Machine Learning Methods for Neural Data Analysis

Mixture Models, and Hidden Markov Models

Scott Linderman

STATS 220/320 (NBIO220, CS339N). Winter 2025.

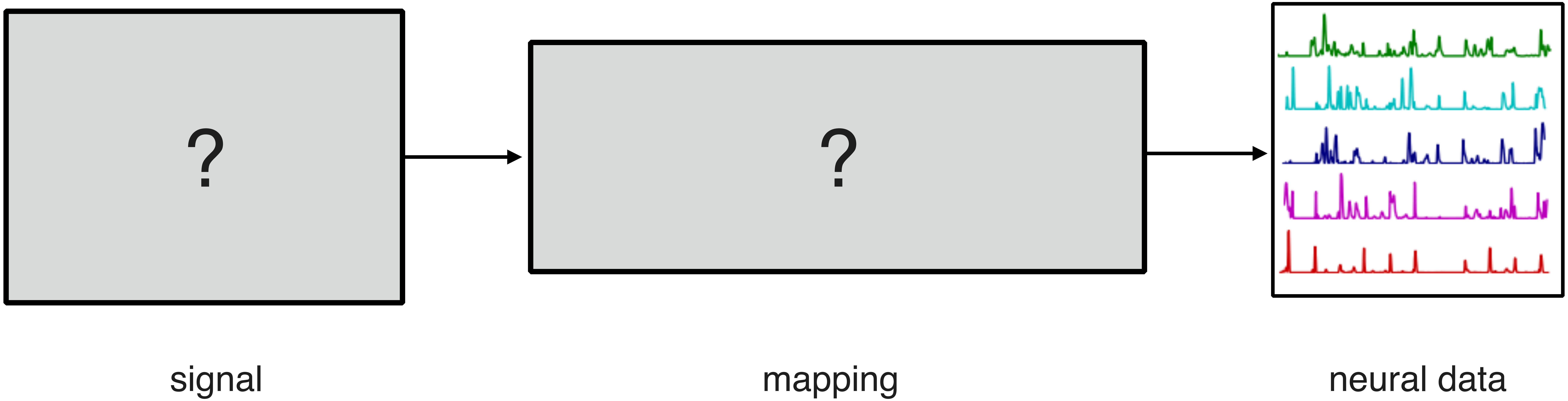
Agenda

- Intro to Unit III: Unsupervised Learning
- Revisiting Gaussian mixture models
- Hidden Markov models and the forward-backward algorithm

Unit III: Unsupervised learning

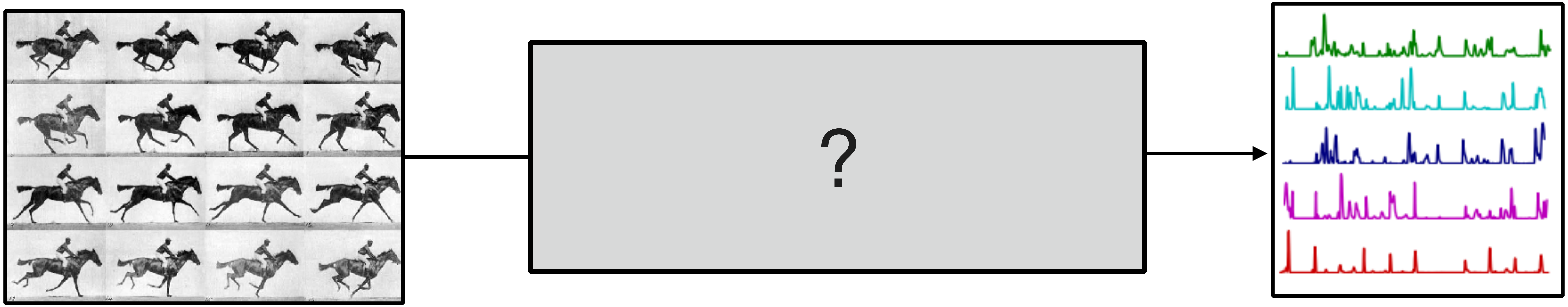
Data-driven modeling

Searching for signals to explain neural activity



Data-driven modeling

Searching for signals to explain neural activity

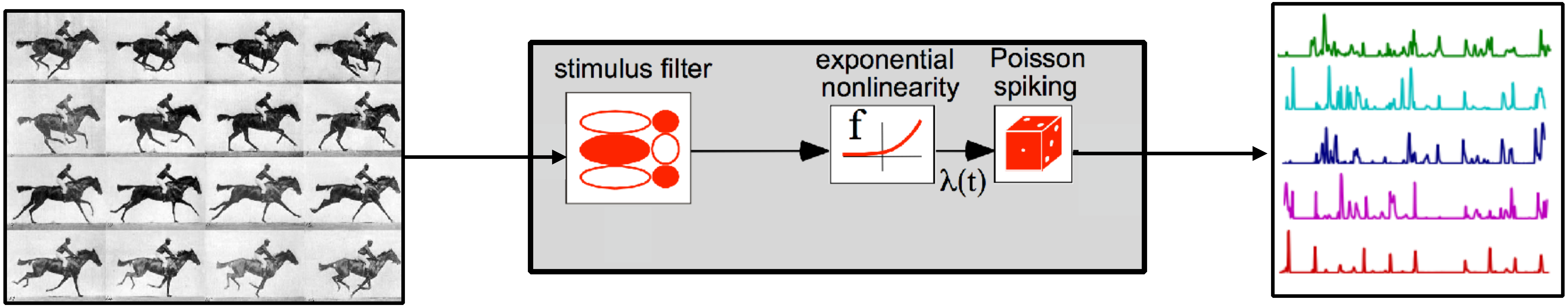


signal mapping neural data

Encoding models: given stimulus (covariates) and response, find mapping.

Data-driven modeling

Searching for signals to explain neural activity



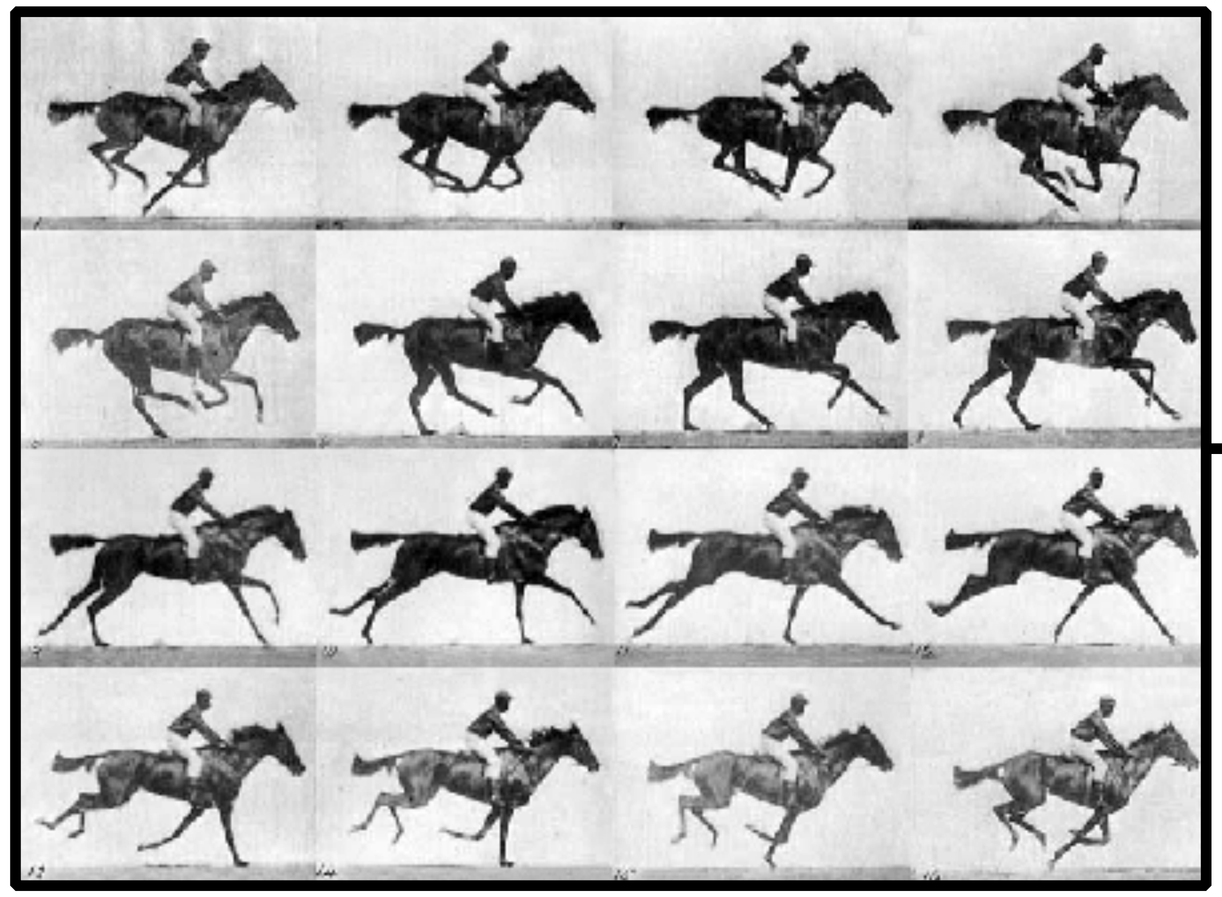
signal mapping neural data

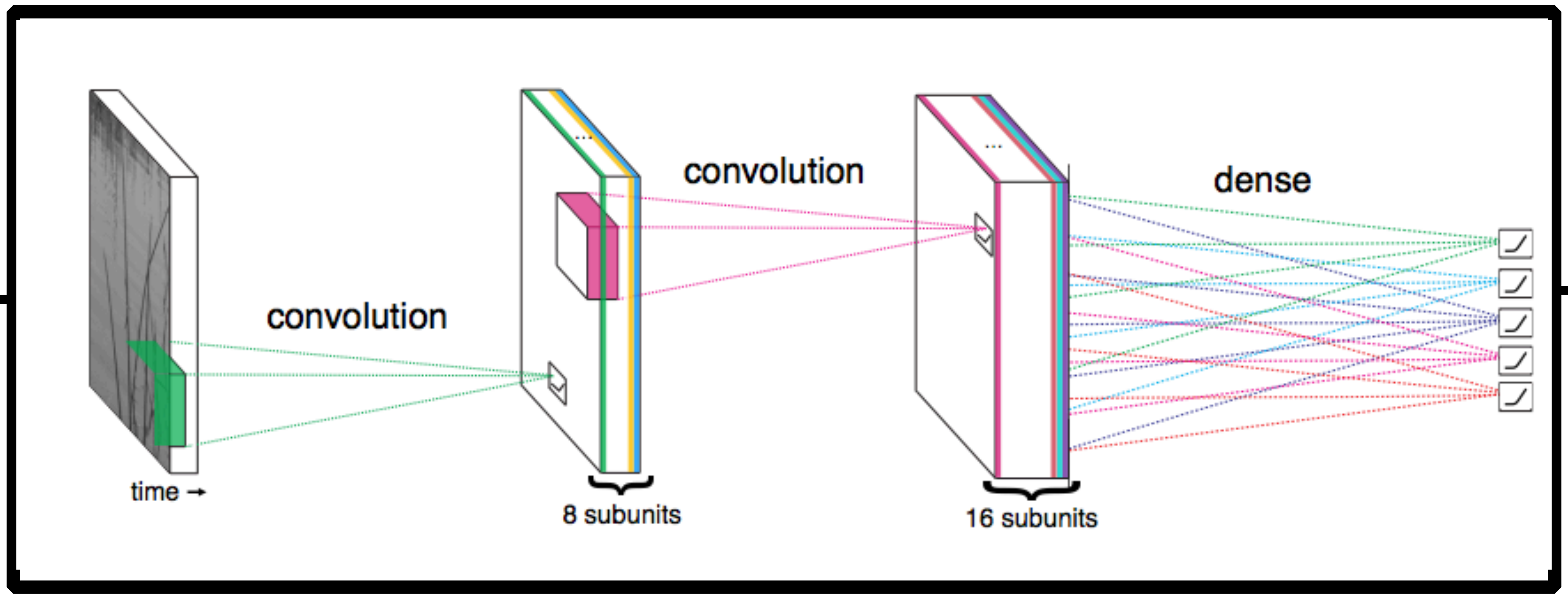
Recent examples: Musall et al (2018), Stringer et al (2018)

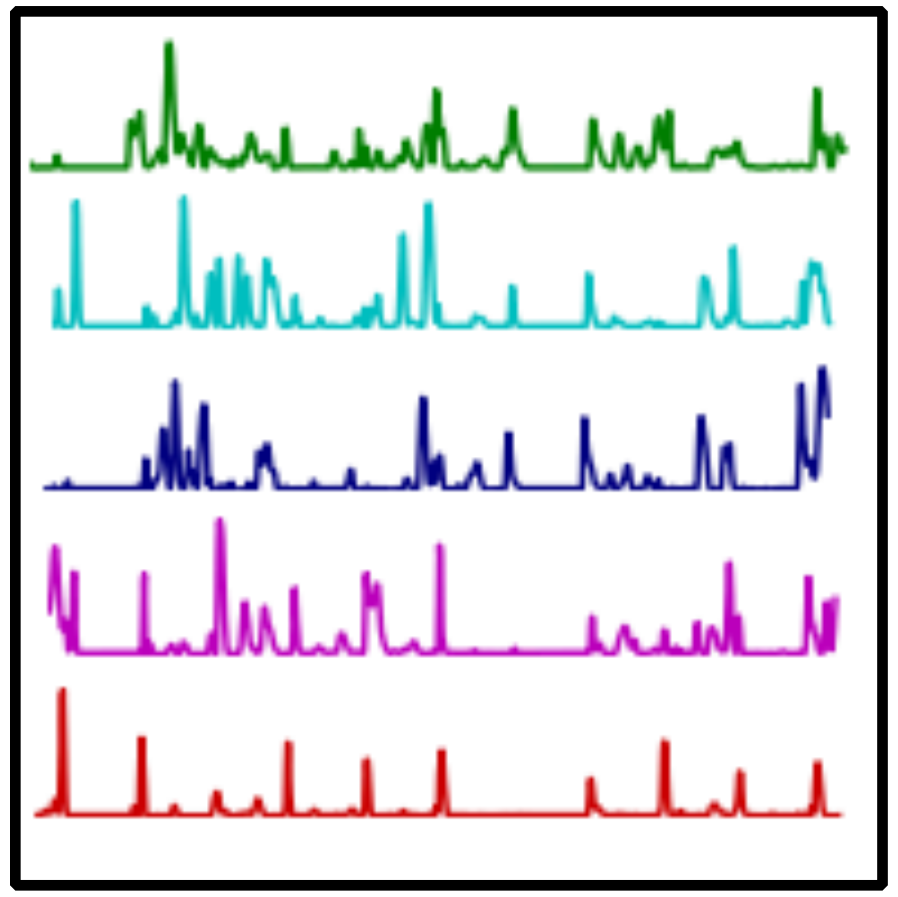
Paninski (2004)
Truccolo et al (2005)
Pillow et al (2008)

Data-driven modeling

Searching for signals to explain neural activity







signal

mapping

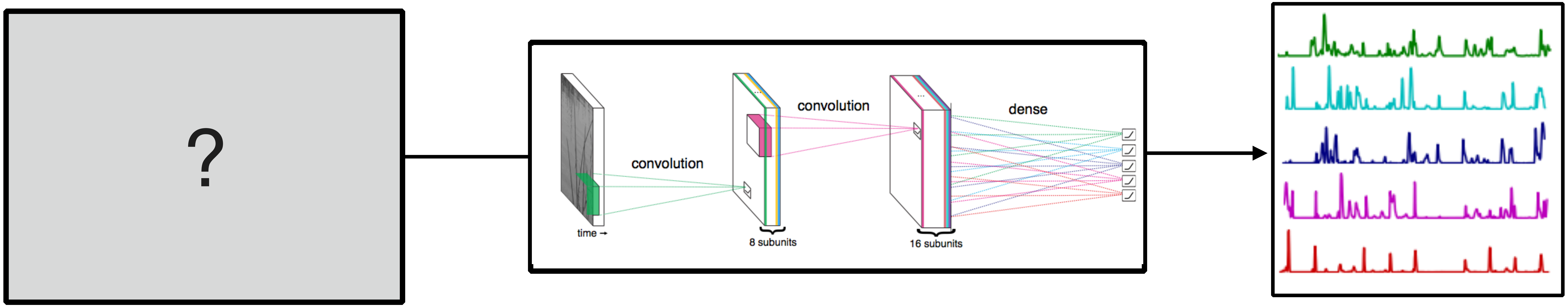
neural data

Toward nonlinear and/or more biophysically plausible mappings.

McIntosh et al (2017)

Data-driven modeling

Searching for signals to explain neural activity



latent signal

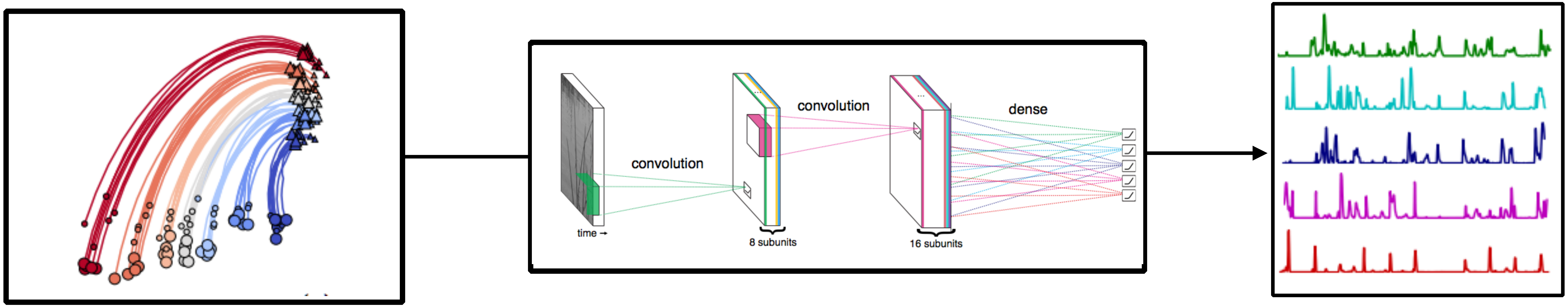
mapping

neural data

Alternative: try to infer latent signals from the data

Data-driven modeling

Searching for signals to explain neural activity



latent signal

mapping

neural data

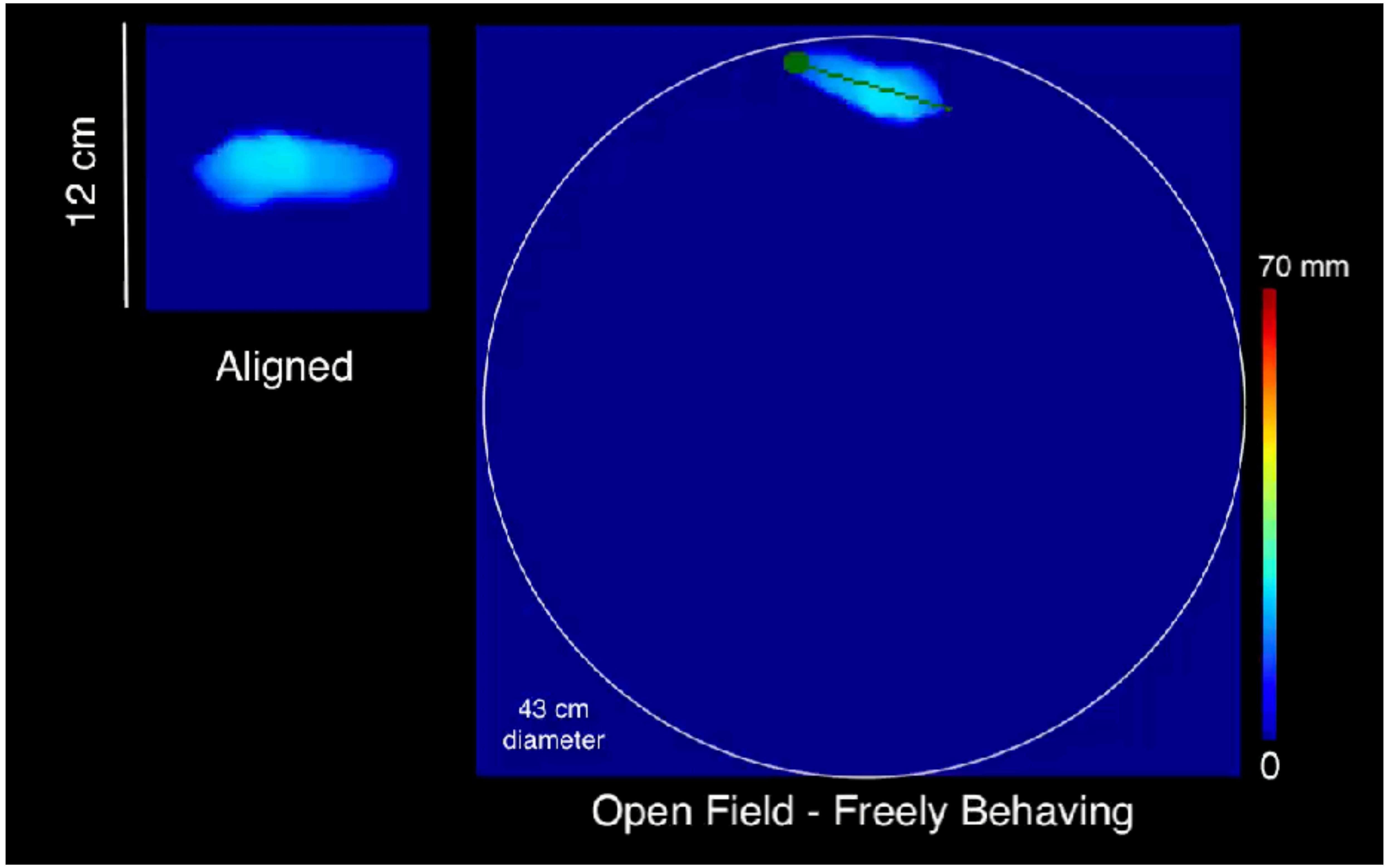
Alternative: try to infer latent signals from the data, *subject to constraints.*

Latent variable modeling is all about constraints

The five D's

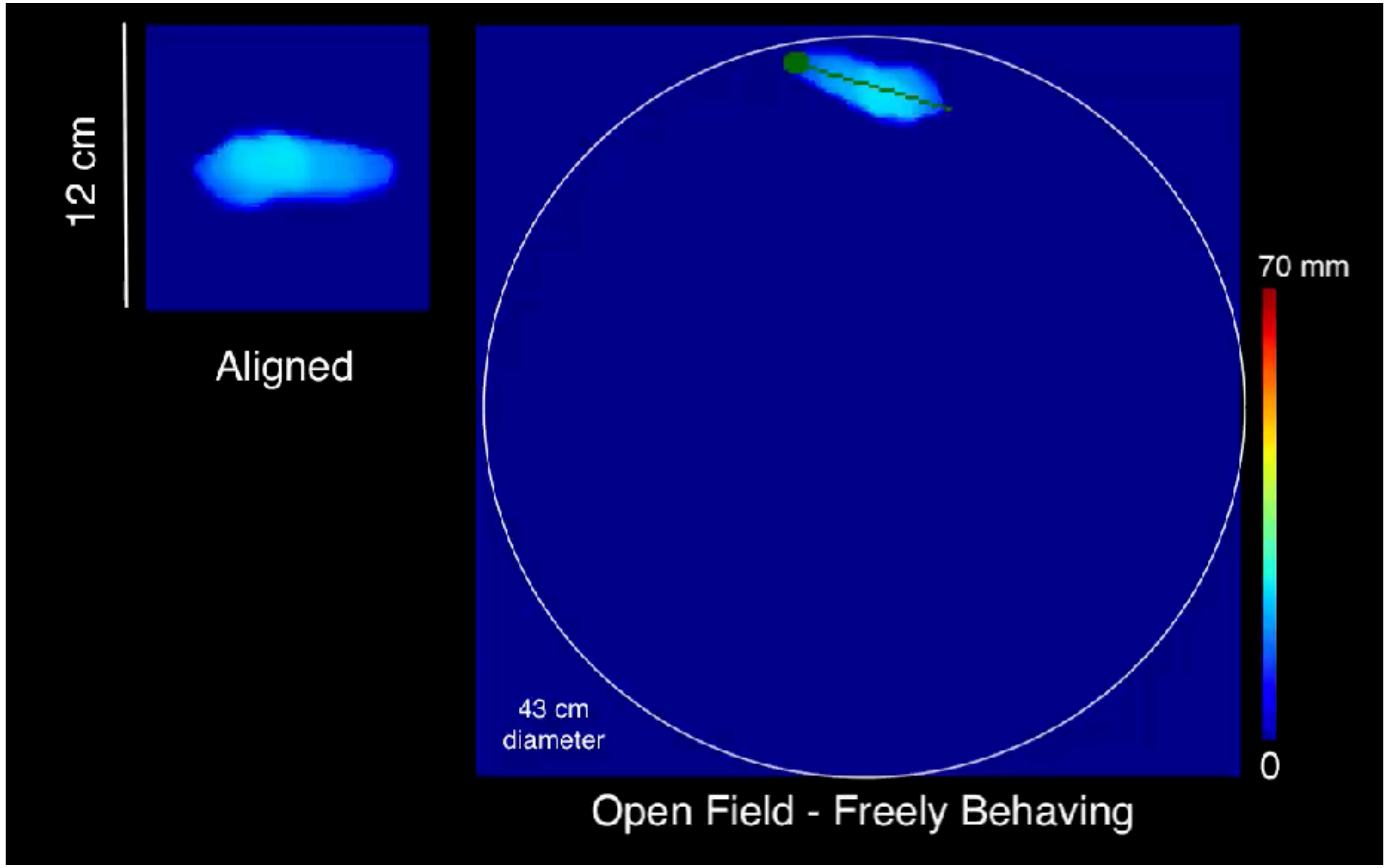
- *Dimensionality*: how many latent clusters, factors, etc.?
- *Domain*: are the latent variables discrete, continuous, bounded, sparse, etc.?
- *Dynamics*: how do the latent variables change over time?
- *Dependencies*: how do the latent variables relate to the observed data?
- *Distribution*: do we have prior knowledge about the variables' probability?

- We've already seen some examples in Unit 1!



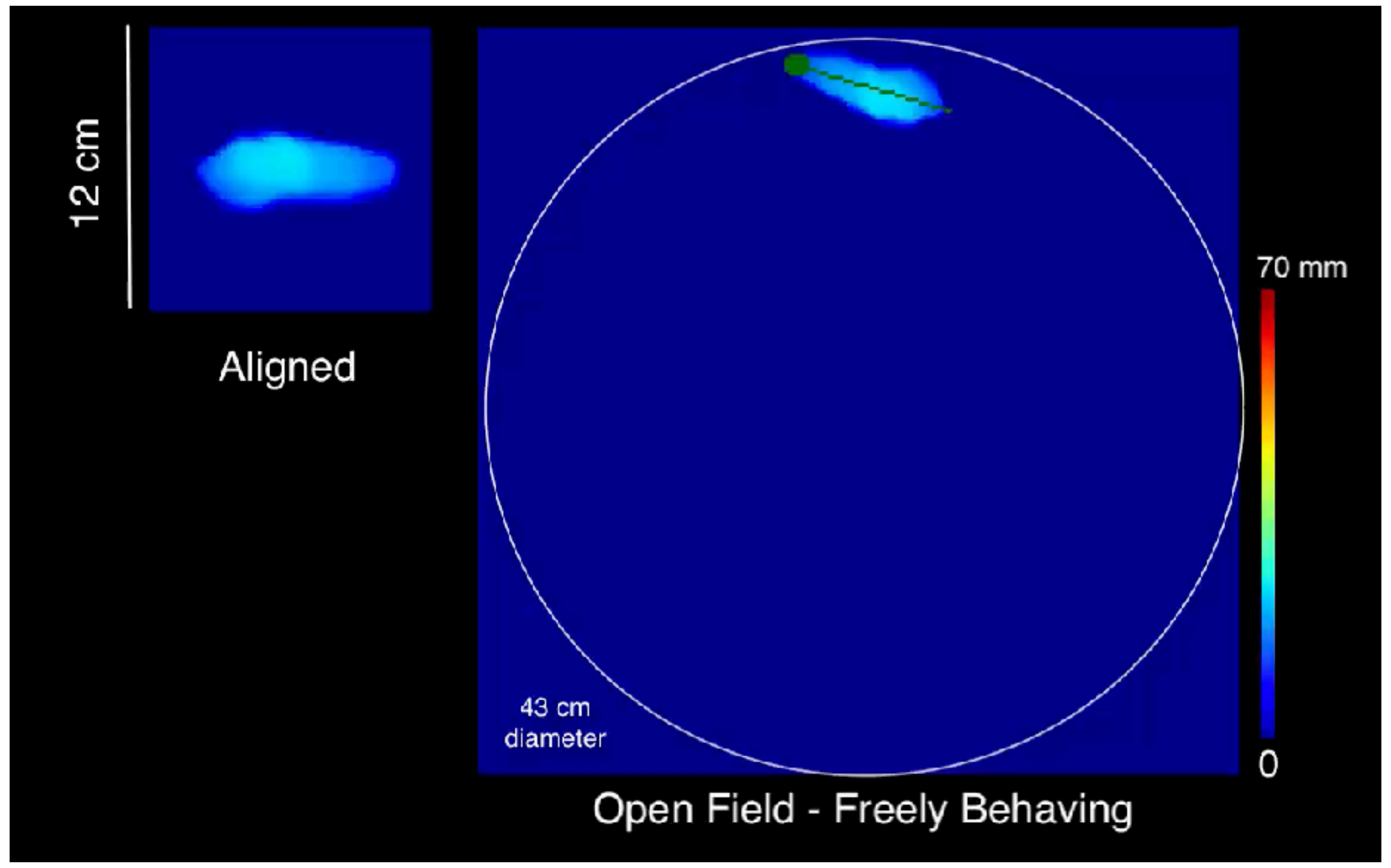
Wiltschko et al 2015

- Data: depth camera video of mouse exploring a circular arena
- Question: how does the brain produce spontaneous behavior?
  - Specifically interested in the neurotransmitter **dopamine**, which is implicated in movement/timing deficits in Parkinson's
  - How does dopamine impact both the speed and occurrence of different behaviors?

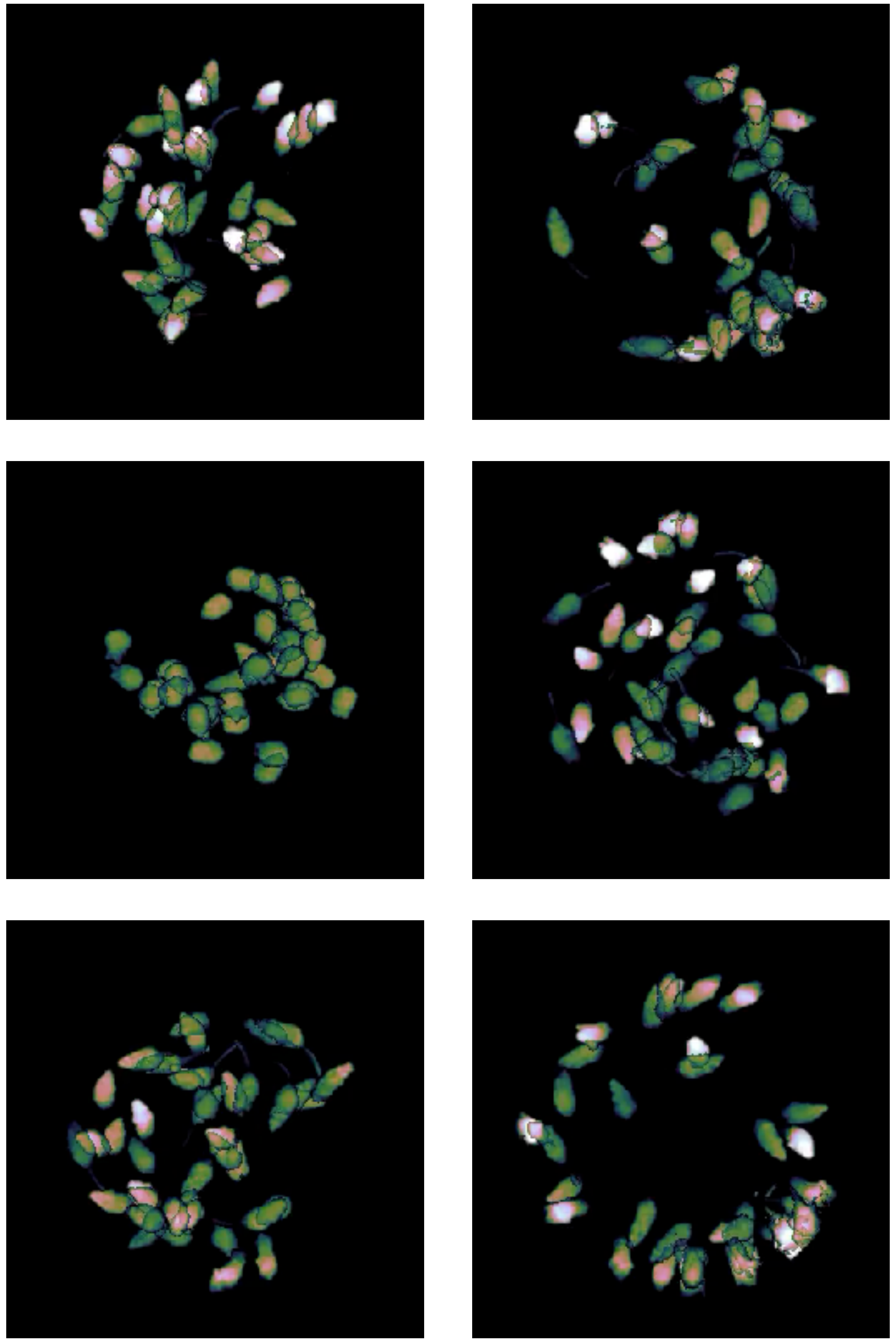


Wiltschko et al 2015

- To answer these questions, we need a behavioral description of what’s going on in this video
- Hard to do by hand: time-consuming and biased
- Latent variable models can give us a latent state summary of what’s going on in the video!



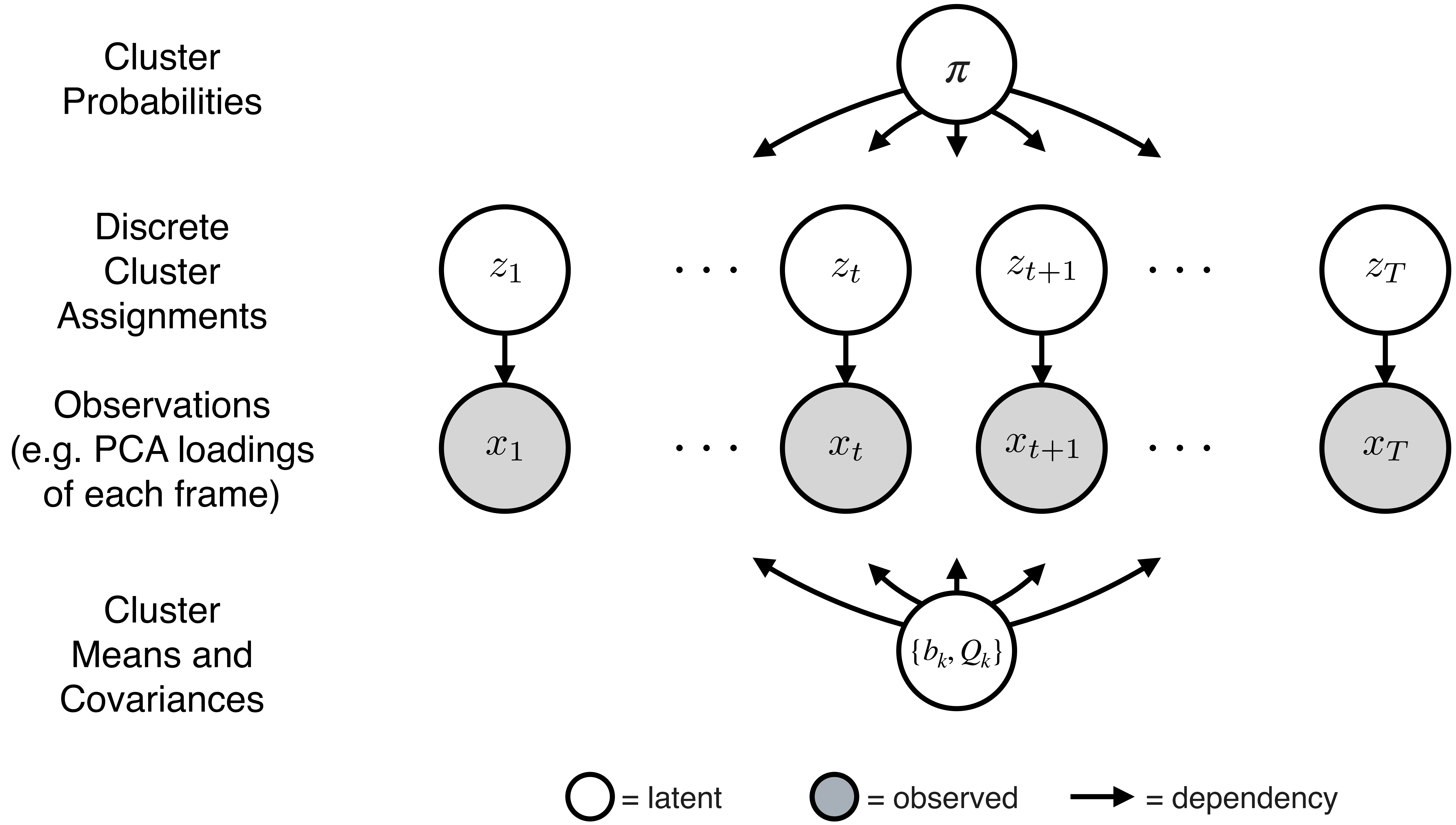
Wiltschko et al 2015



This result comes from a Hidden Markov Model.
Let’s learn about them!

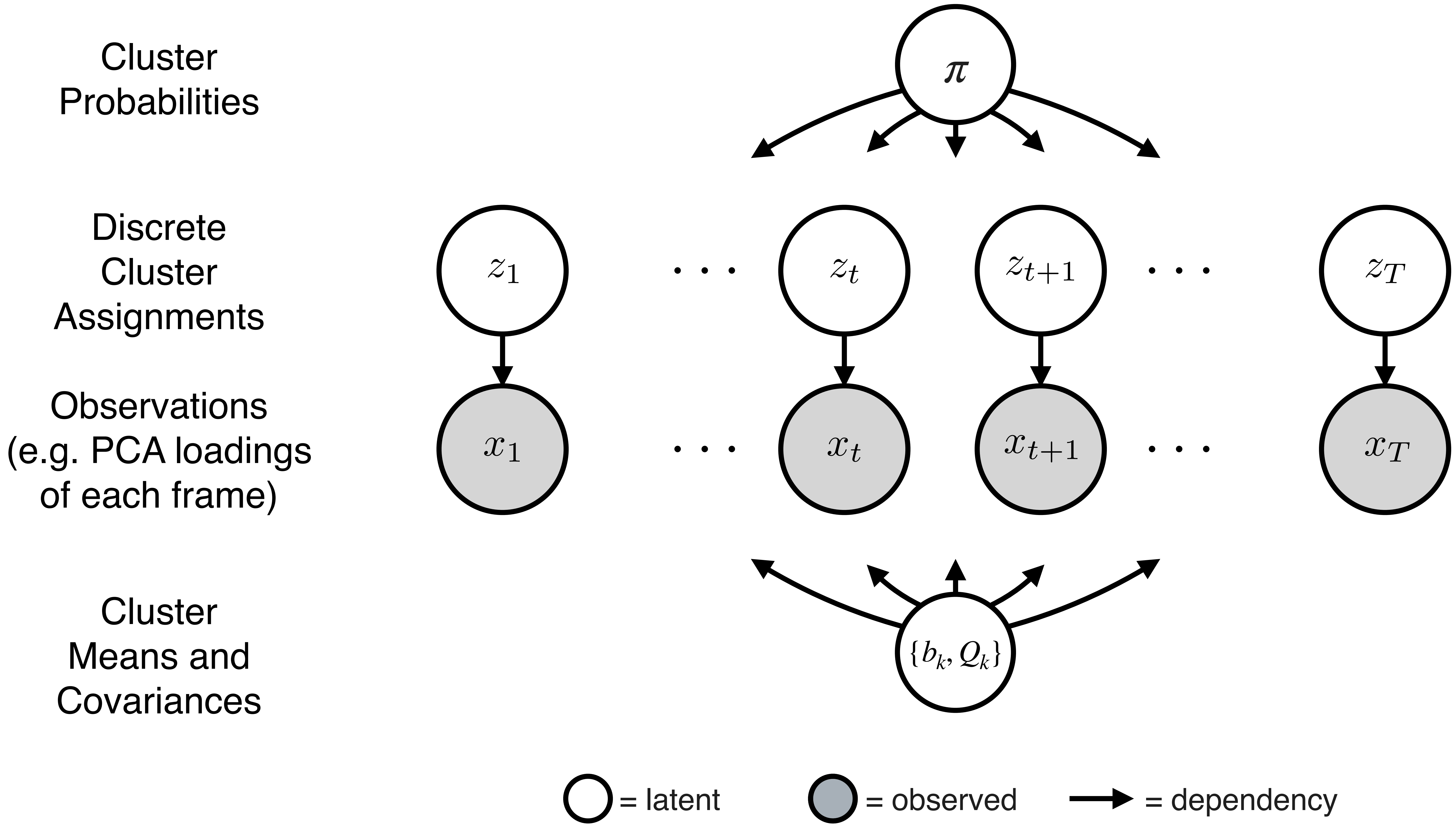
What We've Seen: The Gaussian Mixture Model

Graphical Model



What We've Seen: The Gaussian Mixture Model

Graphical Model

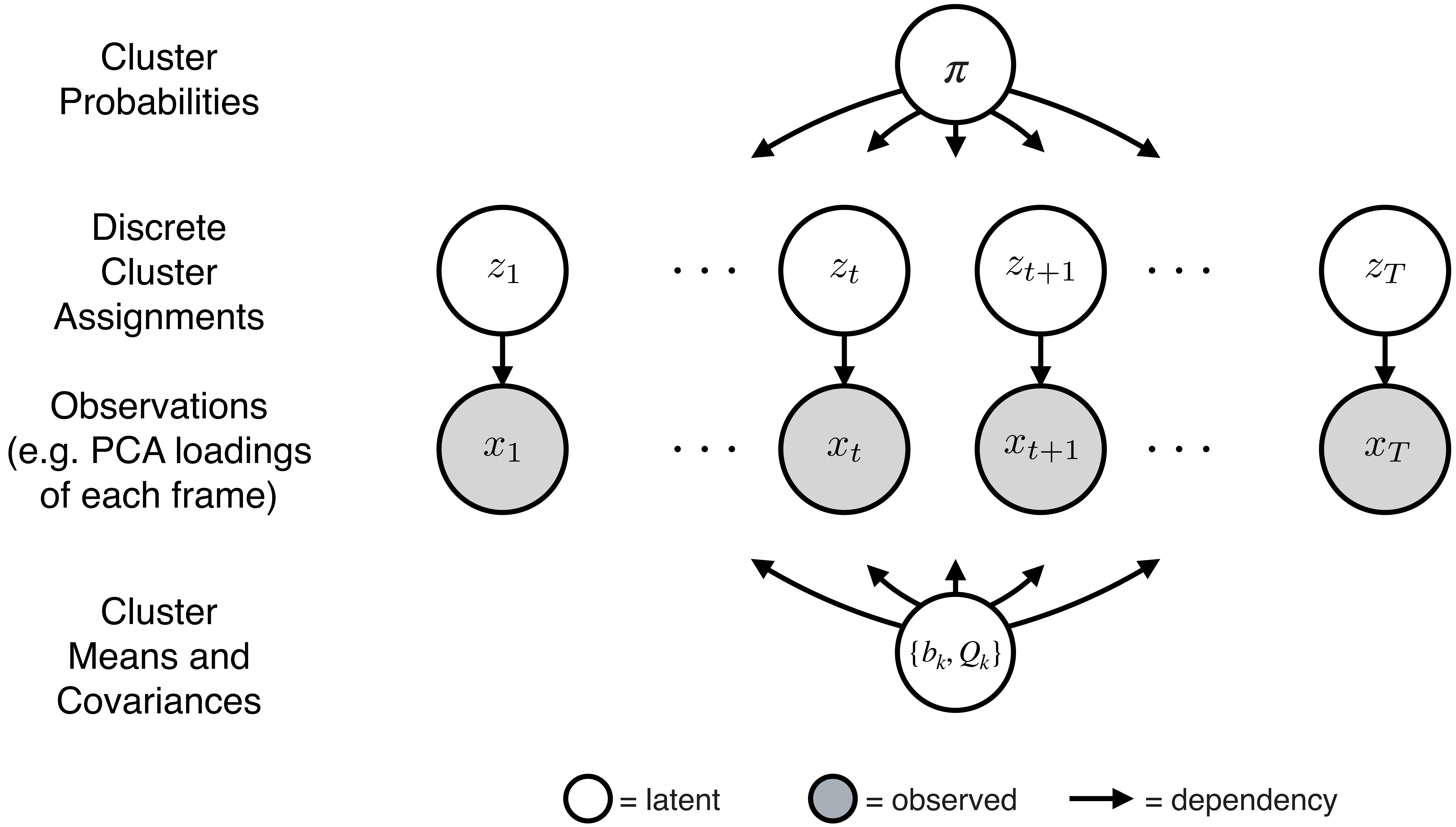


Questions:

- **Inference:** Given parameters and observations, what are most likely $\{z_t\}_{t=1}^{T}$?
- **Learning:** How do we estimate the parameters given our observations?
- Relatively easy to answer since timesteps are independent

What We've Seen: The Gaussian Mixture Model

Graphical Model



Questions:

- **Inference:** Given parameters and observations, what are most likely $\{z_t\}_{t=1}^{T}$?
- **Learning:** How do we estimate the parameters given our observations?
- Relatively easy to answer since timesteps are independent

What might go wrong if we apply this model to the mouse video data?

Hidden Markov Models

The Gaussian HMM

A Gaussian HMM is just a Gaussian mixture model but where cluster assignments are linked across time!

$$
\begin{aligned}
z_1 &\sim \mathrm{Cat}(\pi), \\
z_t \mid z_{t-1} &\sim \mathrm{Cat}(P_{z_{t-1}}), \qquad \text{for } t = 2,\ldots,T. \\
x_t \mid z_t &\sim \mathcal{N}(b_{z_t}, Q_{z_t}) \qquad \text{for } t = 1,\ldots,T
\end{aligned}
$$

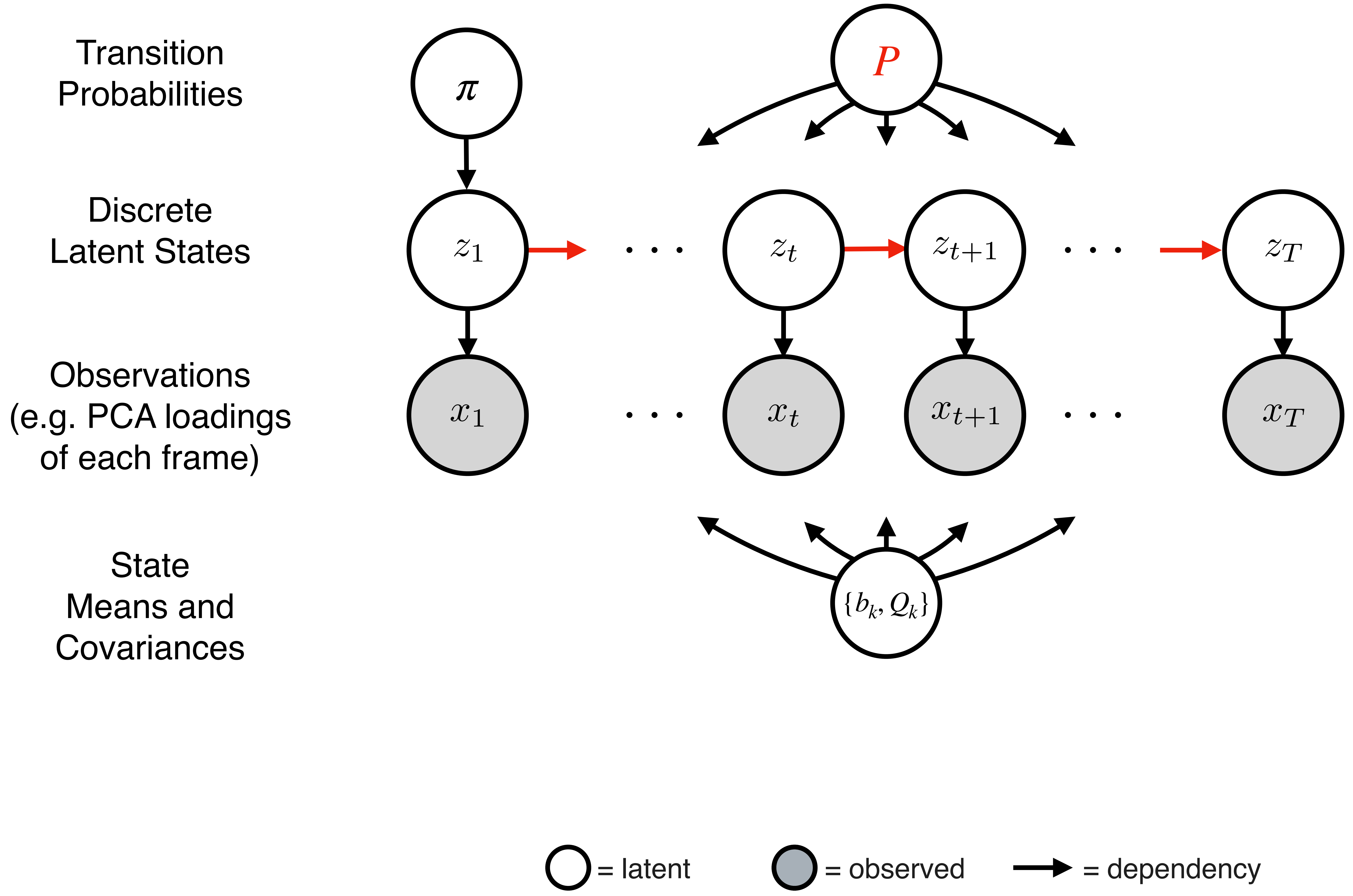
Its parameters are $\Theta = \pi, P, \{b_k, Q_k\}_{k=1}^K$ where $P \in [0,1]^{K \times K}$ is a row-stochastic **transition matrix.**

Under this model, the **joint probability** factors as

$$
p(x, z, \Theta) = p(z_1) \prod_{t=1}^{T-1} p(z_{t+1} \mid z_t) \prod_{t=1}^{T} p(x_t \mid z_t)
$$

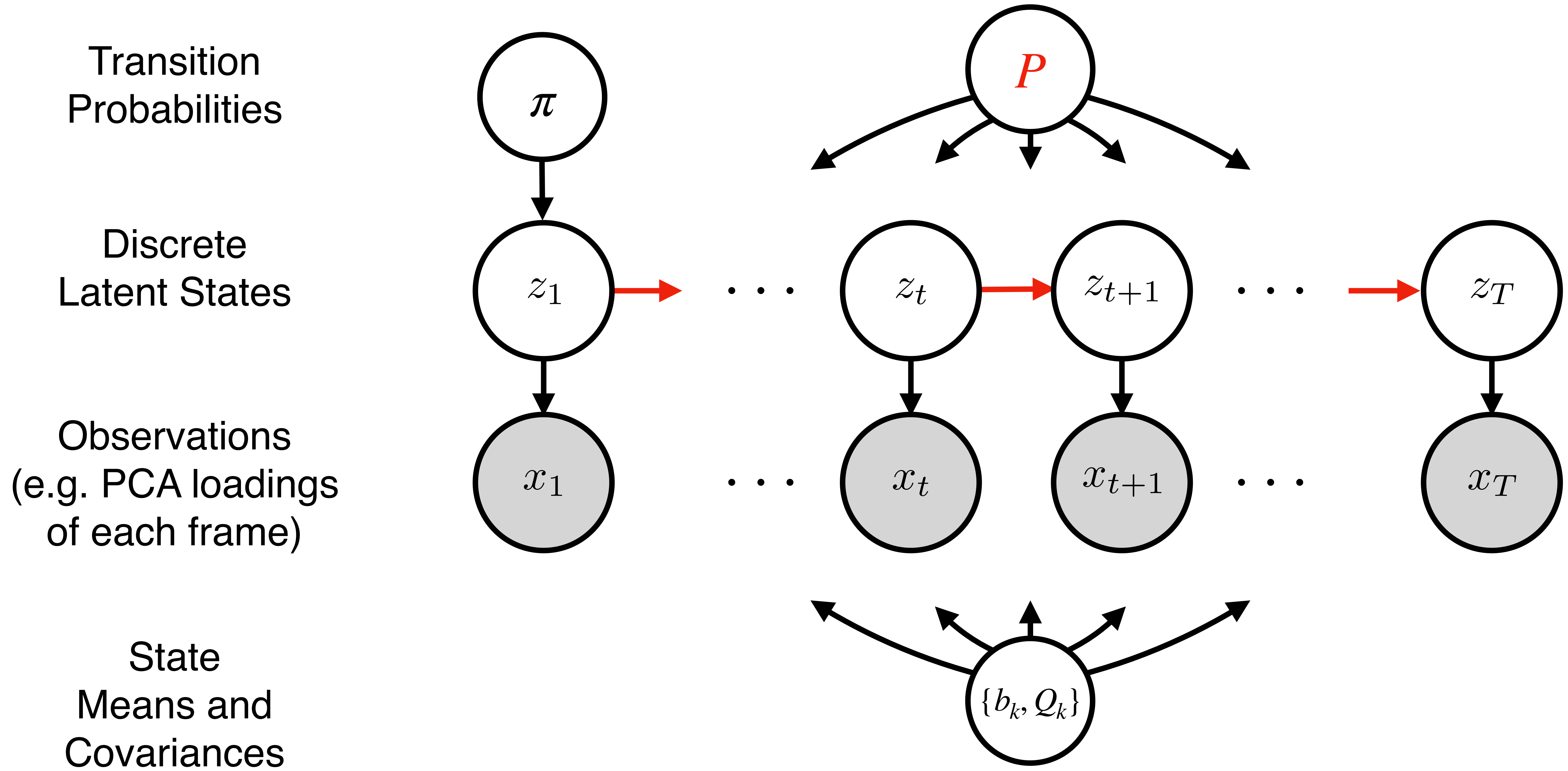
The Gaussian HMM

Graphical Model



The Gaussian HMM

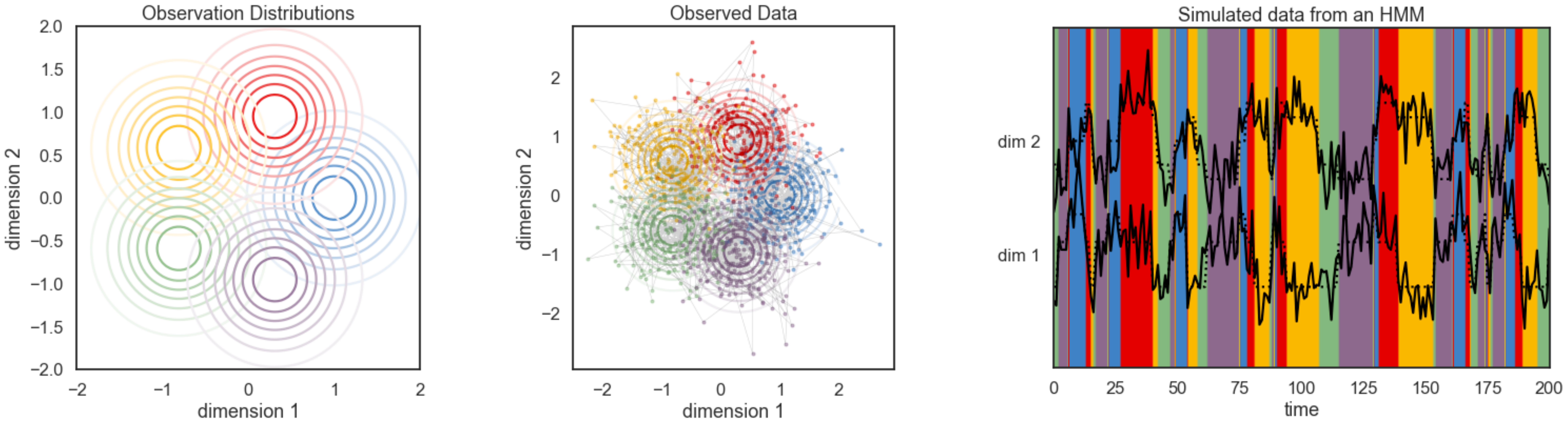
Graphical Model



$$p(x, z, \Theta) = p(z_1) \prod_{t=1}^{T-1} p(z_{t+1} \mid z_t) \prod_{t=1}^{T} p(x_t \mid z_t)$$

The Gaussian HMM

Example draw from a 2D Gaussian HMM with 5 clusters



Inference for the Gaussian HMM

The posterior is a little trickier…

- Update the posterior over latent variables given data and parameters,

$$p(z \mid x, \Theta) \propto p(x, z, \Theta) = p(z_1) \prod_{t=1}^{T-1} p(z_{t+1} \mid z_t) \prod_{t=1}^{T} p(x_t \mid z_t)$$

- The normalized posterior no longer has a simple **closed form** because states depend on each other!

- However, we can still **efficiently compute** the **marginal probabilities.**

Inference for the Gaussian HMM

Computing the marginal likelihood

- Consider the marginal probability of state k at time t:

$$p(z_t = k \mid x) = \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} p(z_1, \ldots, z_{t-1}, z_t = k, z_{t+1}, \ldots, z_T \mid x)$$

Inference for the Gaussian HMM

Computing the marginal likelihood

- Consider the marginal probability of state k at time t:

$$p(z_t = k \mid x) = \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} p(z_1, \ldots, z_{t-1}, z_t = k, z_{t+1}, \ldots, z_T \mid x)$$

Inference for the Gaussian HMM

Computing the marginal likelihood

- Consider the marginal probability of state k at time t:

$$p(z_t = k \,|\, x) = \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} p(z_1, \ldots, z_{t-1}, z_t = k, z_{t+1}, \ldots, z_T \,|\, x)$$

$$\propto \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(z_{s+1} \mid z_s) p(x_s \mid z_s) p(x_t \mid z_t = k)$$

$$\times \prod_{u=t+1}^{T} p(z_u \mid z_{u-1}) p(x_u \mid z_u)$$

Inference for the Gaussian HMM

Computing the marginal likelihood

- Consider the marginal probability of state k at time t:

$$p(z_t = k \mid x) = \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} p(z_1, \ldots, z_{t-1}, z_t = k, z_{t+1}, \ldots, z_T \mid x)$$

$$\propto \left[\sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(x_s \mid z_s)\, p(z_{s+1} \mid z_s)\right] \times \left[p(x_t \mid z_t = k)\right]$$

$$\times \left[\sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} \prod_{u=t+1}^{T} p(z_u \mid z_{u-1})\, p(x_u \mid z_u)\right]$$

Inference for the Gaussian HMM

Computing the marginal likelihood

- Consider the marginal probability of state k at time t:

$$p(z_t = k \mid x) = \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} p(z_1, \ldots, z_{t-1}, z_t = k, z_{t+1}, \ldots, z_T \mid x)$$

$$\propto \left[\sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(x_s \mid z_s)\, p(z_{s+1} \mid z_s)\right] \times \left[p(x_t \mid z_t = k)\right]$$

$$\times \left[\sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} \prod_{u=t+1}^{T} p(z_u \mid z_{u-1})\, p(x_u \mid z_u)\right]$$

$$\triangleq \alpha_t(z_t) \times p(x_t \mid z_t) \times \beta_t(z_t)$$

Inference for the Gaussian HMM

Computing the forward messages $\alpha_t(z_t)$

- Consider the **forward messages**:

$$\alpha_t(z_t) \triangleq \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(x_s \mid z_s)\, p(z_{s+1} \mid z_s)$$

Inference for the Gaussian HMM

Computing the forward messages $\alpha_t(z_t)$

- Consider the forward messages:

$$\alpha_t(z_t) \triangleq \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(x_s \mid z_s)\, p(z_{s+1} \mid z_s)$$

$$= \sum_{z_{t-1}=1}^{K} \left[\left(\sum_{z_1=1}^{K} \cdots \sum_{z_{t-2}=1}^{K} p(z_1) \prod_{s=1}^{t-2} p(x_s \mid z_s) p(z_{s+1} \mid z_s) \right) p(x_{t-1} \mid z_{t-1})\, p(z_t \mid z_{t-1}) \right]$$

Inference for the Gaussian HMM

Computing the forward messages $\alpha_t(z_t)$

- Consider the forward messages:

$$\alpha_t(z_t) \triangleq \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(x_s \mid z_s)\, p(z_{s+1} \mid z_s)$$

$$= \sum_{z_{t-1}=1}^{K} \left[\left(\sum_{z_1=1}^{K} \cdots \sum_{z_{t-2}=1}^{K} p(z_1) \prod_{s=1}^{t-2} p(x_s \mid z_s) p(z_{s+1} \mid z_s) \right) p(x_{t-1} \mid z_{t-1})\, p(z_t \mid z_{t-1}) \right]$$

Inference for the Gaussian HMM

Computing the forward messages $\alpha_t(z_t)$

- Consider the forward messages:

$$\alpha_t(z_t) \triangleq \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(x_s \mid z_s)\, p(z_{s+1} \mid z_s)$$

$$= \sum_{z_{t-1}=1}^{K} \left[\left(\sum_{z_1=1}^{K} \cdots \sum_{z_{t-2}=1}^{K} p(z_1) \prod_{s=1}^{t-2} p(x_s \mid z_s) p(z_{s+1} \mid z_s) \right) p(x_{t-1} \mid z_{t-1})\, p(z_t \mid z_{t-1}) \right]$$

$$= \sum_{z_{t-1}=1}^{K} \alpha_{t-1}(z_{t-1})\, p(x_{t-1} \mid z_{t-1})\, p(z_t \mid z_{t-1})$$

- We can compute these messages **recursively!**

Inference for the Gaussian HMM

Computing the forward messages $\alpha_t(z_t)$. Vectorized.

- Let $\alpha_t = [\alpha_t(z_t = 1), \ldots, \alpha_t(z_t = K)]^\top$ denote the column vector of forward messages. Then,

$$\alpha_t = P^\top(\alpha_{t-1} \odot \ell_{t-1})$$

  where

  - $\ell_{t-1} = [p(x_{t-1} \mid z_{t-1} = 1), \ldots, p(x_{t-1} \mid z_{t-1} = K)]^\top$ is the vector of likelihoods,
  - $\odot$ denotes the element-wise product, and
  - P is the transition matrix with $P_{ij} = p(z_t = j \mid z_{t-1} = i)$.

- For the base case, let $\alpha_1(z_1) = p(z_1)$.

Inference for the Gaussian HMM

Computing the forward messages $\alpha_t(z_t)$

- Take a step back: what are we actually computing anyway?

$$\alpha_t(z_t) = \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(z_1) \prod_{s=1}^{t-1} p(x_s \mid z_s)\, p(z_{s+1} \mid z_s)$$

$$= \sum_{z_1=1}^{K} \cdots \sum_{z_{t-1}=1}^{K} p(\{z_s\}_{s=1}^{t-1}, \{x_s\}_{s=1}^{t-1}) p(z_t \mid z_{t-1})$$

- we can normalize this to get the conditional distribution $p(z_t \mid \{x_s\}_{s=1}^{t-1})$!

Inference for the Gaussian HMM

Computing the backward messages $\beta_t(z_t)$

- Now take the **backward messages**:

$$\beta_t(z_t) \triangleq \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} \prod_{u=t+1}^{T} p(z_u \mid z_{u-1})\, p(x_u \mid z_u)$$

Inference for the Gaussian HMM

Computing the backward messages $\beta_t(z_t)$

- Now take the **backward messages**:

$$\beta_t(z_t) \triangleq \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} \prod_{u=t+1}^{T} p(z_u \mid z_{u-1})\, p(x_u \mid z_u)$$

$$= \sum_{z_{t+1}=1}^{K} p(z_{t+1} \mid z_t)\, p(x_{t+1} \mid z_{t+1}) \sum_{z_{t+2}=1}^{K} \cdots \sum_{z_T=1}^{K} \prod_{u=t+2}^{T} p(z_u \mid z_{u-1})\, p(x_u \mid z_u)$$

Inference for the Gaussian HMM

Computing the backward messages $\beta_t(z_t)$

- Now take the **backward messages**:

$$\beta_t(z_t) \triangleq \sum_{z_{t+1}=1}^{K} \cdots \sum_{z_T=1}^{K} \prod_{u=t+1}^{T} p(z_u \mid z_{u-1})\, p(x_u \mid z_u)$$

$$= \sum_{z_{t+1}=1}^{K} p(z_{t+1} \mid z_t)\, p(x_{t+1} \mid z_{t+1}) \sum_{z_{t+2}=1}^{K} \cdots \sum_{z_T=1}^{K} \prod_{u=t+2}^{T} p(z_u \mid z_{u-1})\, p(x_u \mid z_u)$$

$$= \sum_{z_{t+1}=1}^{K} p(z_{t+1} \mid z_t)\, p(x_{t+1} \mid z_{t+1})\, \beta_{t+1}(z_{t+1})$$

- Again, we can compute the backward messages recursively!

Inference for the Gaussian HMM

Computing the backward messages $\beta_t(z_t)$. Vectorized.

- Let $\beta_t = [\beta_t(z_t = 1), \ldots, \beta_t(z_t = K)]^\top$ denote the column vector of backward messages. Then,

$$\beta_t = P(\beta_{t+1} \odot \ell_{t+1})$$

- For the base case, let $\beta_T(z_T) = 1$.

Inference for the Gaussian HMM

Combining the forward and backward messages

- The posterior marginal probability of state k at time t is,

$$
\begin{aligned}
p(z_t = k \mid x) &\propto \alpha_t(z_t = k) \times p(x_t \mid z_t = k) \times \beta_t(z_t = k) \\
&= \alpha_{tk} \ell_{tk} \beta_{tk}
\end{aligned}
$$

- The probabilities need to sum to one. Normalizing yields,

$$
p(z_t = k \mid x) = \frac{\alpha_{tk} \ell_{tk} \beta_{tk}}{\sum_{j=1}^{K} \alpha_{tj} \ell_{tj} \beta_{tj}}
$$

- Finally, note the marginal is invariant to multiplying α_t and/or β_t by a constant.

Inference for the Gaussian HMM

Normalizing the messages to prevent underflow

- The messages involve **products of probabilities**, which quickly underflow.
- We can leverage the scale invariance to renormalize the messages. I.e. replace:

$$\alpha_t = P^\top(\alpha_{t-1} \odot \ell_{t-1}) \quad \text{with} \quad \begin{aligned} A_{t-1} &= \textstyle\sum_k \tilde{\alpha}_{t-1,k}\, \ell_{t-1,k} \\ \tilde{\alpha}_t &= \frac{1}{A_{t-1}} P^\top(\tilde{\alpha}_{t-1} \odot \ell_{t-1}) \end{aligned}$$

where $\tilde{\alpha}_t$ are normalized for numerical stability. As before, $\tilde{\alpha}_1 = p(z_1)$.

Inference for the Gaussian HMM

Computing the marginal likelihood

- Finally, we can compute the marginal likelihood alongside the forward messages

$$\log p(x \mid \Theta) = \log \sum_{z_1=1}^{K} \cdots \sum_{z_T=1}^{K} \left[p(z_1) \prod_{t=1}^{T-1} p(z_{t+1} \mid z_t) \prod_{t=1}^{T} p(x_t \mid z_t) \right]$$

$$= \log \sum_{z_T=1}^{K} \alpha_T(z_T)\, p(x_T \mid z_T)$$

Conclusion

- Hidden Markov models (HMMs) are just mixture models with dependencies across time.
- We use the **forward-backward algorithm** to compute latent state probabilities and expected sufficient stats.
- Next time: we'll see how to update parameters