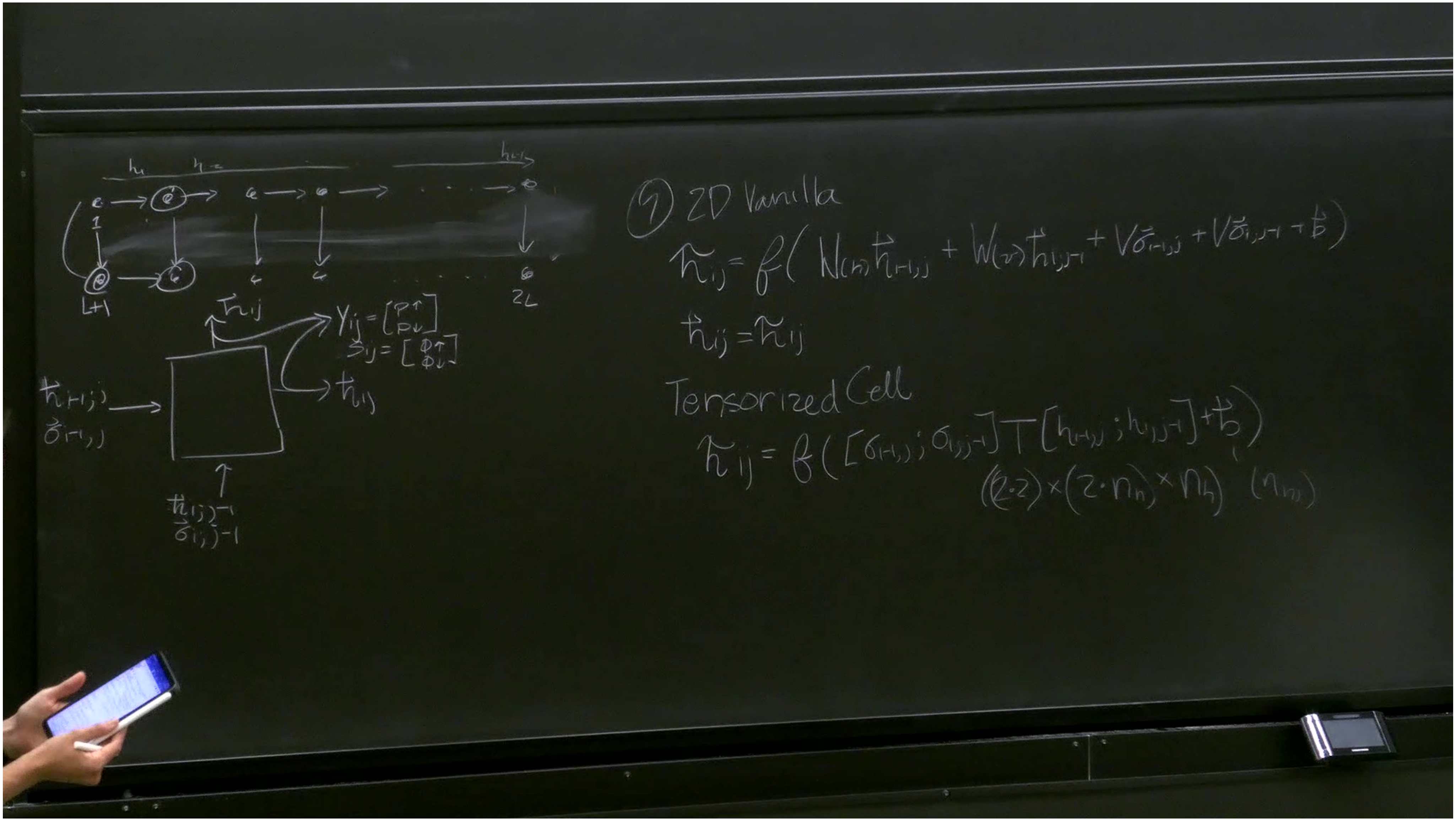
$[-1, 1]$

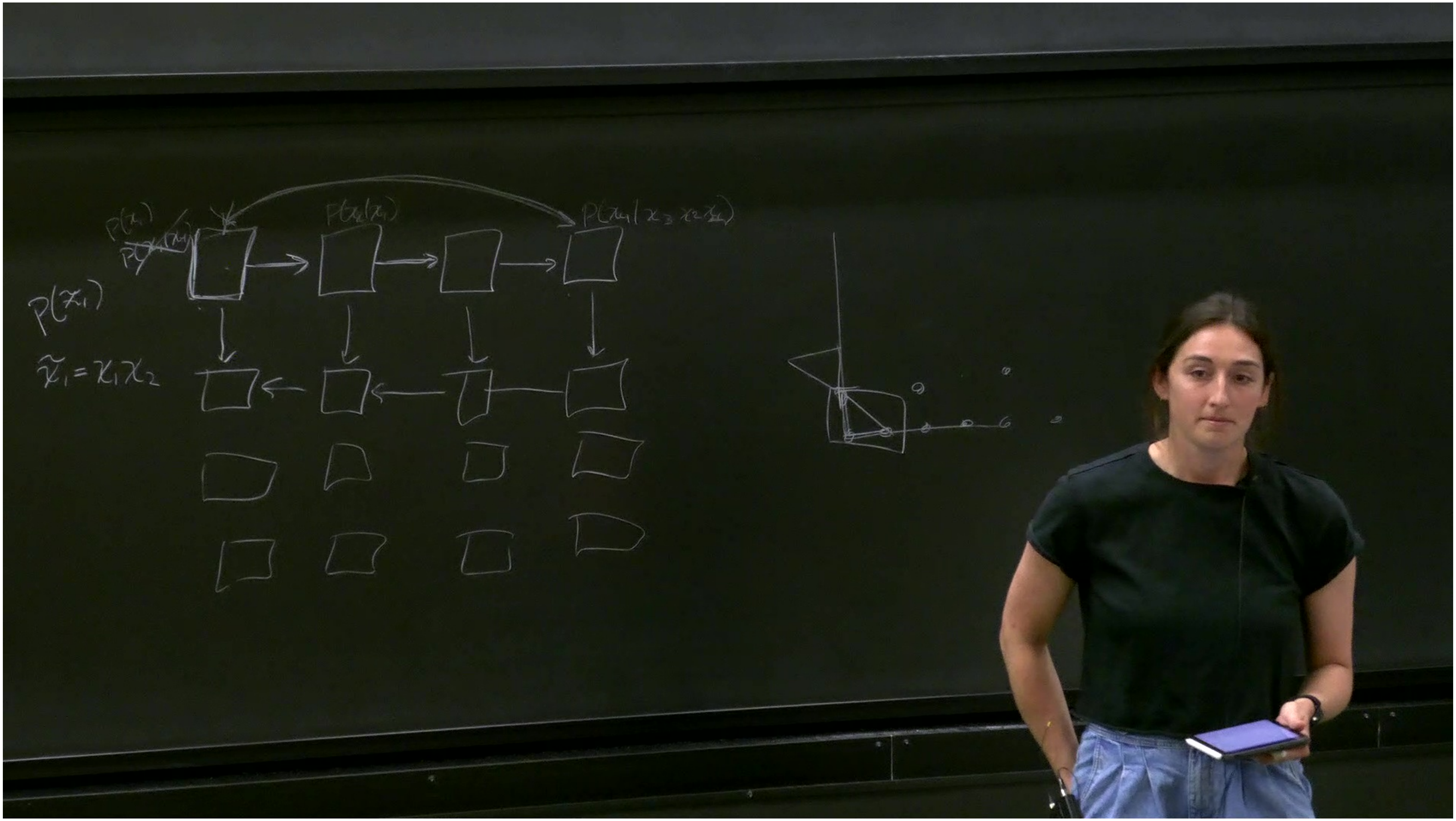
$$③ \; P(\sigma_n | \sigma_{<n}) \; \forall n \longrightarrow P_\lambda(\vec{\sigma}) = \prod_i^N p_\lambda(\sigma_i | \sigma_{<i})$$

$$\psi_\lambda = \sqrt{P_\lambda(\vec{\sigma})}$$


$\vec{y} = \begin{bmatrix} p(\sigma_n=\downarrow \mid \sigma_{<n}) \\ p(\sigma_n=\uparrow \mid \sigma_{<n}) \end{bmatrix}$
$\vec{s} = \begin{bmatrix} \phi_n(\sigma_n=\uparrow) \\ \phi_n(\sigma_n=\downarrow) \end{bmatrix}$
$\vec{h}_{n-1}$
$\vec{h}_n$
$\vec{\sigma}_{n-1}$
① Vanilla : $\tilde{h}_n = f(W \vec{h}_{n-1} + V \vec{\sigma}_{n-1} + \vec{b})$
$(n_h \times n_h)$ $(n_h \times 2)$ (n_h)
$\vec{h}_n = \tilde{h}_n$
GRU : $f = \tanh \Rightarrow (\tilde{h}_n), \in [-1,1]$
"gate" $g_n = \text{sigmoid}(W_g \vec{h}_{n-1} +$
$\vec{h}_n = g_n \odot \tilde{h}_n + (1-g$
① $\vec{h}_{n-1} \longrightarrow \vec{h}_n$
2 $\vec{h}_n \rightarrow p(\sigma_n)$; σ_n
③ $p(\vec{\sigma}) \rightarrow \psi(\vec{\sigma})$


② y = [p(σn = ↓) p(σn = ↑)] = Softmax(U hn + c)
(2 x nh) (2,)
s = π softsign(Uss hn + Css)
Softmax(v) = [e^v1/(e^v1 + e^v2) e^v2/(e^v1 + e^v2)]
σn = {0} = [1 0] but
(gn)i ∈ [0,1]
P(σn) = y · σn
[-1, 1]
③ p(σn | σ<n) ∀
p(σi | σ<i)


7) 2D Vanilla
$\tilde{h}_{ij} = f(W_{(h)} \vec{h}_{i-1,j} + W_{(v)} \vec{h}_{i,j-1} + V \vec{\sigma}_{i-1,j} + V \vec{\sigma}_{i,j-1} + \vec{b})$
$\vec{h}_{ij} = \tilde{h}_{ij}$
Tensorized Cell
$\tilde{h}_{ij} = f([\sigma_{i-1,j} ; \sigma_{i,j-1}] T [h_{i-1,j} ; h_{i,j-1}] + \vec{b})$
$((2\cdot 2) \times (2\cdot n_h) \times n_h) \quad (n_h)$
$y_{ij} = [P\uparrow ; P\downarrow]$
$\vec{h}_{i-1,j}$ $\vec{\sigma}_{i-1,j}$
$\vec{h}_{i,j-1}$ $\vec{\sigma}_{i,j-1}$
L+1
2L


$p(x_1)$
$\tilde{x}_1 = x_1 x_2$