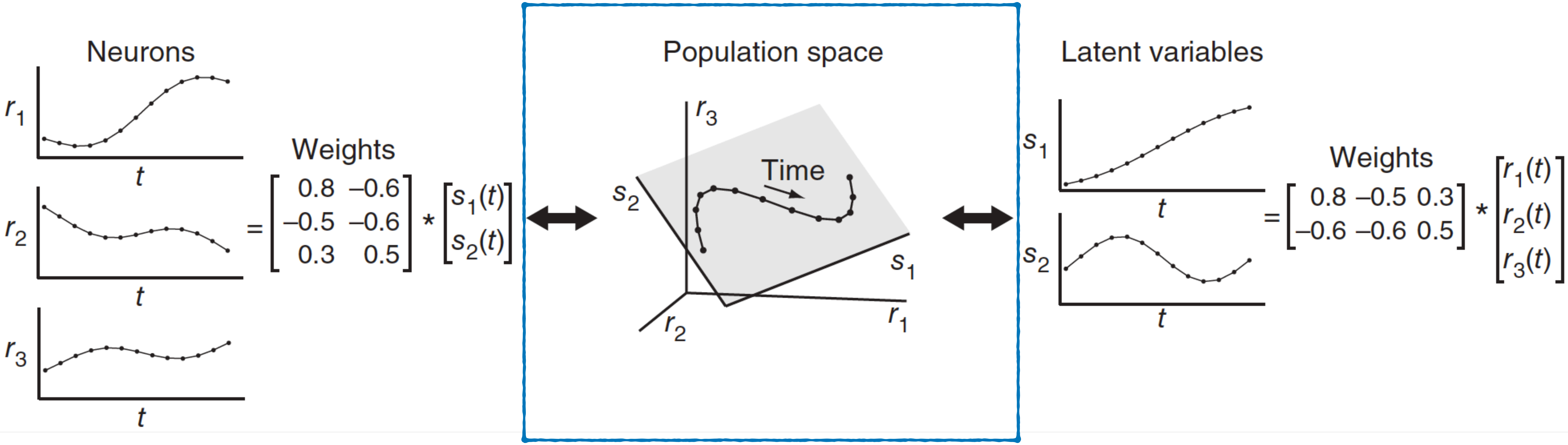
# Neural Dynamics Discovery via Gaussian Process Recurrent Neural Networks


Qi She Intel Labs China

Anqi Wu Princeton University

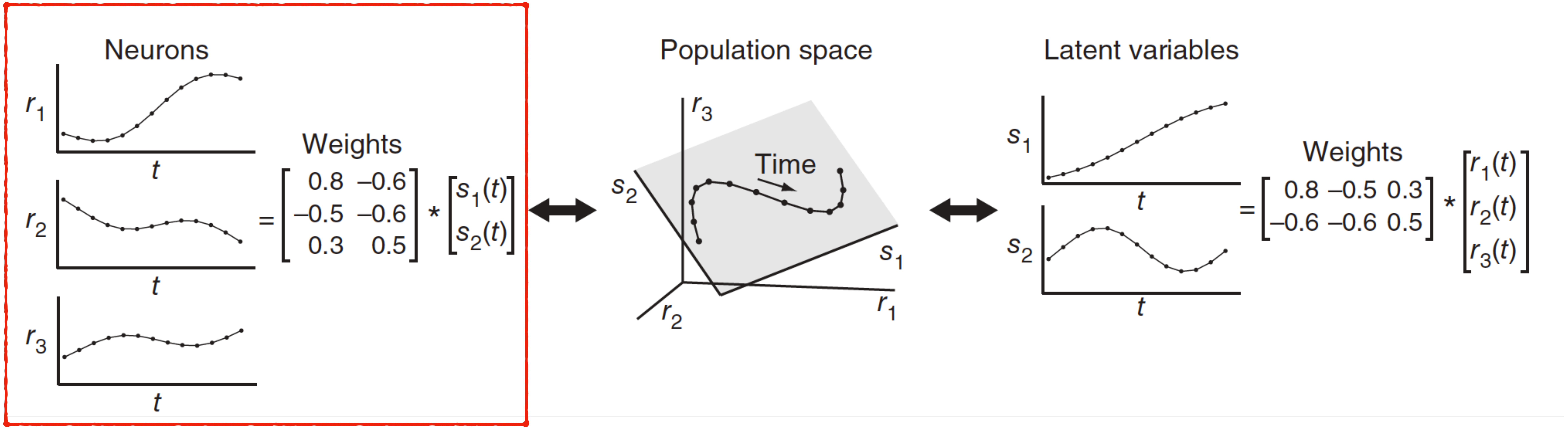

July 23rd, UAI 2019

# Conceptual Illustration



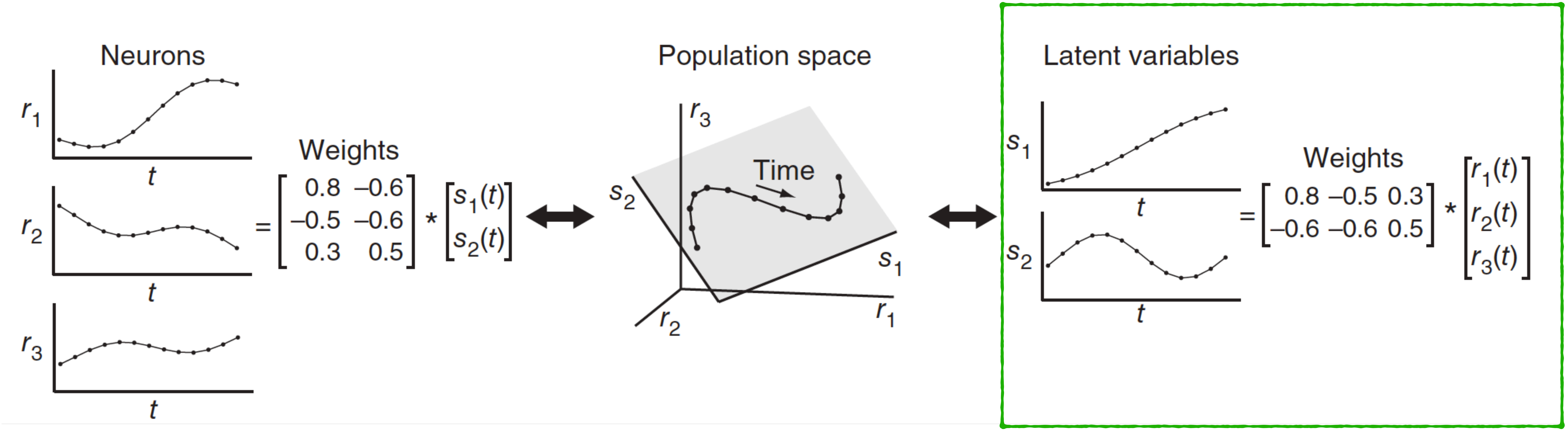

Cunningham, John P., and M. Yu Byron. "Dimensionality reduction for large-scale neural recordings." *Nature neuroscience* 17.11 (2014): 1500.

# Conceptual Illustration



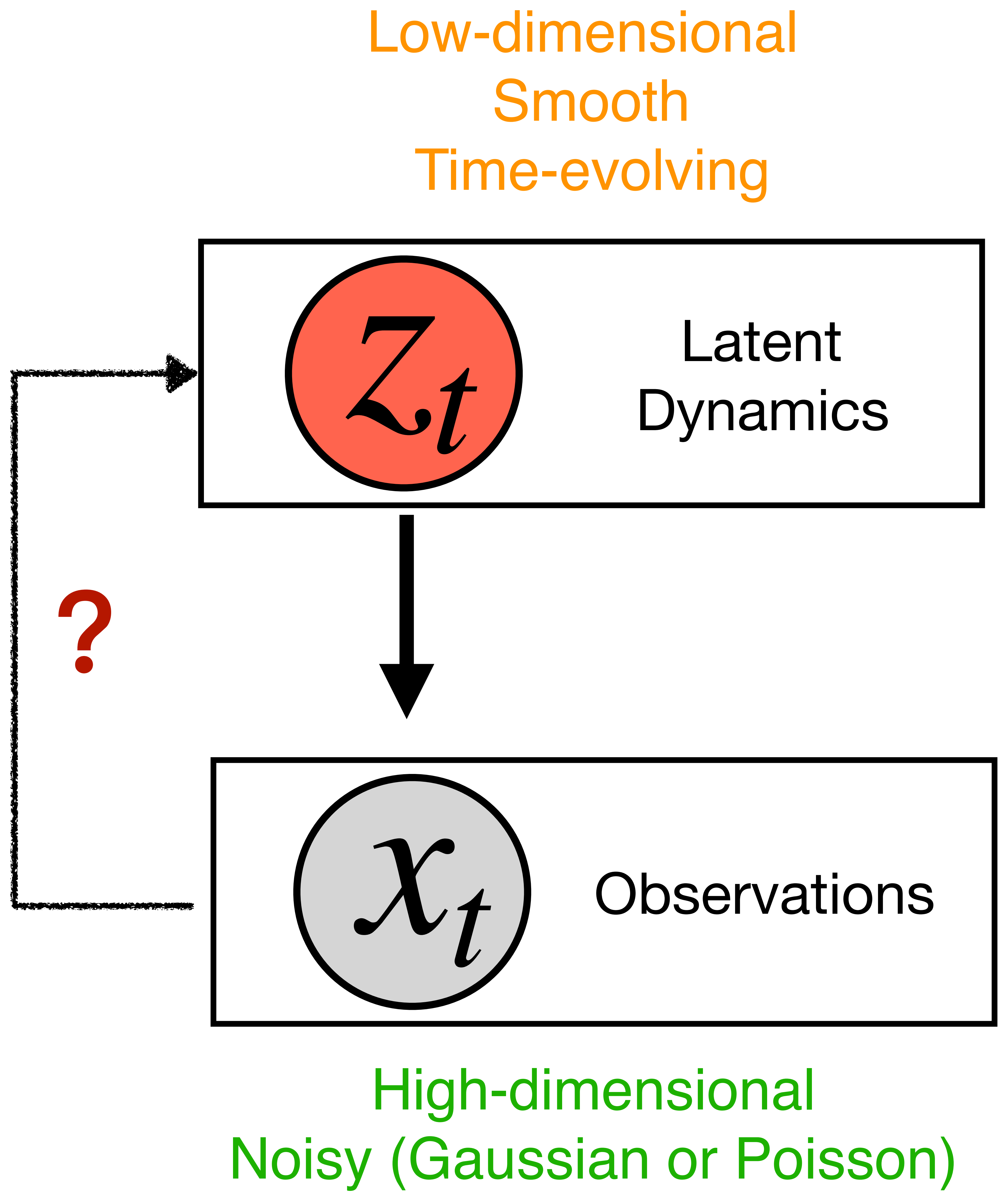

Cunningham, John P., and M. Yu Byron. "Dimensionality reduction for large-scale neural recordings." *Nature neuroscience* 17.11 (2014): 1500.

# Conceptual Illustration



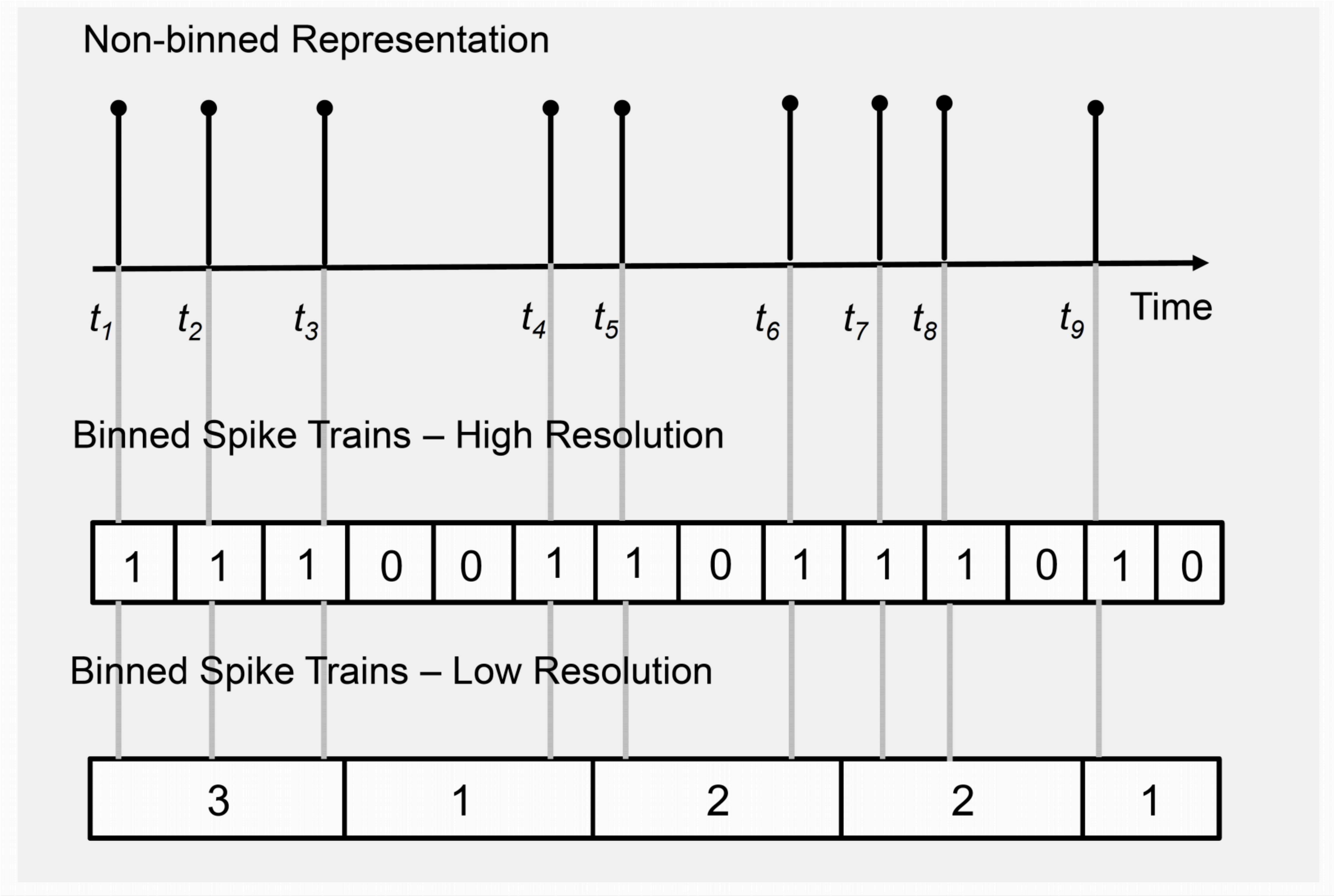

Cunningham, John P., and M. Yu Byron. "Dimensionality reduction for large-scale neural recordings." *Nature neuroscience* 17.11 (2014): 1500.

# Neural Dynamics

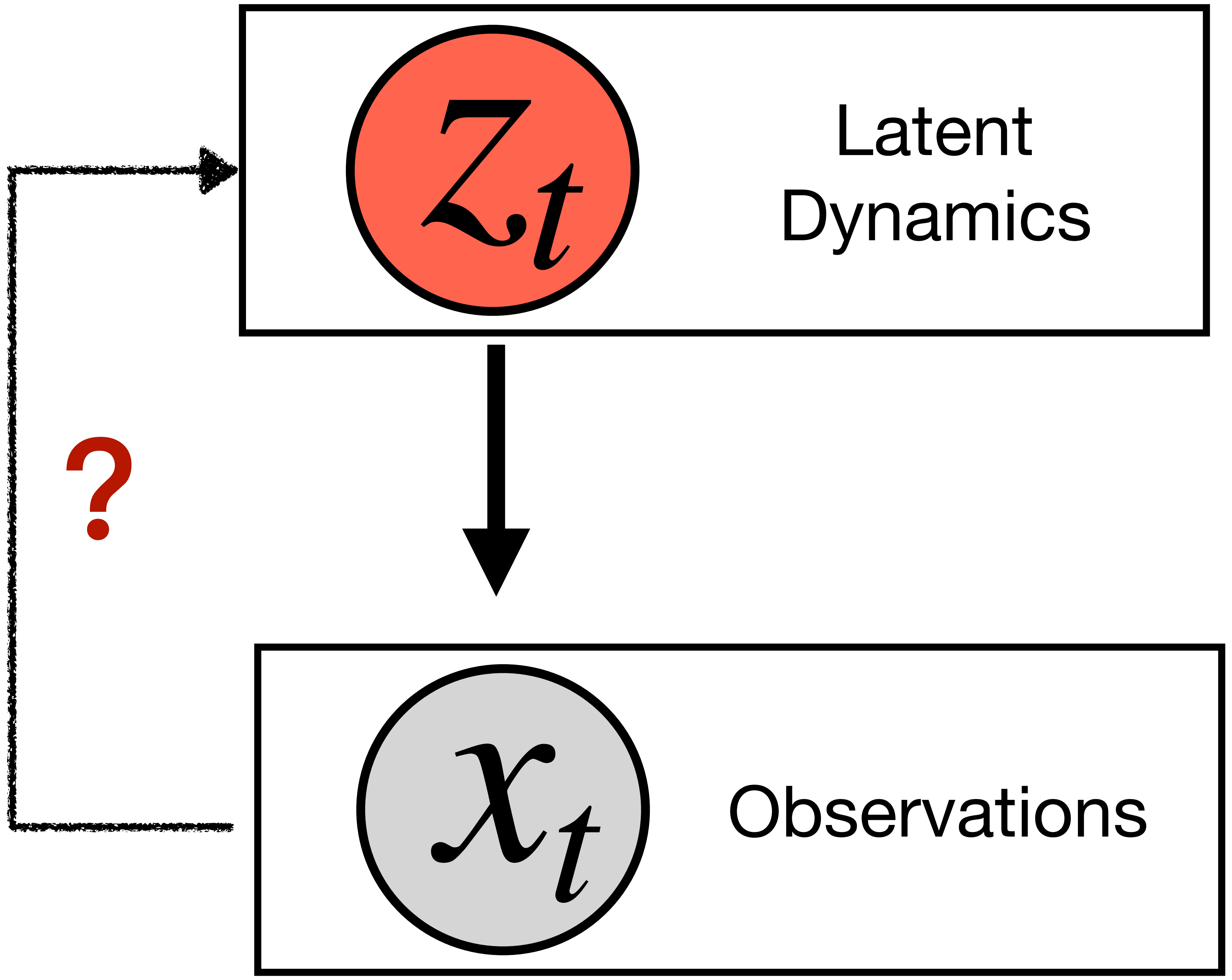


# Neural Dynamics

$x_t$ **Neural spiking activity**

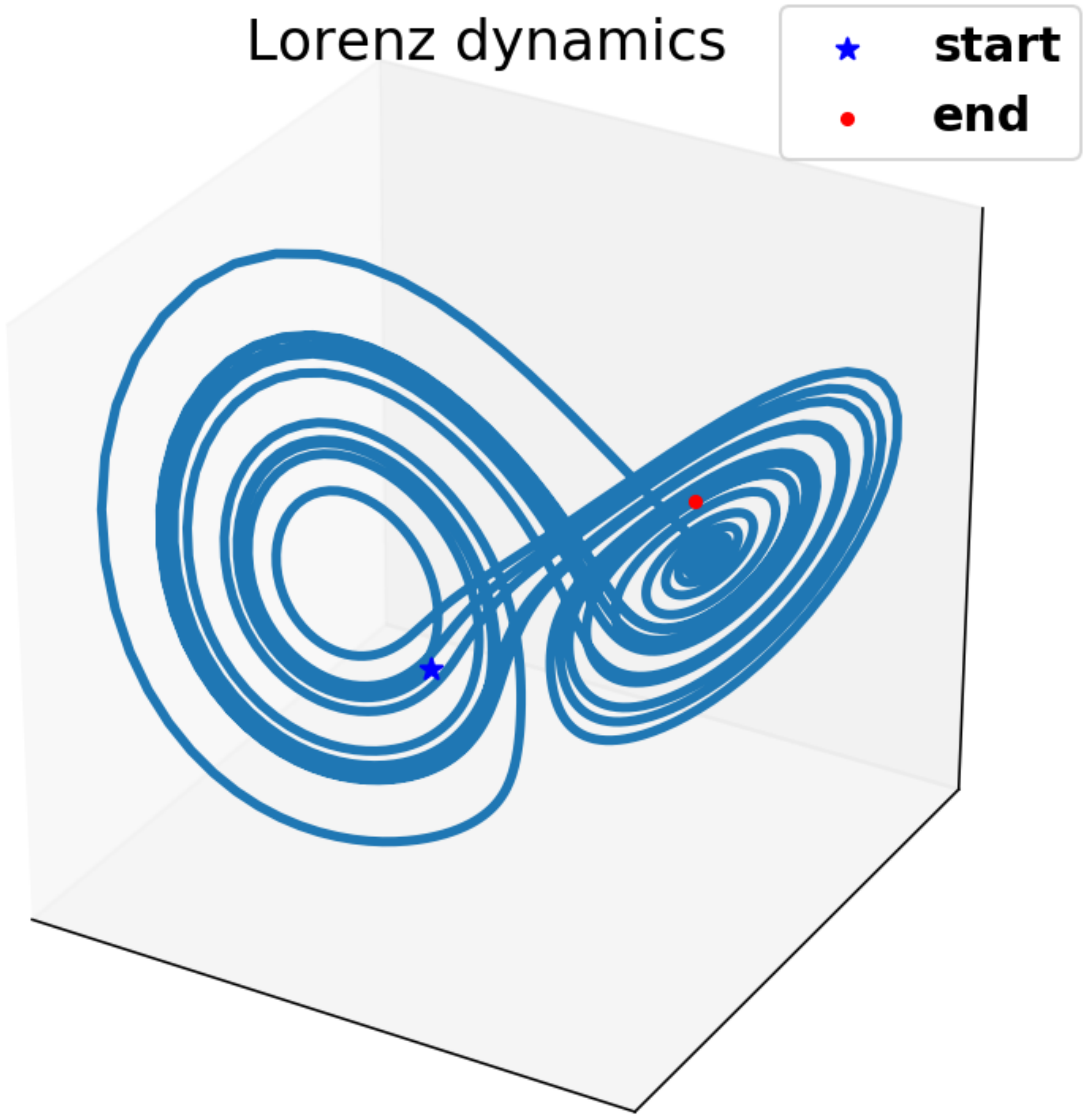



Low-dimensional
Smooth
Time-evolving

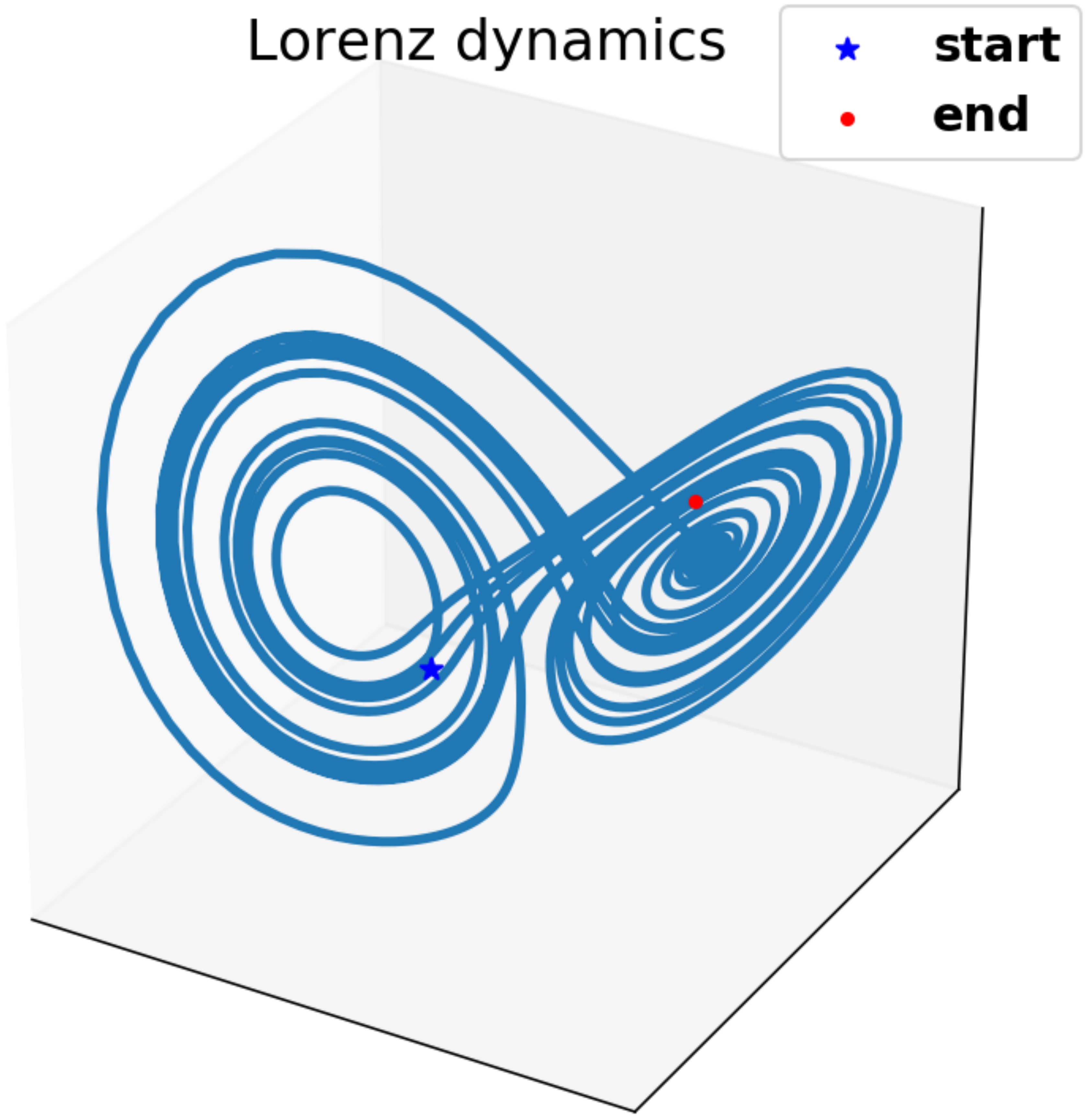



High-dimensional
Noisy (Gaussian or Poisson)

# Latent Dynamics $z_t$



# Latent Dynamics $z_t$




1. Linear Dynamic: $z_t = \mathrm{A}z_{t-1} + \mathrm{b}$

**Pros**: Efficient, tractable

**Cons**: Nonlinear complex dynamics

2. Gaussian Process: $z_t \sim gp(0,\mathrm{k_t})$

**Pros**: Smooth, Nonlinear dynamics

**Cons**: Pairwise dependence, implicit dynamics

3. RNNs: $z_t = \mathrm{RNN}_\theta(z_{t-1}, q_{t-1})$

**Pros**: Long short-term memory

**Cons**: Deterministic dynamics, rely on initial state

# Latent Dynamics $z_t$

1. Linear Dynamic:

$$z_t = \mathrm{A}z_{t-1} + \mathrm{b}$$

2. Gaussian Process:

$$z_t \sim gp(0, \mathrm{k_t})$$

3. Recurrent Neural Networks:

$$z_t = \mathrm{RNN}_\theta(z_{t-1}, q_{t-1})$$

Ours: stochastic RNNs

Prior with RNN structure

$$\nu_t = \mathrm{RNN}(\nu_{t-1}, z_{t-1})$$

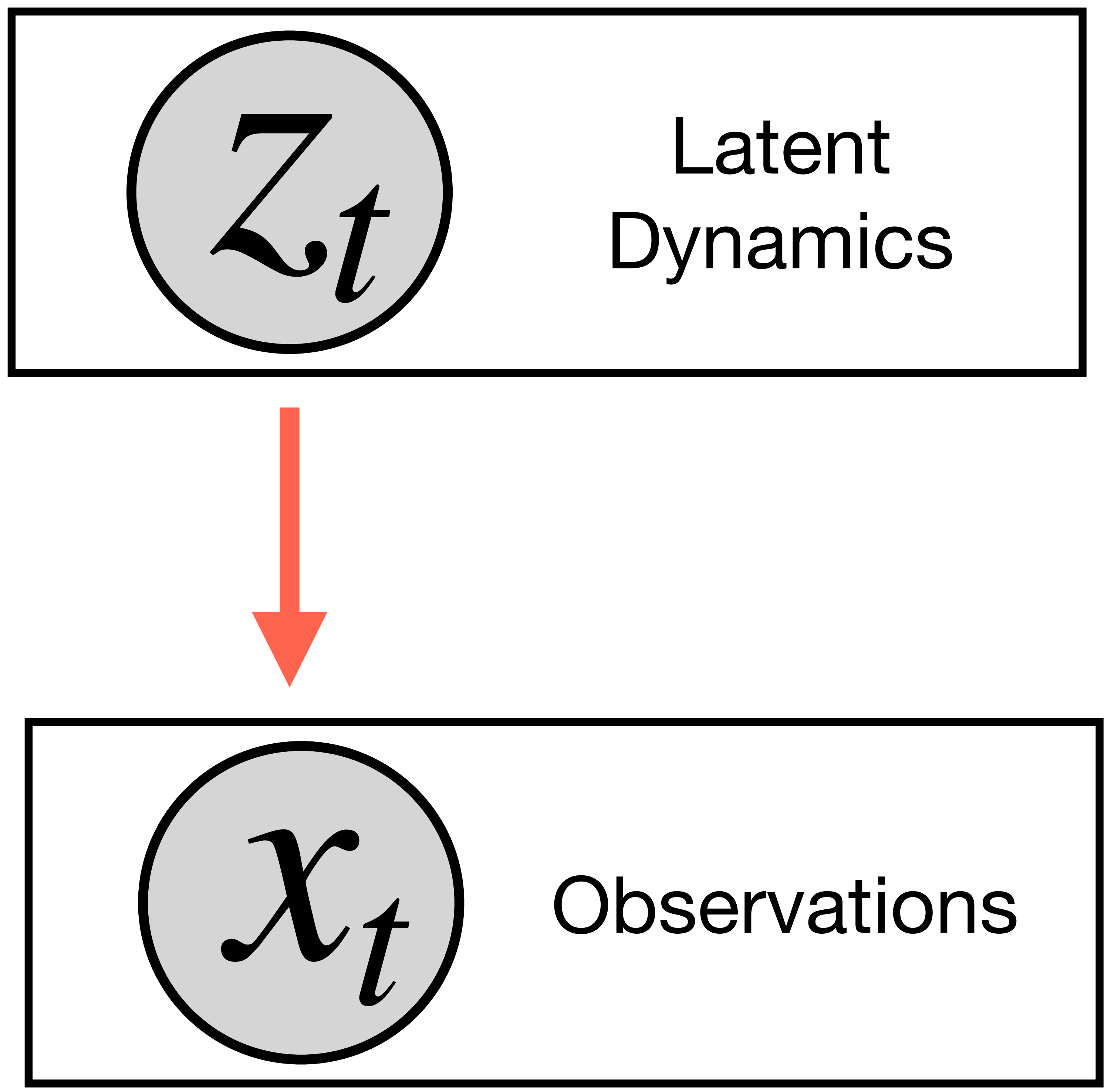
$$z_t \sim \mathrm{Gaussian}(\mu_{\nu_t}, \sigma^2_{\nu_t})$$

① complex nonlinear dynamics

② better uncertainty propagation

# Mapping function

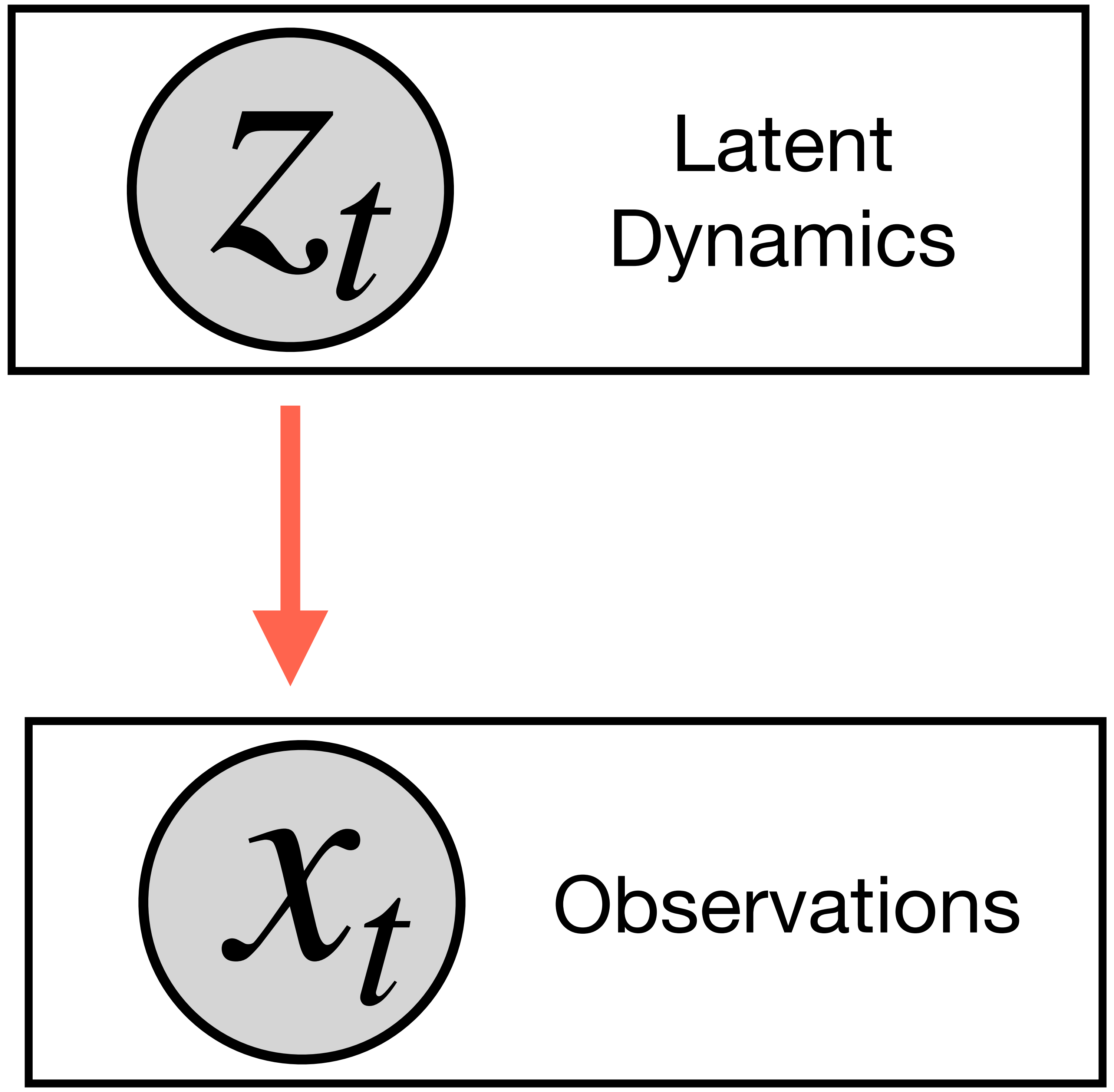
Low-dimensional
Smooth
Time-evolving




High-dimensional
Noisy (Gaussian or Poisson)

# Mapping function

Low-dimensional
Smooth
Time-evolving




High-dimensional
Noisy (Gaussian or Poisson)

$$x_t = \text{Poisson}\Big(\lambda_t = \exp\big(\boxed{\text{NN}(\text{z}_\text{t})}\big)\Big)$$

- Smooth turning curve
- Better modelling uncertainty

$$\boxed{\mathbf{f}_i \,|\, z_{1:T} \sim gp(0, \mathbf{K}_z)}$$

$$x_t \,|\, f_i, z_t \sim \text{Poisson}\Big(\lambda_t = \exp\big(\boxed{f_i(z_t)}\big)\Big)$$

# Observation

Simultaneous recording of many hundreds or thousands of neurons

## Dynamics

$$\nu_t = \mathrm{RNN}(\nu_{t-1}, z_{t-1})$$

$$z_t \sim \mathrm{Gaussian}(\mu_{\nu_t}, \sigma^2_{\nu_t})$$

## Mapping function

$$f_i \,|\, z_t \sim gp(0, k_z)$$

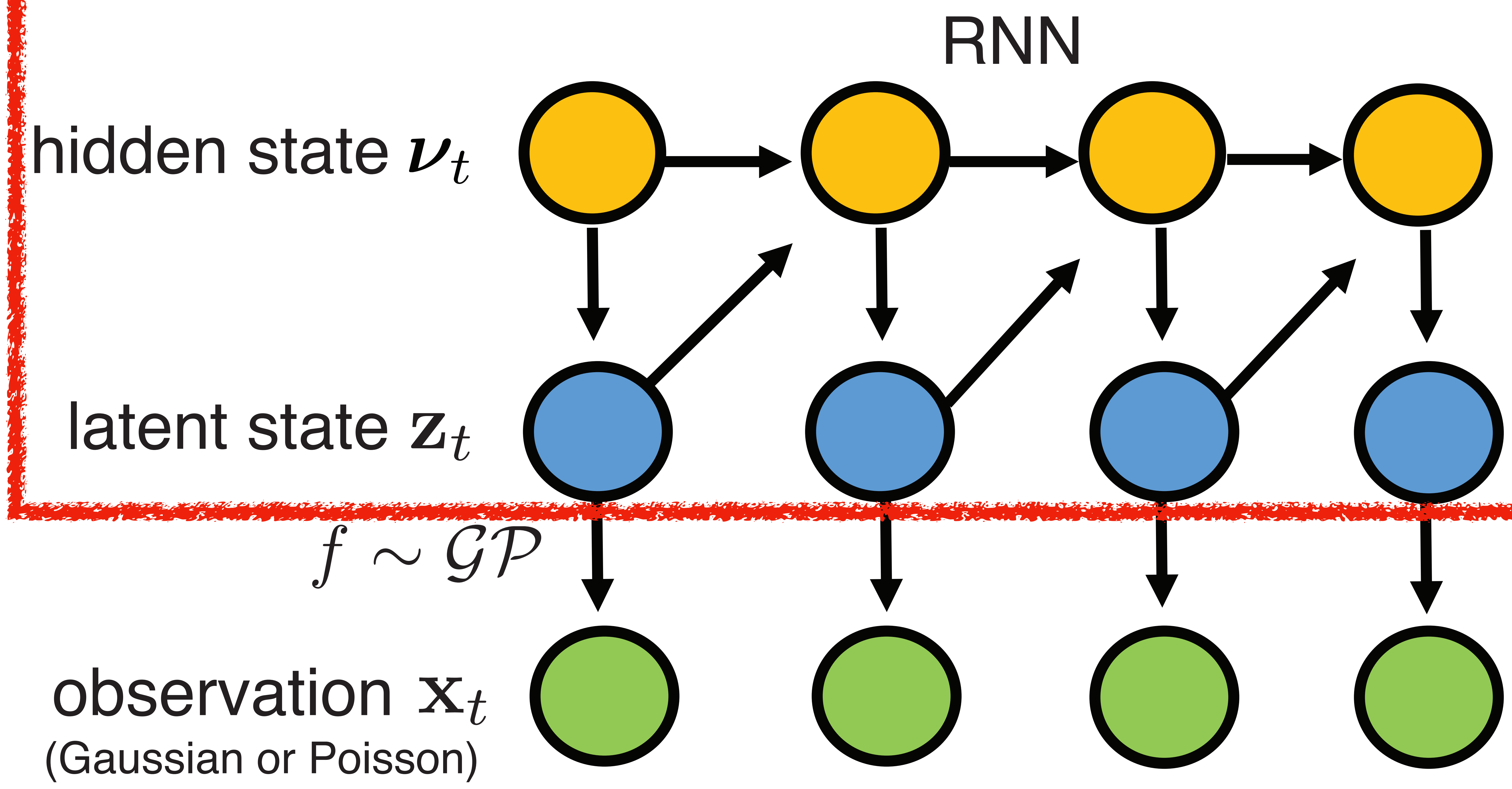
## 1. Gaussian response

$$x_t \,|\, f_i, z_t \sim \mathrm{Gaussian}\big(f_i(z_t), l\big)$$

## 2. Poisson response

$$x_t \,|\, f_i, z_t \sim \mathrm{Poisson}\Big(\lambda_t = \exp\big(f_i(z_t)\big)\Big)$$

# Gaussian Process Recurrent Neural Networks

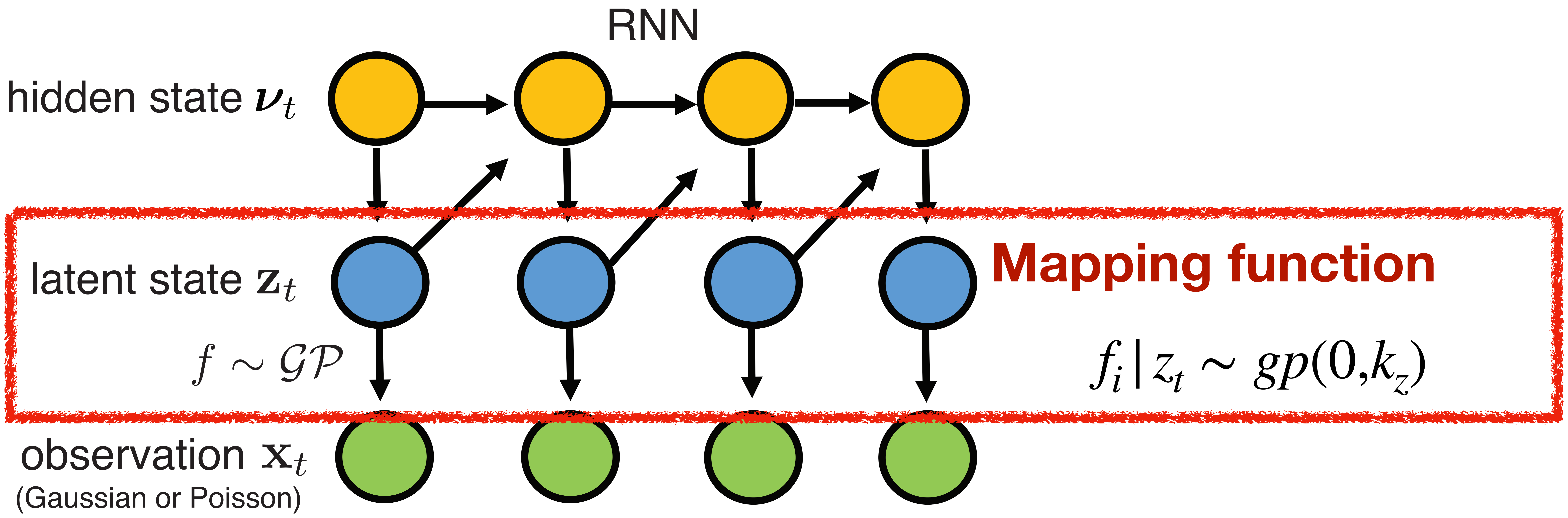



**Dynamics**

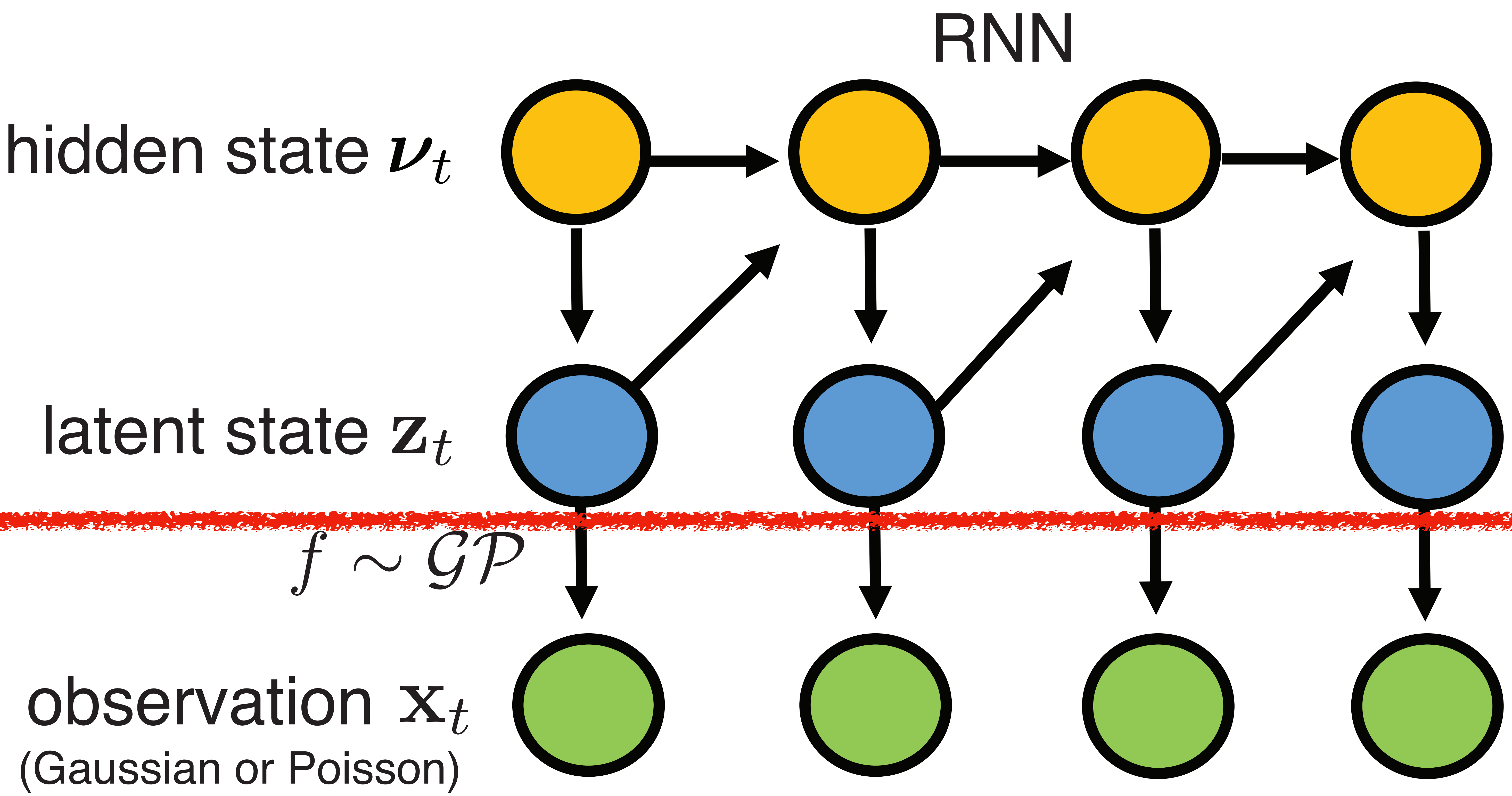
$$\nu_t = \mathrm{RNN}(\nu_{t-1}, z_{t-1})$$

$$z_t \sim \mathrm{Gaussian}(\mu_{\nu_t}, \sigma^2_{\nu_t})$$

# Gaussian Process Recurrent Neural Networks



# Gaussian Process Recurrent Neural Networks




**Observation**

$$x_t \mid f_i, z_t \sim \mathrm{Gaussian}\big(f_i(z_t), l\big)$$

**or**

$$x_t \mid f_i, z_t \sim \mathrm{Poisson}\Big(\lambda_t = \exp\big(f_i(z_t)\big)\Big)$$

# Challenges

Simultaneous recording of many hundreds or thousands of neurons

## Dynamics

$$\nu_t = \mathrm{RNN}(\nu_{t-1}, z_{t-1})$$

$$z_t \sim \mathrm{Gaussian}(\mu_{\nu_t}, \sigma^2_{\nu_t})$$

## Mapping function

$$f_i \,|\, z_t \sim gp(0, k_z)$$

$$p(x\,|\,z) = \int p(x\,|\,f, z)p(f\,|\,z)df \quad \textbf{?}$$

## 1. Gaussian response

$$x_t \,|\, f_i, z_t \sim \mathrm{Gaussian}\big(f_i(z_t), I\big)$$

# Challenges

Simultaneous recording of many hundreds or thousands of neurons

## Dynamics

$$\nu_t = \mathrm{RNN}(\nu_{t-1}, z_{t-1})$$

$$z_t \sim \mathrm{Gaussian}(\mu_{\nu_t}, \sigma^2_{\nu_t})$$

## Mapping function

$$f_i \,|\, z_t \sim gp(0, k_z)$$

$$p(x \,|\, z) = \int p(x \,|\, f, z) p(f \,|\, z) df$$ ?

## 1. Gaussian response

$$x_t \,|\, f_i, z_t \sim \mathrm{Gaussian}\big(f_i(z_t), l\big)$$

$$x_t \,|\, z_t \sim \mathrm{Gaussian}(0, k_z + l)$$ ✔

# Challenges

Simultaneous recording of many hundreds or thousands of neurons

**Dynamics**

$$\nu_t = \mathrm{RNN}(\nu_{t-1}, z_{t-1})$$

$$z_t \sim \mathrm{Gaussian}(\mu_{\nu_t}, \sigma^2_{\nu_t})$$

**Mapping function**

$$f_i \,|\, z_t \sim gp(0, k_z)$$

$$p(x \,|\, z) = \int p(x \,|\, f, z) p(f \,|\, z) df$$ ?

**1. Gaussian response**

$$x_t \,|\, f_i, z_t \sim \mathrm{Gaussian}\big(f_i(z_t), l\big)$$

$$x_t \,|\, z_t \sim \mathrm{Gaussian}(0, k_z + l)$$ ✔

**2. Poisson response**

$$x_t \,|\, f_i, z_t \sim \mathrm{Poisson}\Big(\lambda_t = \exp\big(f_i(z_t)\big)\Big)$$

✘

# Approximate Inference

## 1. Gaussian response

$$x_t \,|\, z_t \sim \mathrm{Gaussian}(0, k_z + l) \qquad\qquad p(z\,|\,x) \propto p(x\,|\,z)p(z)$$

### Laplace approximation

$$q^\star(z) \approx p^\star(z_{\mathrm{MAP}})\det(2\pi\Sigma)\mathrm{Gaussian}(z; z_{\mathrm{MAP}}, \Sigma)$$

### Moment matching

$$q^\star(z) = \mathrm{argmin}_{q^\star(z)}\mathrm{KL}\left(p^\star(z)\,||\,q^\star(z)\right)$$

### Importance sampling

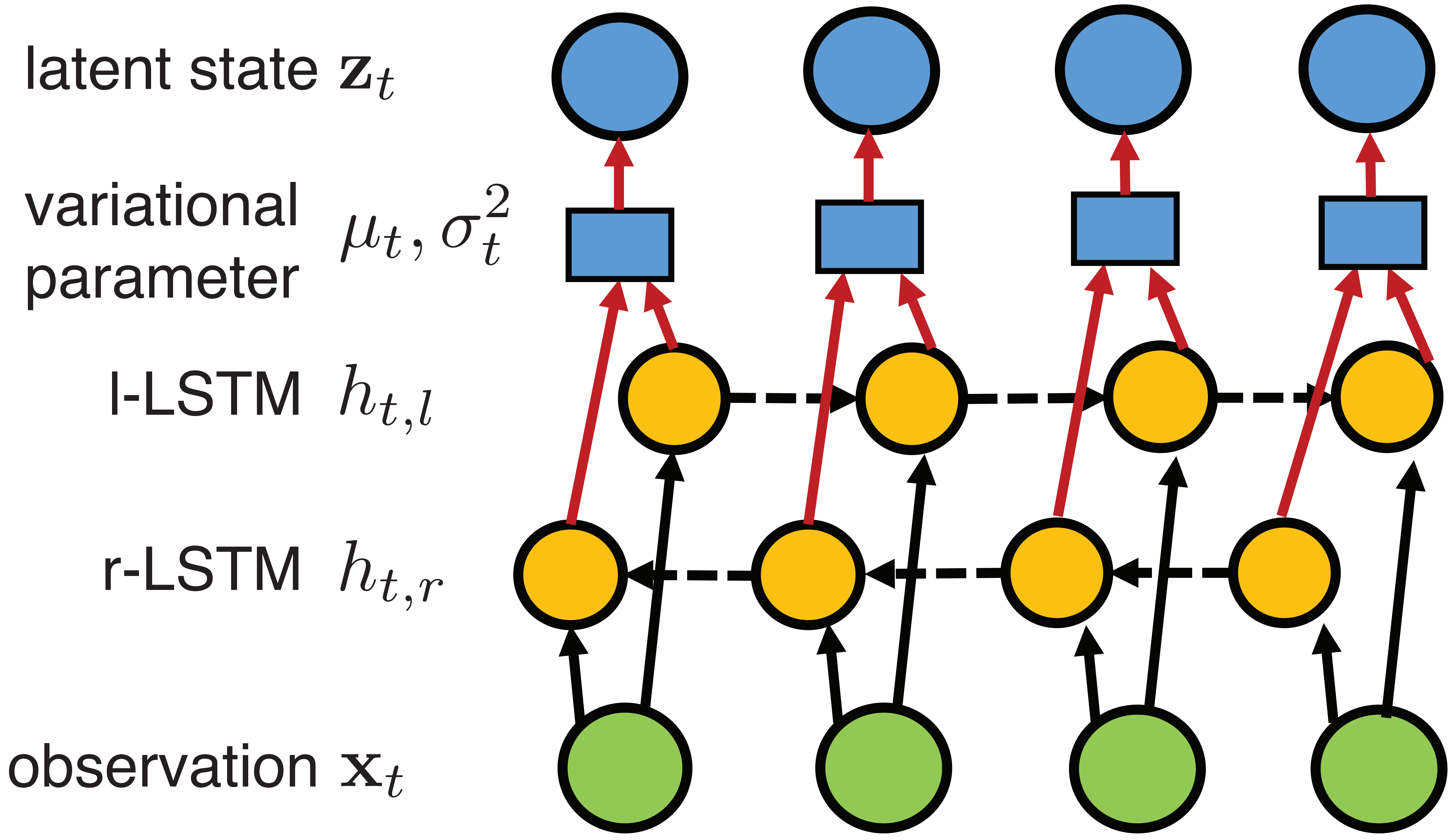
$$q^\star(z) = \sum_k \frac{1}{K}\frac{p^\star(z_m)}{q^\star(z_m)}\delta(z - z_m), \quad z_m \sim q(z)$$

### Variational free energy

$$\boxed{q^\star(z) = \mathrm{argmin}_{q^\star(z)}\mathrm{KL}\left(q^\star(z)\,||\,p^\star(z)\right)}$$

# Inference Network

| Inference Network | Vanilla MF | VAE | r-LSTM | l-LSTM | bi-LSTM |
|---|---|---|---|---|---|
| Variational Approximation | $q(\mathbf{z}_t)$ | $q(\mathbf{z}_t \| \mathbf{x}_t)$ | $q(\mathbf{z}_t \| \mathbf{x}_{t:T})$ | $q(\mathbf{z}_t \| \mathbf{x}_{1:t})$ | $q(\mathbf{z}_t \| \mathbf{x}_{1:T})$ |



# Challenges

Simultaneous recording of many hundreds or thousands of neurons

**Dynamics**

$$\nu_t = \mathrm{RNN}(\nu_{t-1}, z_{t-1})$$

$$z_t \sim \mathrm{Gaussian}(\mu_{\nu_t}, \sigma^2_{\nu_t})$$

**Mapping function**

$$f_i \mid z_t \sim gp(0, k_z)$$

$$p(x \mid z) = \int p(x \mid f, z) p(f \mid z) df$$ ?

**1. Gaussian response**

$$x_t \mid f_i, z_t \sim \mathrm{Gaussian}\left(f_i(z_t), l\right)$$

$$x_t \mid z_t \sim \mathrm{Gaussian}(0, k_z + l)$$ ✔

**2. Poisson response**

$$\boxed{x_t \mid f_i, z_t \sim \mathrm{Poisson}\left(\lambda_t = \exp\left(f_i(z_t)\right)\right)}$$ ✘

# Inference on $z_t$ and $\mathrm{f_i}$

$f_i : \mathrm{R}^L \to \mathrm{R}$ $\qquad f_i | z_t \sim gp(0, k_z)$

$T \times T$

$\mathrm{f_i} \in \mathrm{R}^T$ $\qquad \mathrm{f}_i | z_{1:T} \sim \mathrm{Gaussian}(0, \mathrm{K}_z)$

**Laplace approximation**

$$\hat{\mathrm{f}}_\mathrm{i} = \mathrm{argmax}_{\mathrm{f_i}} \log p(\mathrm{f_i} | \mathrm{x}_i, z_{1:T})$$

$$\Sigma = -\nabla\nabla \log p(\mathrm{f_i} | \mathrm{x}_i, z_{1:T}), \quad \mathrm{f_i} = \hat{\mathrm{f}}_\mathrm{i}$$

# Inference on $z_t$ and $\mathrm{f}_\mathrm{i}$

**Laplace approximation**

$$q(\mathrm{f}_\mathrm{i} \,|\, \mathrm{x}_\mathrm{i}, z_{1:T}) = \mathrm{Gaussian}(\hat{\mathrm{f}}_\mathrm{i}, \Sigma^{-1})$$

$$\hat{\mathrm{f}}_\mathrm{i} = \mathrm{argmax}_{\mathrm{f}_\mathrm{i}} \log p(\mathrm{f}_\mathrm{i} \,|\, \mathrm{x}_i, z_{1:T}) \qquad \Sigma = -\nabla\nabla \log p(\mathrm{f}_\mathrm{i} \,|\, \mathrm{x}_i, z_{1:T}), \quad \mathrm{f}_\mathrm{i} = \hat{\mathrm{f}}_\mathrm{i}$$

# Inference on $z_t$ and $\mathrm{f_i}$

**Laplace approximation**

$$q(\mathrm{f_i} \,|\, \mathrm{x_i}, z_{1:T}) = \mathrm{Gaussian}(\hat{\mathrm{f}}_\mathrm{i}, \Sigma^{-1})$$

$$\hat{\mathrm{f}}_\mathrm{i} = \mathrm{argmax}_{\mathrm{f_i}} \log p(\mathrm{f_i} \,|\, \mathrm{x}_i, z_{1:T}) \qquad \Sigma = -\nabla\nabla \log p(\mathrm{f_i} \,|\, \mathrm{x}_i, z_{1:T}), \quad \mathrm{f_i} = \hat{\mathrm{f}}_\mathrm{i}$$

$$= \mathrm{argmax}_{\mathrm{f_i}} \boxed{\log p(\mathrm{x}_i, \mathrm{f_i} \,|\, z_{1:T})}$$

**Optimize** $\Psi(\mathrm{f_i}) = \log p(\mathrm{x_i} \,|\, \mathrm{f_i}) - \frac{1}{2}\mathrm{f_i}^\top \mathrm{K_z} \mathrm{f}_i - \frac{1}{2}\log|\mathrm{K_z}|$

$$p(\mathrm{x}_i \,|\, z_{1:T}) = \int p(\mathrm{x}_i, \mathrm{f}_i \,|\, z_{1:T}) d\mathrm{f}_i \quad ?$$

# Inference on $z_t$ and $\mathrm{f_i}$

$$\hat{\mathrm{f}}_\mathrm{i} = \mathrm{argmax}_{\mathrm{f_i}} \log p(\mathrm{f_i} \,|\, \mathrm{x}_i, z_{1:T})$$

$$= \mathrm{argmax}_{\mathrm{f_i}} \boxed{\log p(\mathrm{x}_i, \mathrm{f_i} \,|\, z_{1:T})}$$

**Optimize**

$$\Psi(\mathrm{f_i}) = \log p(\mathrm{x}_i, \mathrm{f_i} \,|\, z_{1:T}) = \log p(\mathrm{x_i} \,|\, \mathrm{f_i}) - \frac{1}{2}\mathrm{f_i}^\top \mathrm{K_z} \mathrm{f}_i - \frac{1}{2}\log|\mathrm{K_z}|$$

$$p(\mathrm{x}_i \,|\, z_{1:T}) = \int p(\mathrm{x}_i, \mathrm{f}_i \,|\, z_{1:T}) d\mathrm{f}_i$$ ?

**Taylor expansion**

$$\Psi(\mathrm{f_i}) \approx \Psi(\hat{\mathrm{f}}_\mathrm{i}) - \frac{1}{2}(\mathrm{f}_i - \hat{\mathrm{f}}_\mathrm{i})^\top \Sigma (\mathrm{f}_i - \hat{\mathrm{f}}_\mathrm{i})$$

$$p(\mathrm{x}_i \,|\, z_{1:T}) = \int p(\mathrm{x}_i, \mathrm{f}_i \,|\, z_{1:T}) d\mathrm{f}_i \approx \exp(\hat{\mathrm{f}}_\mathrm{i}) \int \exp\Big(-\frac{1}{2}(\mathrm{f}_i - \hat{\mathrm{f}}_\mathrm{i})^\top \Sigma (\mathrm{f}_i - \hat{\mathrm{f}}_\mathrm{i})\Big) d\mathrm{f}_i$$

# Inference on $z_t$ and $\mathrm{f_i}$

$$p(\mathrm{x}_i \mid z_{1:T}) = \int p(\mathrm{x}_i, \mathrm{f}_i \mid z_{1:T}) d\mathrm{f}_i \approx \boxed{\exp(\hat{\mathrm{f}}_\mathrm{i}) \int \exp\left(-\frac{1}{2}(\mathrm{f}_i - \hat{\mathrm{f}}_\mathrm{i})^\top \Sigma (\mathrm{f}_i - \hat{\mathrm{f}}_\mathrm{i})\right) d\mathrm{f}_i}$$

$$\log p(\mathrm{x}_i \mid z_{1:T}) \approx \log p(\mathrm{x}_i \mid \hat{\mathrm{f}}_i) - \frac{1}{2}\left(\hat{\mathrm{f}}_i^\top \mathrm{K}_z \mathrm{f}_i + \log |\mathrm{A}|\right)$$

$$\mathrm{A} = |\mathrm{K}_z| \, |\mathrm{K}_z^{-1} \nabla \nabla \log p(\mathrm{x}_i \mid \hat{\mathrm{f}}_i)|$$

$$p(\mathrm{x}_i \mid z_{1:T}) = \int p(\mathrm{x}_i, \mathrm{f}_i \mid z_{1:T}) d\mathrm{f}_i$$ ✓

$$p(z \mid x) \propto p(x \mid z) p(z)$$ ✓

# Inference Network and Dynamical Model Analysis.

| Gaussian | AR1-GPLVM | | | | | GP-RNN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MF | VAE | r-LSTM | l-LSTM | bi-LSTM | MF | VAE | r-LSTM | l-LSTM | bi-LSTM |
| linear | 4.12 | 4.10 | 4.01 | 3.27 | **1.64** | 2.17 | 2.17 | 1.98 | 1.54 | **0.96** |
| tanh | 3.20 | 3.22 | 3.01 | 2.46 | **1.17** | 2.01 | 2.01 | 1.83 | 1.41 | **0.78** |
| sine | 3.12 | 3.12 | 2.74 | 2.33 | **1.02** | 1.81 | 1.78 | 1.34 | 1.12 | **0.56** |

| Poisson | AR1-GPLVM | | | | | GP-RNN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MF | VAE | r-LSTM | l-LSTM | bi-LSTM | MF | VAE | r-LSTM | l-LSTM | bi-LSTM |
| linear | 6.34 | 6.34 | 6.02 | 5.71 | **3.67** | 6.01 | 6.01 | 5.94 | 5.71 | **3.10** |
| tanh | 3.22 | 3.21 | 3.01 | 2.84 | **1.57** | 3.09 | 3.11 | 2.98 | 2.54 | **1.21** |
| sine | 2.80 | 2.79 | 2.77 | 2.51 | **1.49** | 2.67 | 2.67 | 2.43 | 2.33 | **1.14** |

RMSE of latent trajectories reconstructed from various simulated models are presented.

# Mapping Function Analysis

| # Data | linear | | tanh | | sine | |
|---|---|---|---|---|---|---|
| | GP | NN | GP | NN | GP | NN |
| N = 50 | **2.51** | 3.88 | **1.45** | 2.75 | **1.97** | 3.43 |
| N = 100 | **1.27** | 1.65 | **1.15** | 1.45 | **1.03** | 1.31 |
| N = 200 | **0.96** | 1.29 | **0.78** | 1.22 | **0.56** | 0.70 |
| N = 500 | **0.34** | 0.35 | **0.26** | **0.26** | **0.12** | **0.12** |

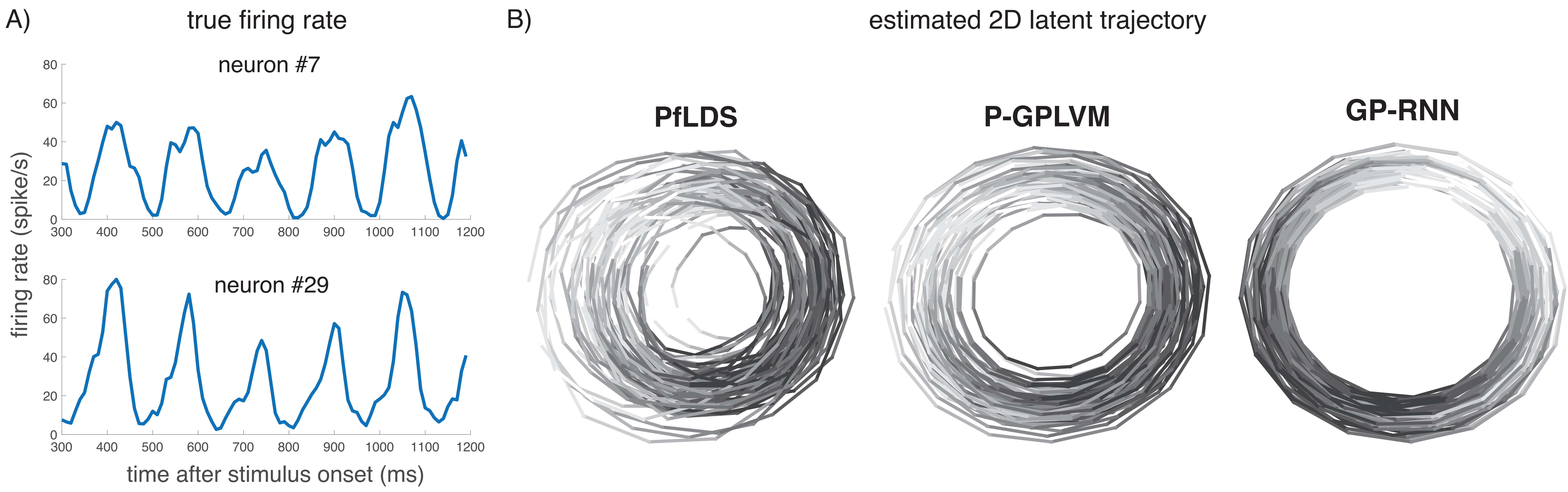
Smaller datasets may affect latent dynamics recovery but a Gaussian process mapping enhances nonlinear embedding recovery

# Related Models

| Model | Dynamics | Mapping function | Link function | Observation | Inference |
|---|---|---|---|---|---|
| PLDS | LDS | Linear | exp | Poisson | LP |
| PfLDS | LDS | NN | exp | Poisson | VI + inference network |
| GCLDS | LDS | Linear | exp | Count | VI |
| LFADS | RNN | Linear | exp | Poisson | VI + inference network |
| P-GPFA | GP | Linear | Identity | Poisson | LP or VI |
| P-GPLVM | GP | GP | exp | Poisson | LP |
| **Ours : GP-RNN** | **RNN** | **GP** | **exp** | **Poisson/Gaussian** | **VI + inference network** |

| Dimension | PLDS | GCLDS | PfLDS | P-GPFA | P-GPLVM | GP-RNN |
|---|---|---|---|---|---|---|
| $z_1$ | 0.641 | 0.435 | 0.698 | 0.733 | 0.784 | **0.869** |
| $z_2$ | 0.547 | 0.364 | 0.659 | 0.720 | 0.785 | **0.873** |
| $z_3$ | 0.903 | 0.755 | 0.797 | 0.960 | 0.966 | **0.971** |

# Smooth and Structured Patterns



Code link:

https://github.com/sheqi/GP-RNN_UAI2019

Full Paper

http://auai.org/uai2019/proceedings/papers/159.pdf

# Thanks !

qi.she@intel.com