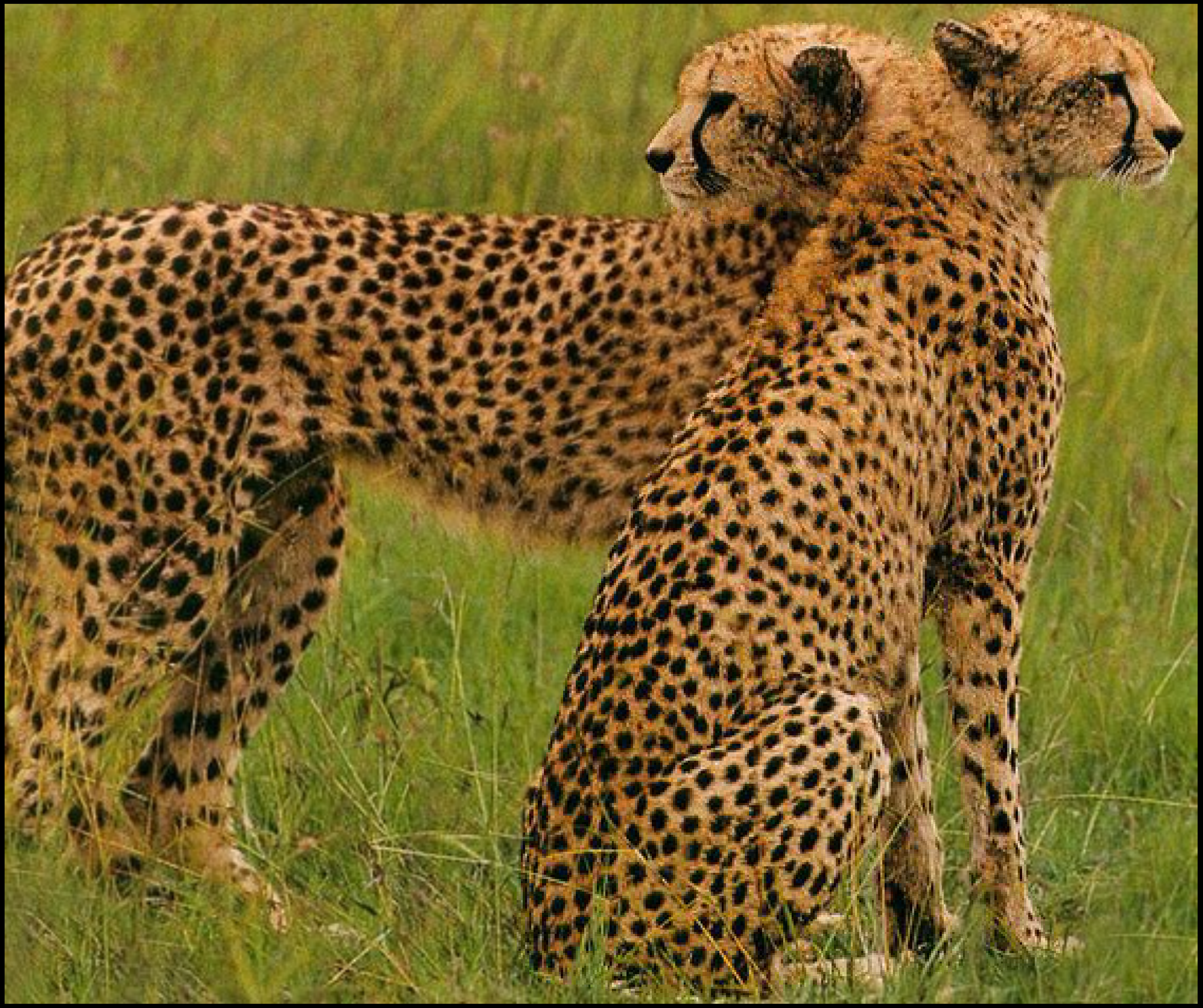
Inference by Reparameterization in Neural Population Codes

Workshop on Theoretical Neuroscience
Janelia Research Campus, Sep 2016

Rajkumar V
Xaq Lab
Rice University, BCM



Probabilistic Inference

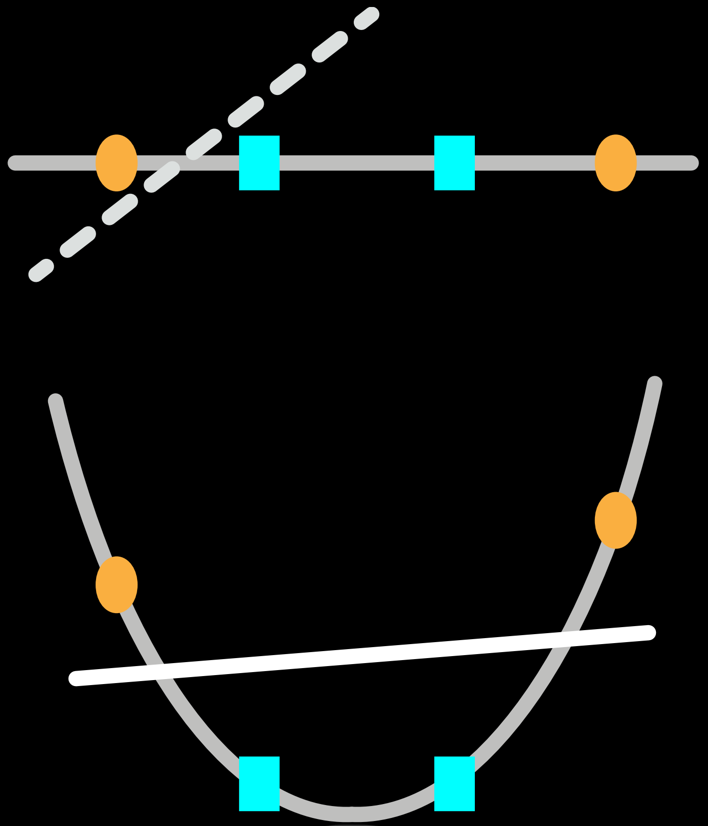
- Natural to perform tasks involving uncertainty using probabilistic computations

- Human behavior is highly consistent with probabilistic reasoning in sensory, motor and cognitive domains

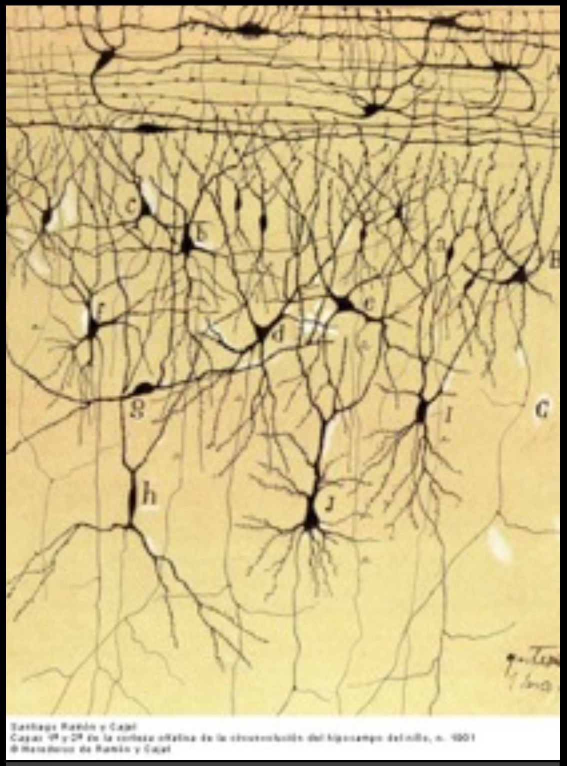

Inference!
How?

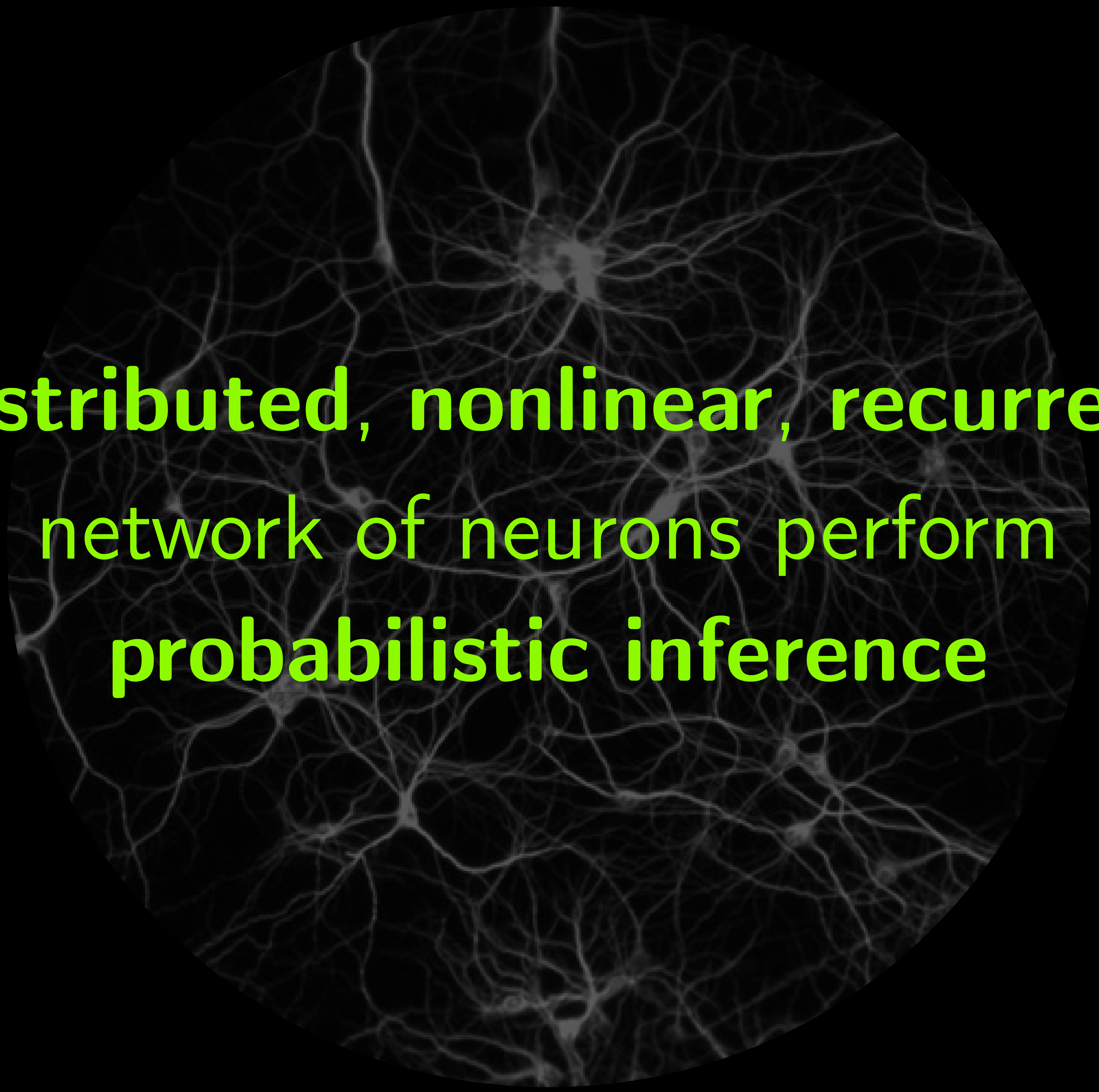
Information is distributed

Nonlinear transformations

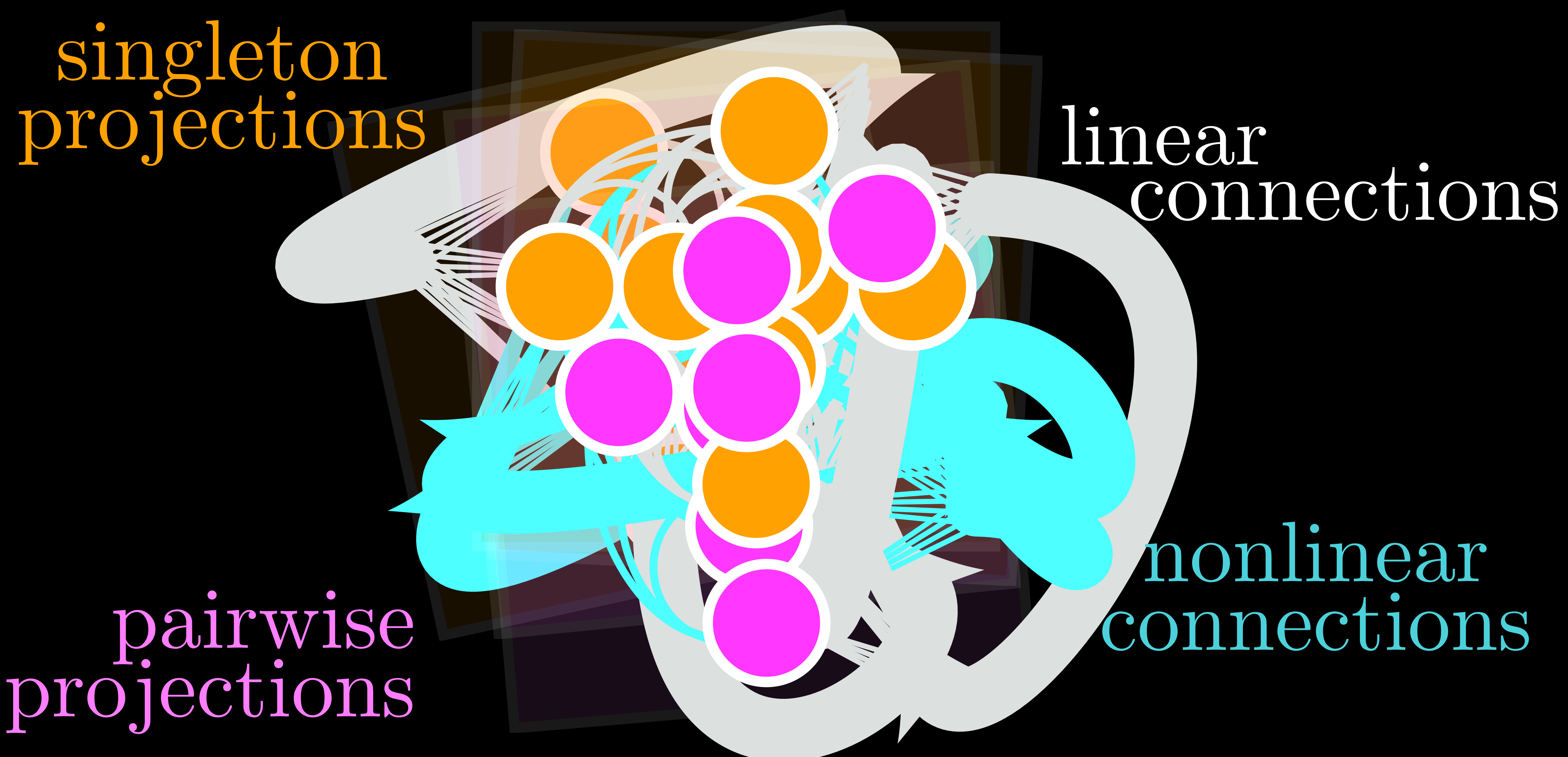


Recurrent connections

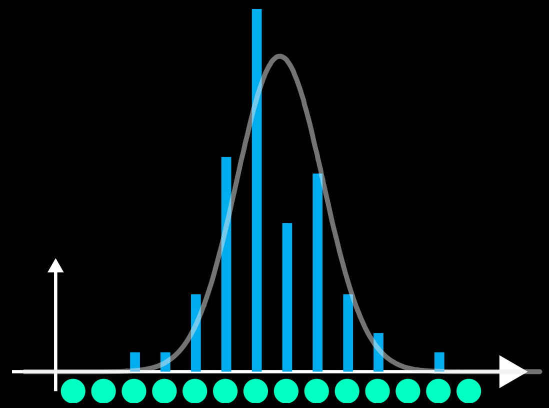




Distributed, **nonlinear**, **recurrent**
network of neurons perform
probabilistic inference

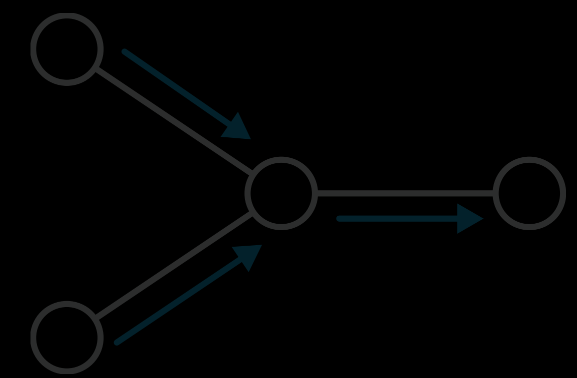

singleton projections
linear connections
nonlinear connections
pairwise projections

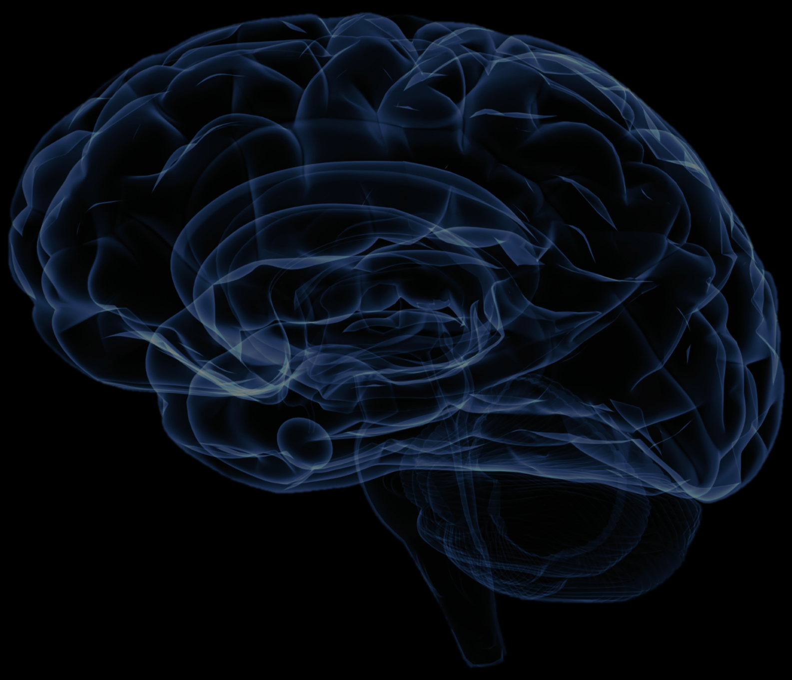
Probabilistic Population Codes



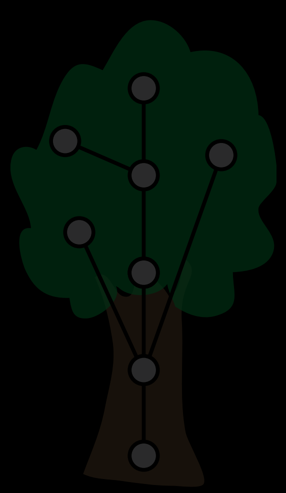
Graphical Models

Belief Propagation

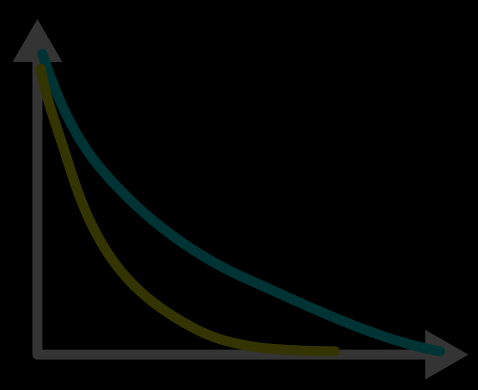




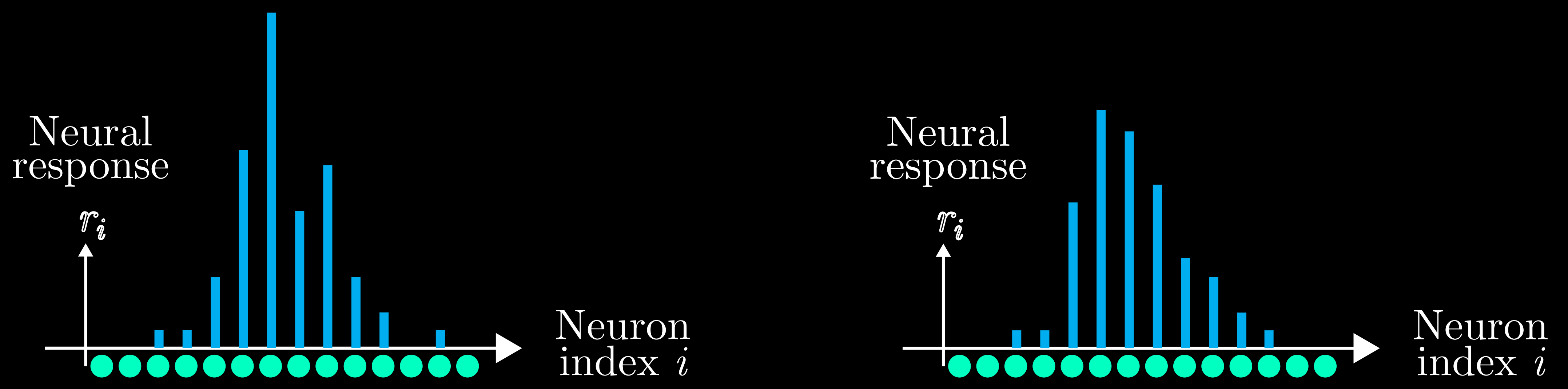
Tree-based Re-parameterization



Dynamical Systems



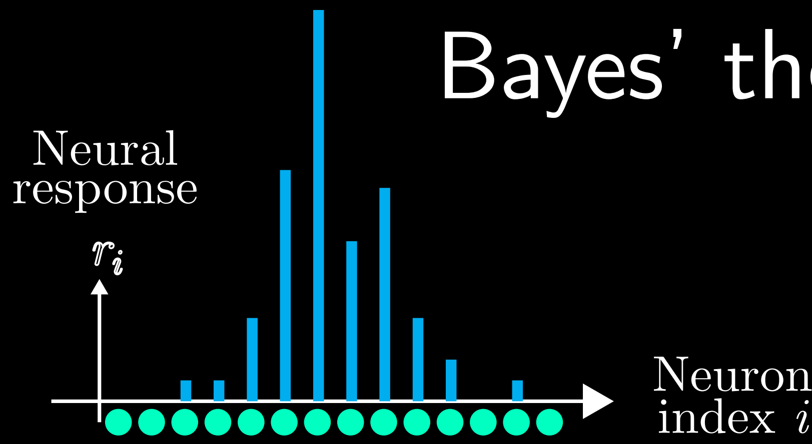
Probabilistic Population Codes

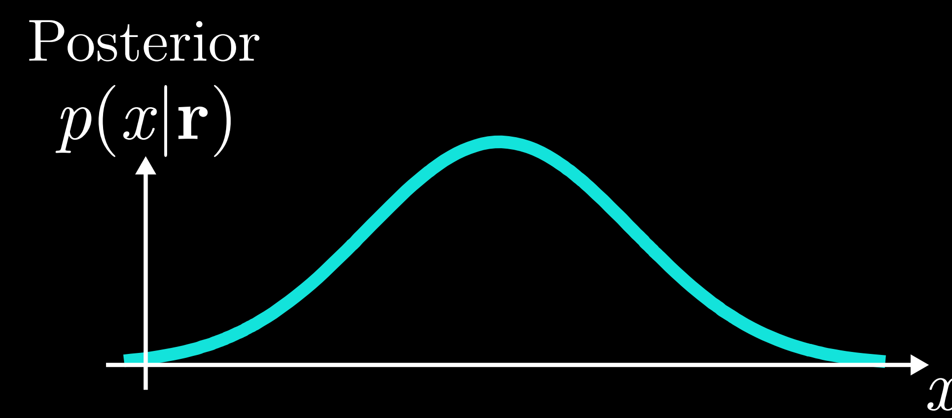


Variability can be captured as $p(\mathbf{r}|x)$

x : stimulus

Using
Bayes' theorem





Linear PPC

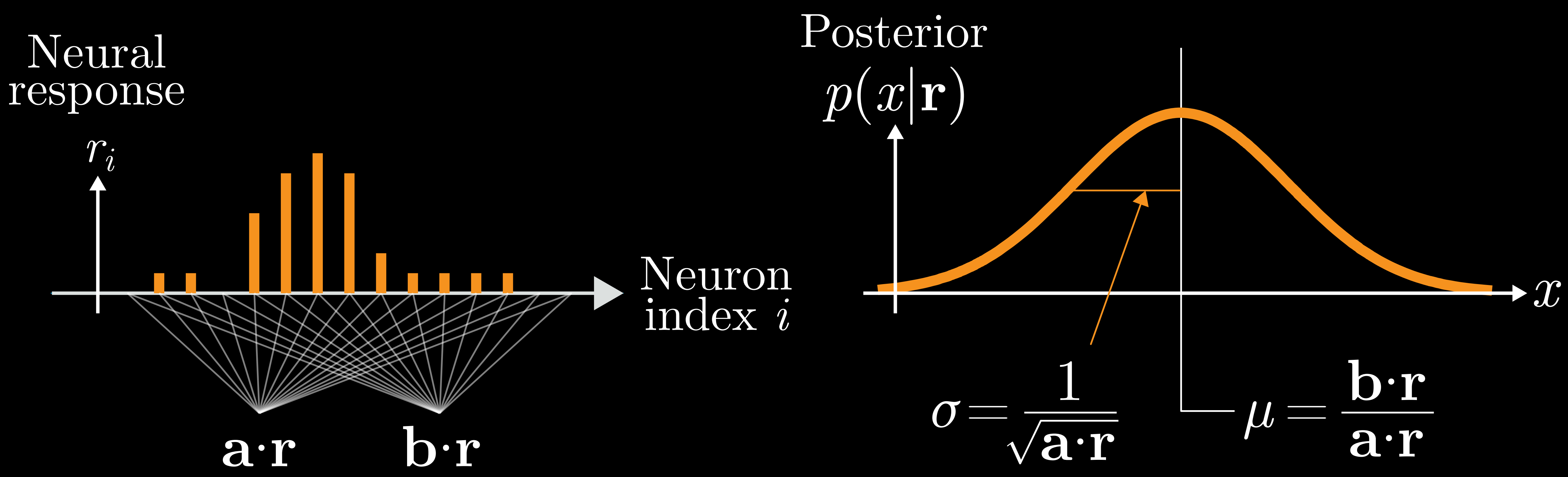
$$p(\mathbf{r}|x) = \phi(\mathbf{r}) \exp(\mathbf{h}(x) \cdot \mathbf{r})$$

Log probability is linear in $\mathbf{r}$

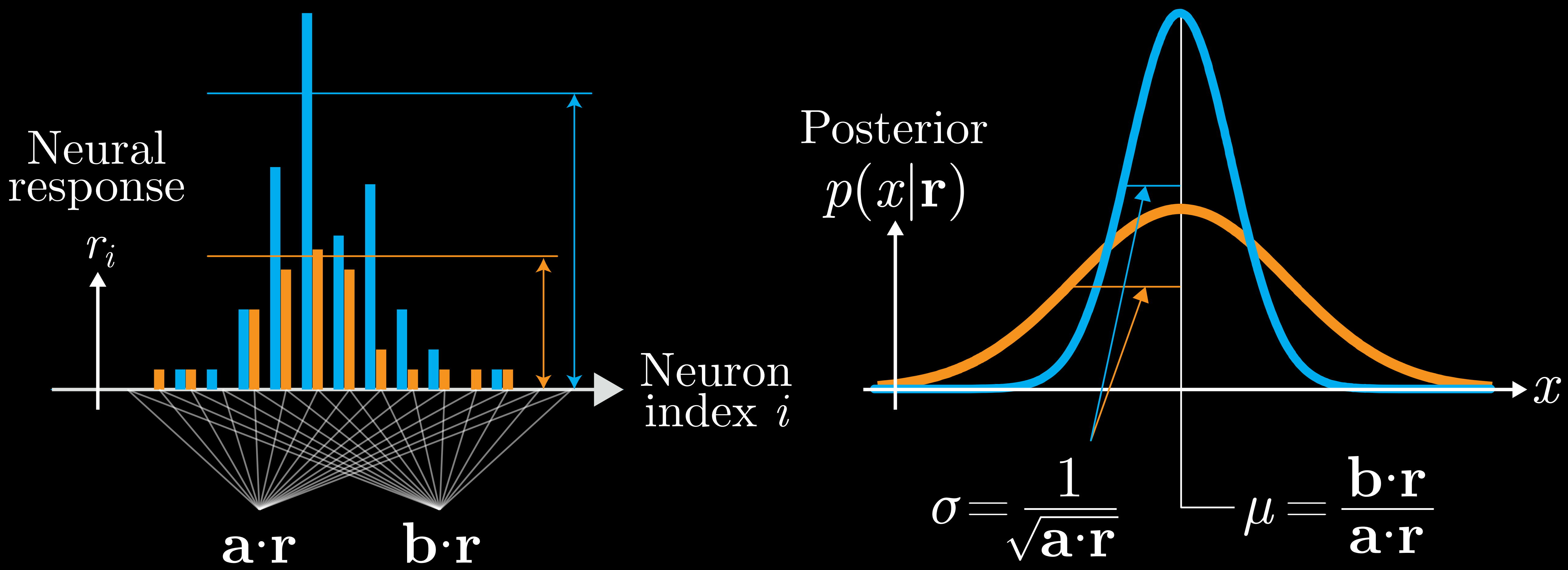
Posterior distribution will also have this general form:

$$p(x|\mathbf{r}) \propto \exp(\mathbf{h}(x) \cdot \mathbf{r})$$

Ma et al. 2006



Different projections of $\mathbf{r}$ encode natural parameters of the underlying distribution



More spikes means more certainty

Probabilistic operations with PPCs

Cue integration

$$p(x|c_1, c_2) \propto p(x|c_1)p(x|c_2)p(x)$$

Linear transformation of $\mathbf{r}$

Marginalization

$$p(x_1) = \int p(x_1, x_2)dx_2$$

Nonlinear

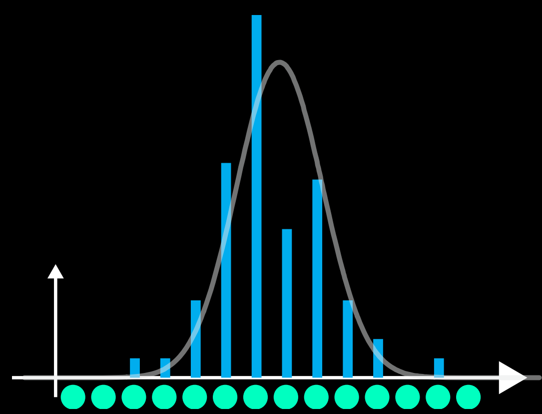
Marginalization

$$p(x_i) = \int p(\mathbf{x}) d\mathbf{x}_{\setminus i}$$

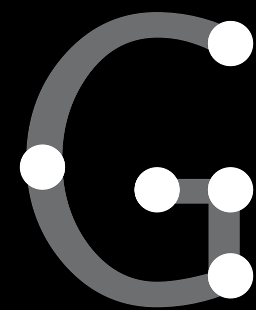
Arises in a wide range of computations in the brain:

causal reasoning, visual tracking, motor control, olfaction, object recognition, decision making ...

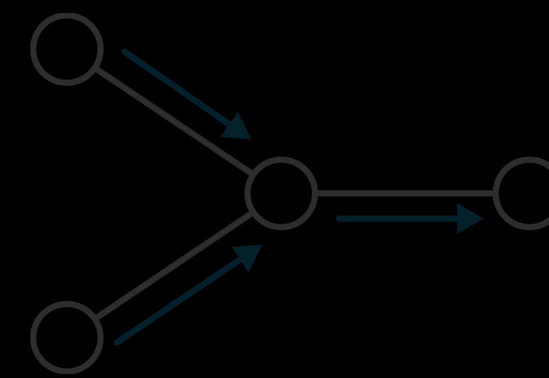
Probabilistic Population Codes

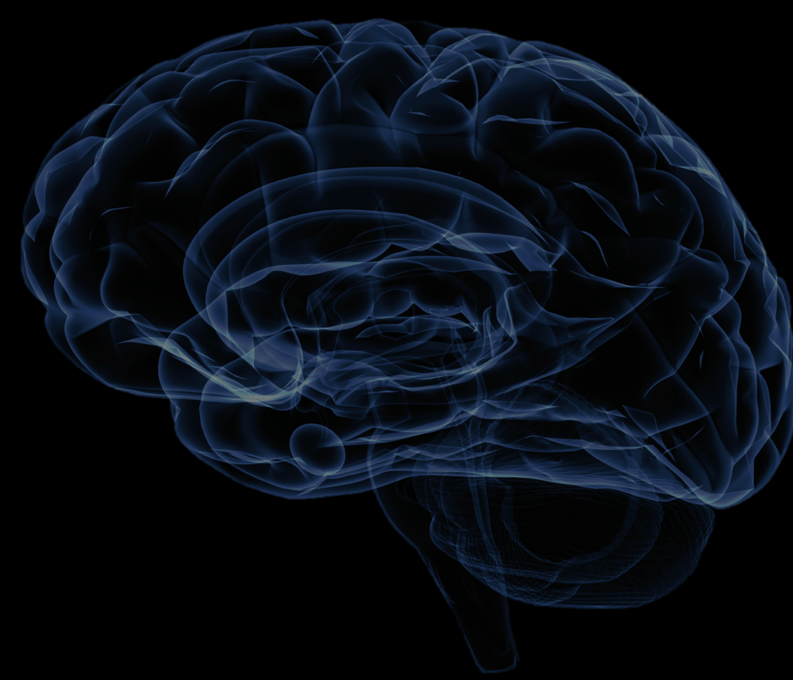


Graphical Models

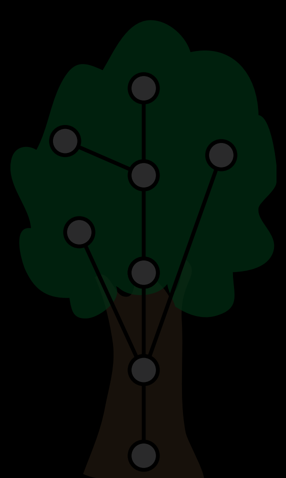


Belief Propagation

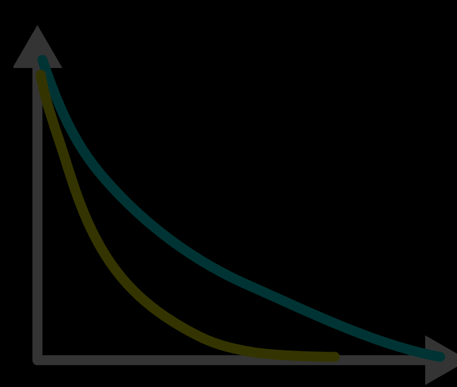




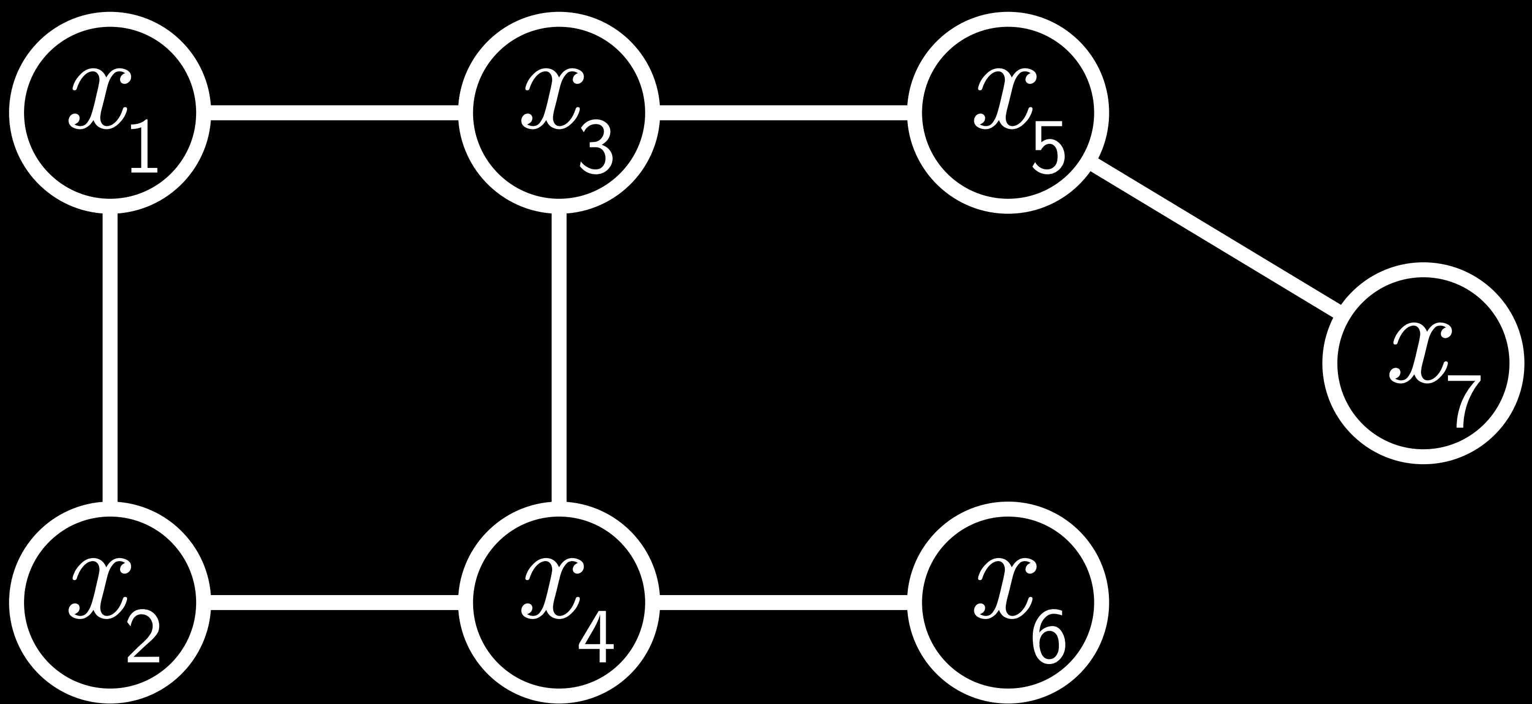
Tree-based Re-parameterization



Dynamical Systems



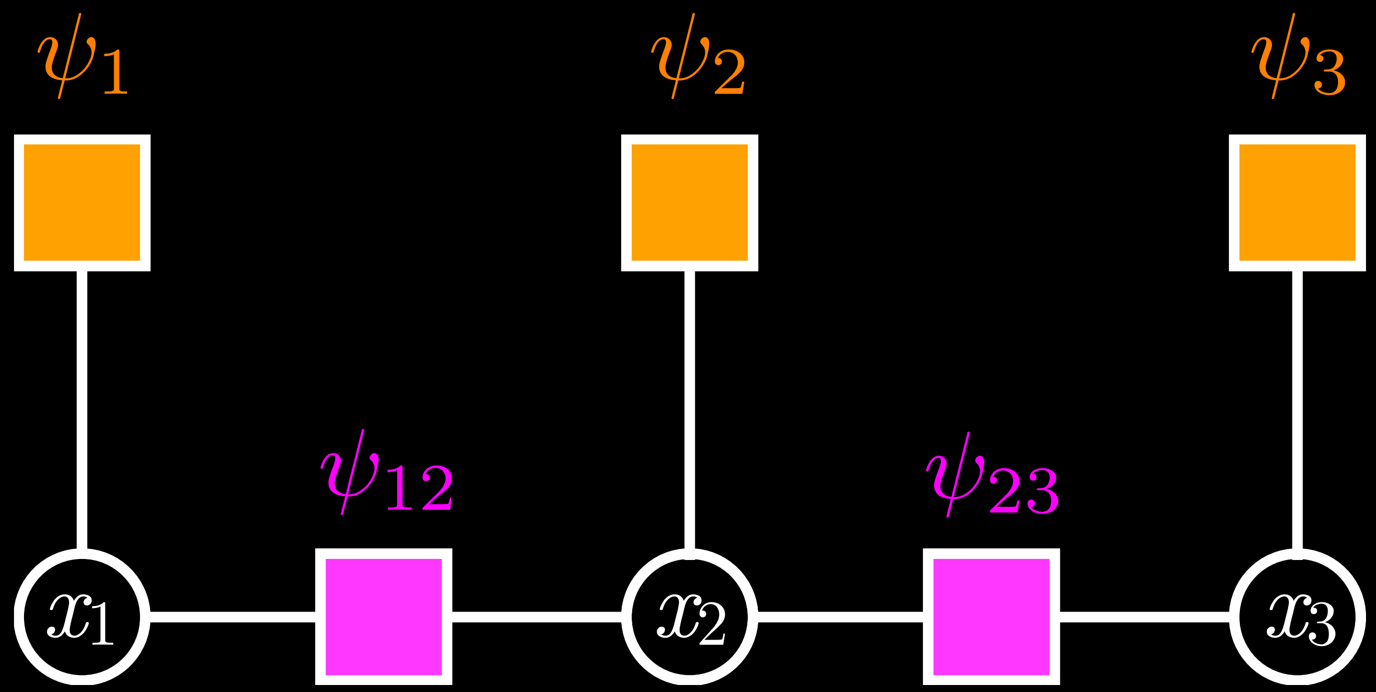
Probabilistic Graphical Models



Graph expresses conditional dependencies between random variables

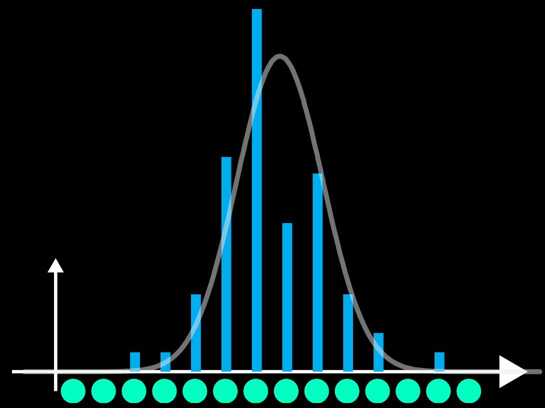
Factorization property: $p(\mathbf{x}) = \frac{1}{Z} \prod_{C \in \mathbf{C}} \psi_C(\mathbf{x}_C)$

Pairwise Markov Random Field

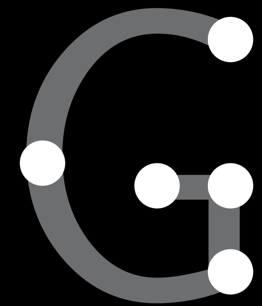


$$p(\mathbf{x}) = \frac{1}{Z} \prod_{s \in V} \boxed{\psi_s(x_s)} \prod_{(s,t) \in E} \boxed{\psi_{st}(x_s, x_t)}$$

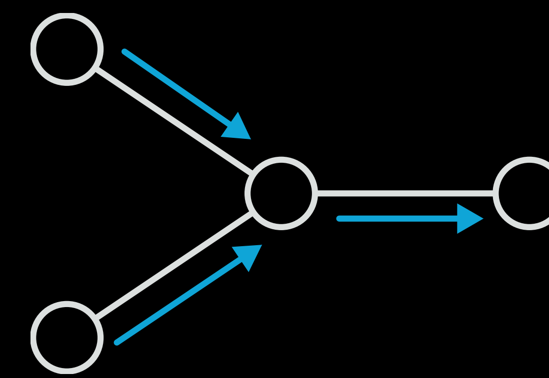
Probabilistic Population Codes

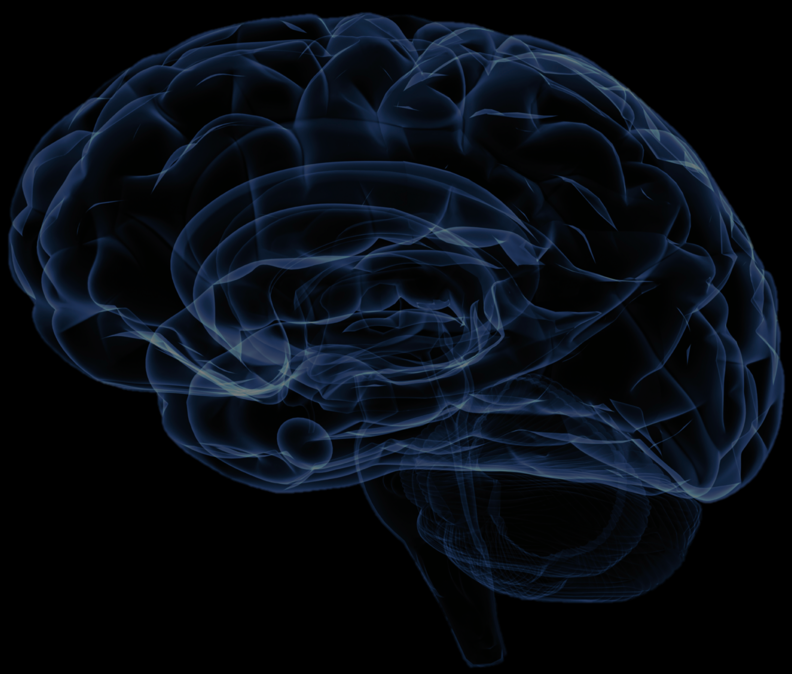


Graphical Models

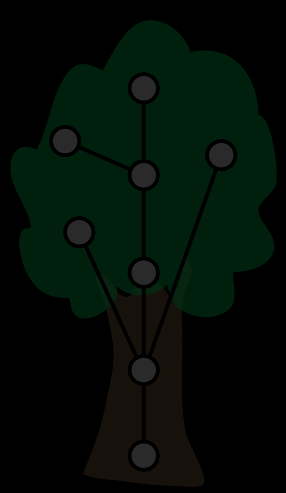


Belief Propagation

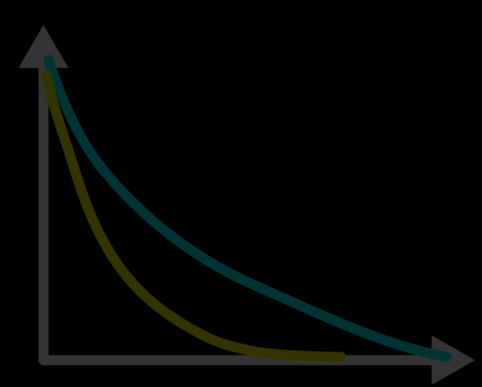




Tree-based Re-parameterization



Dynamical Systems

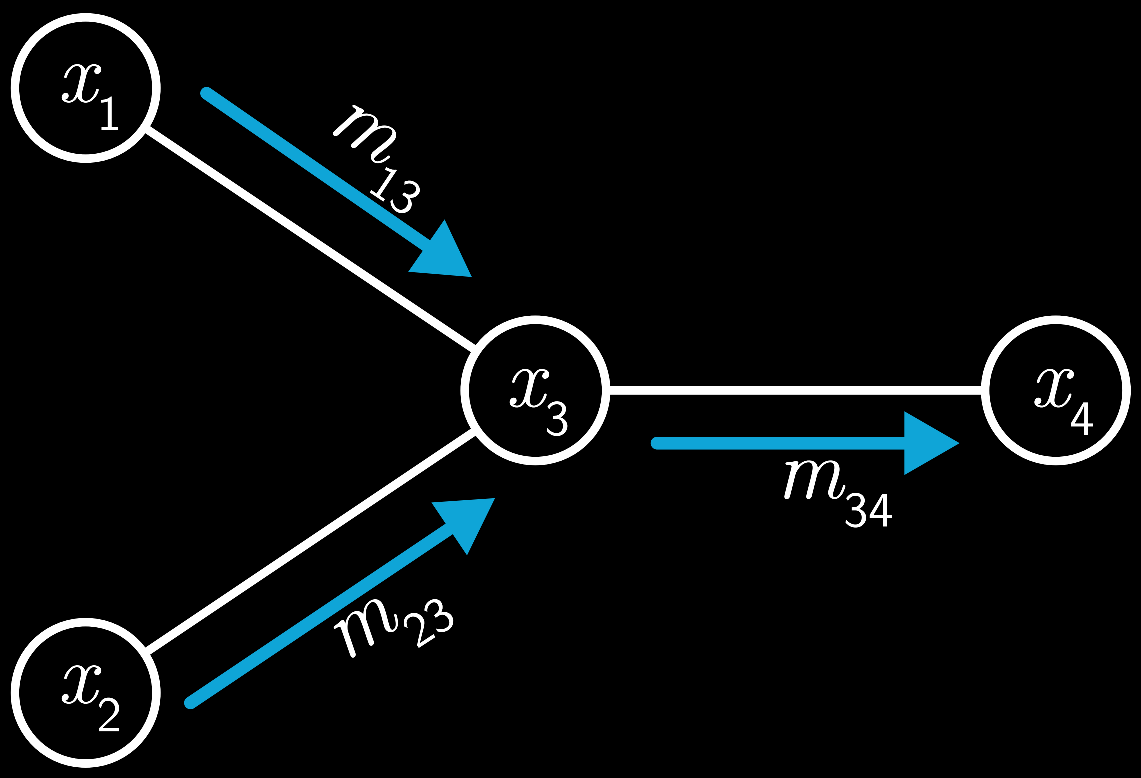


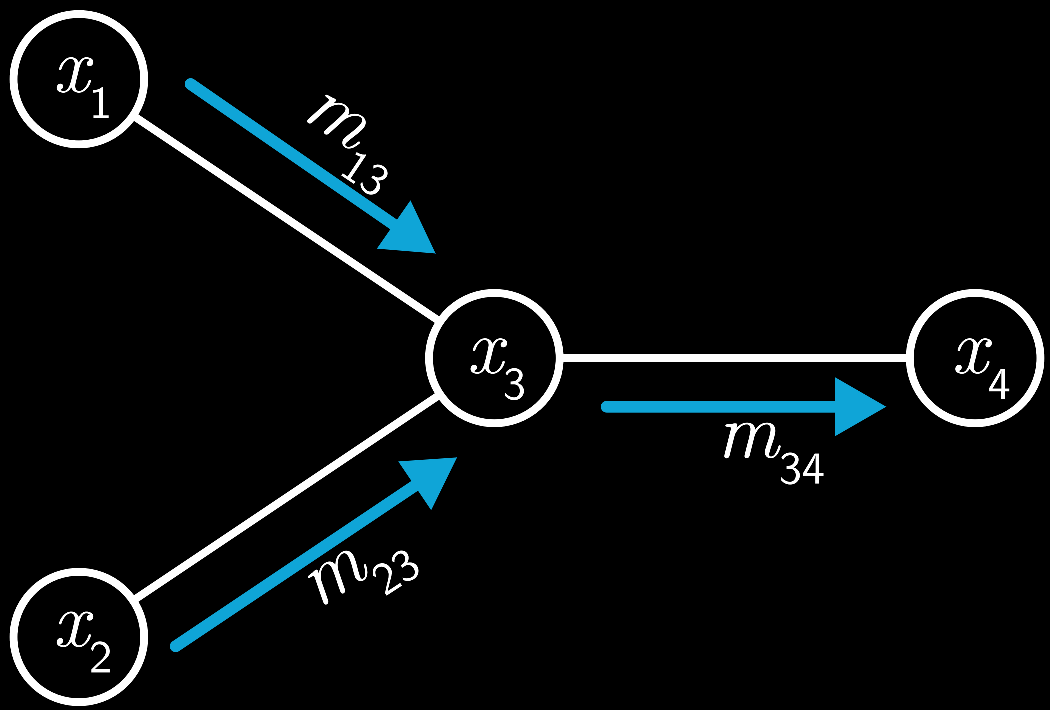
Belief Propagation

Compute marginals:

$$p(x_i) = \int p(\mathbf{x}) d\mathbf{x}_{\setminus i}$$

Standard, efficient inference algorithm which uses **localized computations** and **message passing**





Belief at node s :

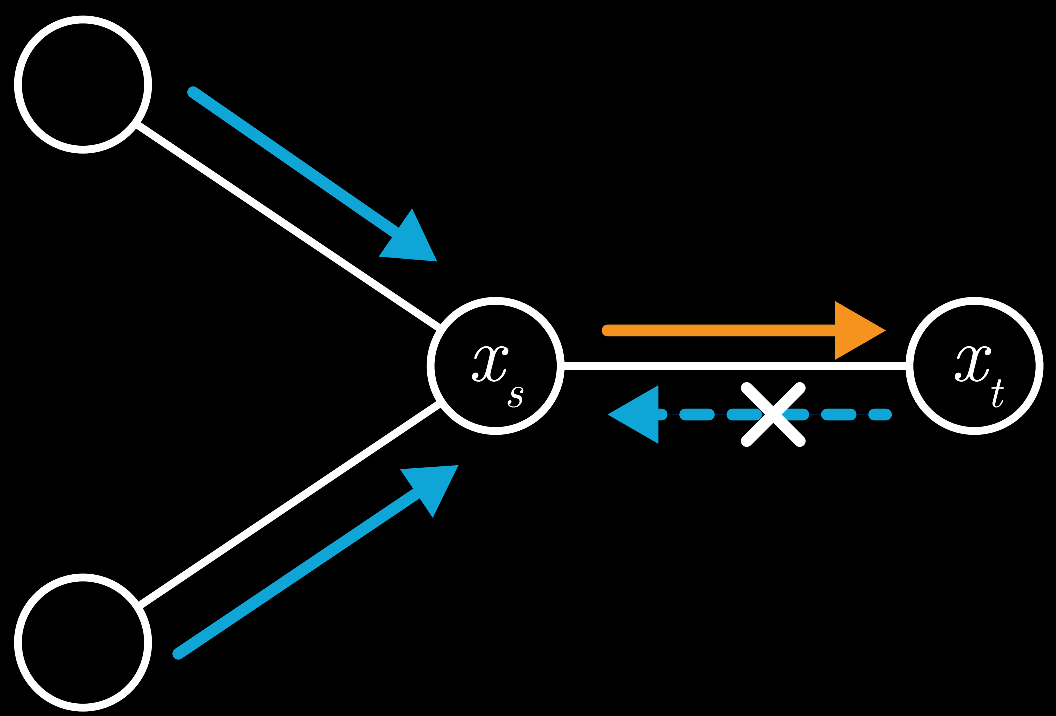
$$b_s(x_s) = \kappa \psi_s(x_s) \prod_{u \in N(s)} m_{us}(x_s)$$

Message update rule:

$$m_{st}^{n+1}(x_t) = \int_{x_s} dx_s \psi_s \psi_{st} \prod_{u \in N(s) \setminus t} m_{us}^{n}(x_s)$$

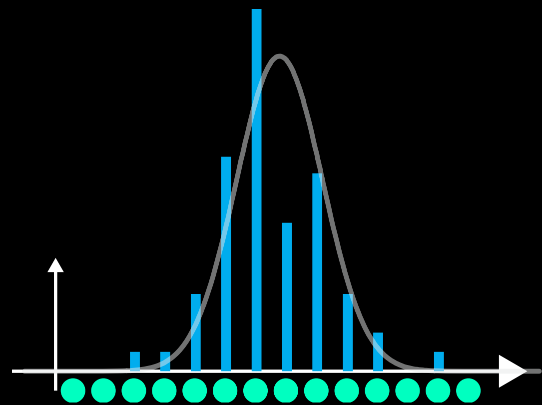
- Exact for trees, approximate for loopy graphs
- E.g. satellite communications

$$\boxed{m_{st}^{n+1}}(x_t) = \int_{x_s} dx_s \psi_s \psi_{st} \prod_{u \in N(s) \setminus t} \boxed{m_{us}^{n}}(x_s)$$

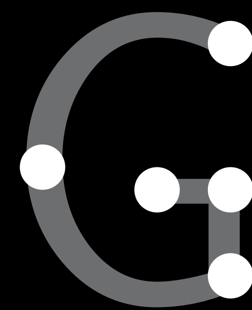


Neurons cannot readily send many different signals to many different target neurons

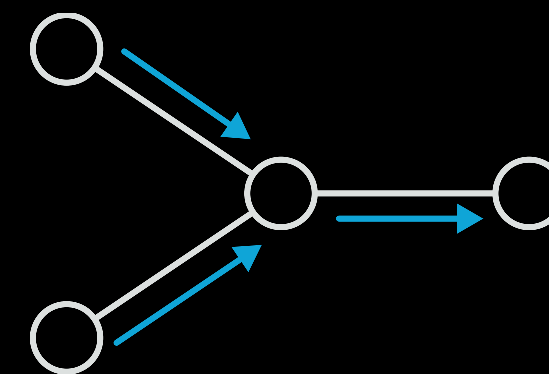
Probabilistic Population Codes

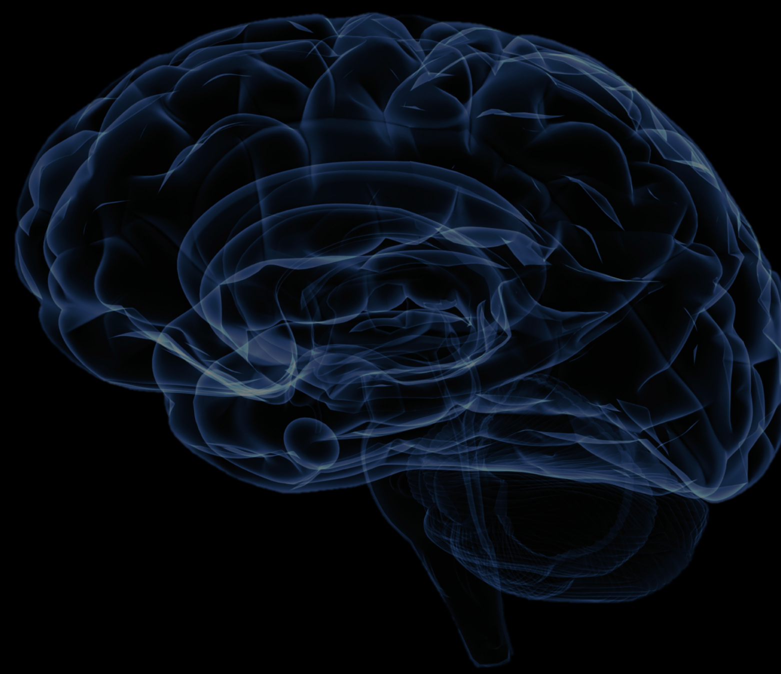


Graphical Models

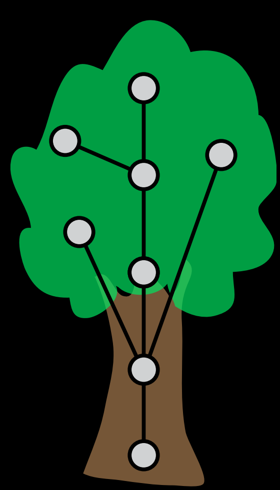


Belief Propagation

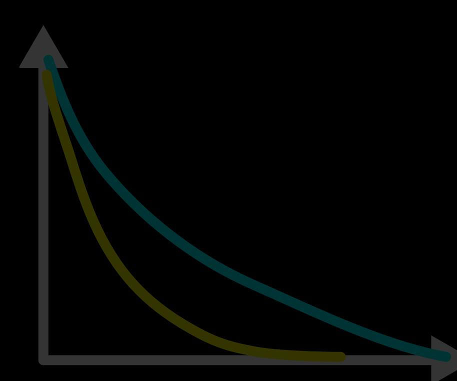




Tree-based Re-parameterization

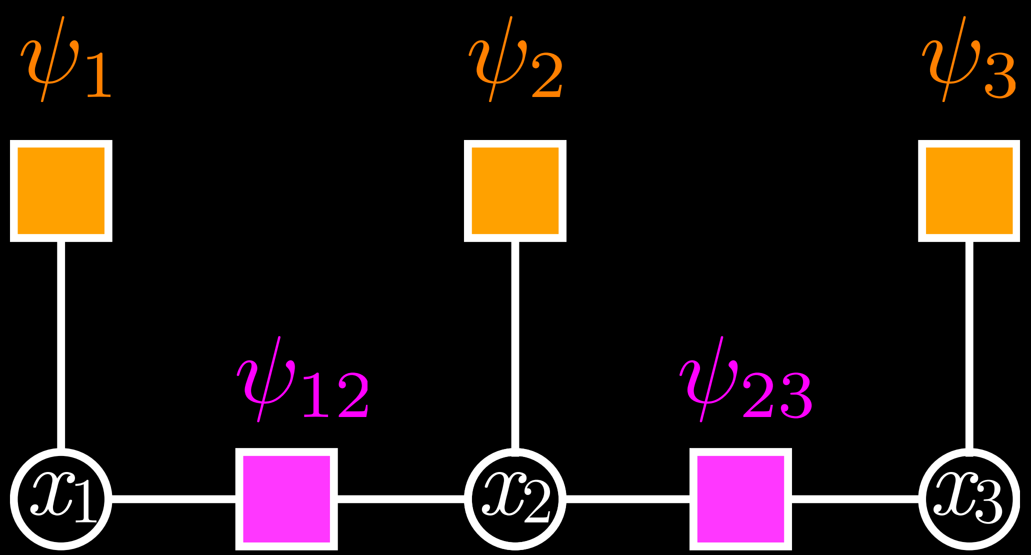


Dynamical Systems



Tree-based Re-parameterization

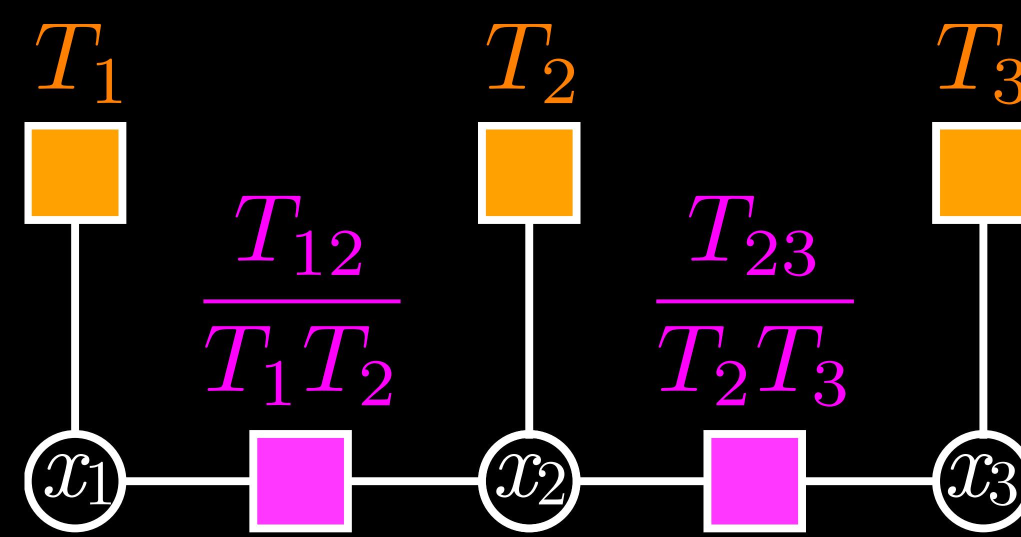
Original



$$\psi_s(x_s), \psi_{st}(x_s, x_t)$$

ψ :Factors

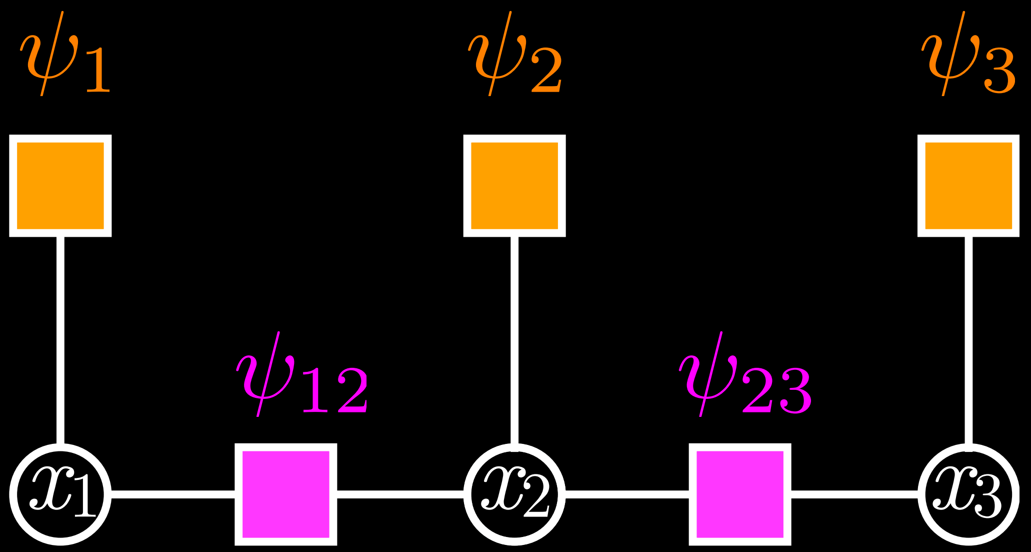
Re-parameterized

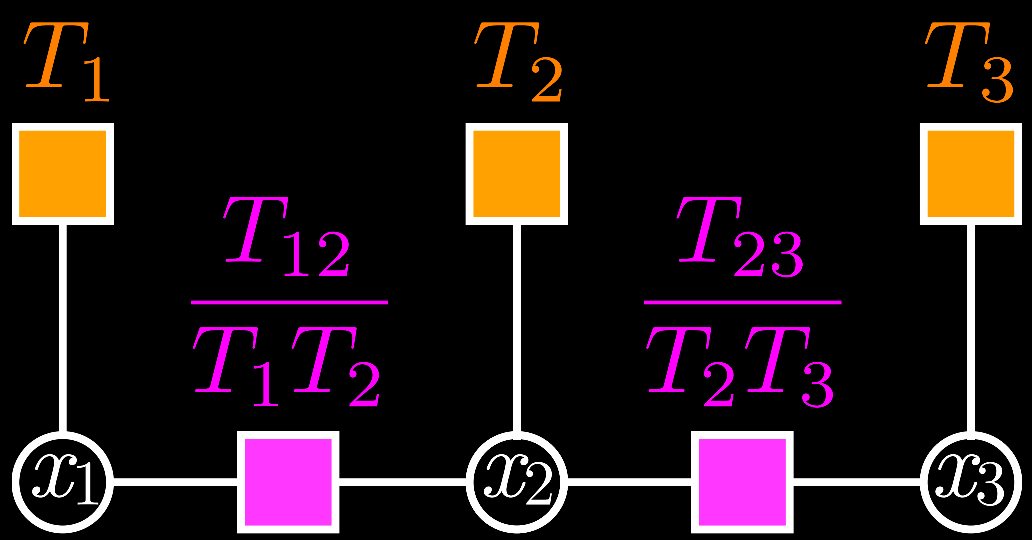


$$T_s(x_s),\ T_{st}(x_s, x_t)$$

T :Marginals

Wainwright et al., 2003.

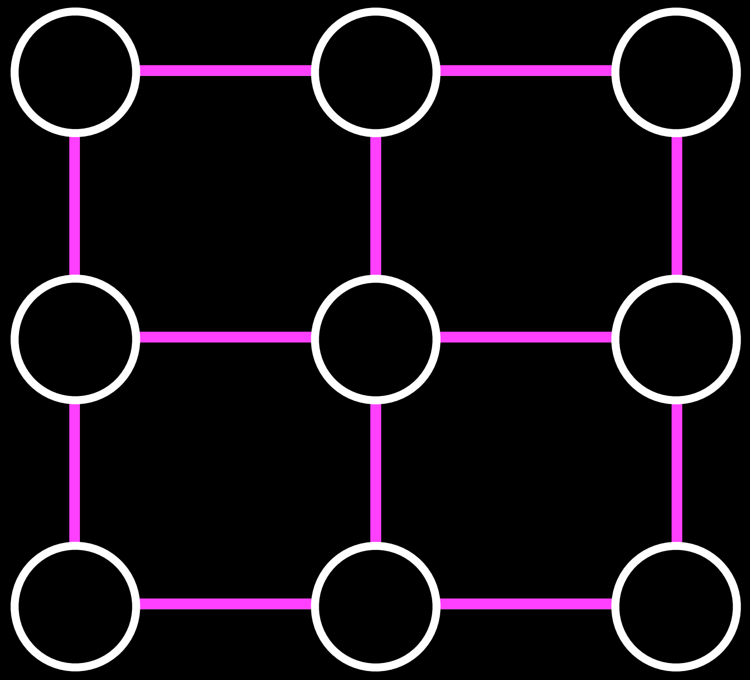




$$p(\mathbf{x}) = \frac{1}{Z} \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

$$p(\mathbf{x}) = \prod_{s \in V} T_s(x_s) \prod_{(s,t) \in E} \frac{T_{st}(x_s, x_t)}{T_s(x_s) T_t(x_t)}$$

For loopy graphs, re-factorize acyclic subgraphs iteratively



$$p(\boldsymbol{x}) = \boxed{p^\tau(\boldsymbol{x})} r^\tau(\boldsymbol{x})$$

- Refactor tree in terms of *pseudomarginals* (T_s, T_{st})
- Pick a different tree and repeat

BP as TRP using extremely simple trees: **edges**

At iteration $n+1$, the pseudomarginals are computed as:

$$T_s^{n+1} \propto T_s^n \prod_{u \in N(s)} \frac{1}{T_s^n} \int T_{su}^n \, dx_u$$

$$T_{st}^{n+1} \propto \frac{T_{st}^n}{\left(\int T_{st}^n \, dx_t\right)\left(\int T_{st}^n \, dx_s\right)} T_s^{n+1} T_t^{n+1}$$

Operations on neighbors do not differentiate between targets

TRP for Exponential Family

Exponential family of distributions:

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \exp(\boldsymbol{\theta}) \cdot \boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{\theta}) + \boldsymbol{B}(\boldsymbol{x}))$$

$\boldsymbol{\theta}$: natural parameters

$\phi(\mathbf{x})$: sufficient statistics

Pseudomarginal for a clique, at iteration n :

$$T_c^n(\boldsymbol{x}_c) = \exp\left(\boldsymbol{\theta}_c^n \cdot \boldsymbol{\phi}_c(\boldsymbol{x}_c)\right)$$

Updates in terms of the natural parameters:

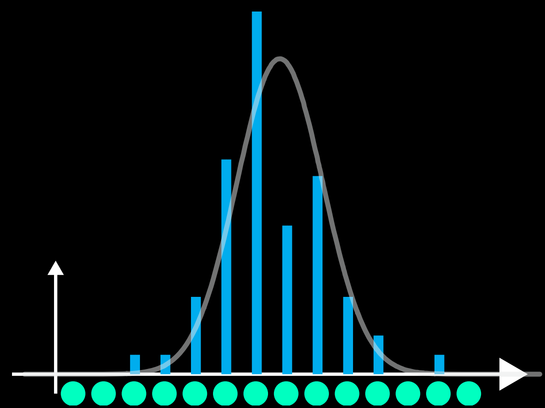
$$T_s^{n+1} \propto T_s^n \prod_{u \in N(s)} \frac{1}{T_s^n} \int T_{su}^n \, dx_u$$

$$\boldsymbol{\theta}_s^{n+1} = (1 - d_s)\boldsymbol{\theta}_s^n + \sum_{u \in N(s)} \mathbf{g}_V(\boldsymbol{\theta}_{su}^n)$$

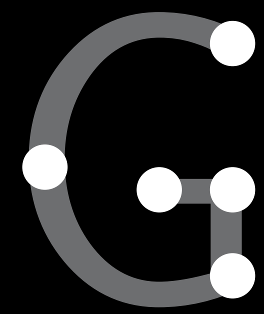
$$T_{st}^{n+1} \propto \frac{T_{st}^n}{\left(\int T_{st}^n \, dx_t\right)\left(\int T_{st}^n \, dx_s\right)} T_s^{n+1} T_t^{n+1}$$

$$\boldsymbol{\theta}_{st}^{n+1} = \boldsymbol{\theta}_{st}^n + Q_s \boldsymbol{\theta}_s^{n+1} + Q_t \boldsymbol{\theta}_t^{n+1} + \mathbf{g}_E(\boldsymbol{\theta}_{st}^n)$$

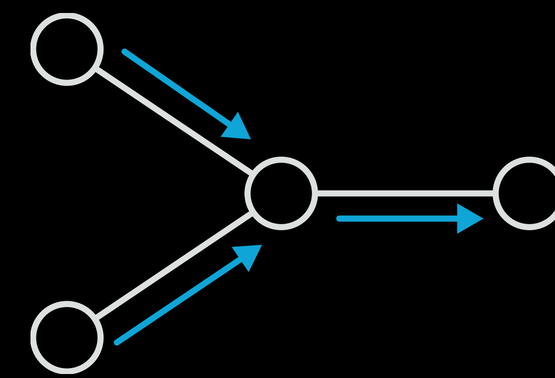
Probabilistic Population Codes

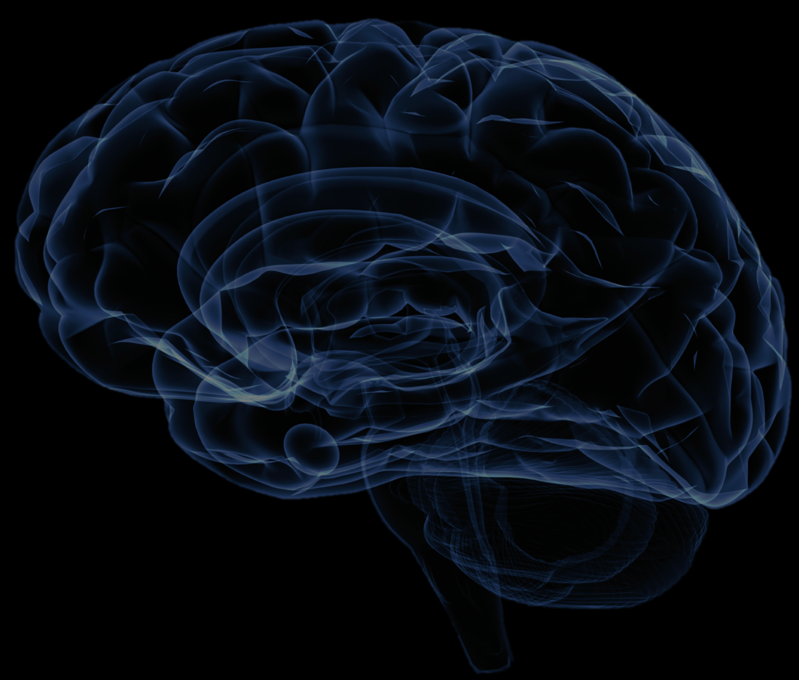


Graphical Models

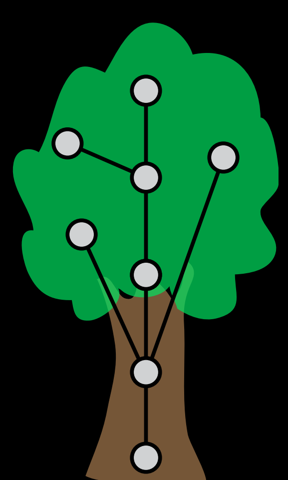


Belief Propagation

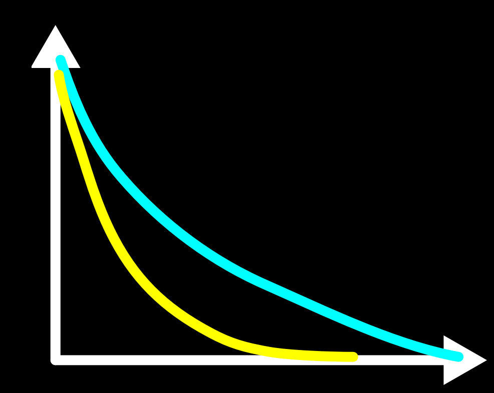




Tree-based Re-parameterization



Dynamical Systems



Separation of Time Scales for TRP updates

TRP updates circumvent the 'message exclusion' problem

$$\boldsymbol{\theta}_s^{n+1} = (1 - d_s)\boldsymbol{\theta}_s^n + \sum_{u \in N(s)} \mathbf{g}_V(\boldsymbol{\theta}_{su}^n)$$

$$\boldsymbol{\theta}_{st}^{n+1} = \boldsymbol{\theta}_{st}^n + Q_s\boldsymbol{\theta}_s^{n+1} + Q_t\boldsymbol{\theta}_t^{n+1} + \mathbf{g}_E(\boldsymbol{\theta}_{st}^n)$$

No free lunch: update for pairwise term is instantaneous

The brain can use fast and slow timescales
instead of instant and delayed signals

Introduce auxiliary variables $\tilde{\boldsymbol{\theta}}$,
convert updates to continuous time

$$\tau_{\text{slow}}\dot{\tilde{\boldsymbol{\theta}}} = -\tilde{\boldsymbol{\theta}} + \boldsymbol{\theta}$$

Nonlinear dynamics are updated on a faster timescale:

$$\tau_{\text{fast}}\dot{\boldsymbol{\theta}}_s = -d_s\tilde{\boldsymbol{\theta}}_s + \sum_{u \in N(s)} \mathbf{g}_V(\tilde{\boldsymbol{\theta}}_{su})$$

$$\tau_{\text{fast}}\dot{\boldsymbol{\theta}}_{st} = Q_s\boldsymbol{\theta}_s + Q_t\boldsymbol{\theta}_t + \mathbf{g}_E(\tilde{\boldsymbol{\theta}}_{st})$$

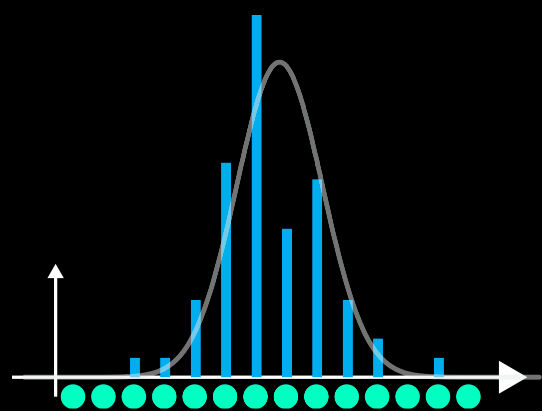
Concatenate the two sets of parameters: $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$

Dynamical system which represents the approximation to the TRP iterations:

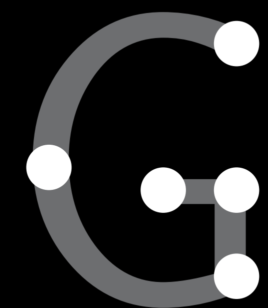
$$\dot{\boldsymbol{\Theta}} = W\boldsymbol{\Theta} + \mathbf{G}(\boldsymbol{\Theta})$$

W and $\mathbf{G}$ inherit their structure from the discrete time updates and the filtering at different timescales

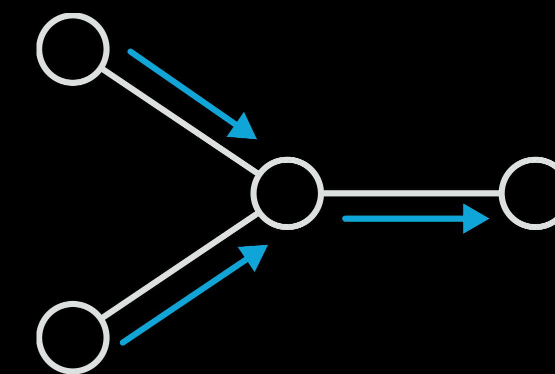
Probabilistic Population Codes



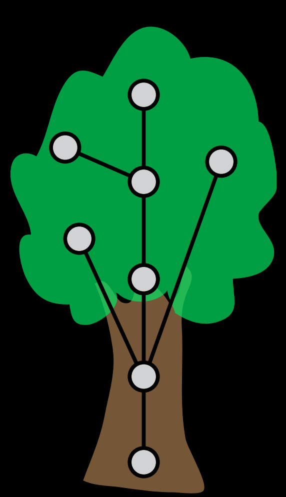
Graphical Models



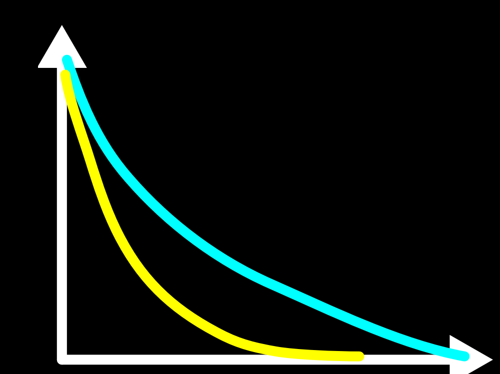
Belief Propagation



Tree-based Re-parameterization



Dynamical Systems



Network Architecture

Map neural activity onto natural parameters:

$$\mathbf{r} = U\mathbf{\Theta}$$

U is a rectangular embedding matrix of size:

$$N_{\mathbf{r}} \times N_{\mathbf{\Theta}}$$

Parameters can be decoded from neural activity:

$$\mathbf{\Theta} = U^{+}\mathbf{r}$$

U^{+} is the pseudoinverse of U

Applying this basis transformation to $\dot{\mathbf{\Theta}} = W\mathbf{\Theta} + \mathbf{G}(\mathbf{\Theta})$ we obtain

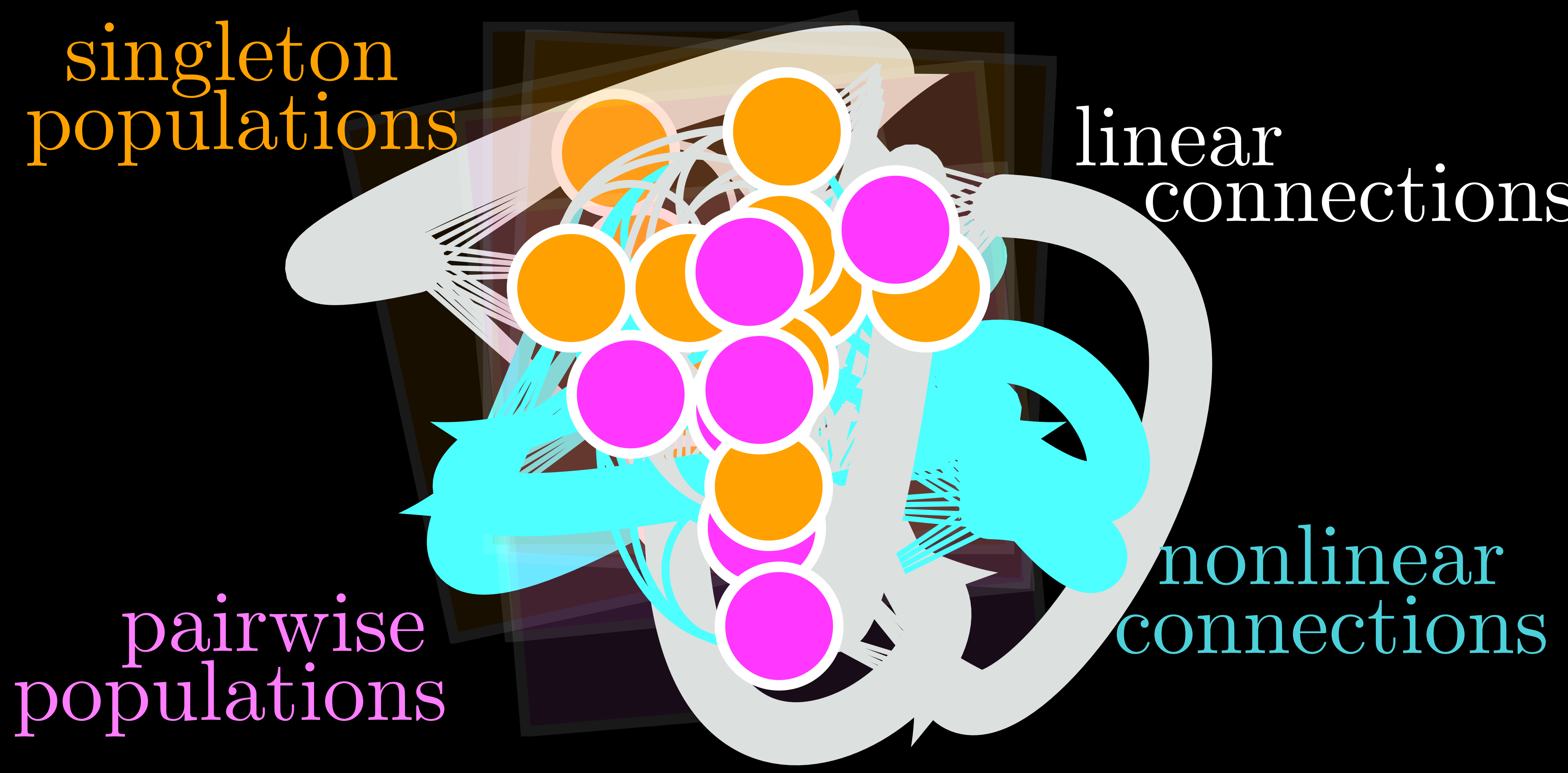
$$\dot{\mathbf{r}} = U\dot{\mathbf{\Theta}} = U(W\mathbf{\Theta} + \mathbf{G}(\mathbf{\Theta})) = UWU^{+}\mathbf{r} + U\mathbf{G}(\mathbf{U}^{+}\mathbf{r})$$

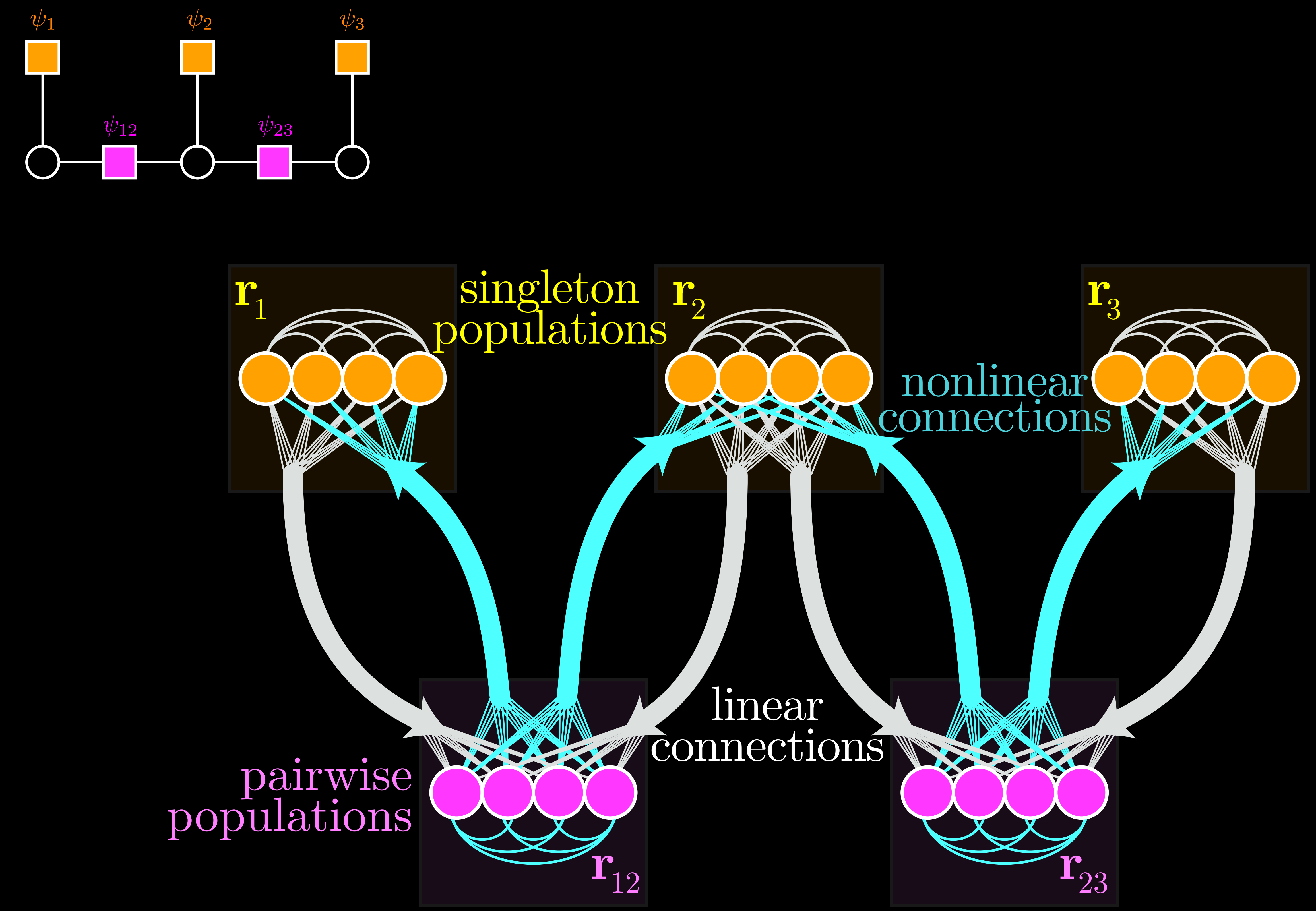
$$\dot{\mathbf{r}} = W_L\mathbf{r} + \mathbf{G}_{NL}(\mathbf{r})$$

Linear computation
Integrate evidence

Nonlinear computation
Marginalize evidence

$$\dot{\mathbf{r}} = W_L \mathbf{r} + \mathbf{G}_{NL}(\mathbf{r})$$



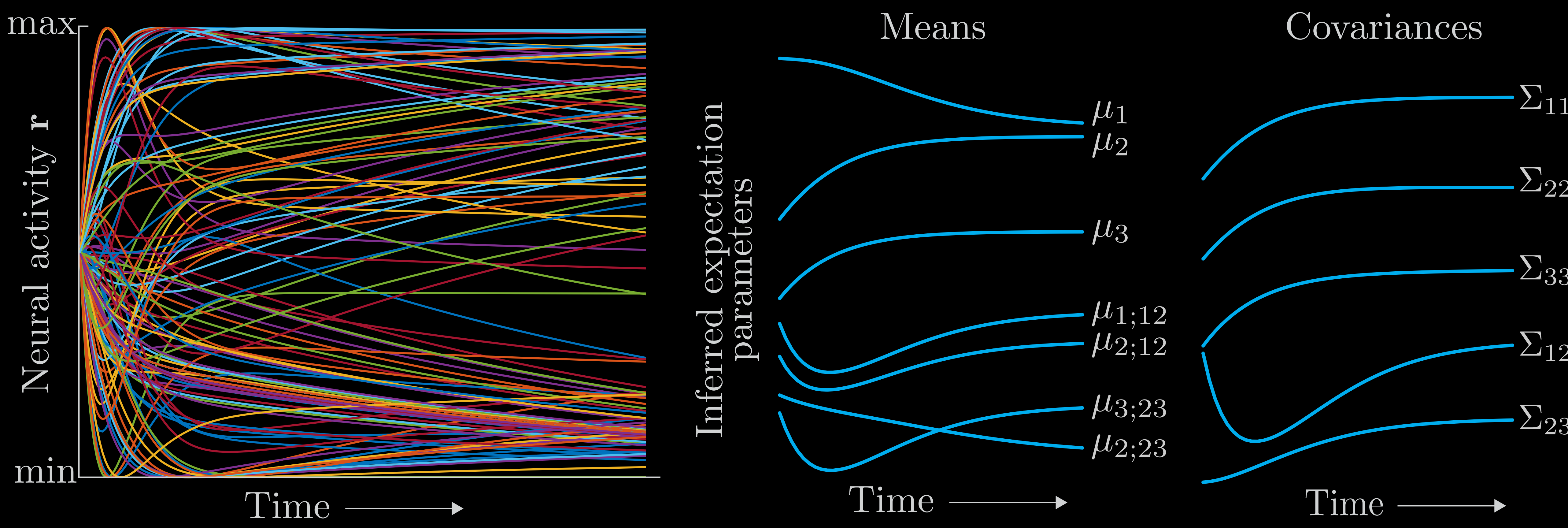


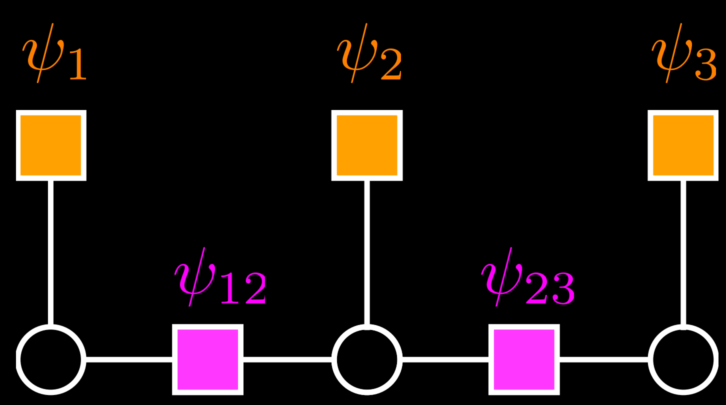
ψ_1
ψ_2
ψ_3
ψ_{12}
ψ_{23}
$\mathbf{r}_1$
singleton populations
$\mathbf{r}_2$
$\mathbf{r}_3$
nonlinear connections
linear connections
pairwise populations
$\mathbf{r}_{12}$
$\mathbf{r}_{23}$

Experiments

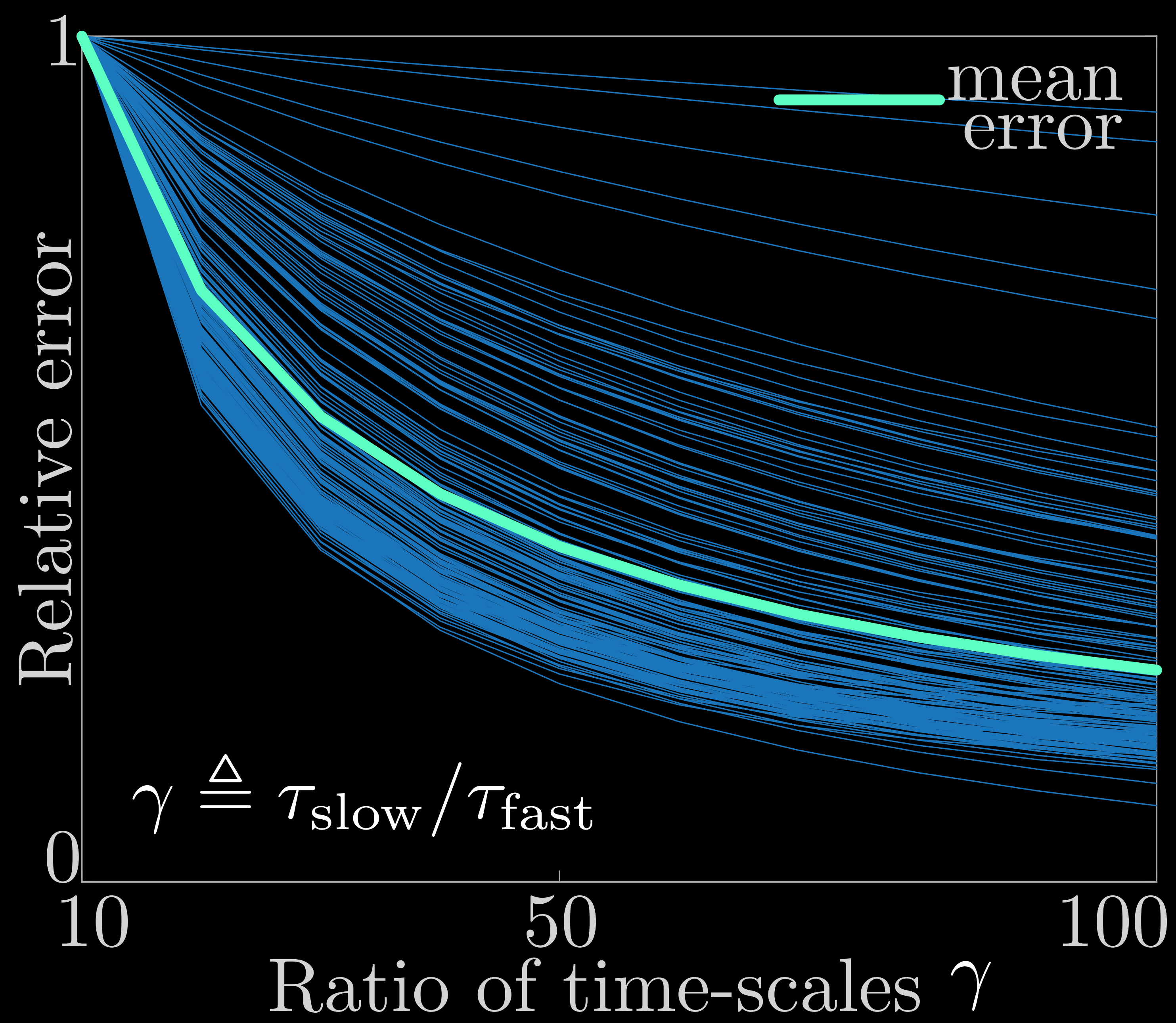
- Performance of the network evaluated on Gaussian graphical models
  - ⋆ of different topologies - chains, loops, grids, dense
  - ⋆ up to 400 variables

- Network time constants were set to have a ratio of $\tau_{\text{slow}}/\tau_{\text{fast}} = 20$

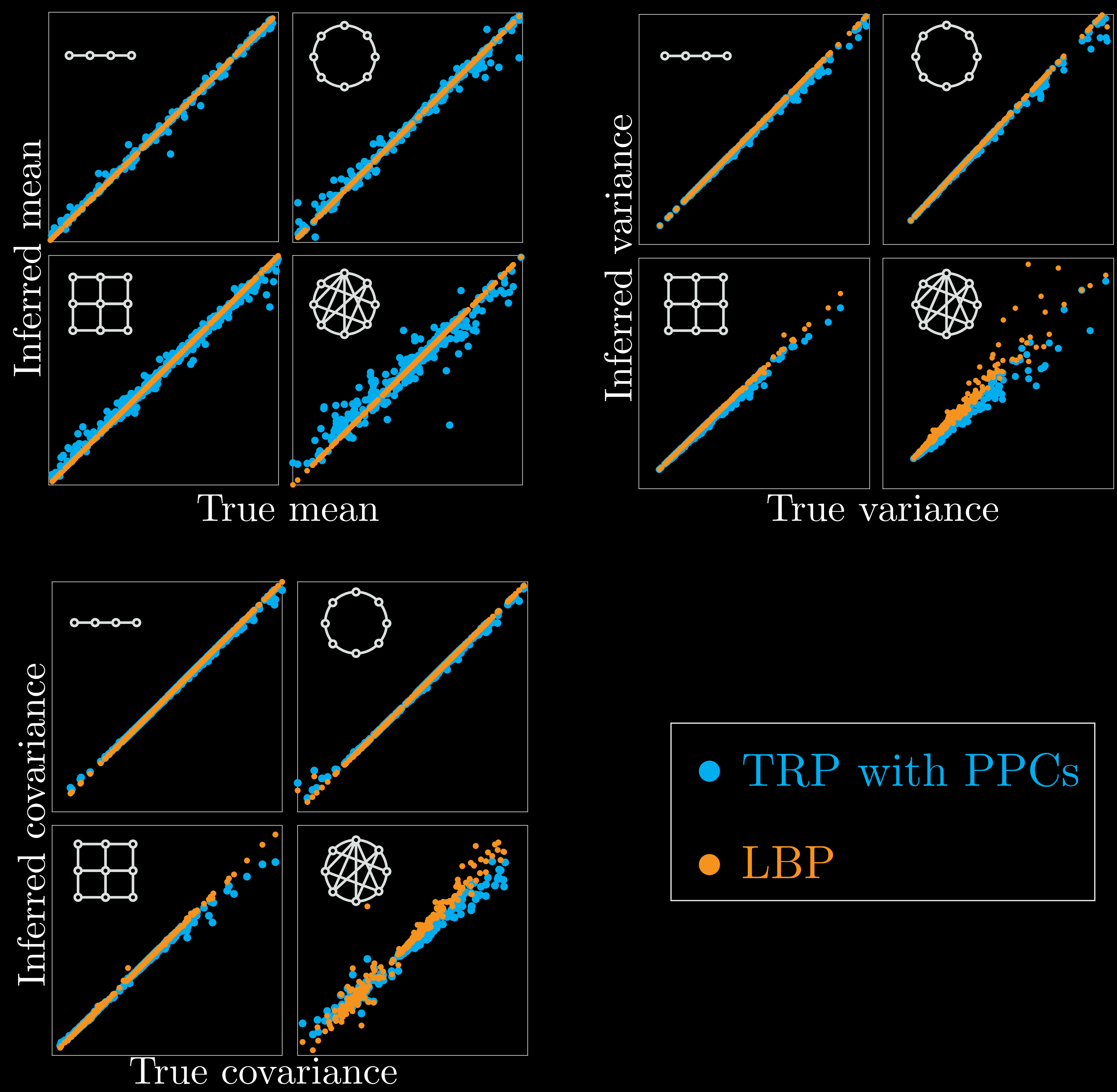




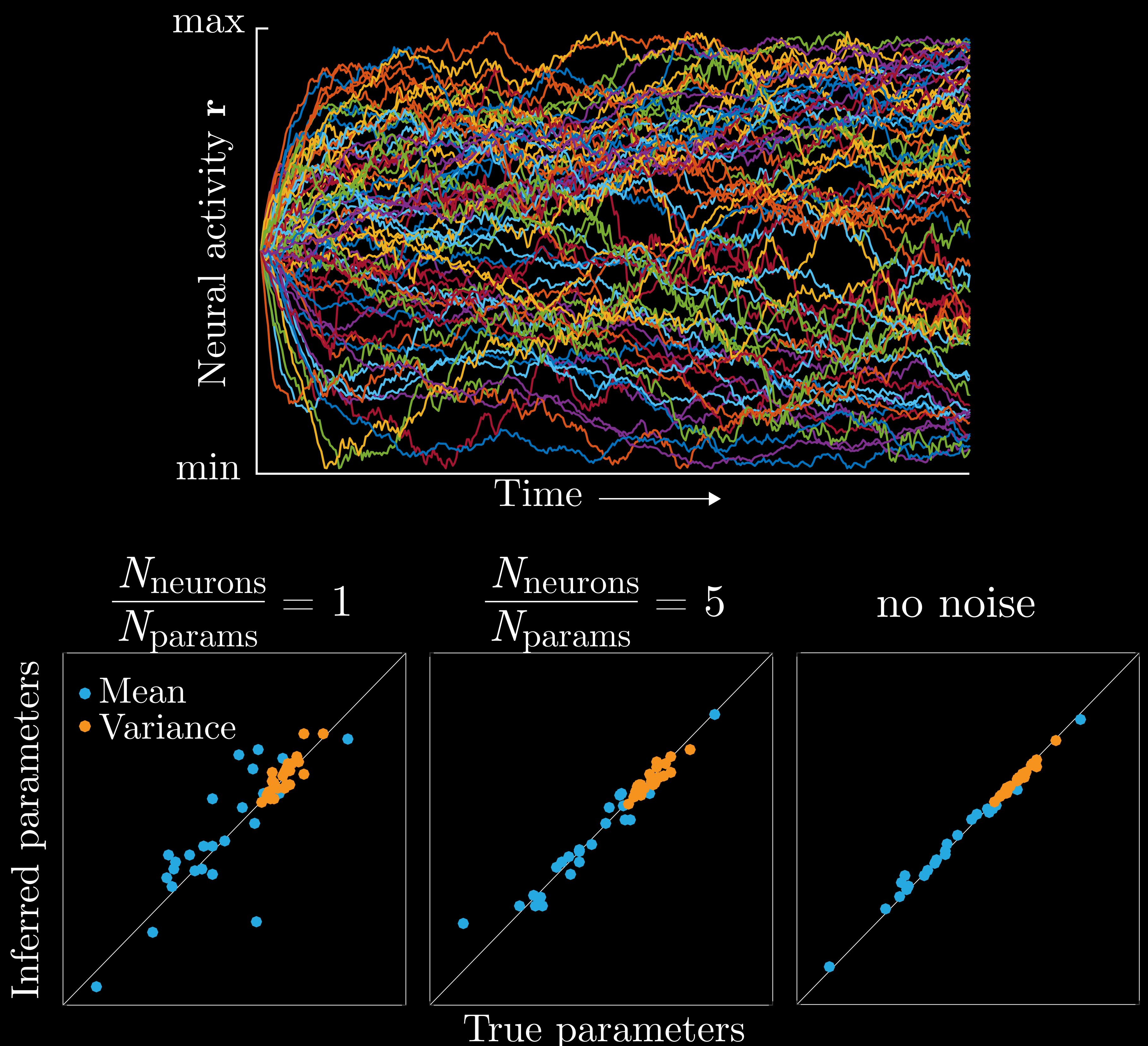
Dynamics of neural population activity and expectation parameters of the underlying distribution for an example graph



Performance is quantified by the mean squared error in inferred parameters for a given γ divided by the error for a reference $\gamma_0 = 10$



Inference performance of our model vs. LBP for various graph topologies



Network performance is robust to noise and improves with more neurons

Conclusion

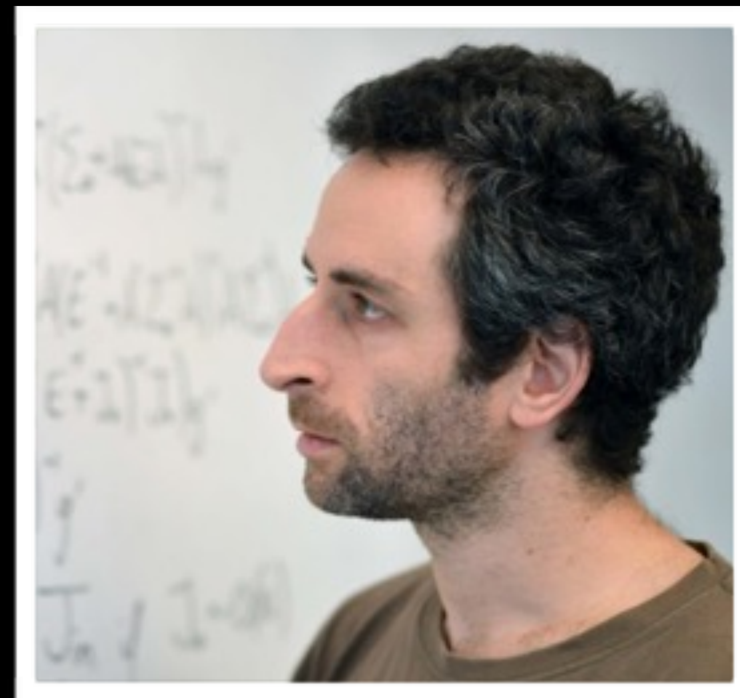
Distributed, **nonlinear**, **recurrent** network of neurons perform **probabilistic inference**

- Neural activity encodes parameters of the distribution
- Nonlinear interactions implement marginalization
- Multiple time scales compensate for message exclusion

Future directions

- Testing detailed biological validity of the model
- Inferring network from data
- Limited to static graphical models
- How can the network learn?

MCNAIR SCHOLAR



xaqlab.com

Baylor College of Medicine