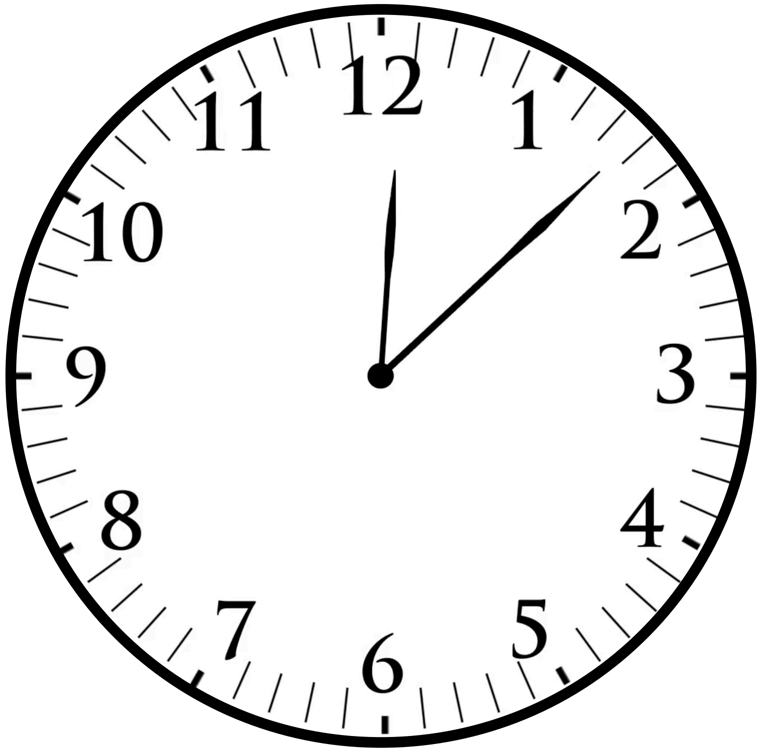
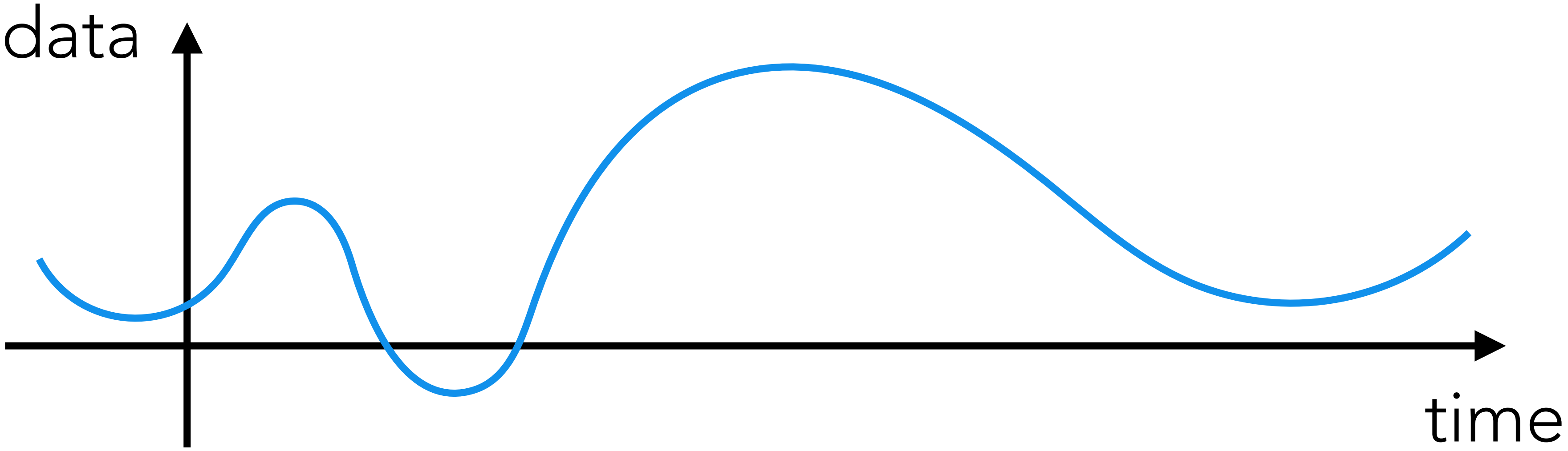
DEEP SEQUENTIAL LATENT VARIABLE MODELS

JOSEPH MARINO
CALTECH

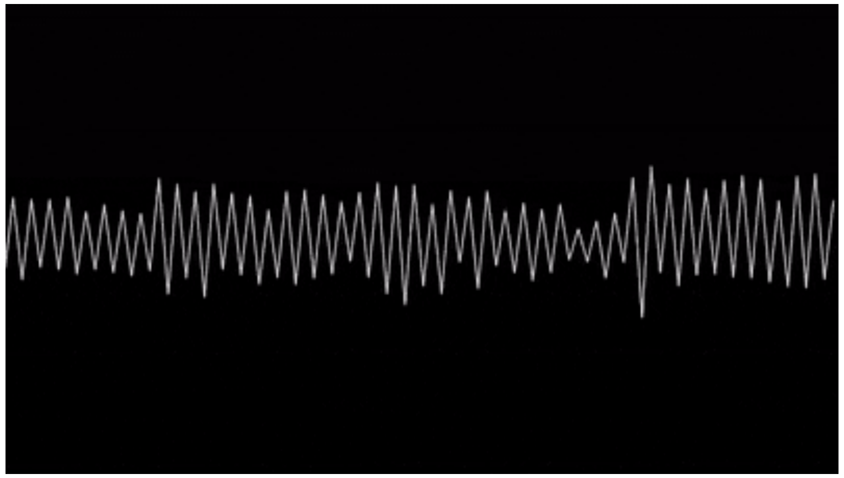


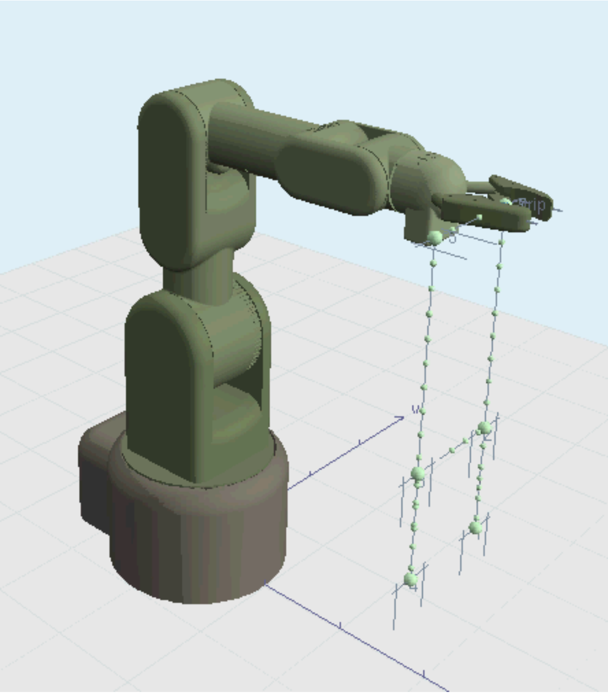
observed data are often sequential



audio

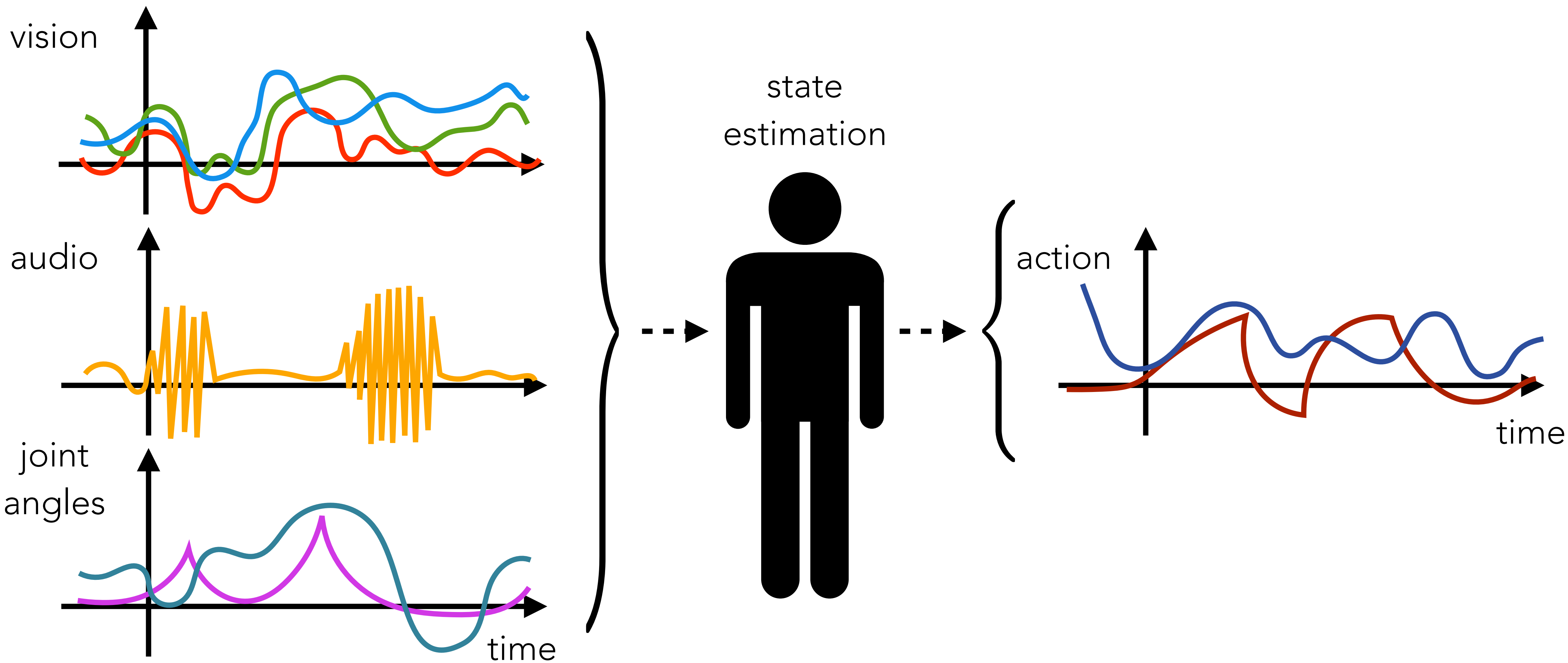
vision



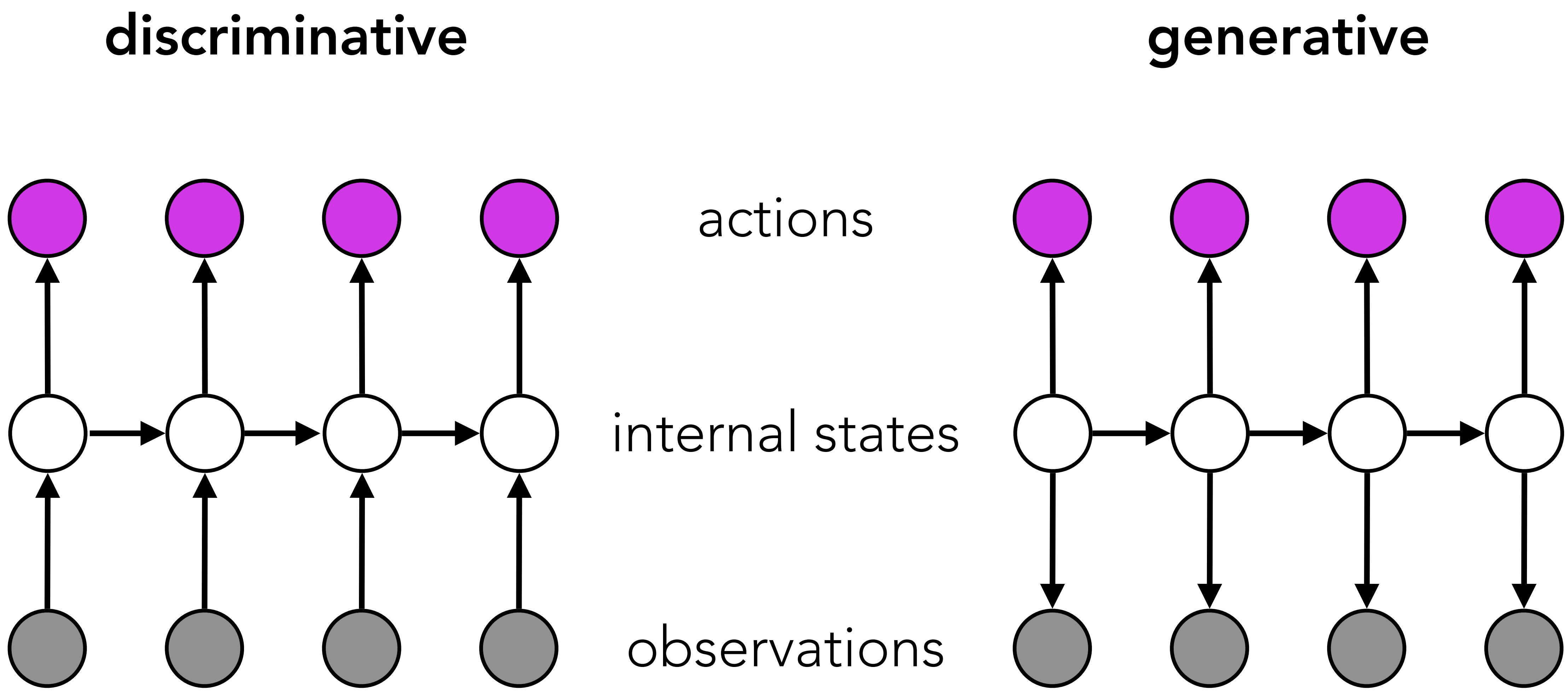


joint angles

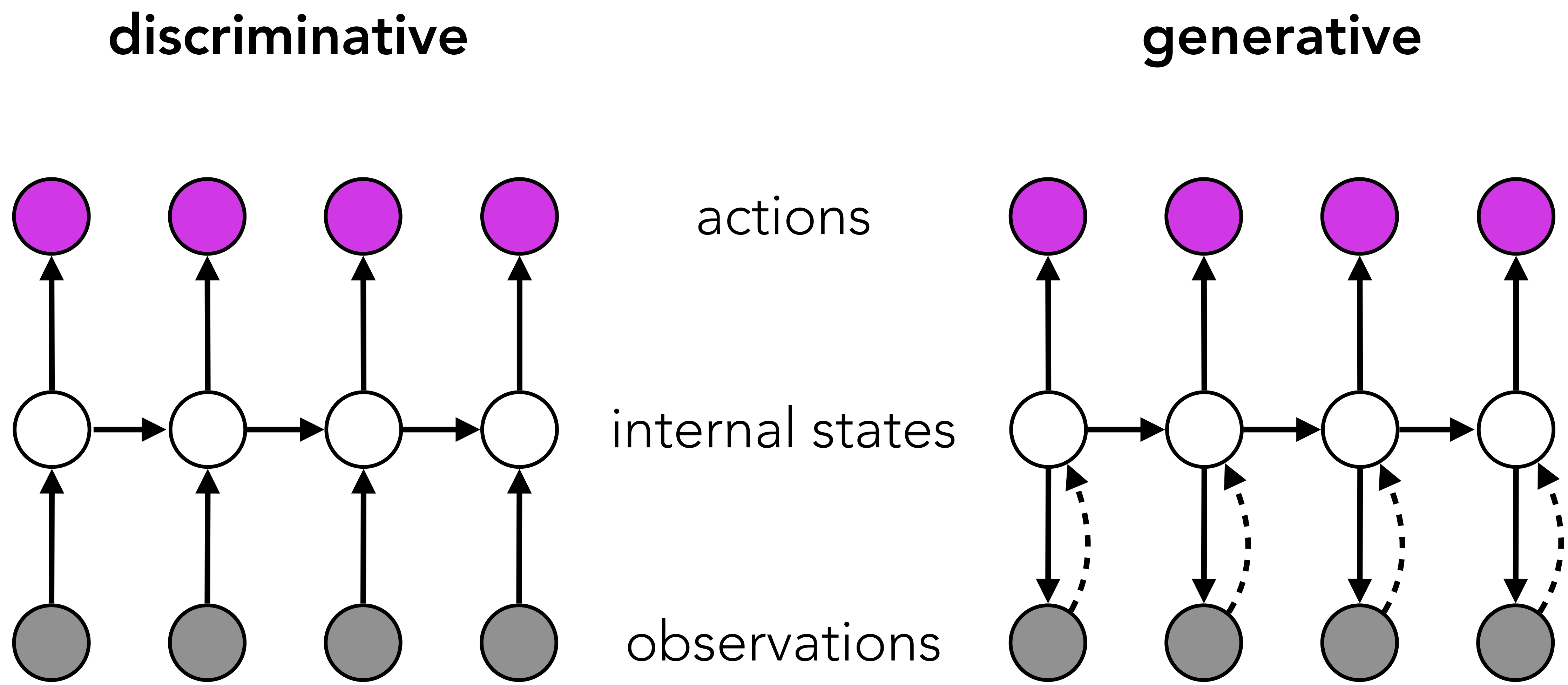
interacting in the world involves processing sequences of data



COMPUTATIONAL APPROACHES TO STATE ESTIMATION



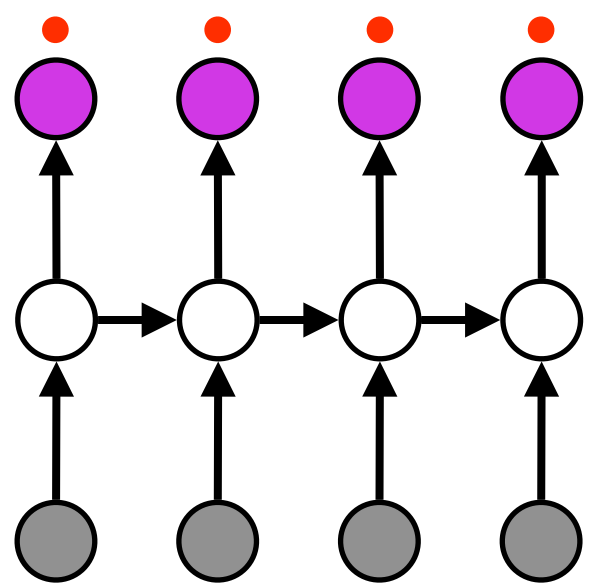
COMPUTATIONAL APPROACHES TO STATE ESTIMATION



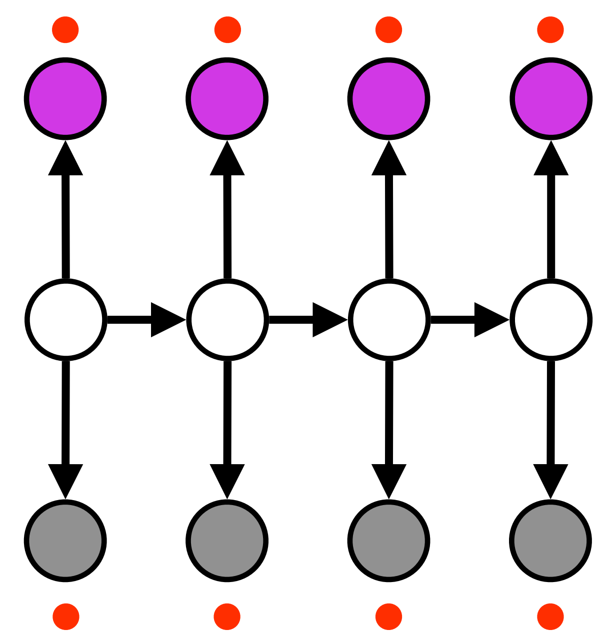
ADVANTAGES OF GENERATIVE MODELING

unsupervised learning: *learn from the data*

discriminative

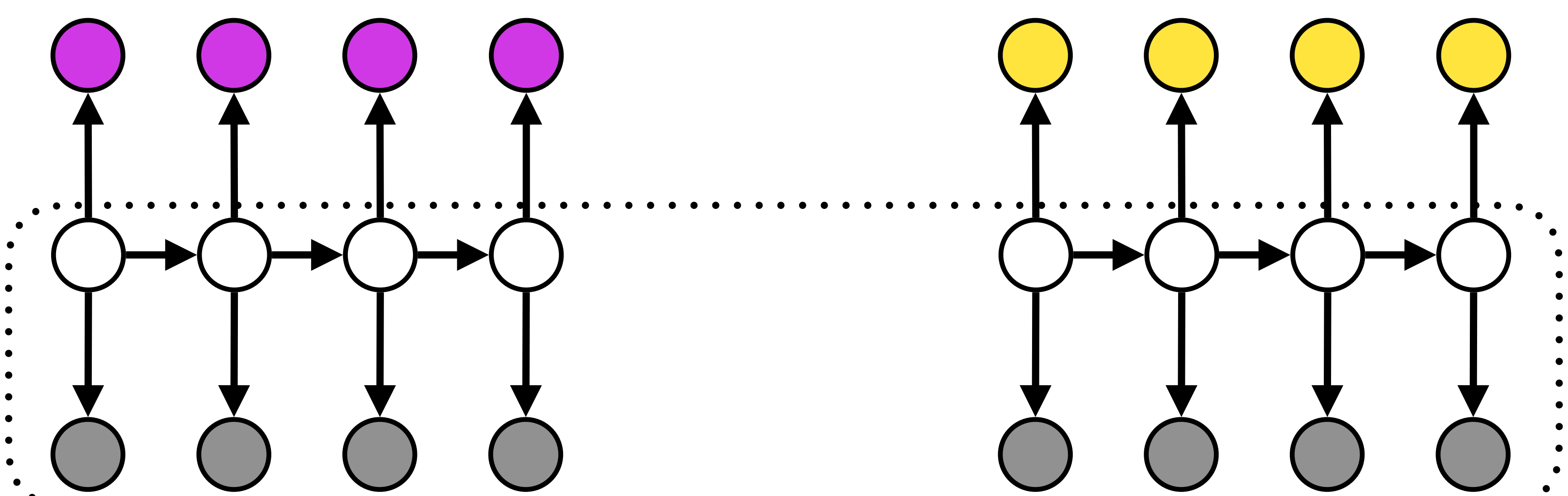


generative

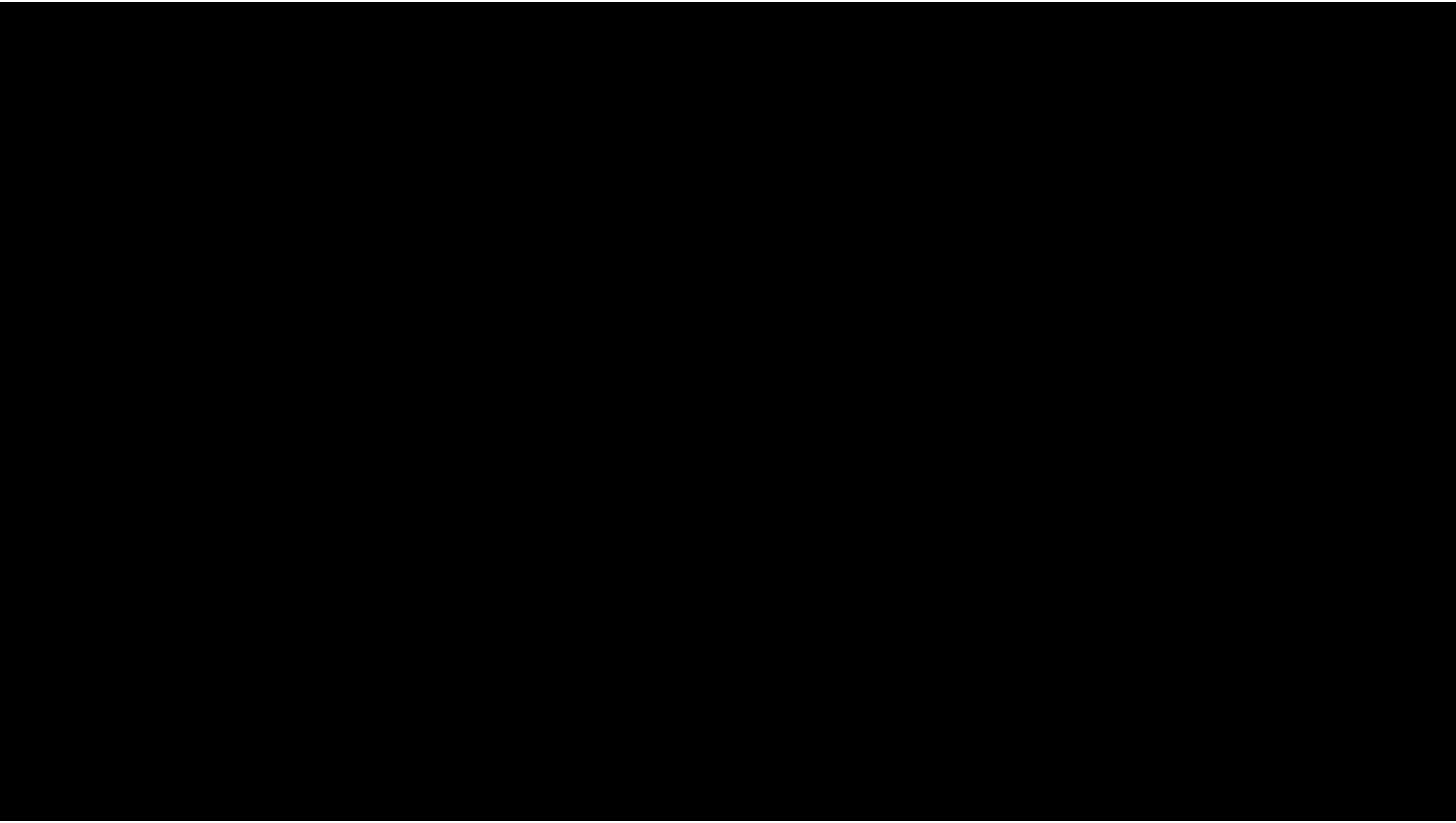


● training signal

generalization: *learn a task-agnostic representation*



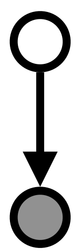
shared



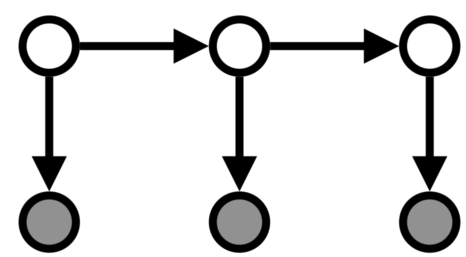
Jaderberg, Minh, Czarnecki et al., 2016

OUTLINE

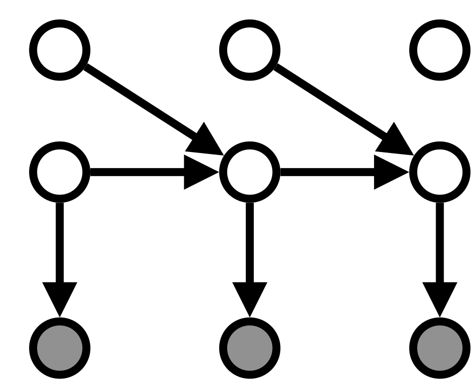
BACKGROUND

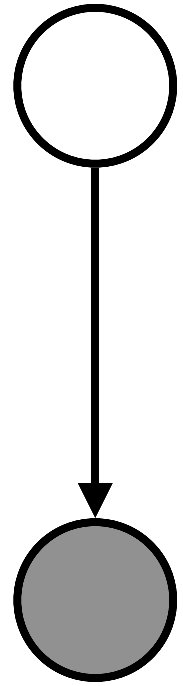


DEEP SEQUENTIAL
LATENT VARIABLE MODELS



MODEL-BASED RL

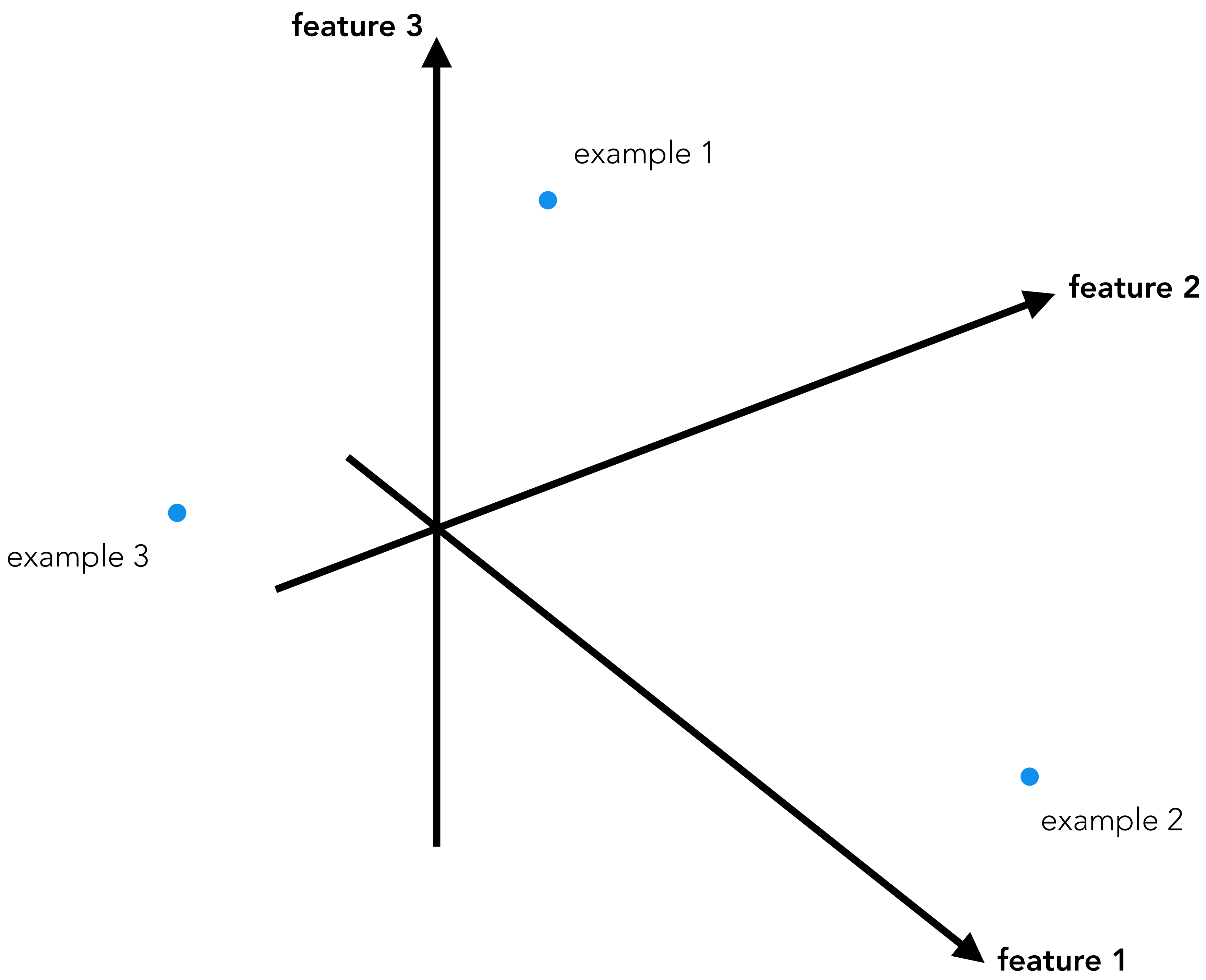


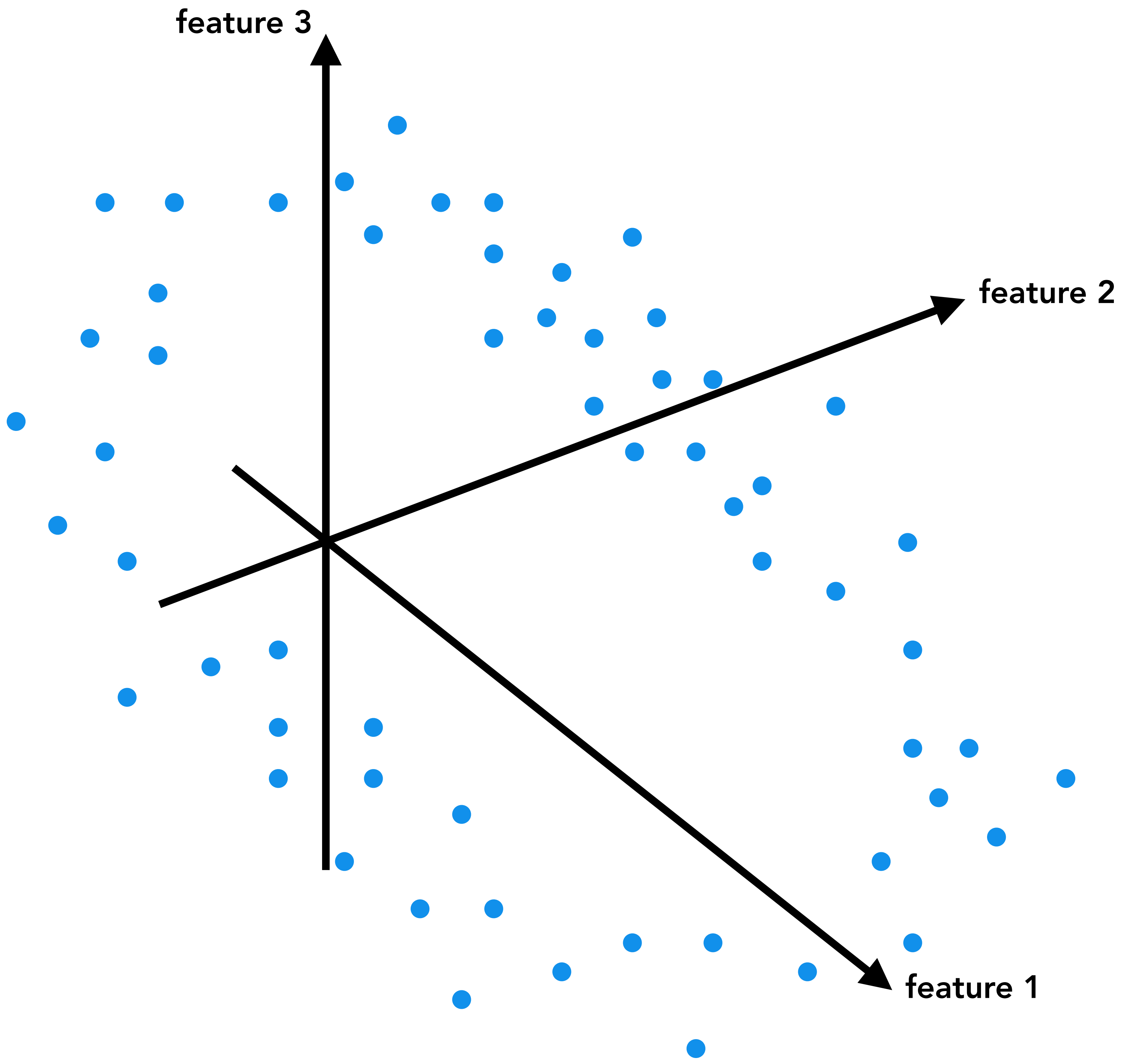


BACKGROUND

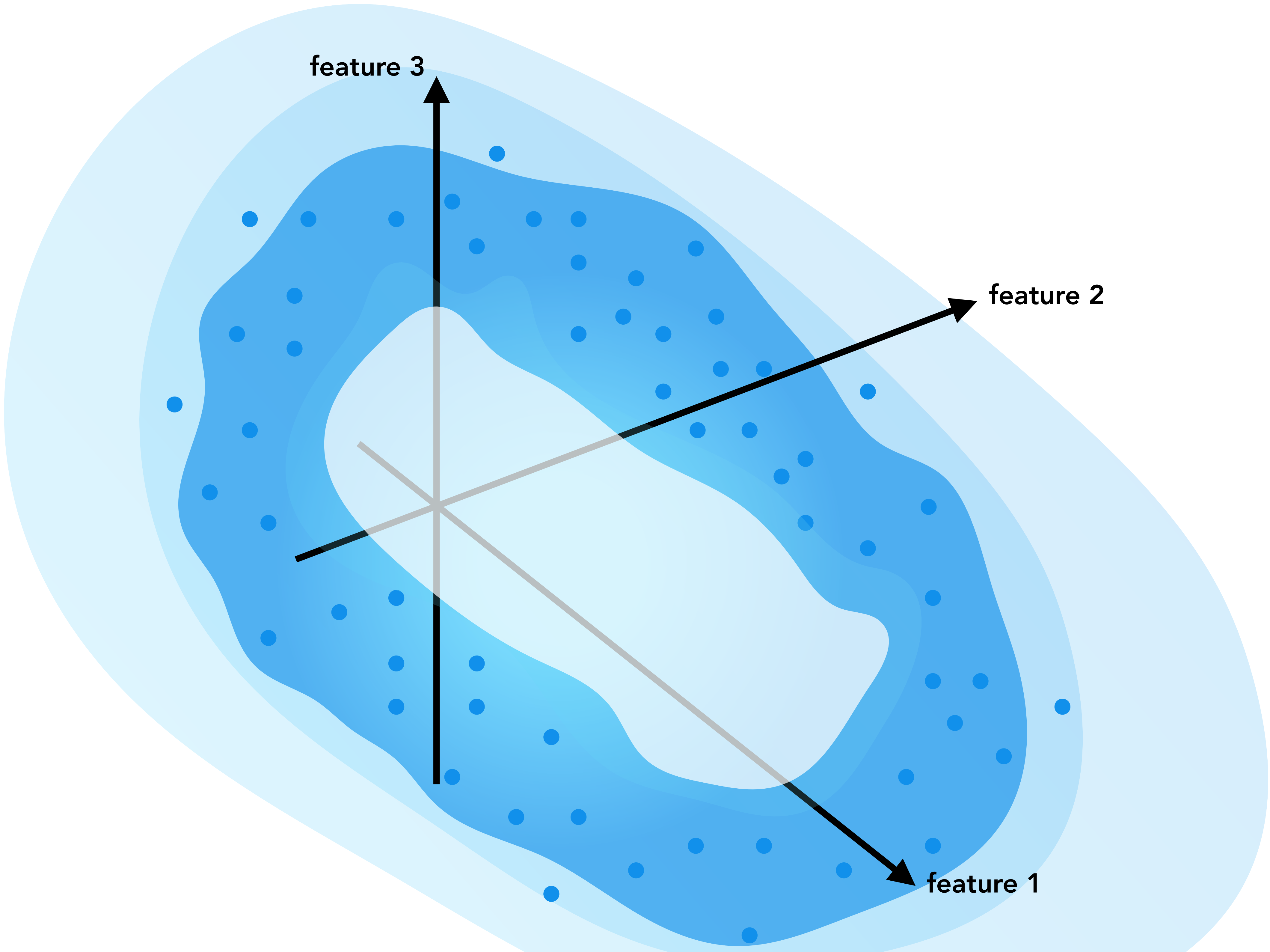
GENERATIVE MODEL

a model of the density of observed data


feature 3
example 1
feature 2
example 3
example 2
feature 1



EMPIRICAL DATA DISTRIBUTION



DENSITY ESTIMATION

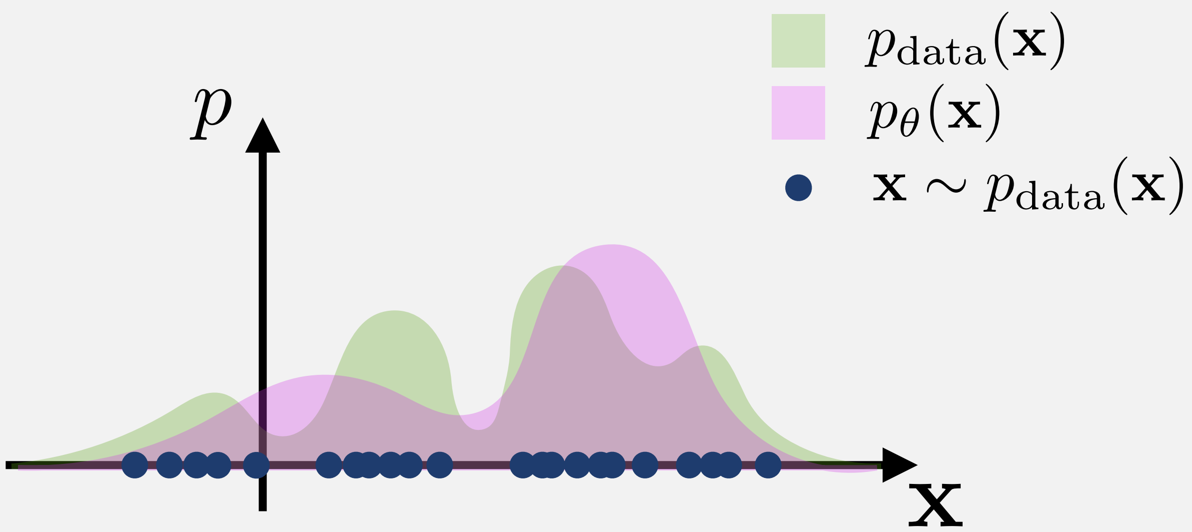
estimating the density of the empirical data distribution

MAXIMUM LIKELIHOOD

data: $p_{\text{data}}(\mathbf{x})$

model: $p_\theta(\mathbf{x})$

parameters: θ



maximum likelihood estimation

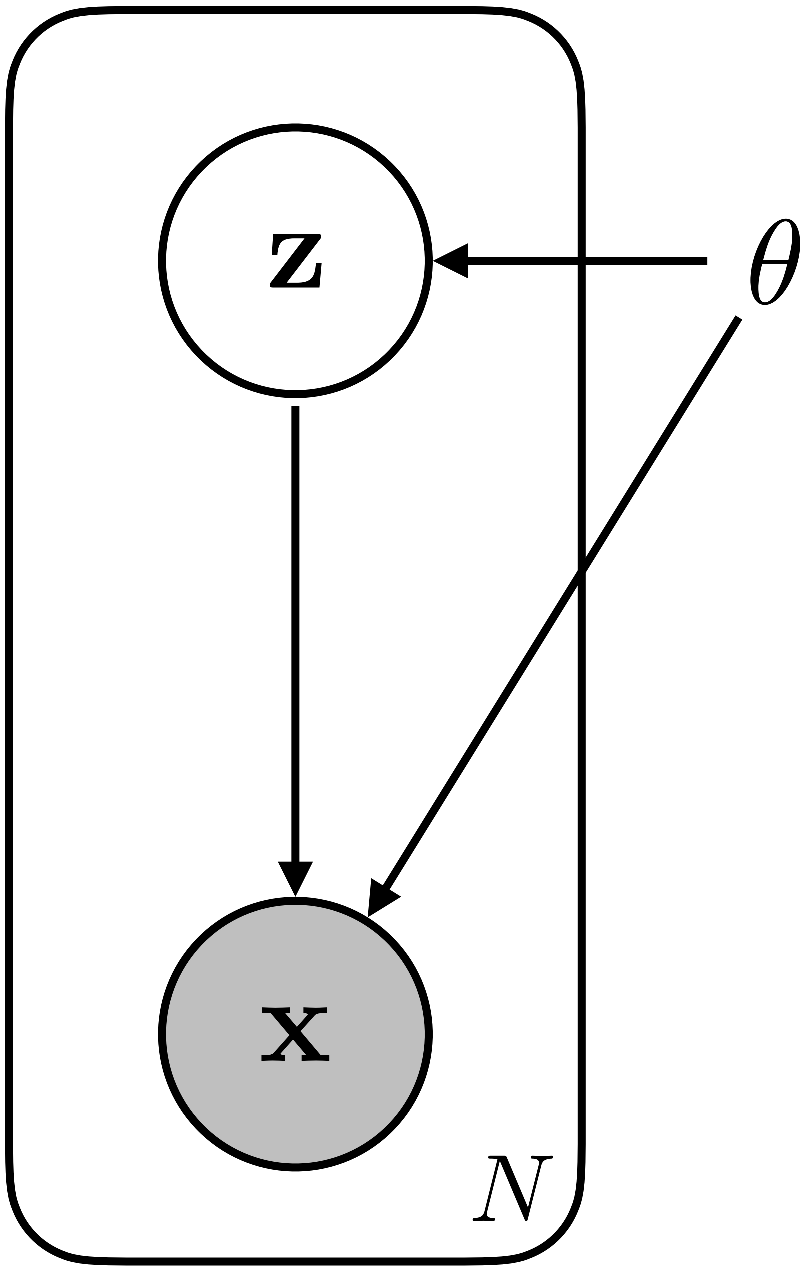
find the model that assigns the *maximum likelihood* to the data

$$\theta^* = \arg\min_\theta \; D_{KL}(p_{\text{data}}(\mathbf{x}) || p_\theta(\mathbf{x}))$$

$$= \arg\min_\theta \; \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \left[\log p_{\text{data}}(\mathbf{x}) - \log p_\theta(\mathbf{x}) \right]$$

$$= \arg\max_\theta \; \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \left[\log p_\theta(\mathbf{x}) \right] \approx \frac{1}{N} \sum_{i=1}^{N} \log p_\theta(\mathbf{x}^{(i)})$$

LATENT VARIABLE MODELS



model:

$$\underbrace{p_\theta(\mathbf{x}, \mathbf{z})}_{\textit{joint}} = \underbrace{p_\theta(\mathbf{x}|\mathbf{z})}_{\substack{\textit{conditional}\\ \textit{likelihood}}}\underbrace{p_\theta(\mathbf{z})}_{\textit{prior}}$$

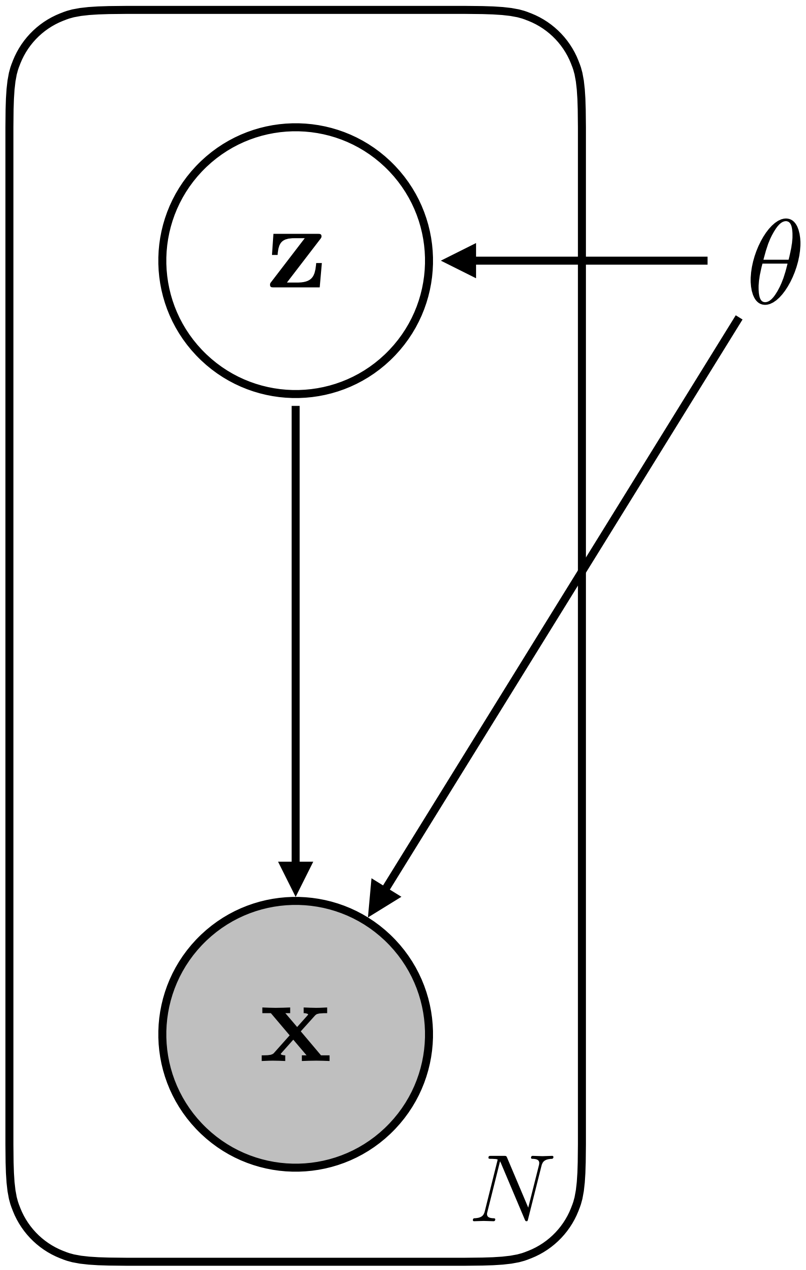
marginalization:

$$\underbrace{p_\theta(\mathbf{x})}_{\substack{\textit{marginal}\\ \textit{likelihood}}} = \int p_\theta(\mathbf{x}, \mathbf{z}) d\mathbf{z}$$

inference:

$$\underbrace{p_\theta(\mathbf{z}|\mathbf{x})}_{\textit{posterior}} = \frac{p_\theta(\mathbf{x}, \mathbf{z})}{p_\theta(\mathbf{x})}$$

LATENT VARIABLE MODELS



maximum likelihood is typically intractable

$$\theta^* = \arg\max_\theta \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \left[\log p_\theta(\mathbf{x})\right]$$

$$\approx \arg\max_\theta \frac{1}{N} \sum_{i=1}^{N} \log p_\theta(\mathbf{x}^{(i)})$$

$$\approx \arg\max_\theta \frac{1}{N} \sum_{i=1}^{N} \log \Bigg[\underbrace{\int p_\theta(\mathbf{x}^{(i)}, \mathbf{z}) d\mathbf{z}}_{\text{intractable integral}} \Bigg]$$

must resort to approximation techniques

VARIATIONAL INFERENCE

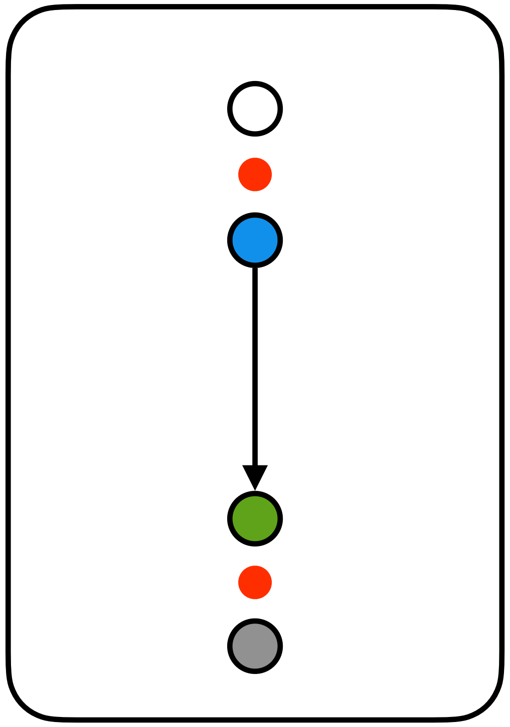
approximate posterior $q(\mathbf{z}|\mathbf{x})$

variational lower bound

$$\log p_\theta(\mathbf{x}) \geq \mathcal{L}(\mathbf{x}; q)$$

where

$$\mathcal{L}(\mathbf{x}; q) = \mathbb{E}_q \Big[\underbrace{\log p_\theta(\mathbf{x}|\mathbf{z})}_{\text{"reconstruction"}} - \underbrace{\log \frac{q(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z})}}_{\text{"regularization"}} \Big]$$



VARIATIONAL INFERENCE

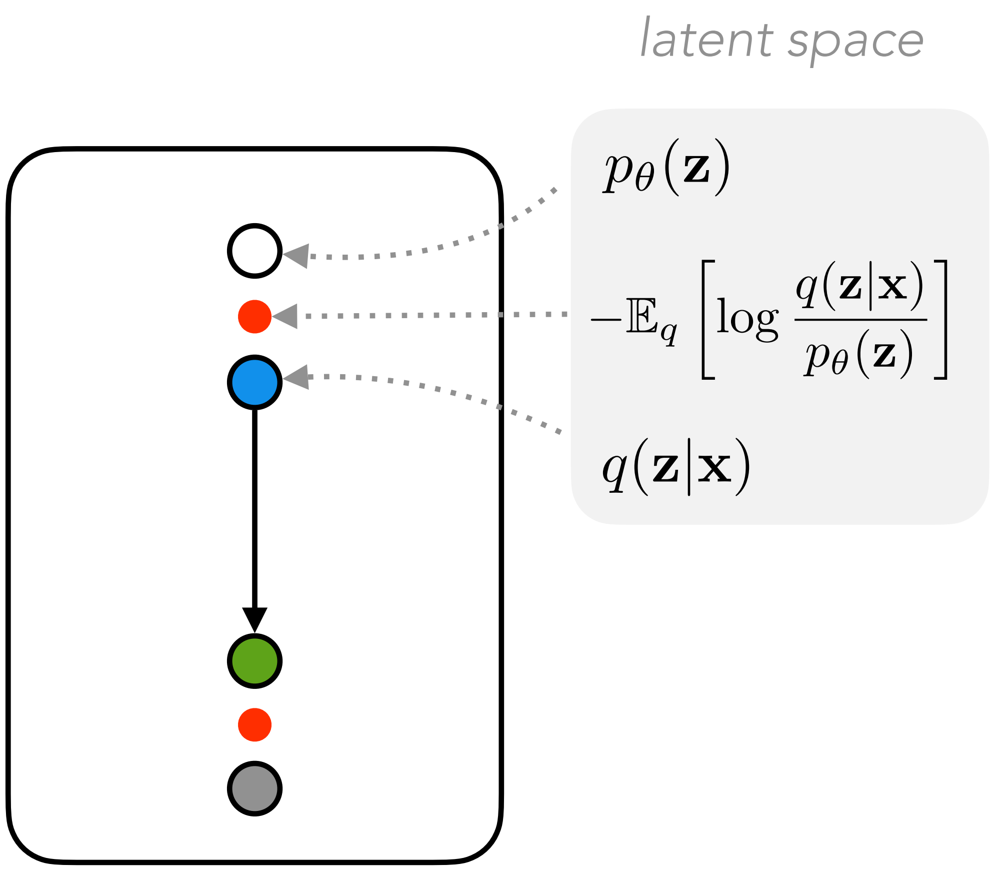
approximate posterior $q(\mathbf{z}|\mathbf{x})$

variational lower bound

$$\log p_\theta(\mathbf{x}) \geq \mathcal{L}(\mathbf{x}; q)$$

where

$$\mathcal{L}(\mathbf{x}; q) = \mathbb{E}_q \left[\underbrace{\log p_\theta(\mathbf{x}|\mathbf{z})}_{\text{“reconstruction”}} - \underbrace{\log \frac{q(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z})}}_{\text{“regularization”}} \right]$$



VARIATIONAL INFERENCE

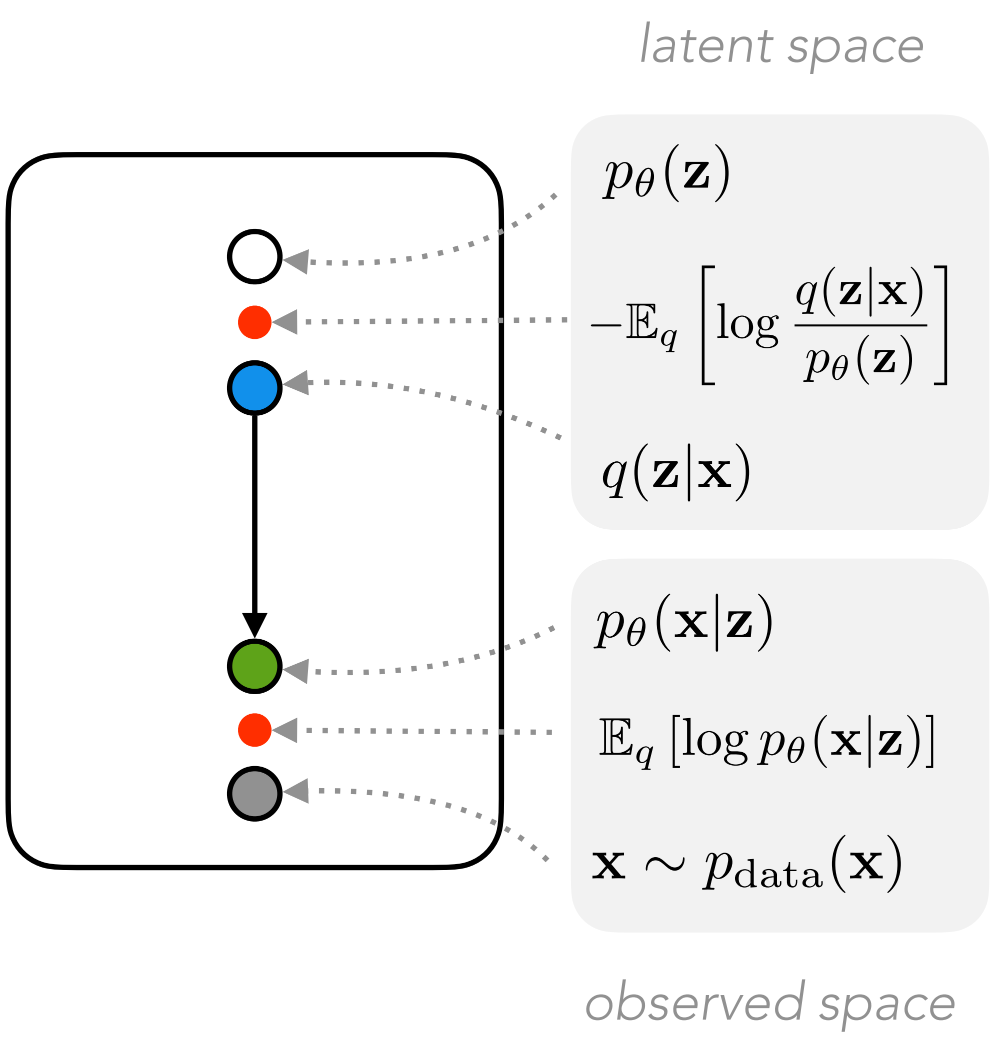
approximate posterior $q(\mathbf{z}|\mathbf{x})$

variational lower bound

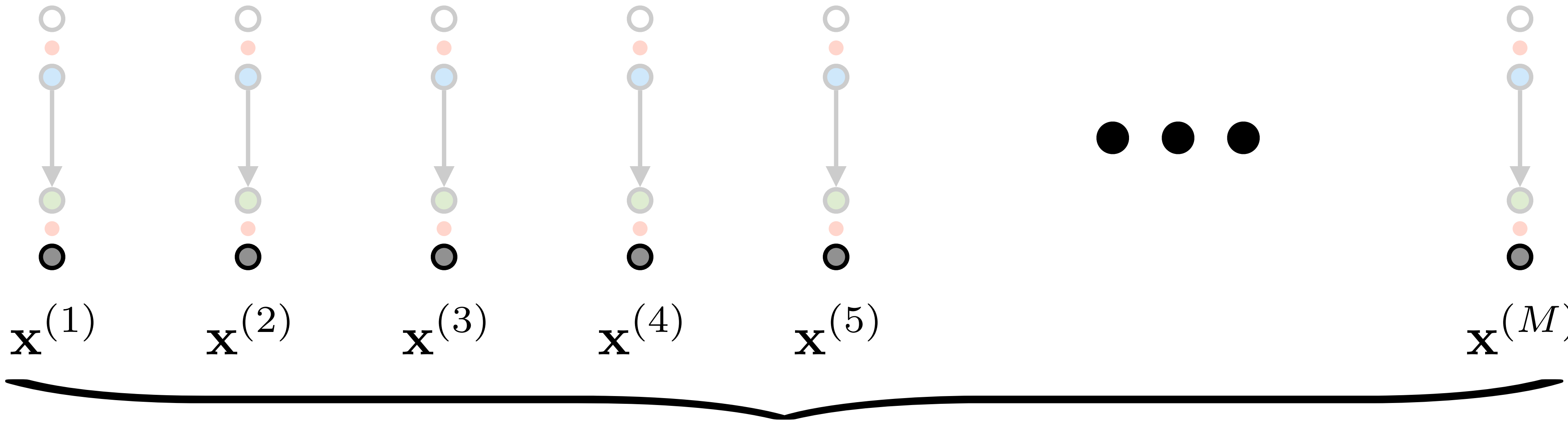
$$\log p_\theta(\mathbf{x}) \geq \mathcal{L}(\mathbf{x}; q)$$

where

$$\mathcal{L}(\mathbf{x}; q) = \mathbb{E}_q \left[\underbrace{\log p_\theta(\mathbf{x}|\mathbf{z})}_{\text{"reconstruction"}} - \underbrace{\log \frac{q(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z})}}_{\text{"regularization"}} \right]$$



VARIATIONAL EXPECTATION MAXIMIZATION

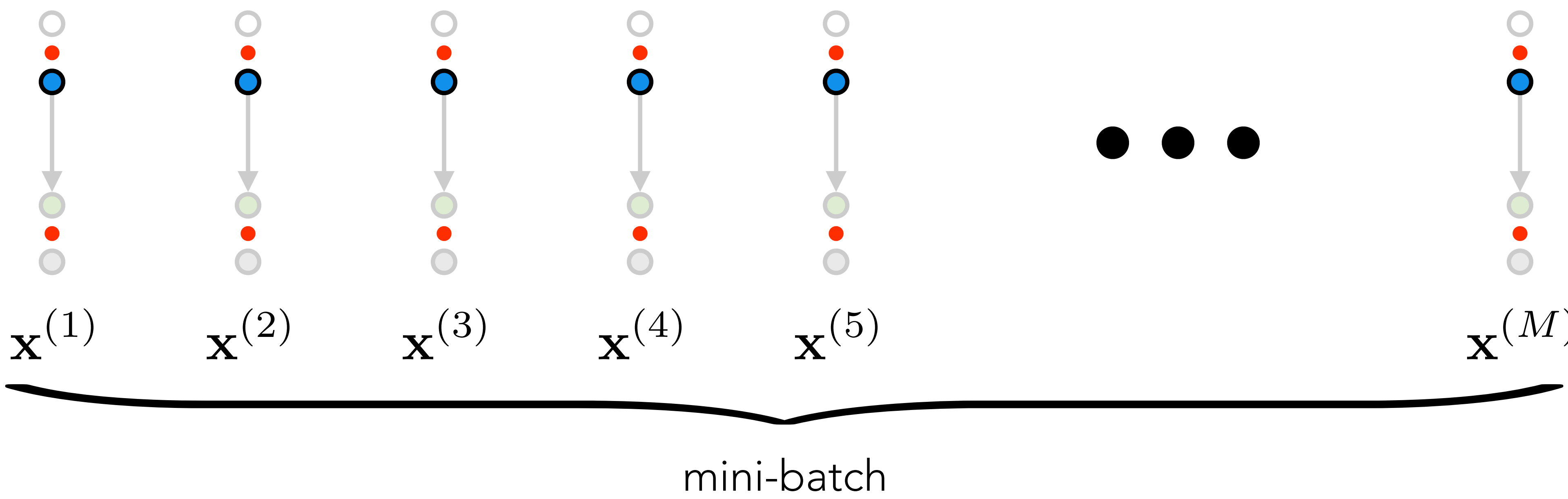


mini-batch

Variational EM (single-step)

`sample` $\mathbf{x}^{(1:M)} \sim p_{\text{data}}(\mathbf{x})$

VARIATIONAL EXPECTATION MAXIMIZATION



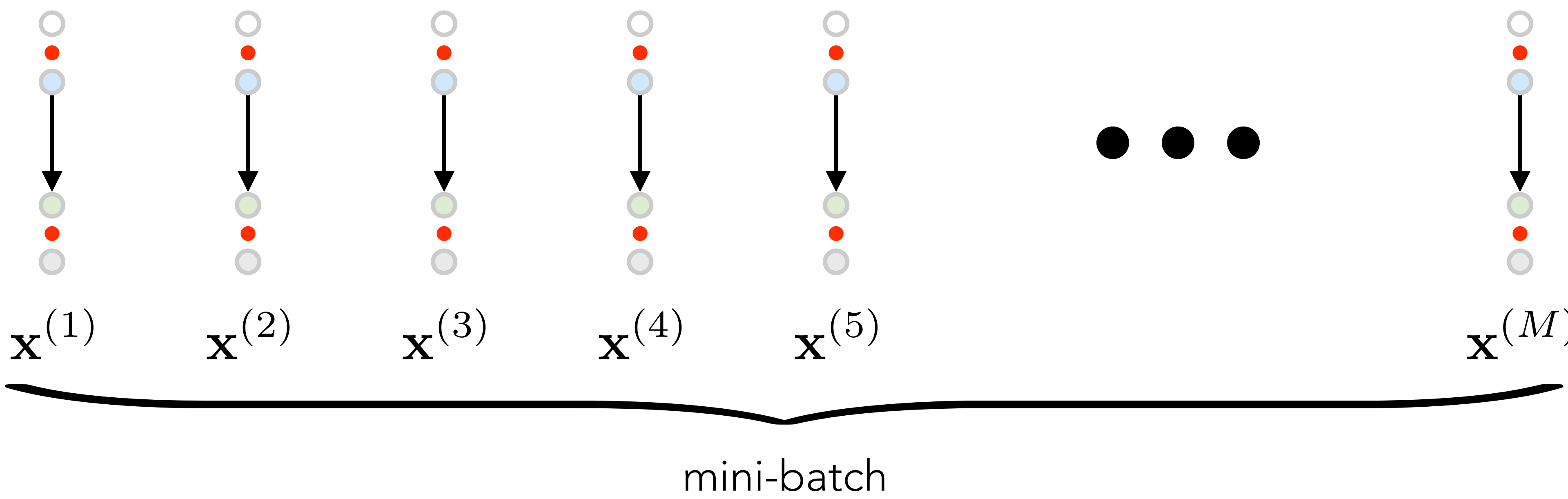
Variational EM (single-step)

sample $\mathbf{x}^{(1:M)} \sim p_{\text{data}}(\mathbf{x})$

for $\mathbf{x}^{(i)}$ in $\mathbf{x}^{(1:M)}$:

    maximize $\mathcal{L}(\mathbf{x}^{(i)}, q^{(i)})$ w.r.t. $q^{(i)}$ *# E-step*

VARIATIONAL EXPECTATION MAXIMIZATION



Variational EM (single-step)

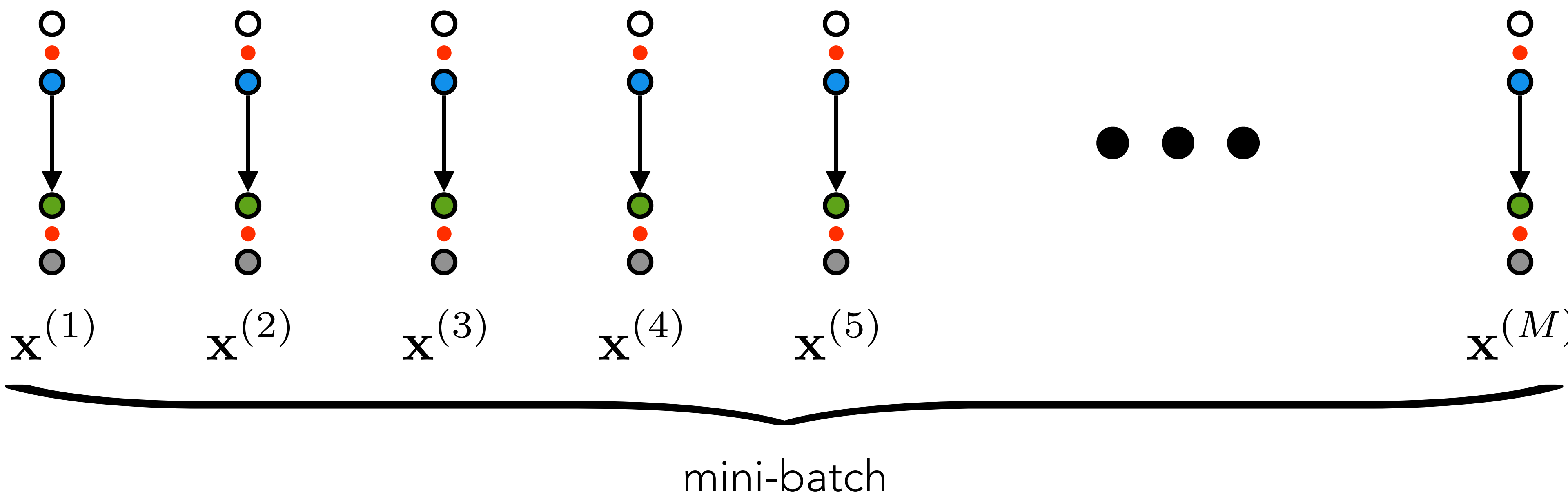
`sample` $\mathbf{x}^{(1:M)} \sim p_{\text{data}}(\mathbf{x})$

`for` $\mathbf{x}^{(i)}$ `in` $\mathbf{x}^{(1:M)}$:

  `maximize` $\mathcal{L}(\mathbf{x}^{(i)}, q^{(i)})$ `w.r.t.` $q^{(i)}$ *# E-step*

`maximize` $\frac{1}{M}\sum_{i=1}^{M} \mathcal{L}(\mathbf{x}^{(i)}, q^{(i)})$ `w.r.t.` θ *# M-step*

VARIATIONAL EXPECTATION MAXIMIZATION



Variational EM (single-step)

sample $\mathbf{x}^{(1:M)} \sim p_{\text{data}}(\mathbf{x})$

for $\mathbf{x}^{(i)}$ in $\mathbf{x}^{(1:M)}$:

  maximize $\mathcal{L}(\mathbf{x}^{(i)}, q^{(i)})$ w.r.t. $q^{(i)}$ *# E-step*

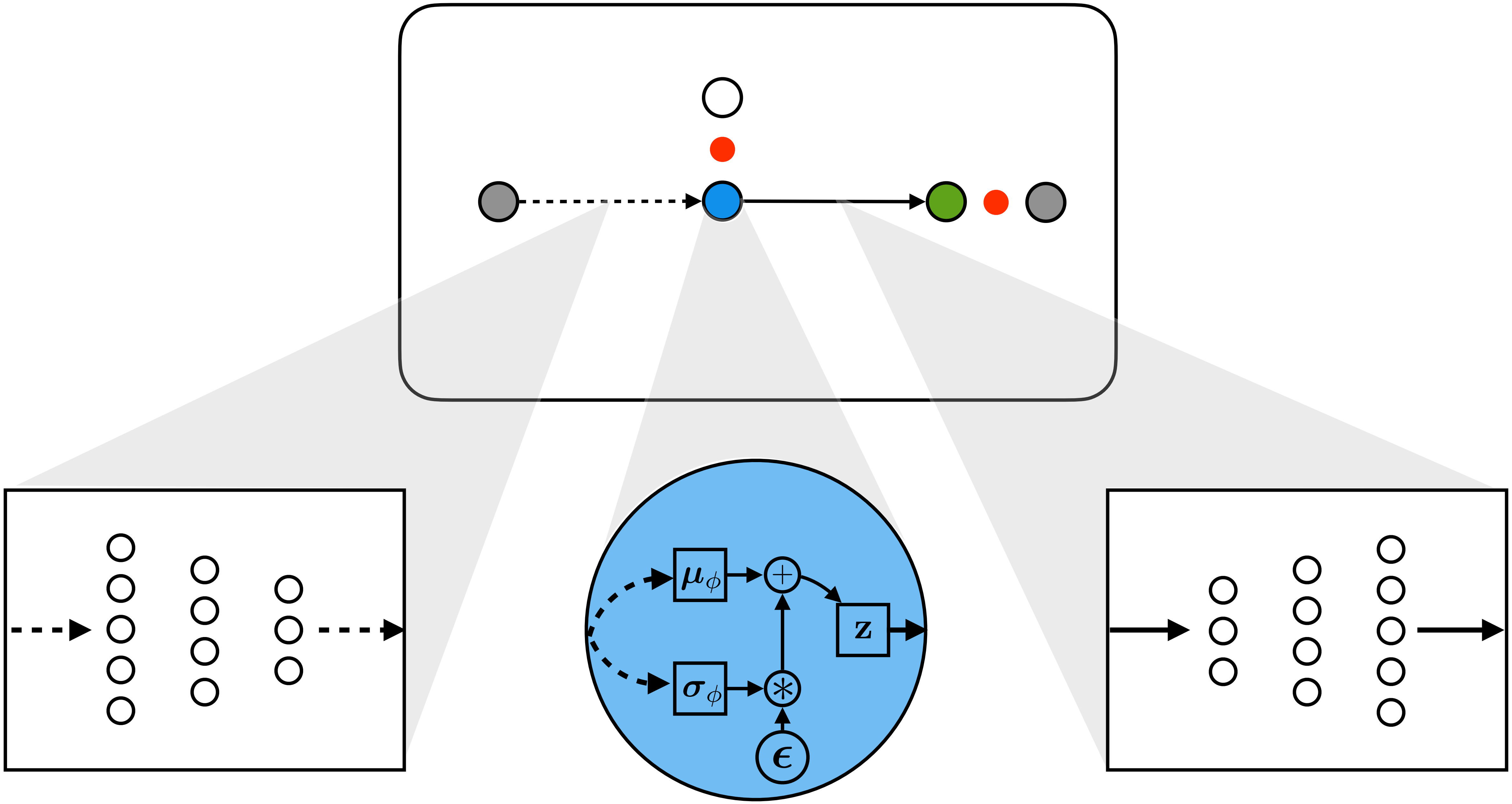
maximize $\frac{1}{M} \sum_{i=1}^{M} \mathcal{L}(\mathbf{x}^{(i)}, q^{(i)})$ w.r.t. θ *# M-step*

expensive

VARIATIONAL AUTOENCODERS

Variational Autoencoder (VAE):

deep latent variable model + variational inference + direct encoder + *reparameterized* Gaussian



Kingma & Welling, 2014

Rezende et al., 2014

AMORTIZED VARIATIONAL INFERENCE

let $\boldsymbol{\lambda}$ be the distribution parameters of $q(\mathbf{z}|\mathbf{x})$, for example, $\boldsymbol{\lambda} = \{\boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$

inference optimization: $q(\mathbf{z}|\mathbf{x}) \leftarrow \arg\max_q \mathcal{L}(\mathbf{x}; q)$

BLACK-BOX VARIATIONAL INFERENCE

gradient-based optimization

$$\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + \eta \nabla_{\boldsymbol{\lambda}} \mathcal{L}$$

DIRECT AMORTIZED INFERENCE

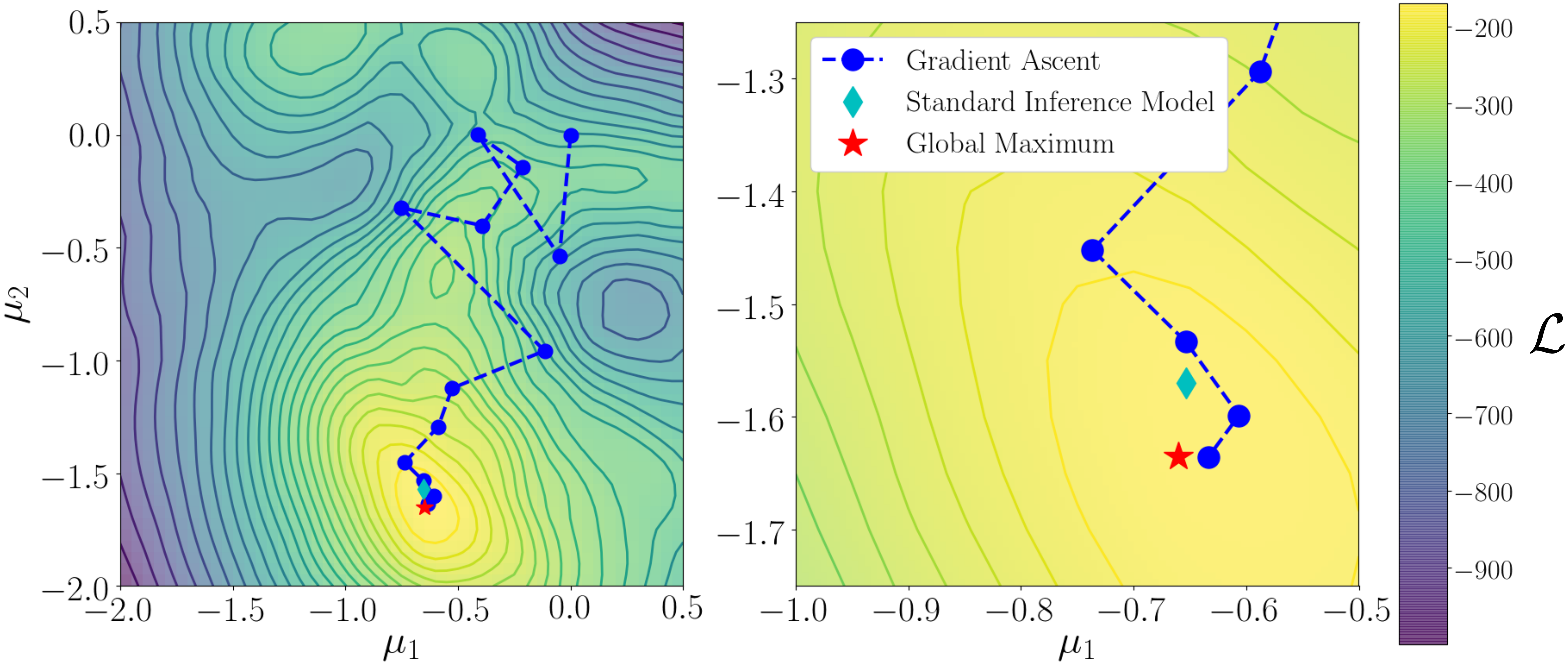
learn a direct mapping

$$\boldsymbol{\lambda} \leftarrow f_\phi(\mathbf{x})$$

efficient, but potentially inaccurate

INFERENCE OPTIMIZATION

2D model, MNIST



inference models may not reach fully optimized estimates

see also: **Inference Suboptimality in Variational Autoencoders**, Cremer *et al.*, 2018

Marino *et al.*, 2018a

ITERATIVE AMORTIZED INFERENCE

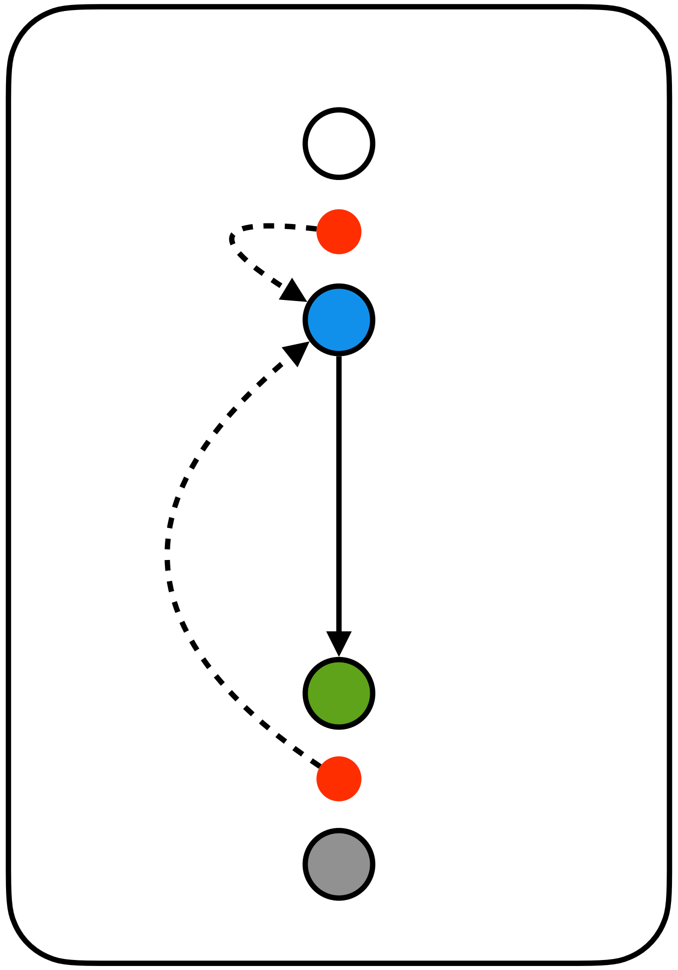
let $\boldsymbol{\lambda}$ be the distribution parameters of $q(\mathbf{z}|\mathbf{x})$, for example, $\boldsymbol{\lambda} = \{\boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$

inference optimization: $q(\mathbf{z}|\mathbf{x}) \leftarrow \arg\max_q \mathcal{L}(\mathbf{x}; q)$

ITERATIVE AMORTIZED INFERENCE

learn an iterative mapping

$$\boldsymbol{\lambda} \leftarrow f_\phi(\boldsymbol{\lambda}, \nabla_{\boldsymbol{\lambda}} \mathcal{L})$$

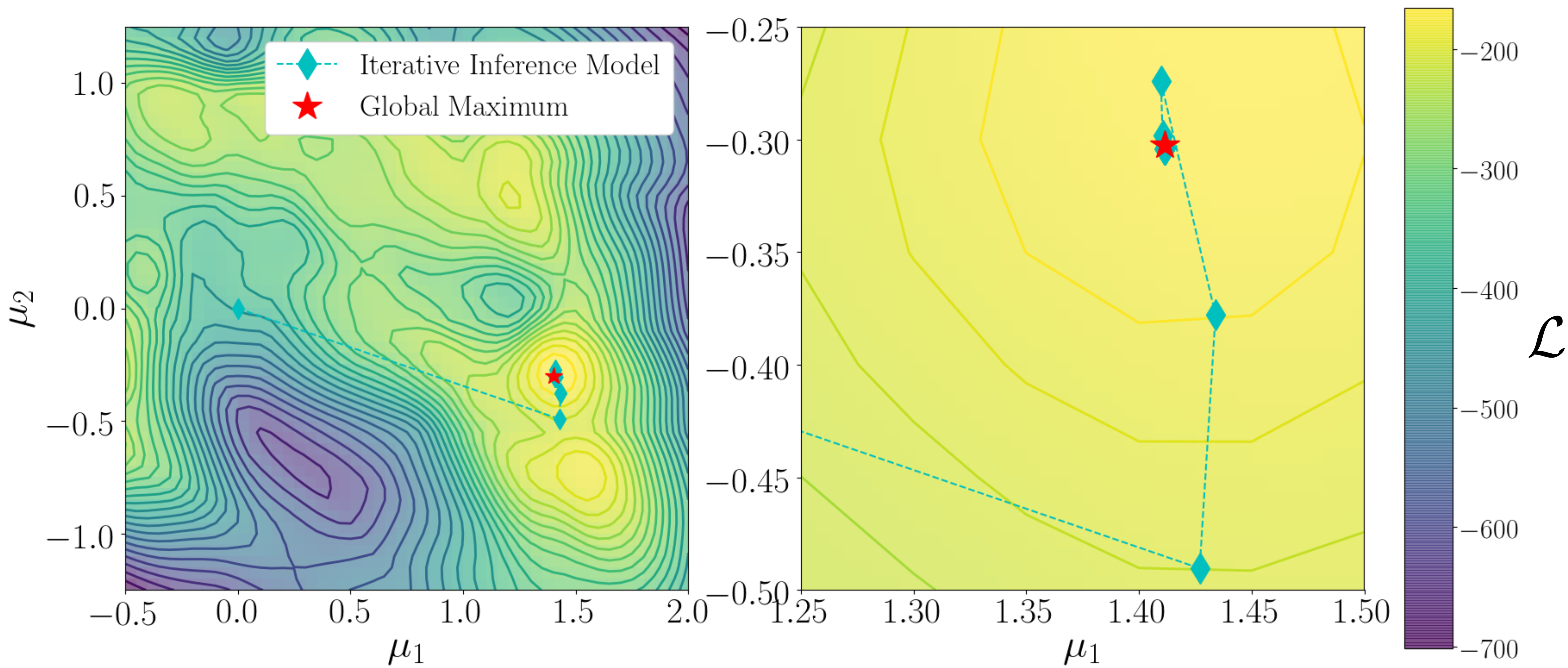


Marino *et al.*, 2018a

INFERENCE OPTIMIZATION

directly visualize inference in the optimization landscape

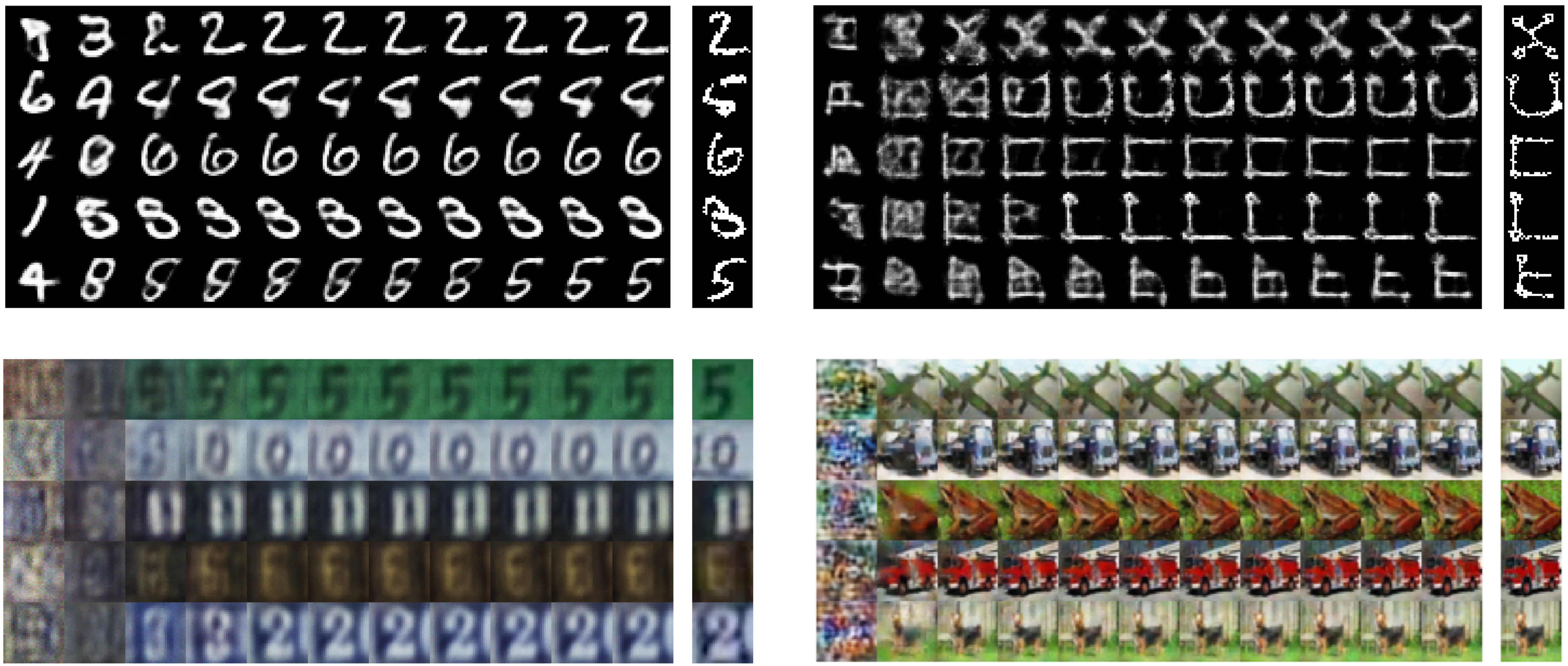
2D model, MNIST



Marino *et al.*, 2018a

INFERENCE OPTIMIZATION

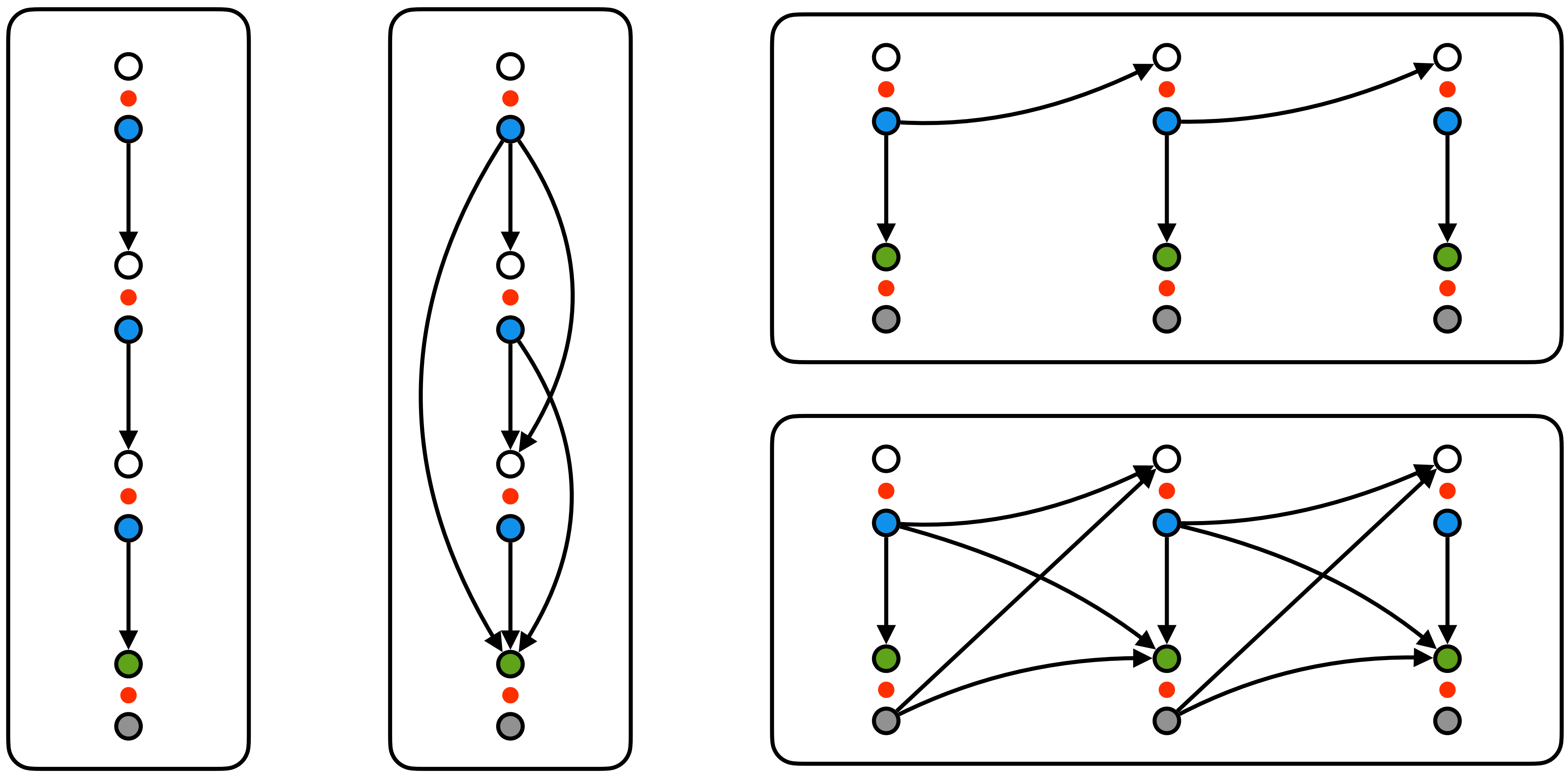
visualize data reconstructions over inference iterations



Marino *et al.*, 2018a

STRUCTURED APPROXIMATE POSTERIORS

structured models

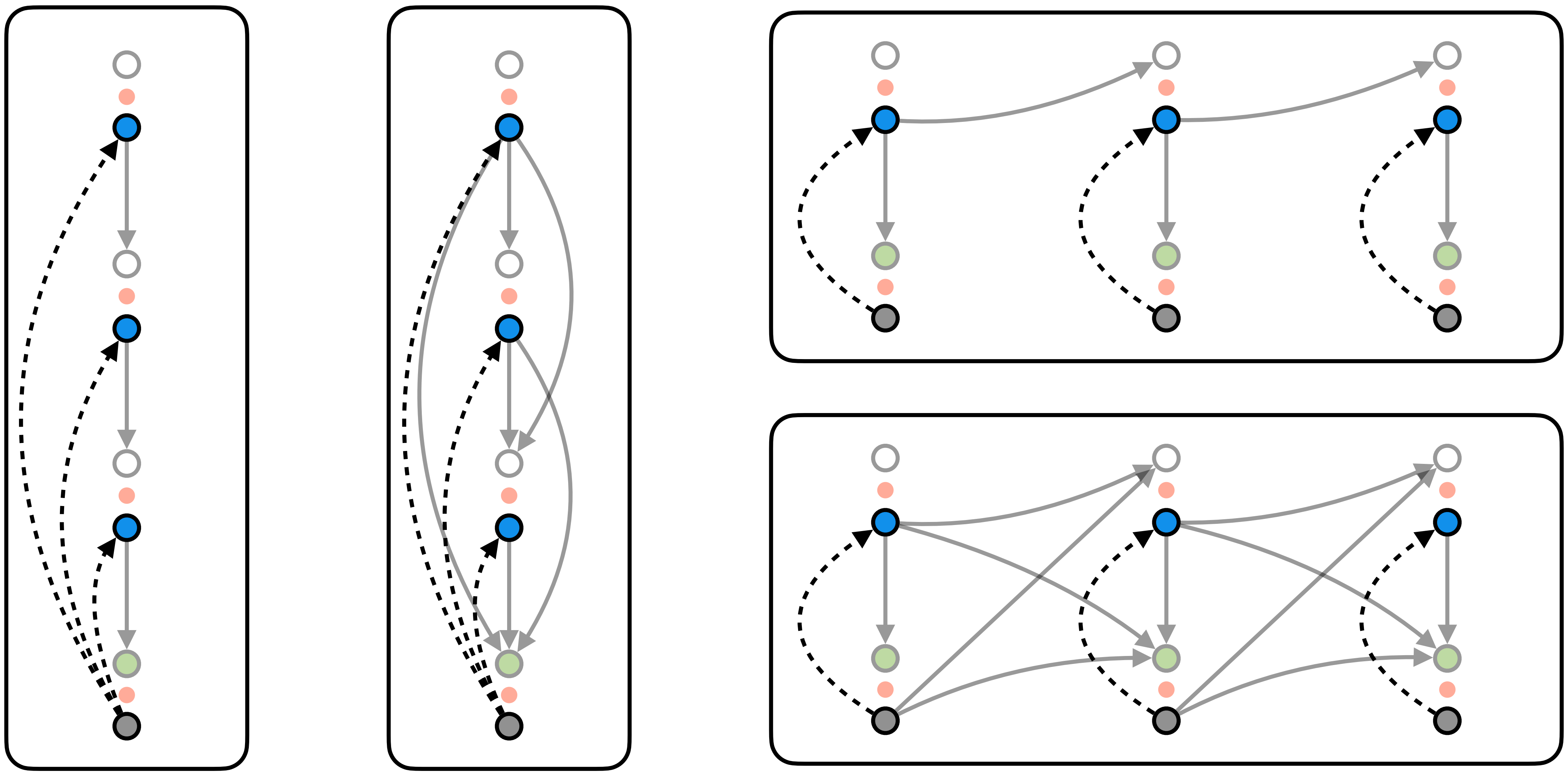


hierarchical

sequential

STRUCTURED APPROXIMATE POSTERIORS

structured models



hierarchical

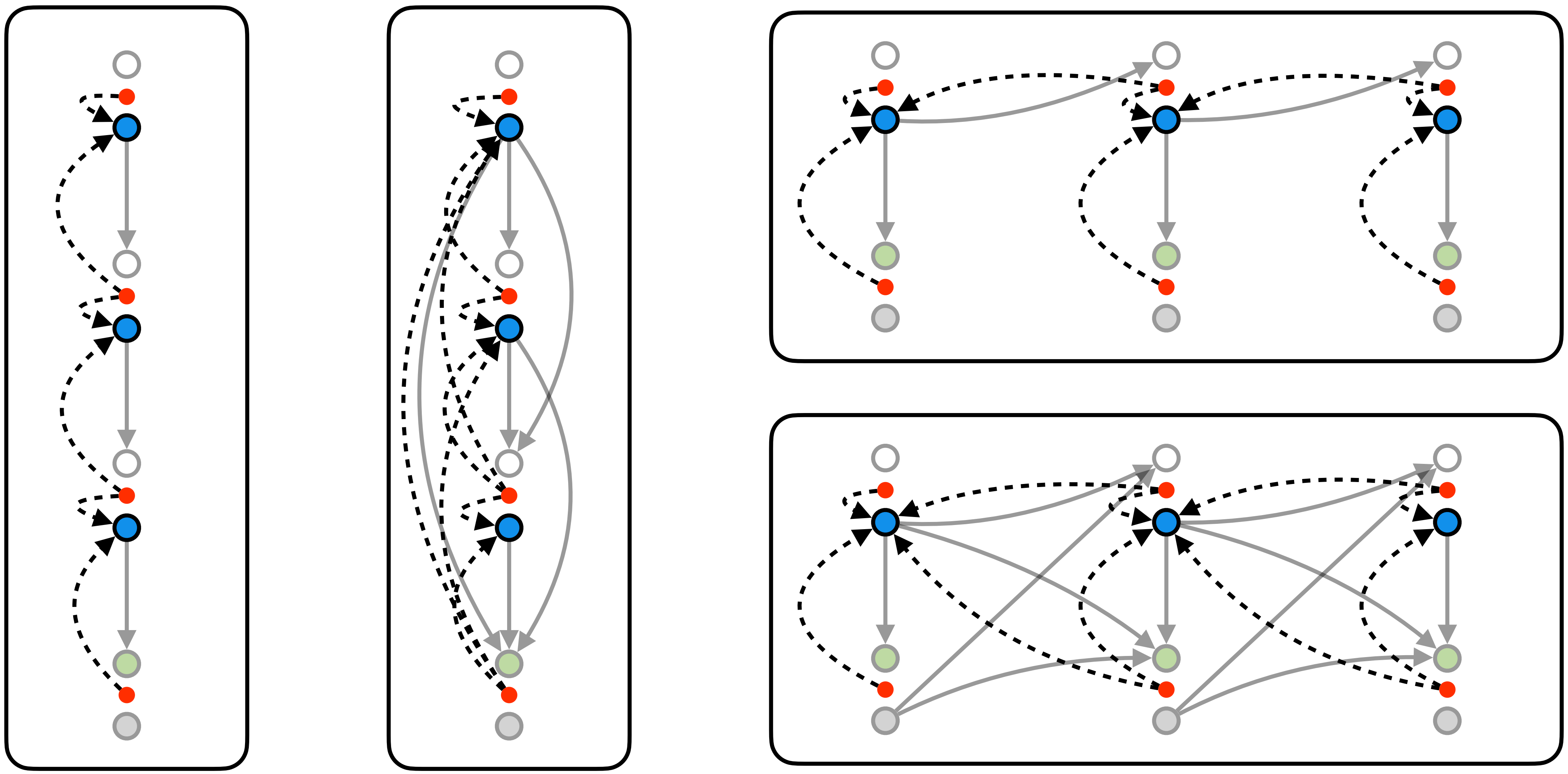
sequential

(naïve) direct encoders cannot account for structured estimates

$\mathbf{z}_k$ depends on $\mathbf{z}_{<k}$, but $q_\phi(\mathbf{z}_k|\mathbf{x})$ does not have access to this information

STRUCTURED APPROXIMATE POSTERIORS

structured models

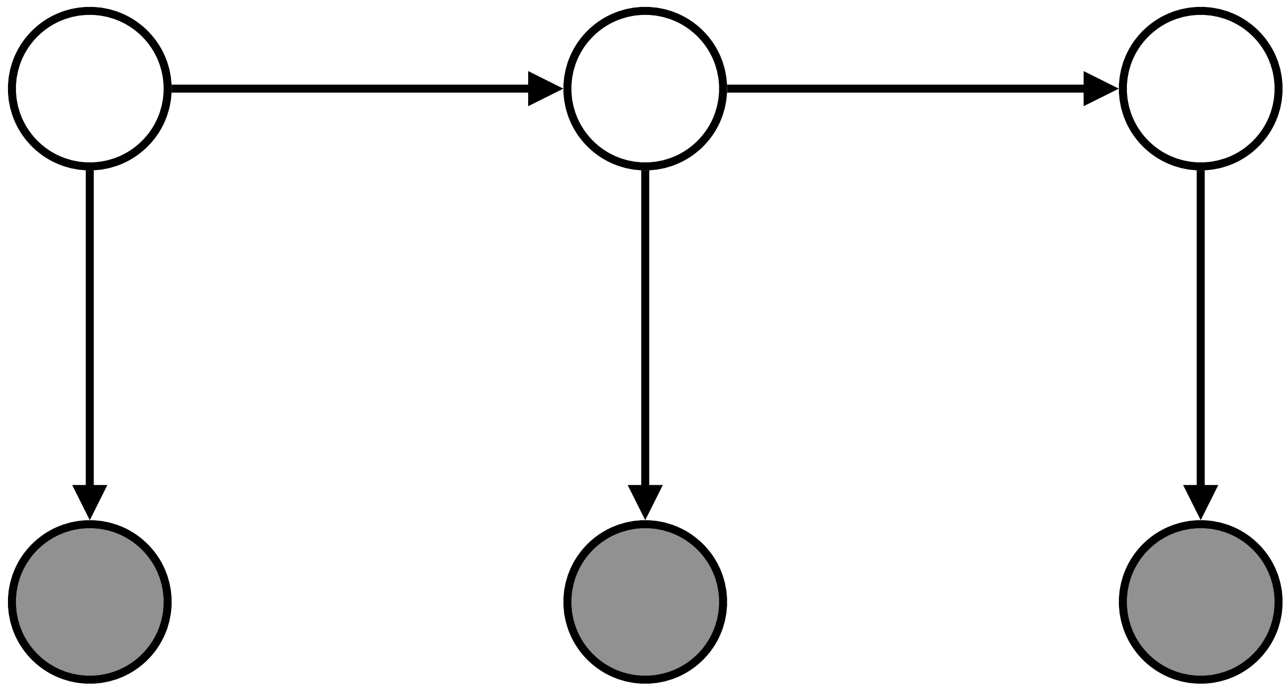


hierarchical

sequential

iterative encoders can be easily extended to structured estimates

structure defines gradients, which define inference

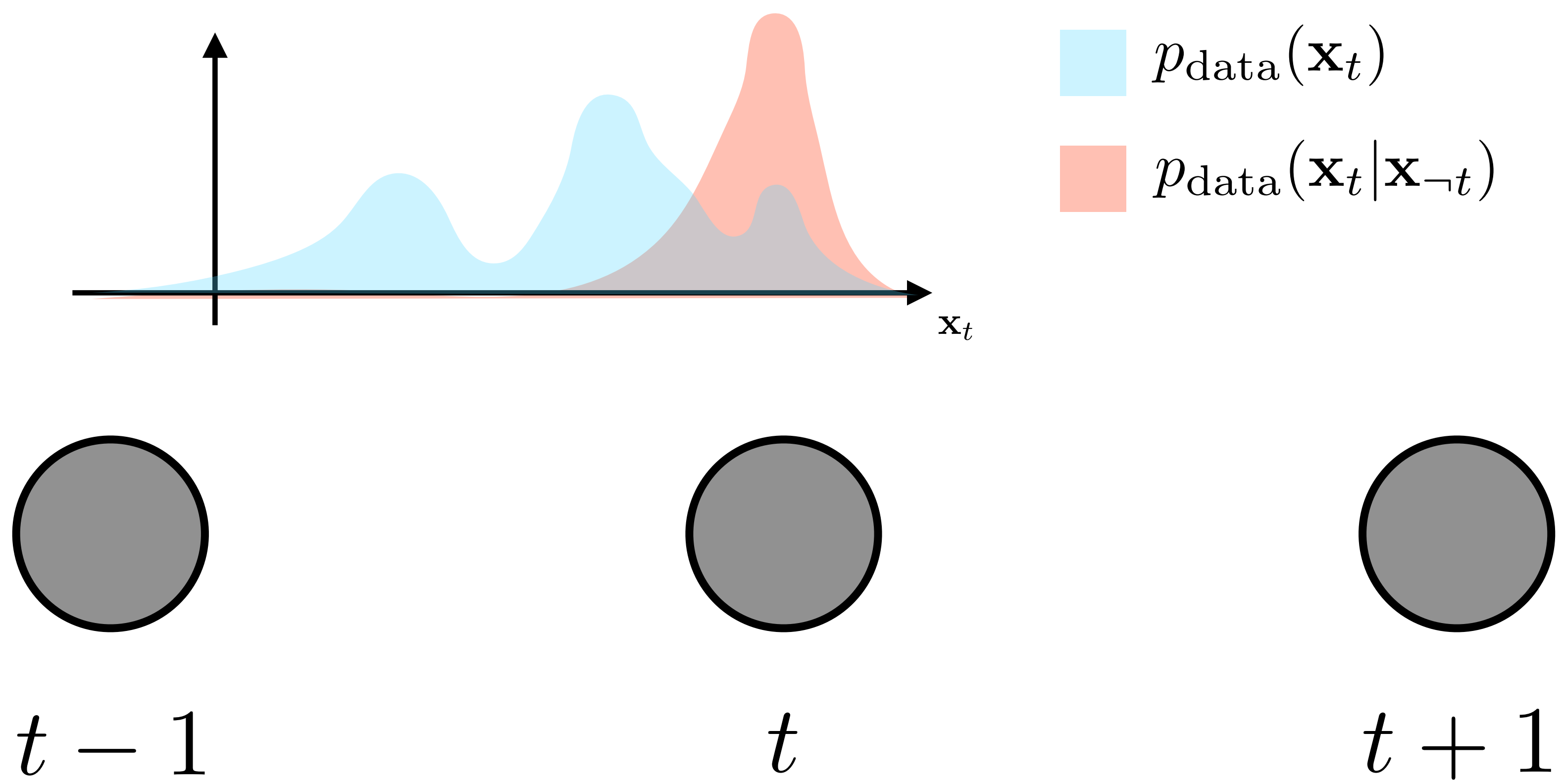


DEEP SEQUENTIAL LATENT VARIABLE MODELS

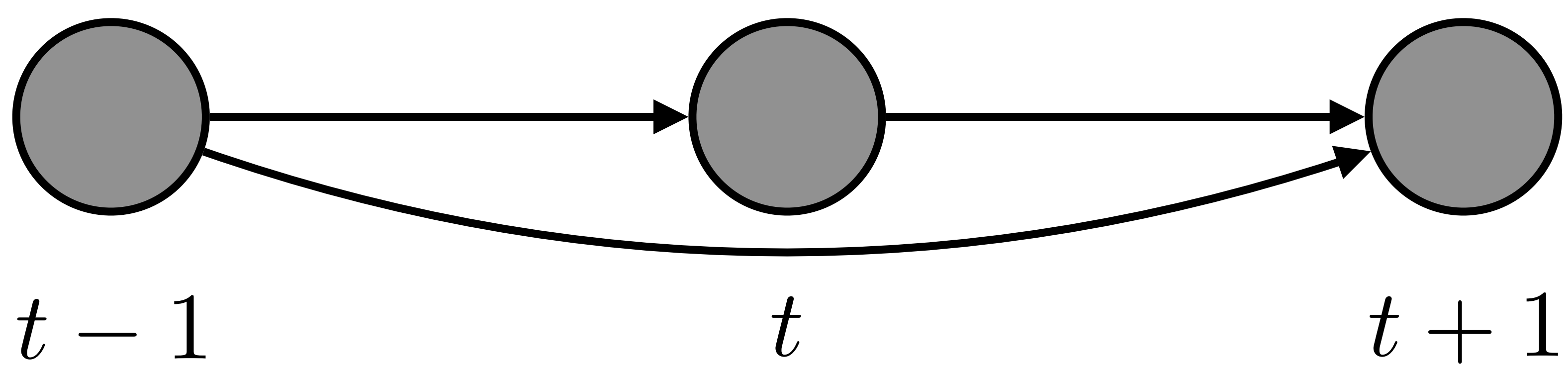
dynamics: dependence in time

multi-information: $\mathcal{I}(\mathbf{x}_{1:T}) = \sum_t \mathcal{H}(\mathbf{x}_t) - \mathcal{H}(\mathbf{x}_{1:T}) \geq 0$

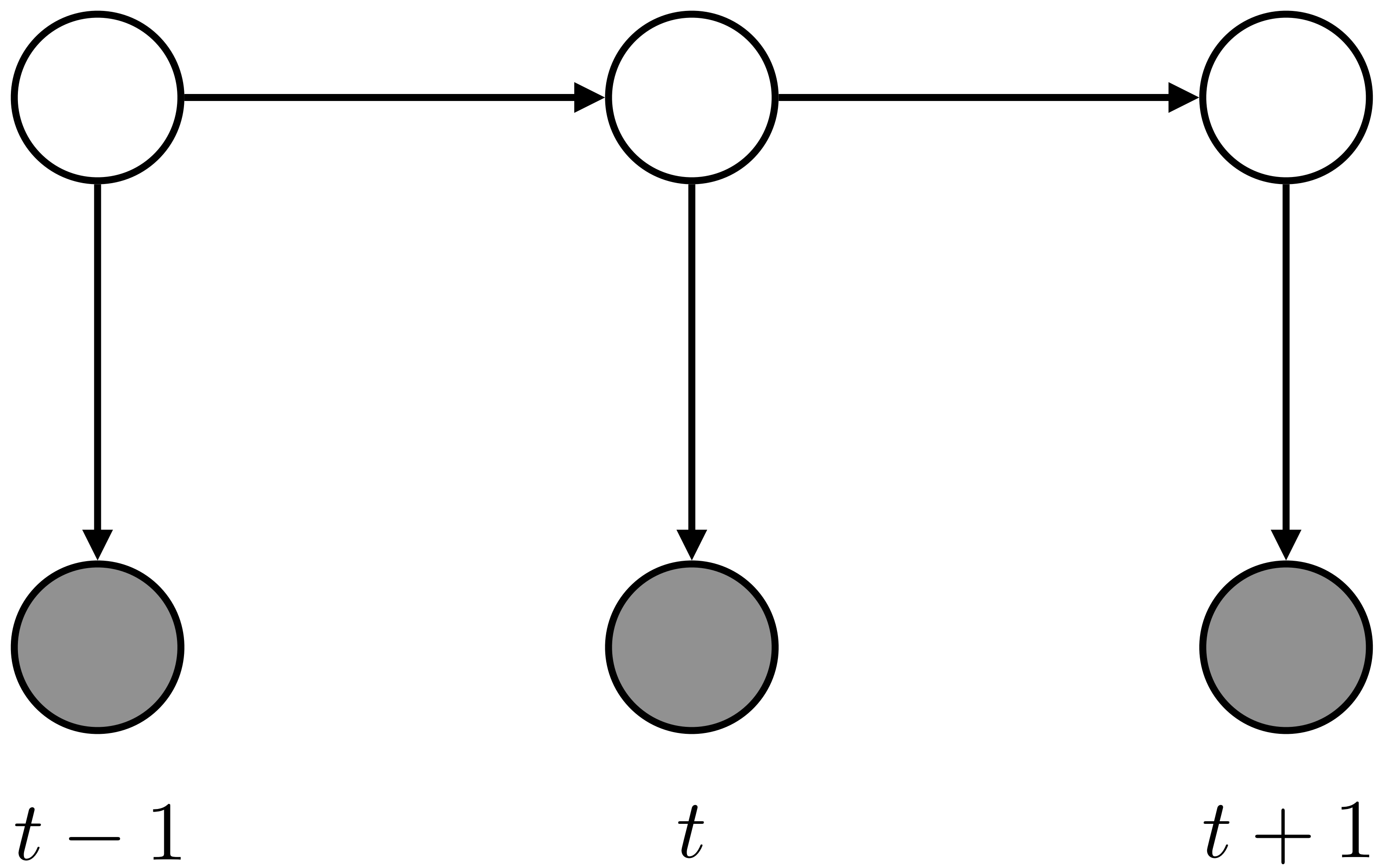
observing $\mathbf{x}_{\neg t}$ *reduces uncertainty* in $\mathbf{x}_t$



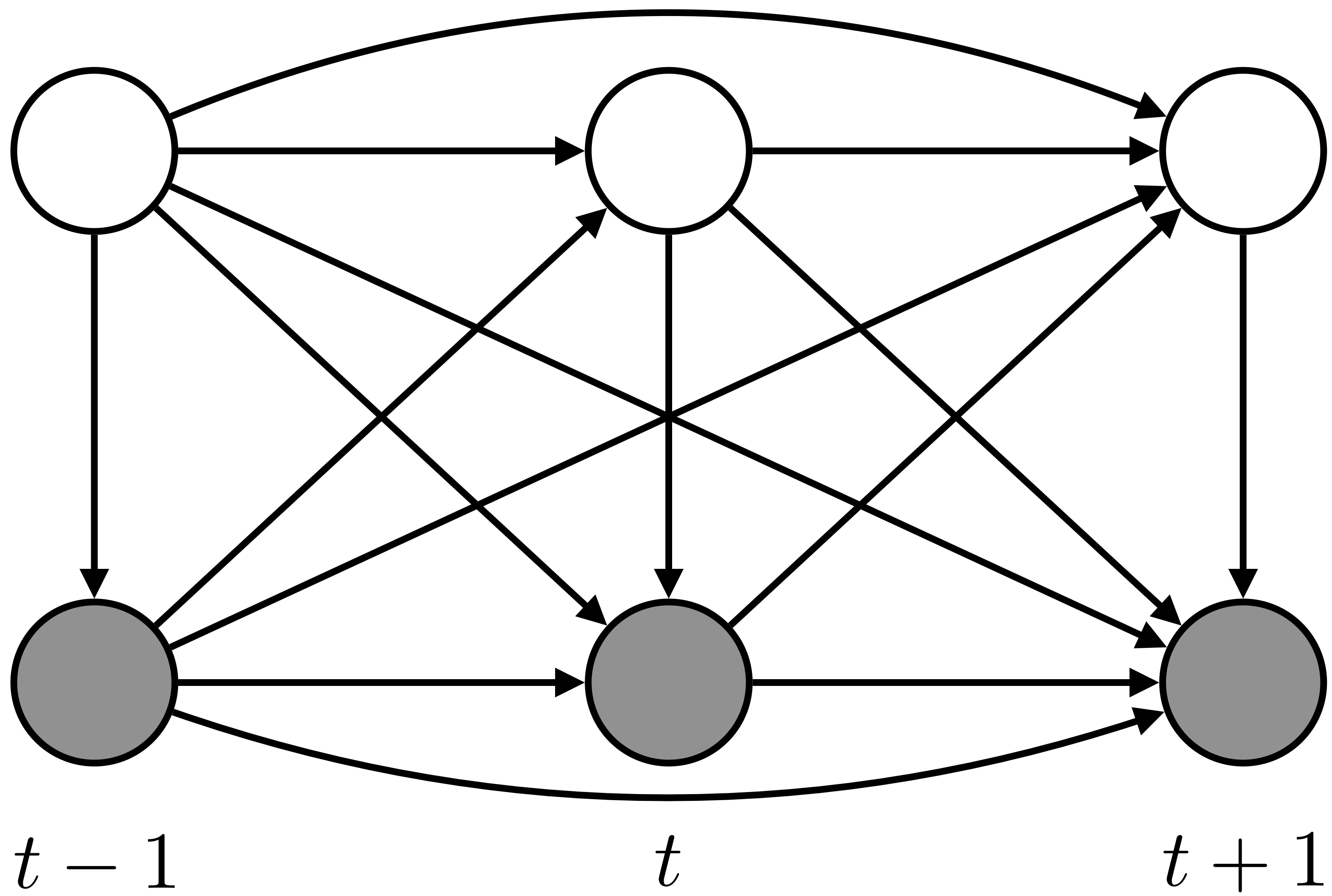
model temporal dependencies



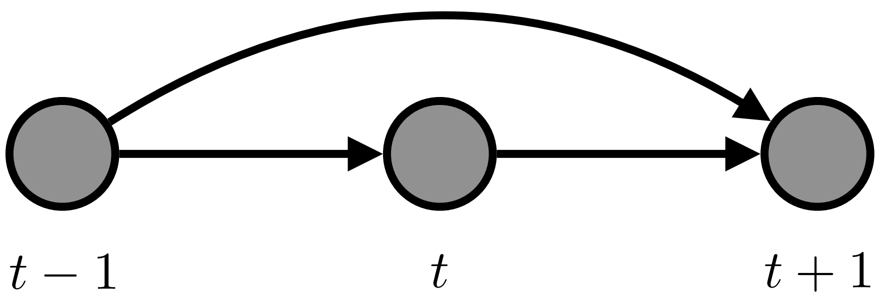
model temporal dependencies



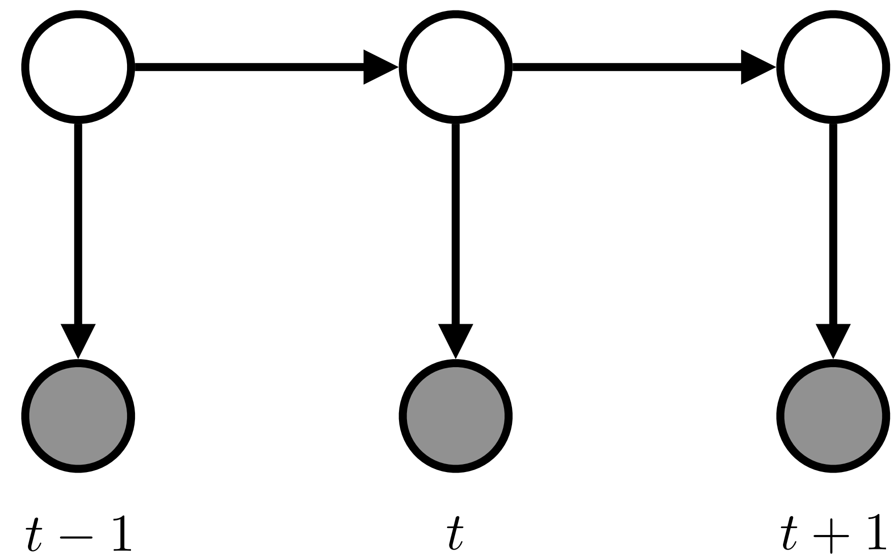
model temporal dependencies



MODELING DYNAMICS



fully-observed



latent

$$p_\theta(\mathbf{x}_t|\mathbf{x}_{<t}) = \int p_\theta(\mathbf{x}_t|\mathbf{z}_t)p_\theta(\mathbf{z}_t|\mathbf{x}_{<t})d\mathbf{z}_t$$

may be more flexible than a fixed-form $p_\theta(\mathbf{x}_t|\mathbf{x}_{<t})$

SEQUENTIAL LATENT VARIABLE MODELS

general form:

$$p_\theta(\mathbf{x}_{\leq T}, \mathbf{z}_{\leq T}) = \prod_{t=1}^{T} \underbrace{p_\theta(\mathbf{x}_t | \mathbf{x}_{<t}, \mathbf{z}_{\leq t})}_{\text{likelihood}} \underbrace{p_\theta(\mathbf{z}_t | \mathbf{x}_{<t}, \mathbf{z}_{<t})}_{\text{prior}}$$

where $\mathbf{x}_{\leq T}$ is a sequence of T observed variables

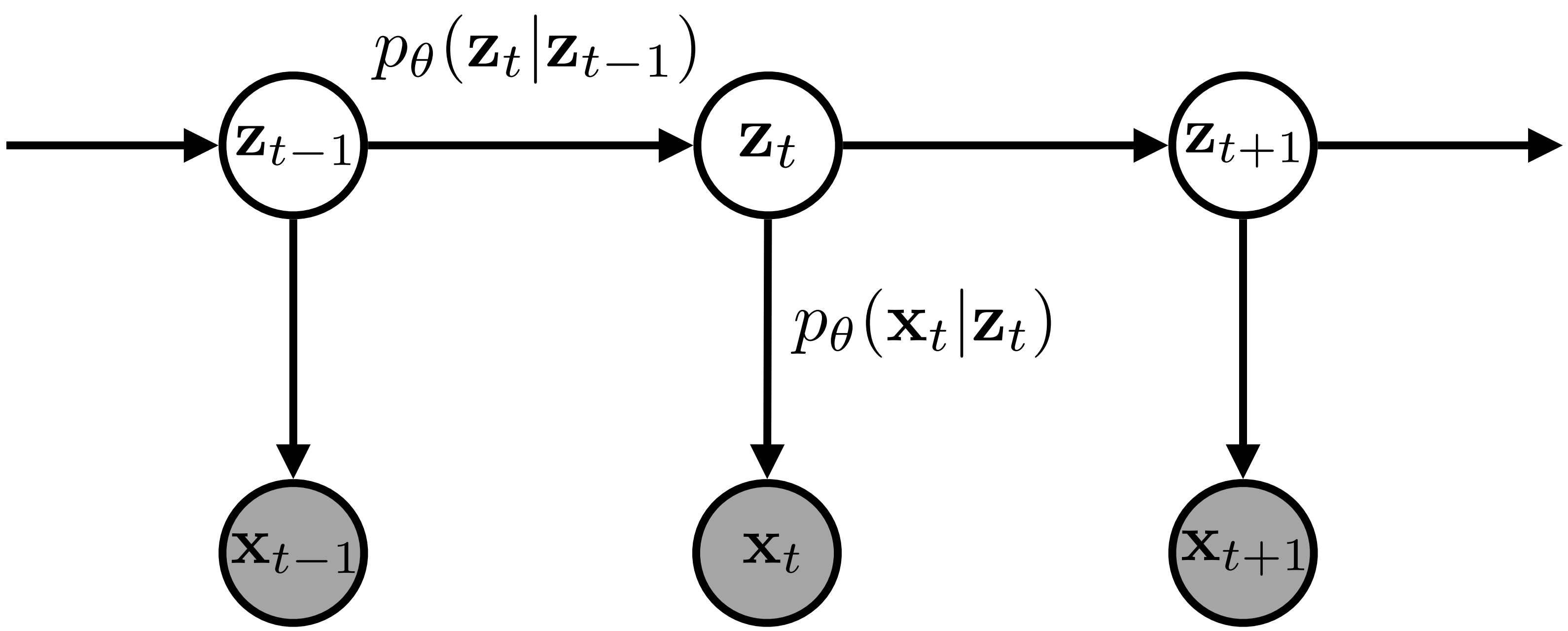
$\mathbf{z}_{\leq T}$ is a sequence of T latent variables

SEQUENTIAL LATENT VARIABLE MODELS

general form:

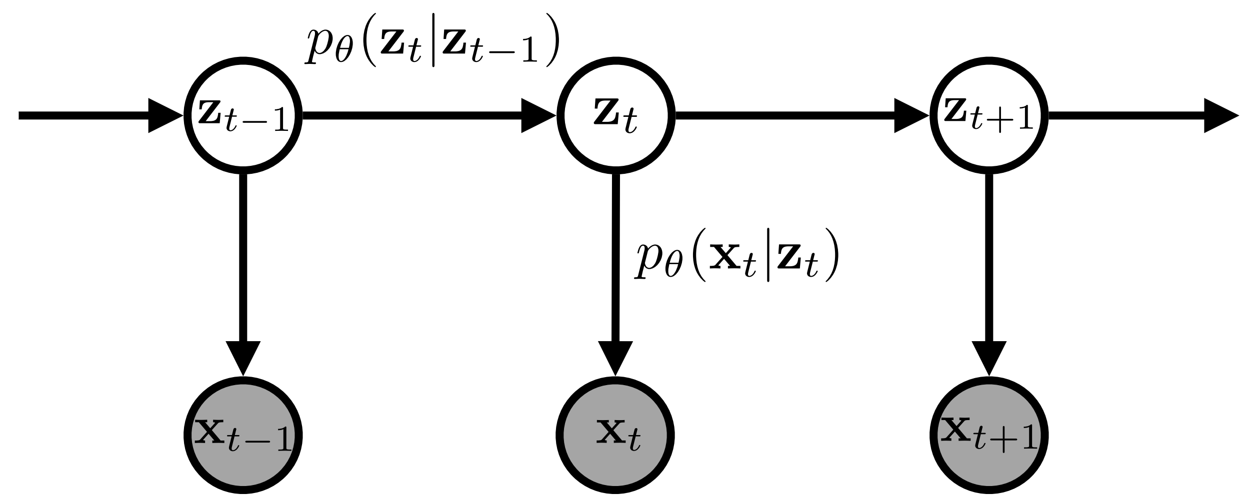
$$p_\theta(\mathbf{x}_{\leq T}, \mathbf{z}_{\leq T}) = \prod_{t=1}^{T} \underbrace{p_\theta(\mathbf{x}_t | \mathbf{x}_{<t}, \mathbf{z}_{\leq t})}_{\text{likelihood}} \underbrace{p_\theta(\mathbf{z}_t | \mathbf{x}_{<t}, \mathbf{z}_{<t})}_{\text{prior}}$$

simplified case (hidden Markov model):



SEQUENTIAL LATENT VARIABLE MODELS

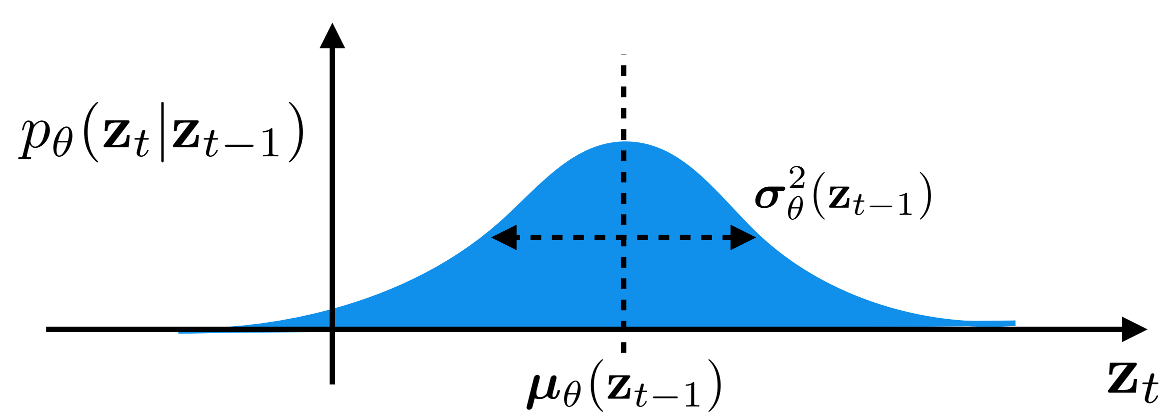
Markov model:



Parameterization:

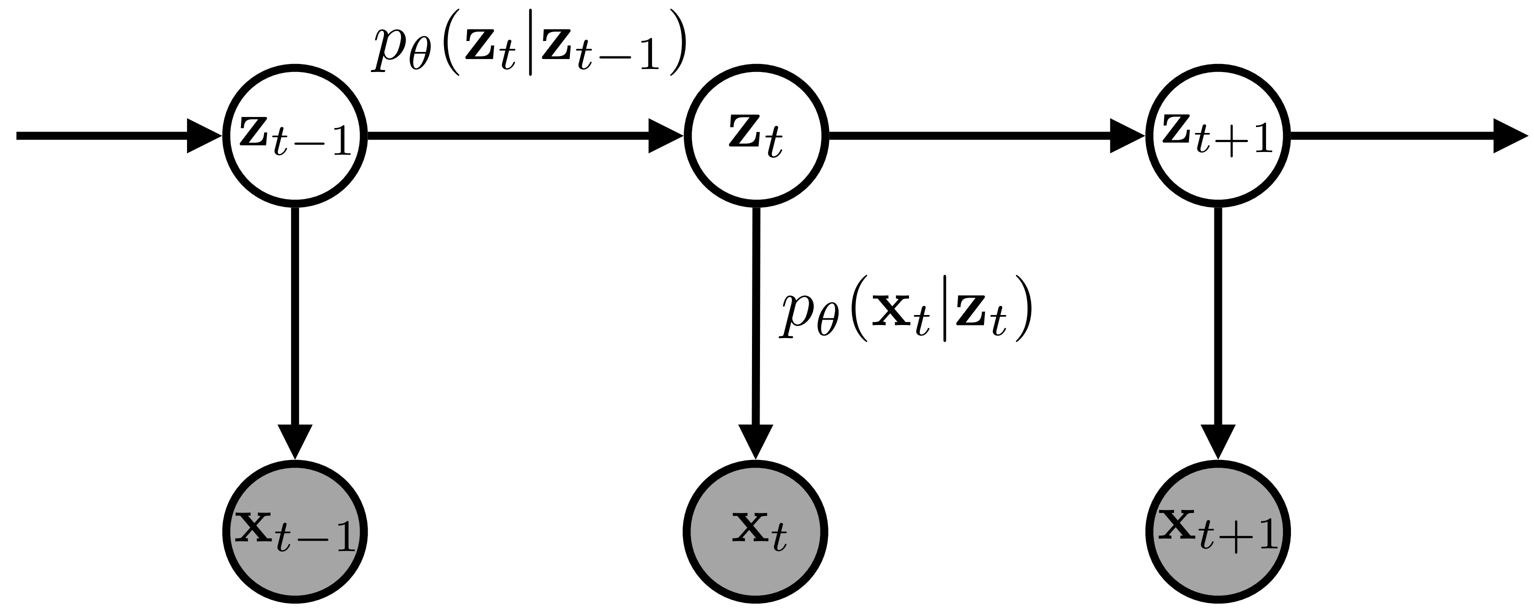
$p_\theta(\mathbf{z}_t|\mathbf{z}_{t-1})$ is typically an analytical distribution

for example, $p_\theta(\mathbf{z}_t|\mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t; \boldsymbol{\mu}_\theta(\mathbf{z}_{t-1}), \mathrm{diag}(\boldsymbol{\sigma}^2_\theta(\mathbf{z}_{t-1})))$



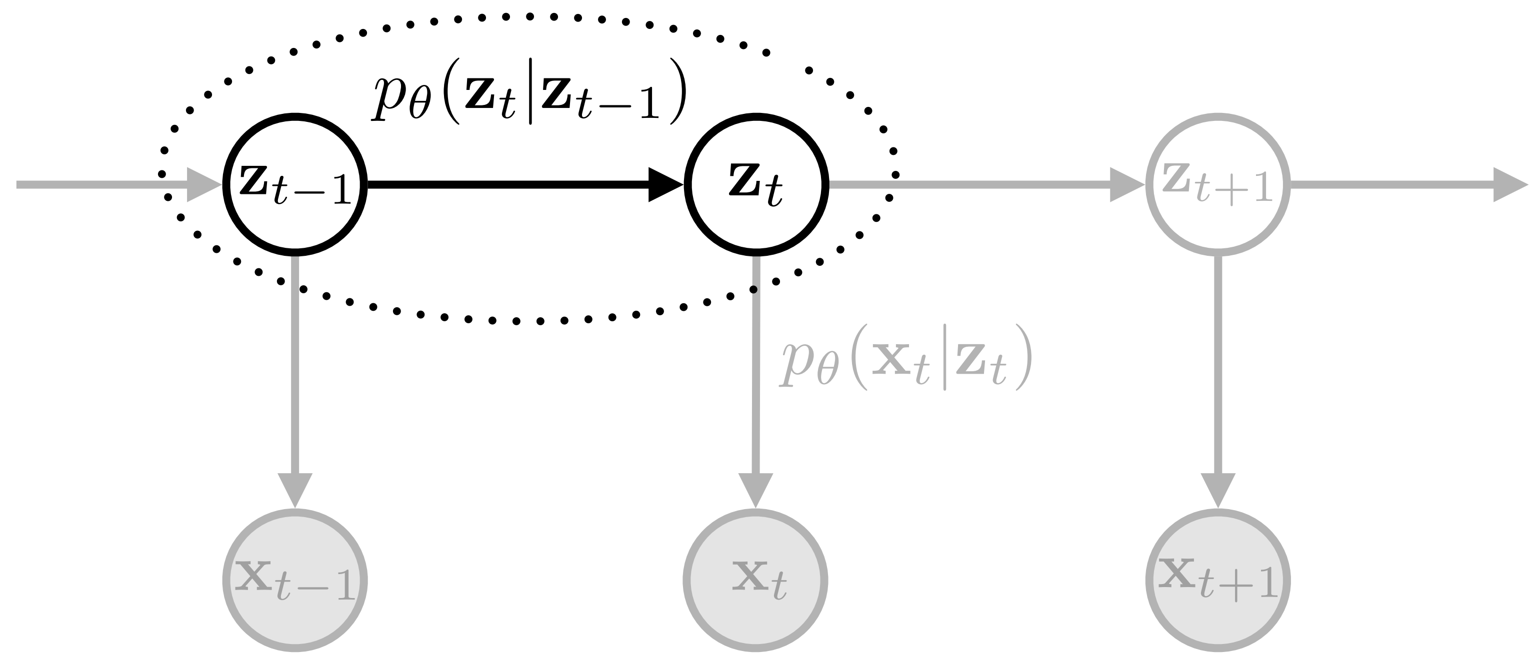
SEQUENTIAL LATENT VARIABLE MODELS

the parameters of these analytical distributions are functions, often *deep networks*



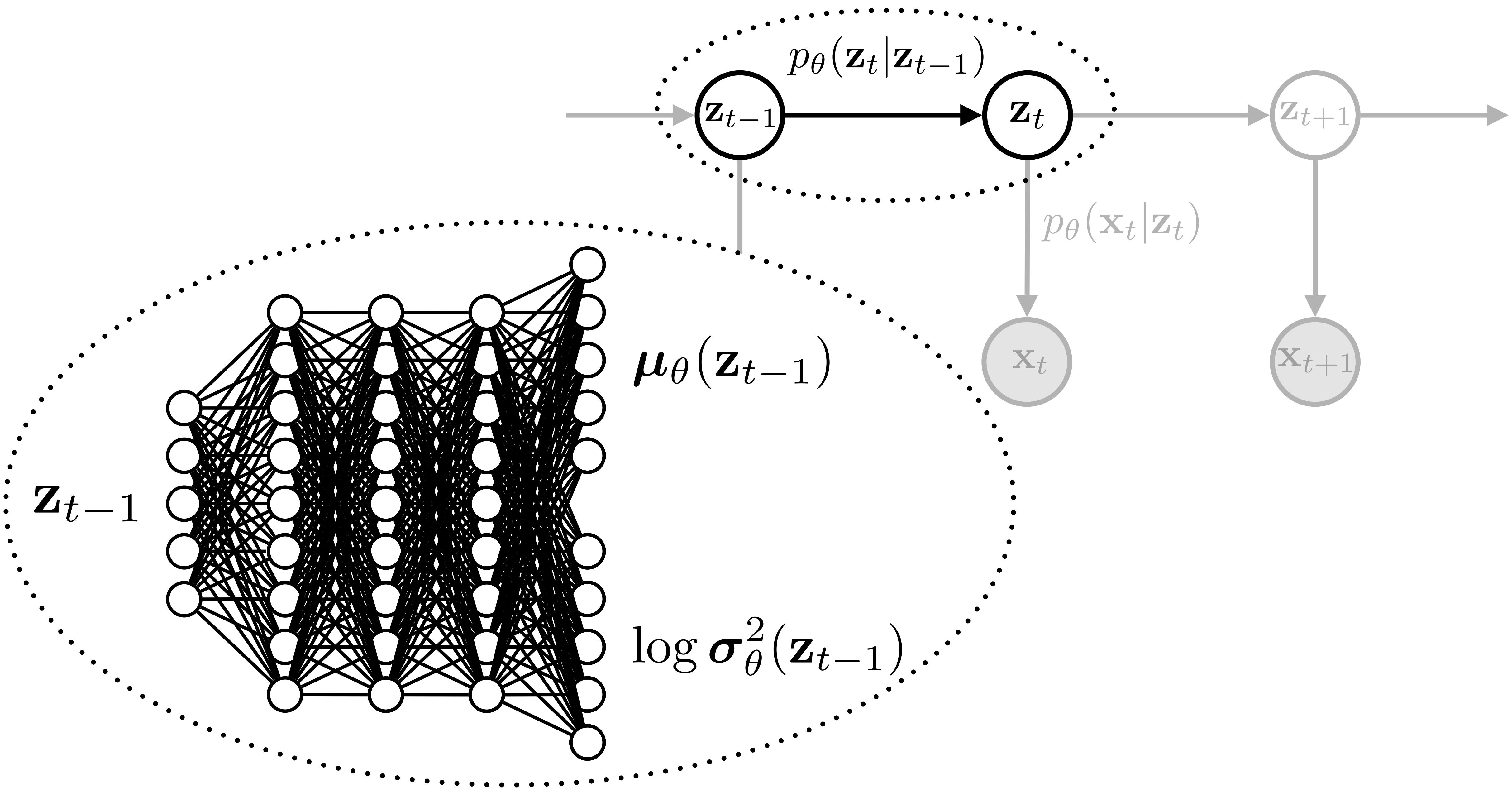
SEQUENTIAL LATENT VARIABLE MODELS

the parameters of these analytical distributions are functions, often *deep networks*



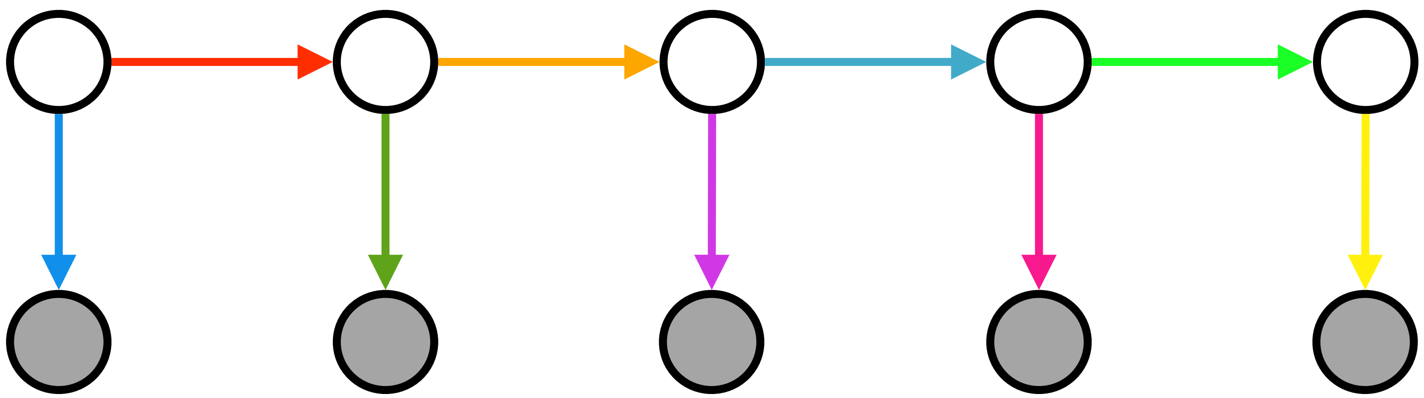
SEQUENTIAL LATENT VARIABLE MODELS

the parameters of these analytical distributions are functions, often *deep networks*



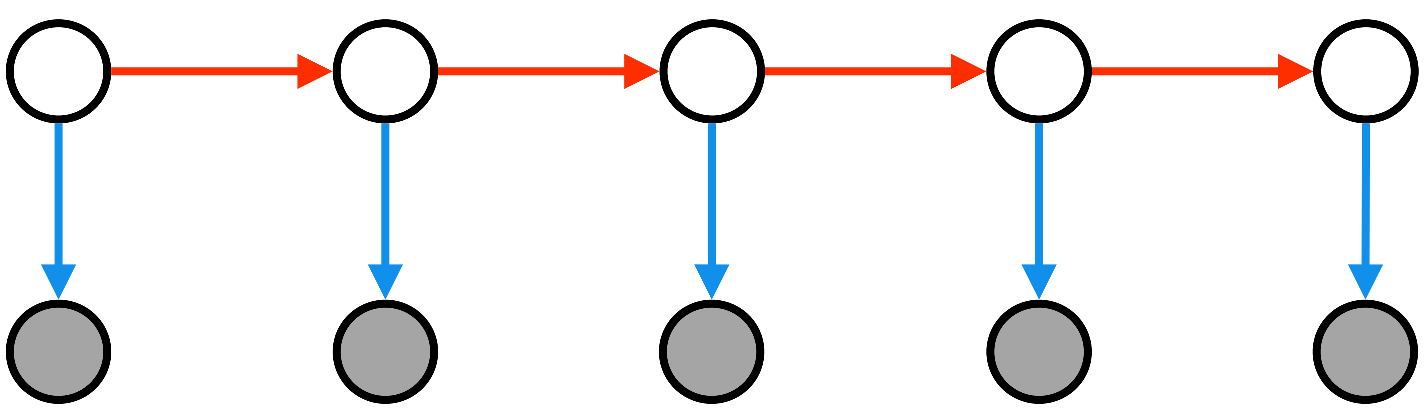
WEIGHT SHARING

could use a separate network for each conditional dependence



number of parameters grows linearly with time

share weights for similar conditional dependencies

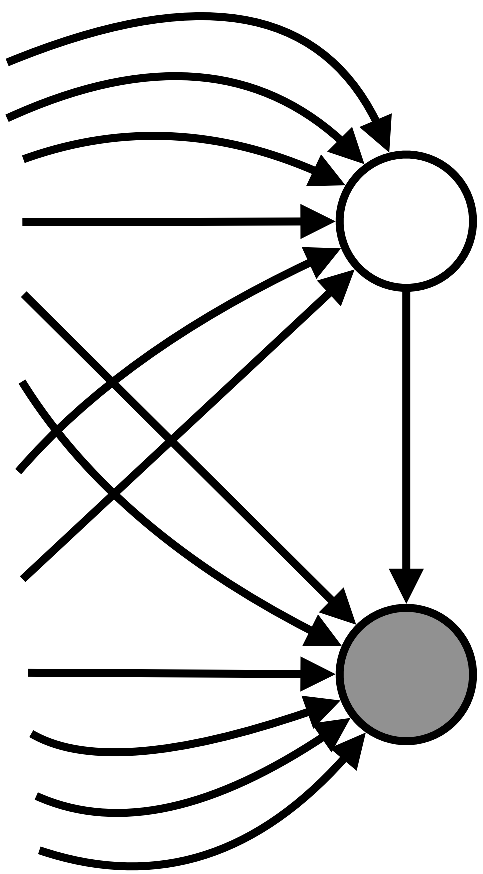


fixed number of parameters

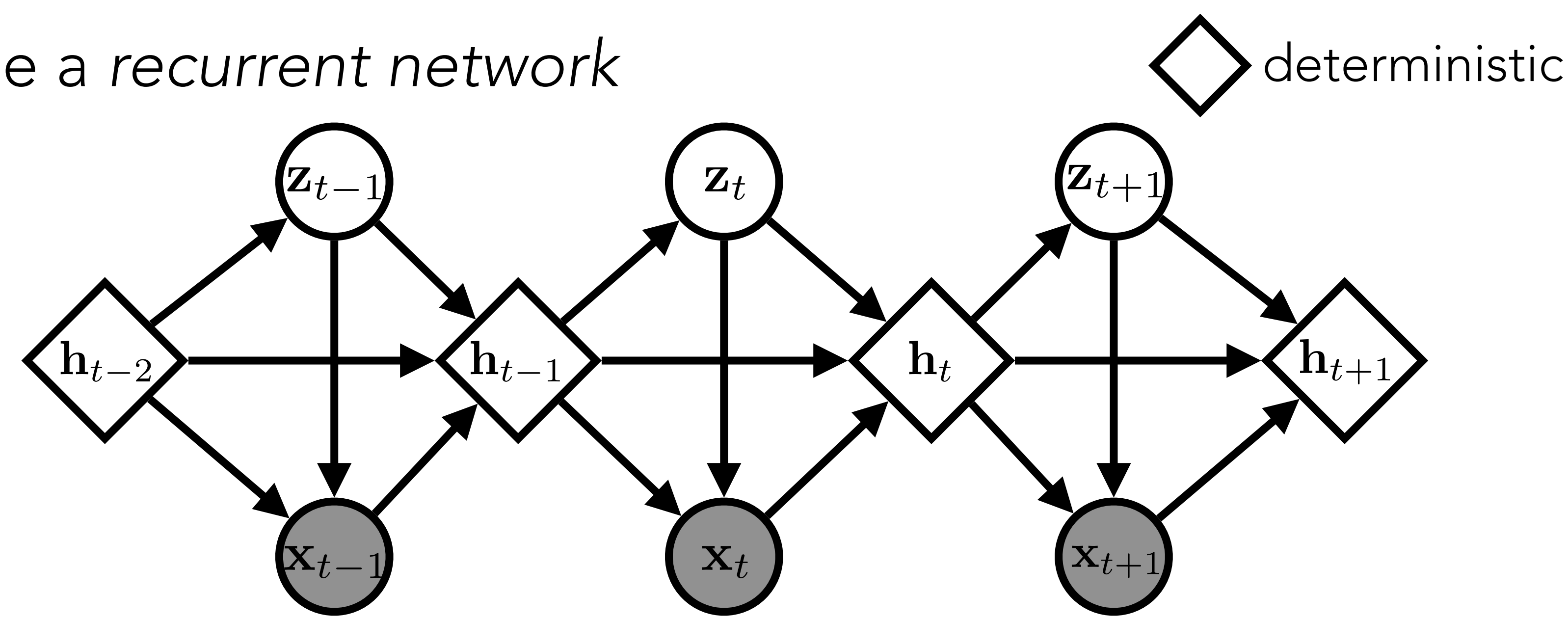
LONG-TERM DEPENDENCIES

general model form $$p_\theta(\mathbf{x}_{\leq T}, \mathbf{z}_{\leq T}) = \prod_{t=1}^{T} p_\theta(\mathbf{x}_t | \mathbf{x}_{<t}, \mathbf{z}_{\leq t}) p_\theta(\mathbf{z}_t | \mathbf{x}_{<t}, \mathbf{z}_{<t})$$

how do we model long-term dependencies?



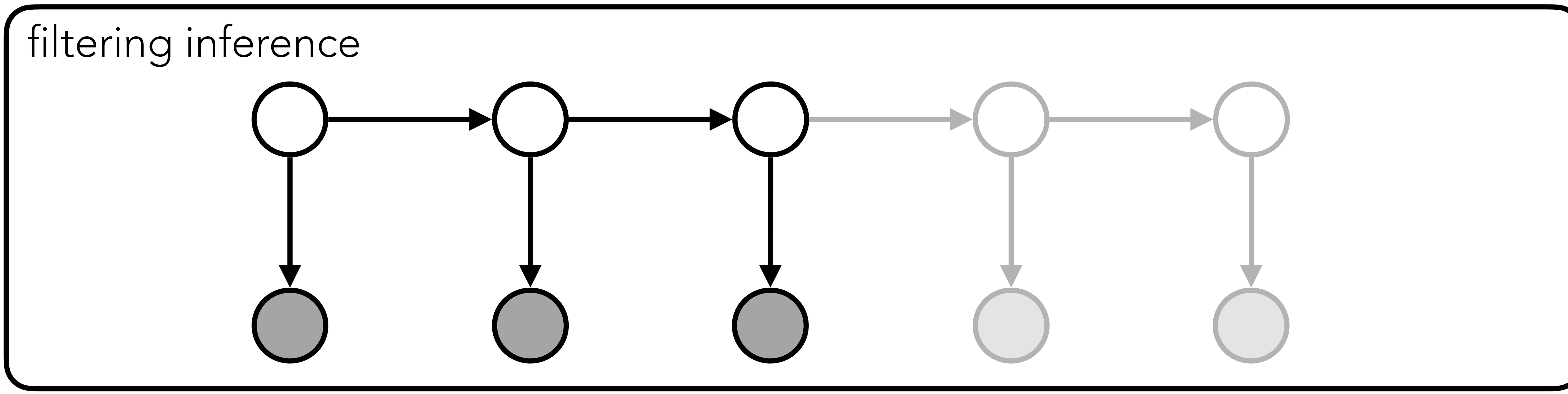
use a *recurrent network*

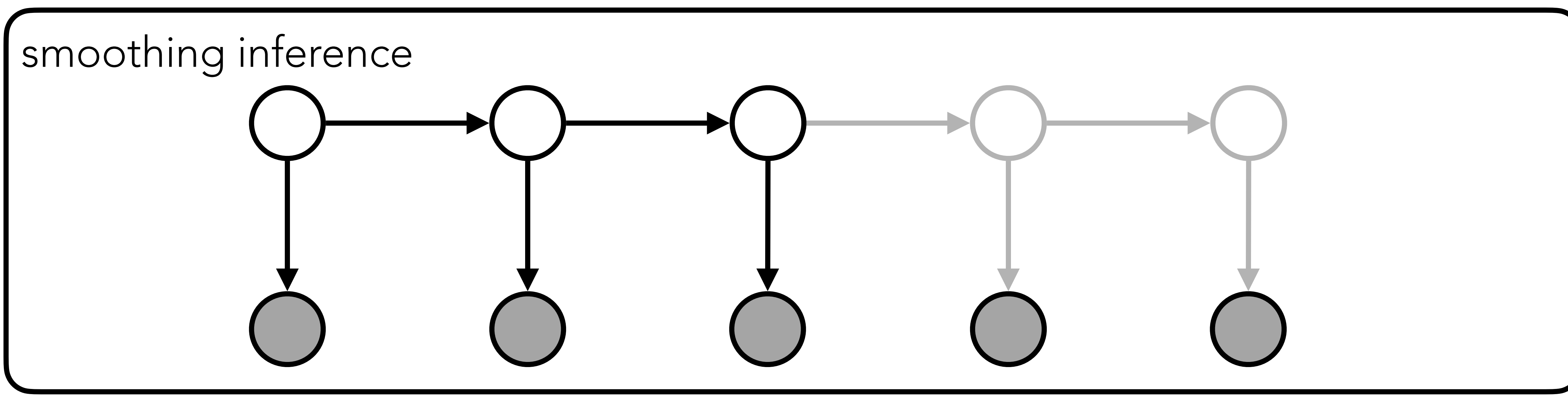


$$p_\theta(\mathbf{x}_{\leq T}, \mathbf{z}_{\leq T}) = \prod_{t=1}^{T} p_\theta(\mathbf{x}_t | \mathbf{h}_{t-1}, \mathbf{z}_t) p_\theta(\mathbf{z}_t | \mathbf{h}_{t-1})$$

INFERENCE

given a sequence of observations, $\mathbf{x}_{\leq T}$, infer $p_\theta(\mathbf{z}_{\leq T}|\mathbf{x}_{\leq T})$





introduce an approximate posterior $q(\mathbf{z}_{\leq T} | \mathbf{x}_{\leq T})$

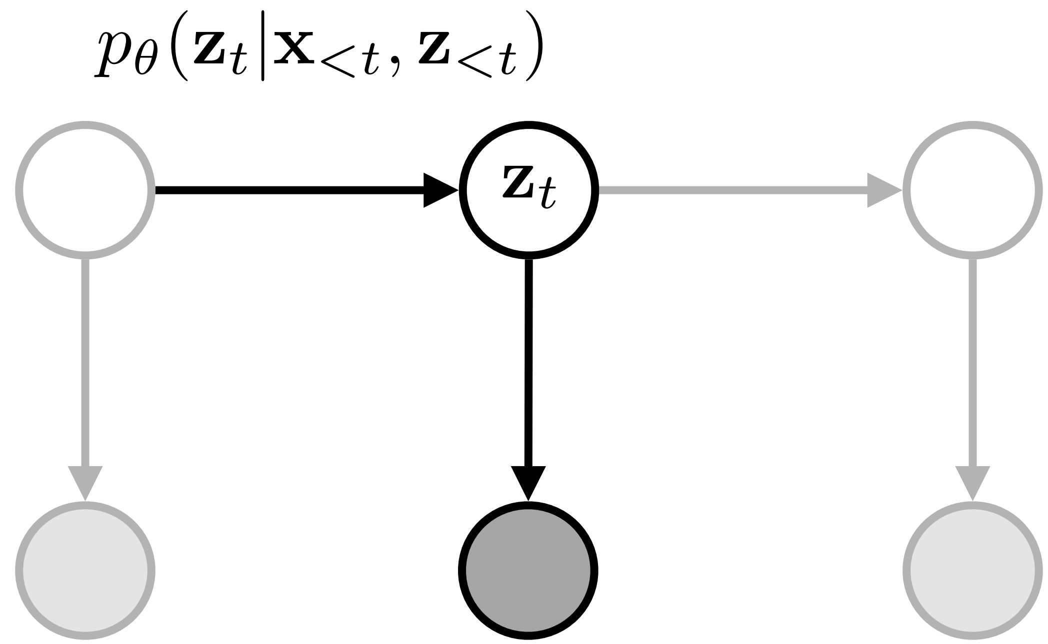
ELBO: $$\mathcal{L}(\mathbf{x}_{\leq T}, q) = \mathbb{E}_q \left[\log \frac{p_\theta(\mathbf{x}_{\leq T}, \mathbf{z}_{\leq T})}{q(\mathbf{z}_{\leq T} | \mathbf{x}_{\leq T})} \right]$$

choices about the form of $q(\mathbf{z}_{\leq T} | \mathbf{x}_{\leq T})$ determine how we evaluate $\mathcal{L}$

→ often $q(\mathbf{z}_{\leq T} | \mathbf{x}_{\leq T})$ is *structured*

STRUCTURED VARIATIONAL INFERENCE

the model contains temporal dependencies



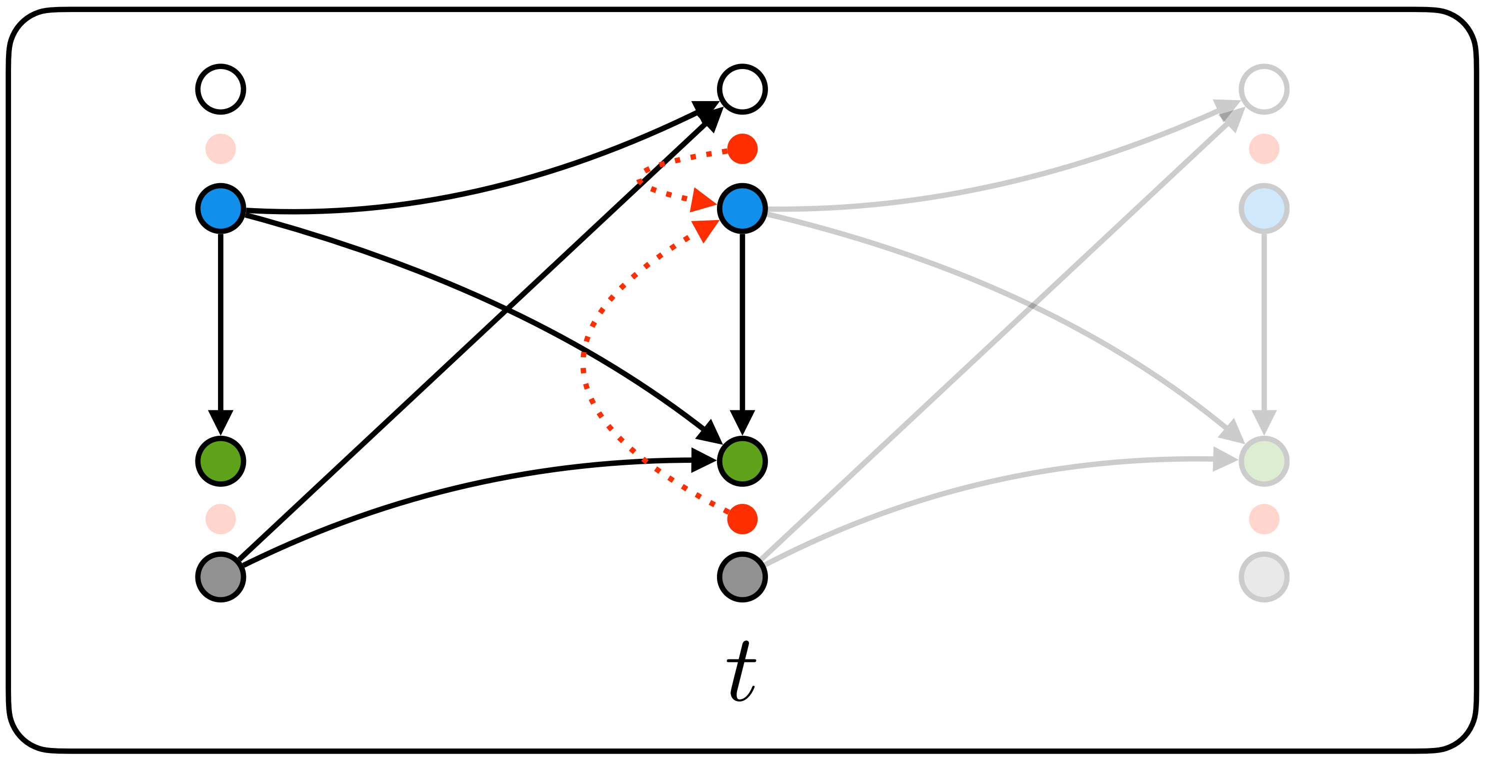
the approximate posterior should account for these dependencies

⟶ if we use $q(\mathbf{z}_t|\mathbf{x}_t)$, we cannot account for $\mathbf{x}_{<t}$ and $\mathbf{z}_{<t}$

FILTERING INFERENCE

filtering approximate posterior

$$q(\mathbf{z}_{\leq T}|\mathbf{x}_{\leq T}) = \prod_{t=1}^{T} q(\mathbf{z}_t|\mathbf{x}_{\leq t}, \mathbf{z}_{<t})$$

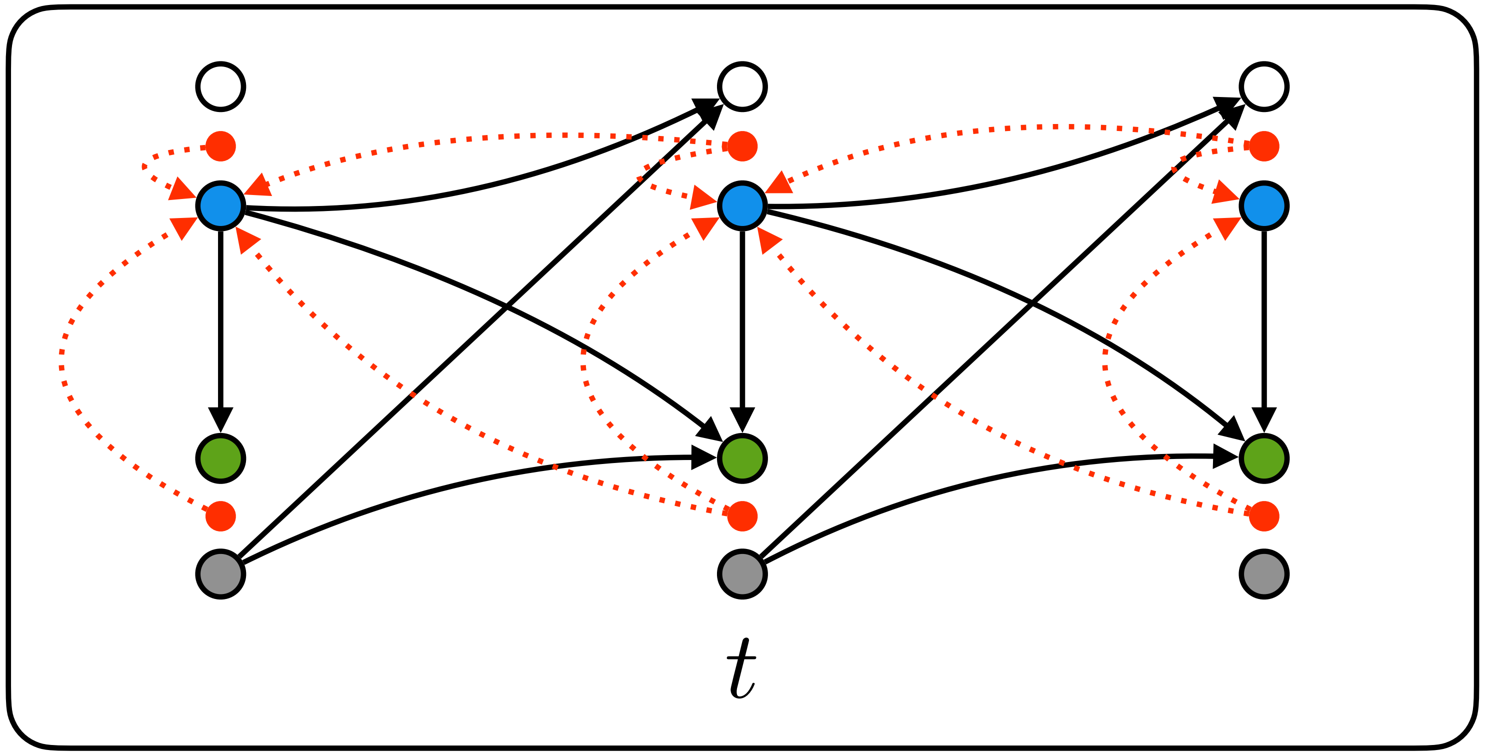


condition on observations at past and present time steps

SMOOTHING INFERENCE

smoothing approximate posterior

$$q(\mathbf{z}_{\leq T} | \mathbf{x}_{\leq T}) = \prod_{t=1}^{T} q(\mathbf{z}_t | \mathbf{x}_{\leq T}, \mathbf{z}_{<t})$$



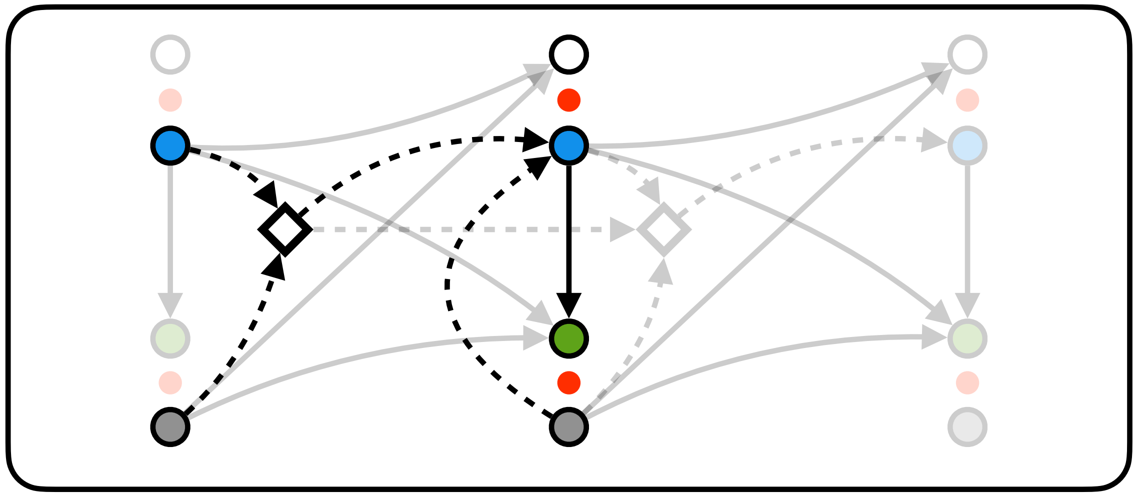
condition on observations at all time steps

AMORTIZED VARIATIONAL INFERENCE

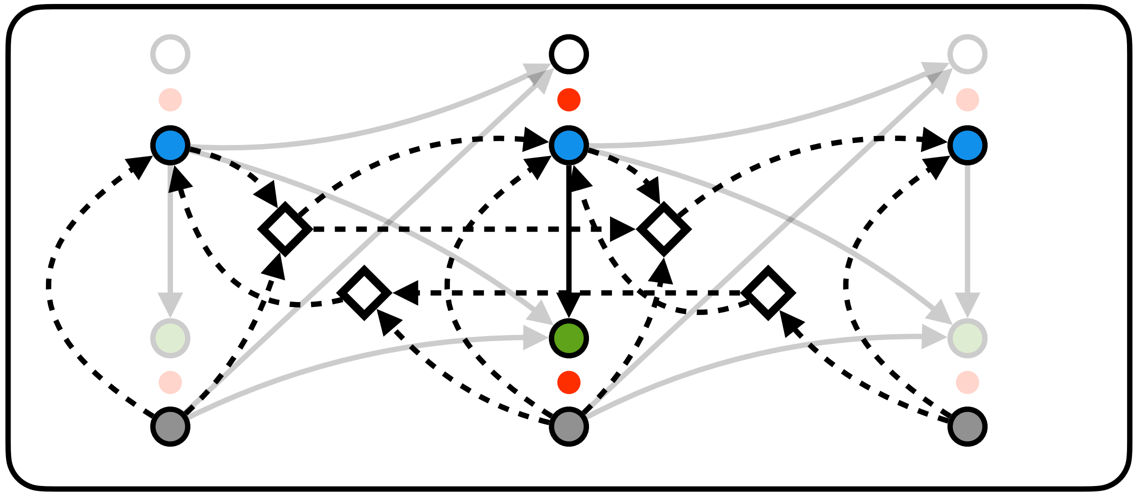
how do we amortize inference in sequential models?

typical approach:

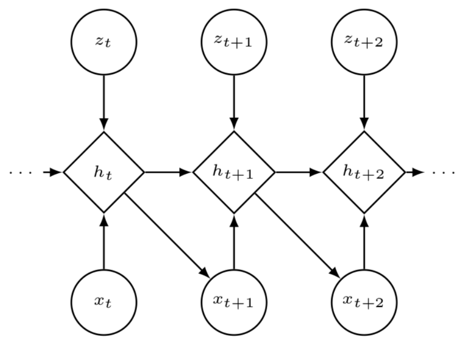
filtering: use a recurrent network



smoothing: use a bi-directional recurrent network

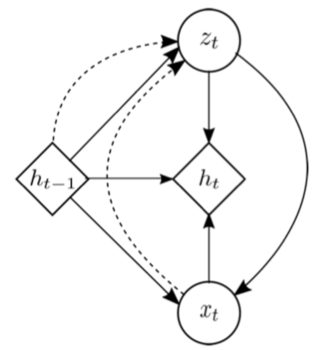


RECENT MODELS



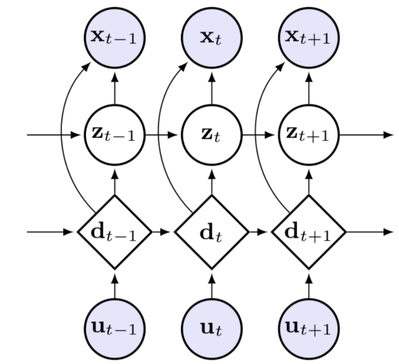
STORN

Bayer & Osendorfer, 2014



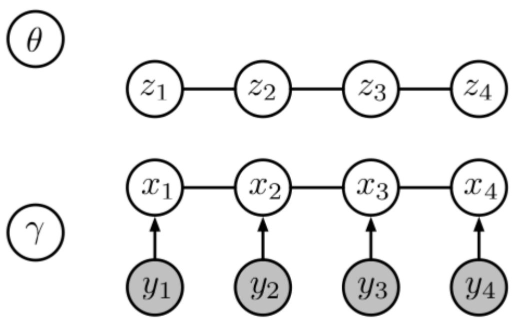
VRNN

Chung *et al.*, 2015



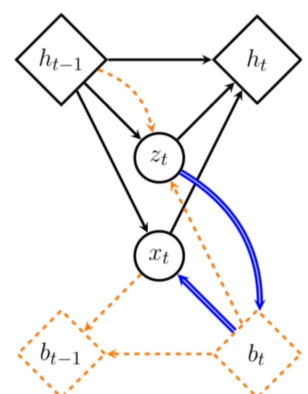
SRNN

Fraccaro *et al.*, 2016



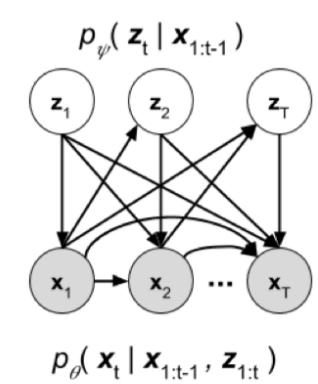
Structured VAE

Johnson *et al.*, 2016



Z-Forcing

Goyal *et al.*, 2017



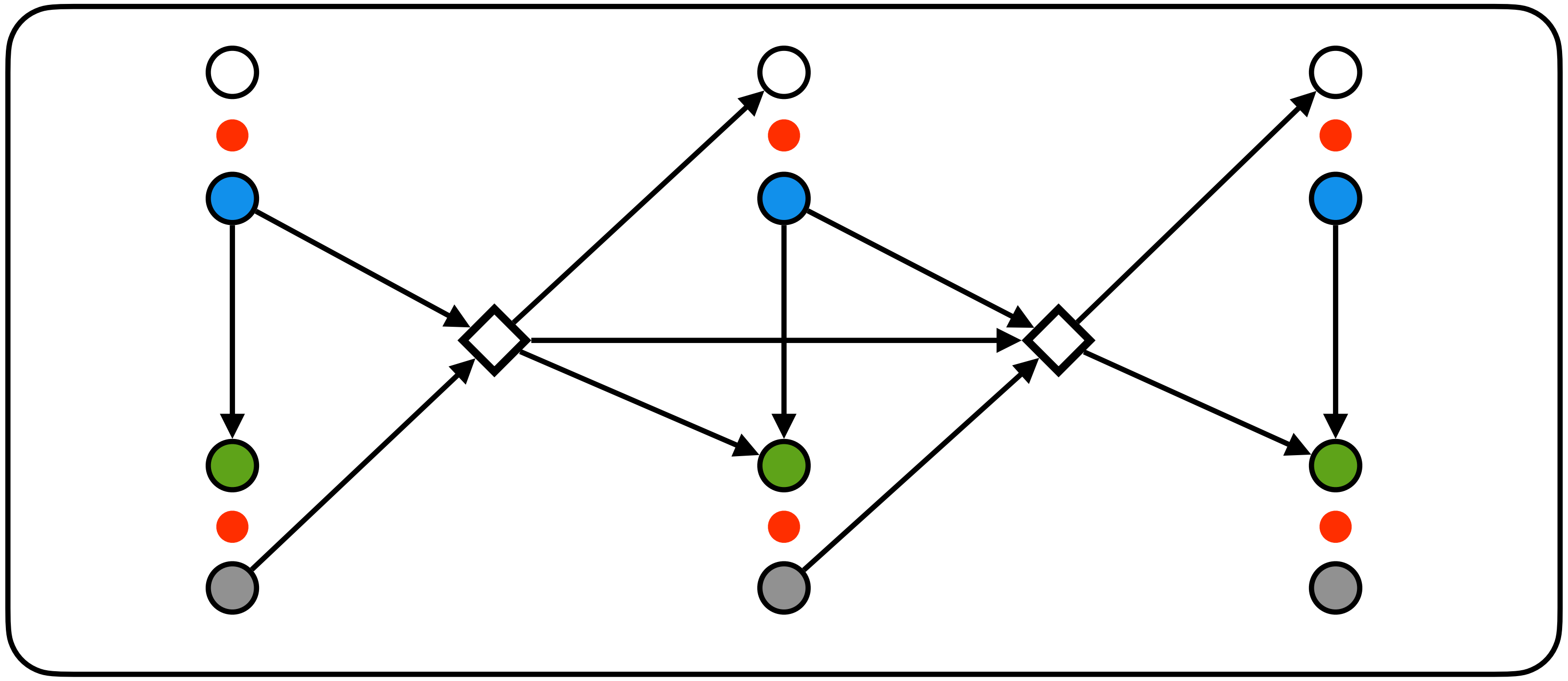
SVG

Denton & Fergus, 2018

● ● ●

VRNN

generative model



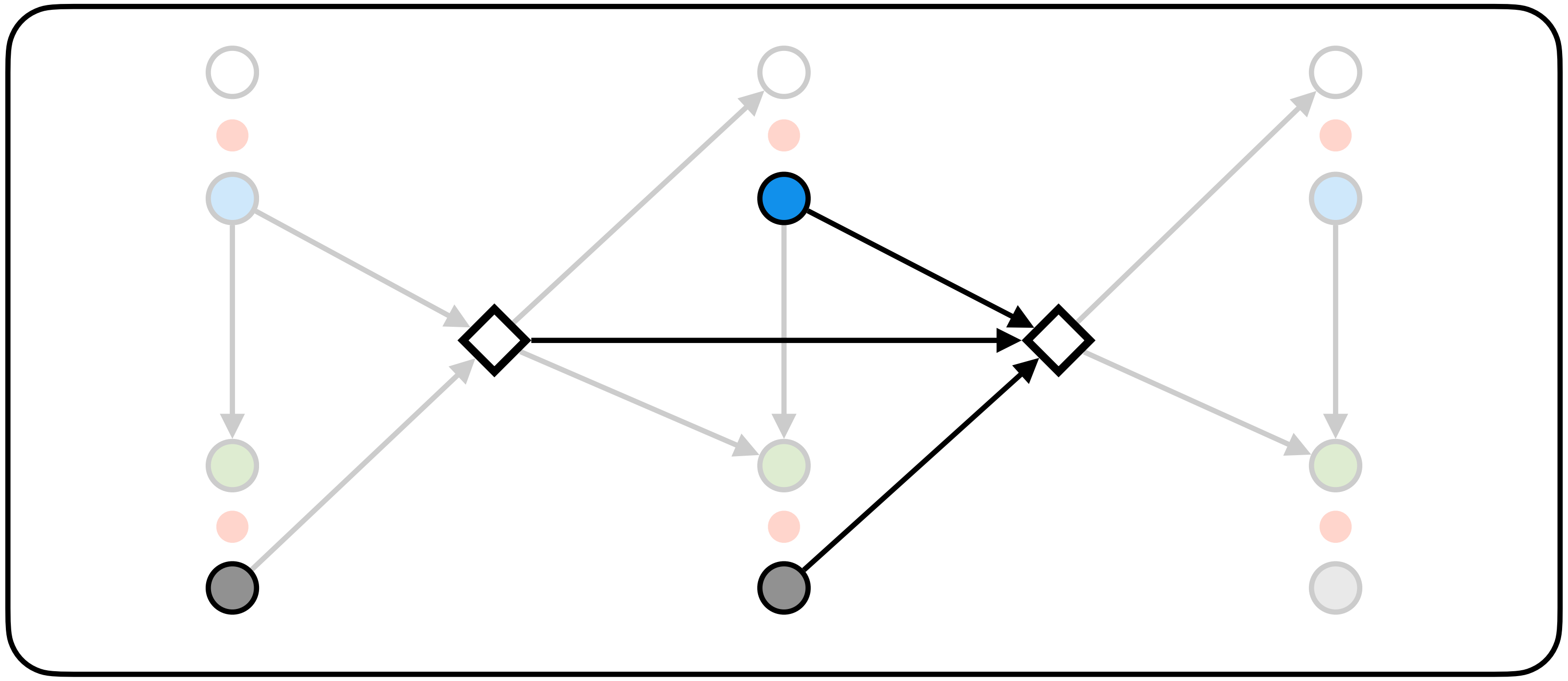
general model form $$p_\theta(\mathbf{x}_{\leq T}, \mathbf{z}_{\leq T}) = \prod_{t=1}^{T} p_\theta(\mathbf{x}_t | \mathbf{x}_{<t}, \mathbf{z}_{\leq t}) p_\theta(\mathbf{z}_t | \mathbf{x}_{<t}, \mathbf{z}_{<t})$$

VRNN model form $$= \prod_{t=1}^{T} p_\theta(\mathbf{x}_t | \mathbf{z}_t, \mathbf{h}_{t-1}) p_\theta(\mathbf{z}_t | \mathbf{h}_{t-1})$$

Chung *et al.*, 2015

VRNN

generative model



recurrence:

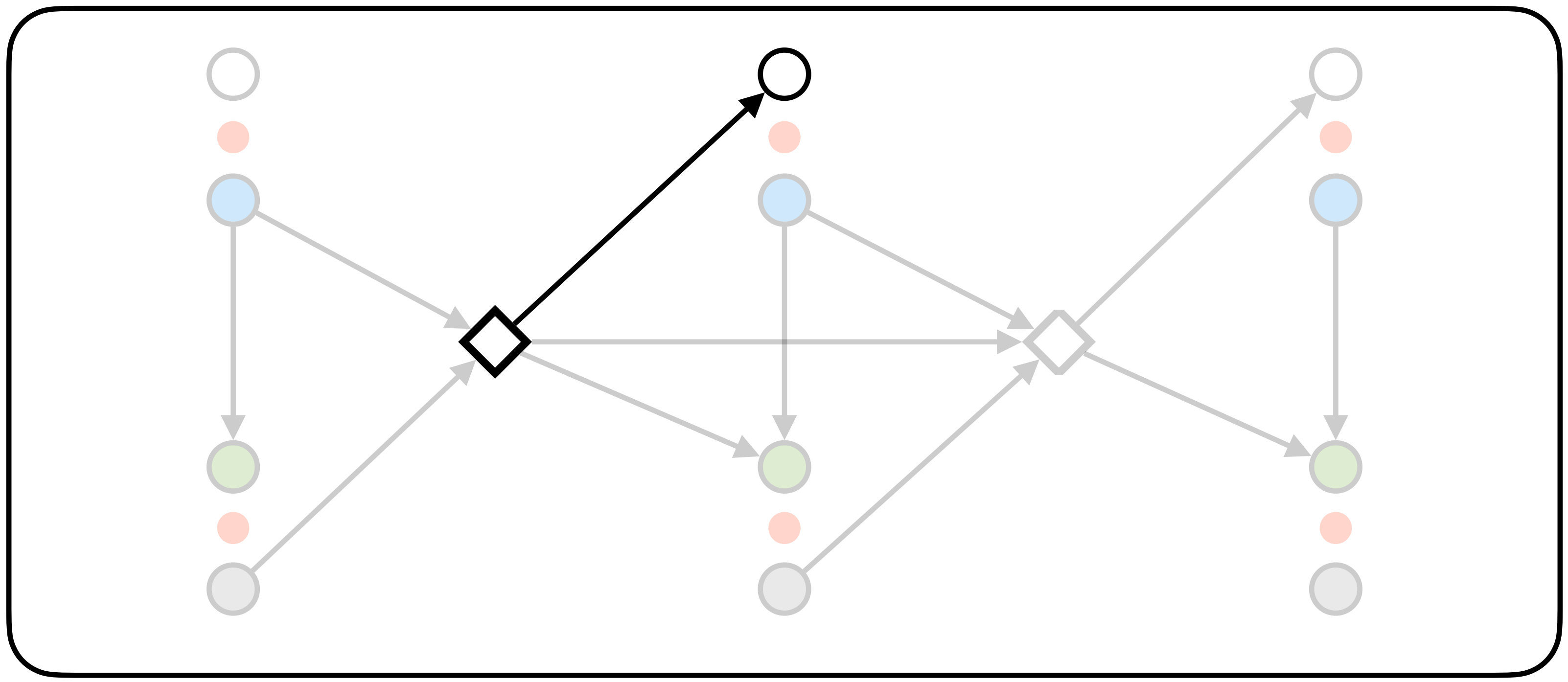
$$\mathbf{h}_t = \mathrm{LSTM}([\varphi_{\mathbf{x}}(\mathbf{x}_t), \varphi_{\mathbf{z}}(\mathbf{z}_t)], \mathbf{h}_{t-1})$$

φ are fully-connected networks

Chung *et al.*, 2015

VRNN

generative model



prior:

$$p_\theta(\mathbf{z}_t | \mathbf{h}_{t-1}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z},t}, \mathrm{diag}(\boldsymbol{\sigma}^2_{\mathbf{z},t}))$$

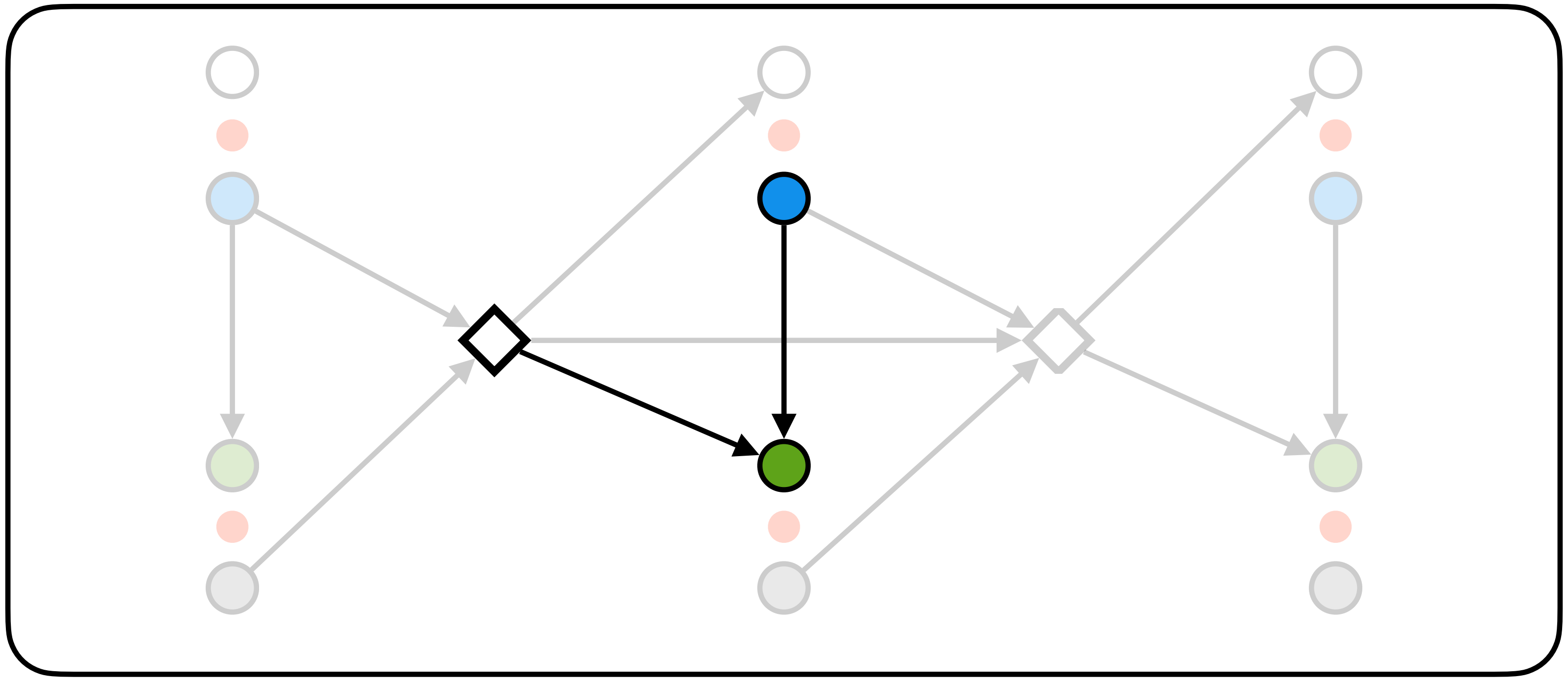
where $[\boldsymbol{\mu}_{\mathbf{z},t}, \boldsymbol{\sigma}_{\mathbf{z},t})] = \varphi_{\text{prior}}(\mathbf{h}_{t-1})$

φ are fully-connected networks

Chung *et al.*, 2015

VRNN

generative model



conditional likelihood:

$$p_\theta(\mathbf{x}_t|\mathbf{z}_t, \mathbf{h}_{t-1}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x},t}, \mathrm{diag}(\boldsymbol{\sigma}^2_{\mathbf{x},t}))$$

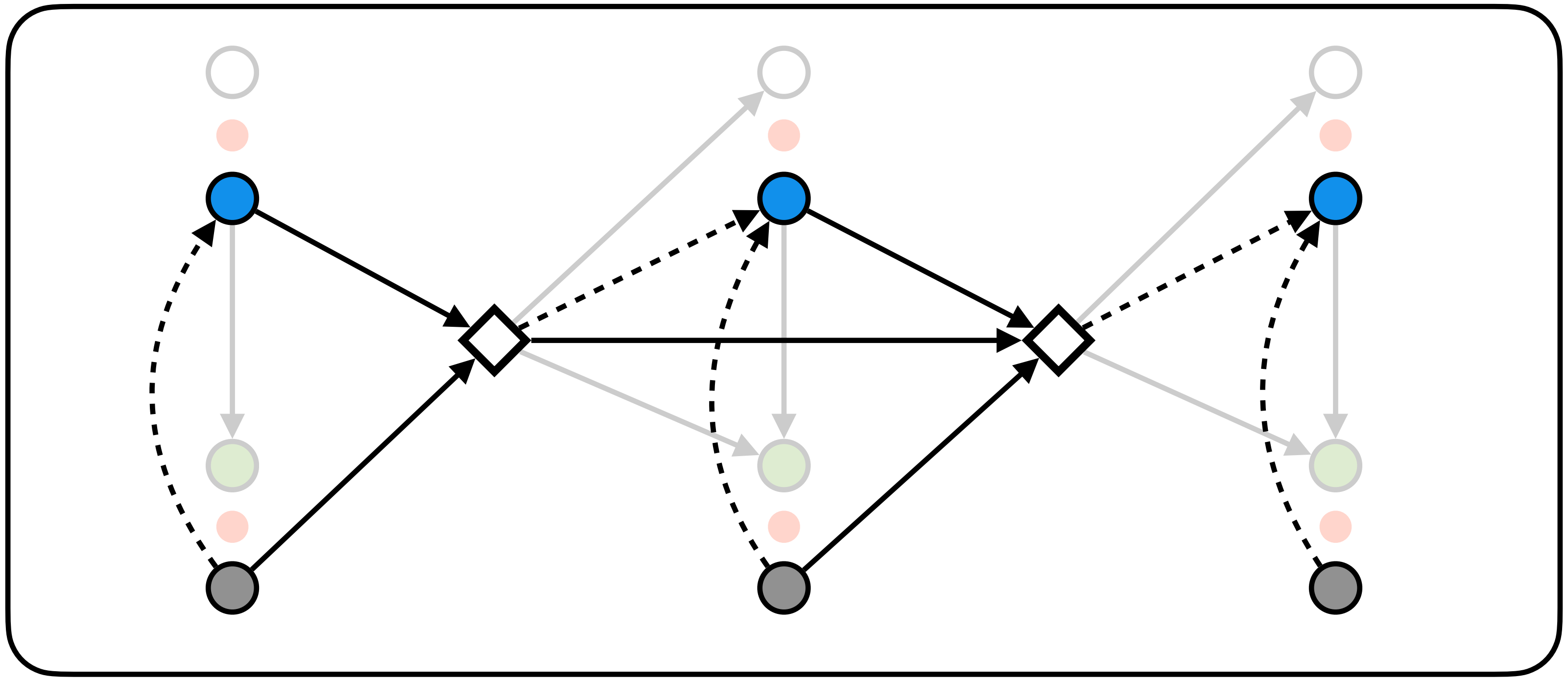
where $[\boldsymbol{\mu}_{\mathbf{x},t}, \boldsymbol{\sigma}_{\mathbf{x},t})] = \varphi_{\mathrm{dec}}(\varphi_{\mathbf{z}}(\mathbf{z}_t), \mathbf{h}_{t-1})$

φ are fully-connected networks

Chung *et al.*, 2015

VRNN

inference model



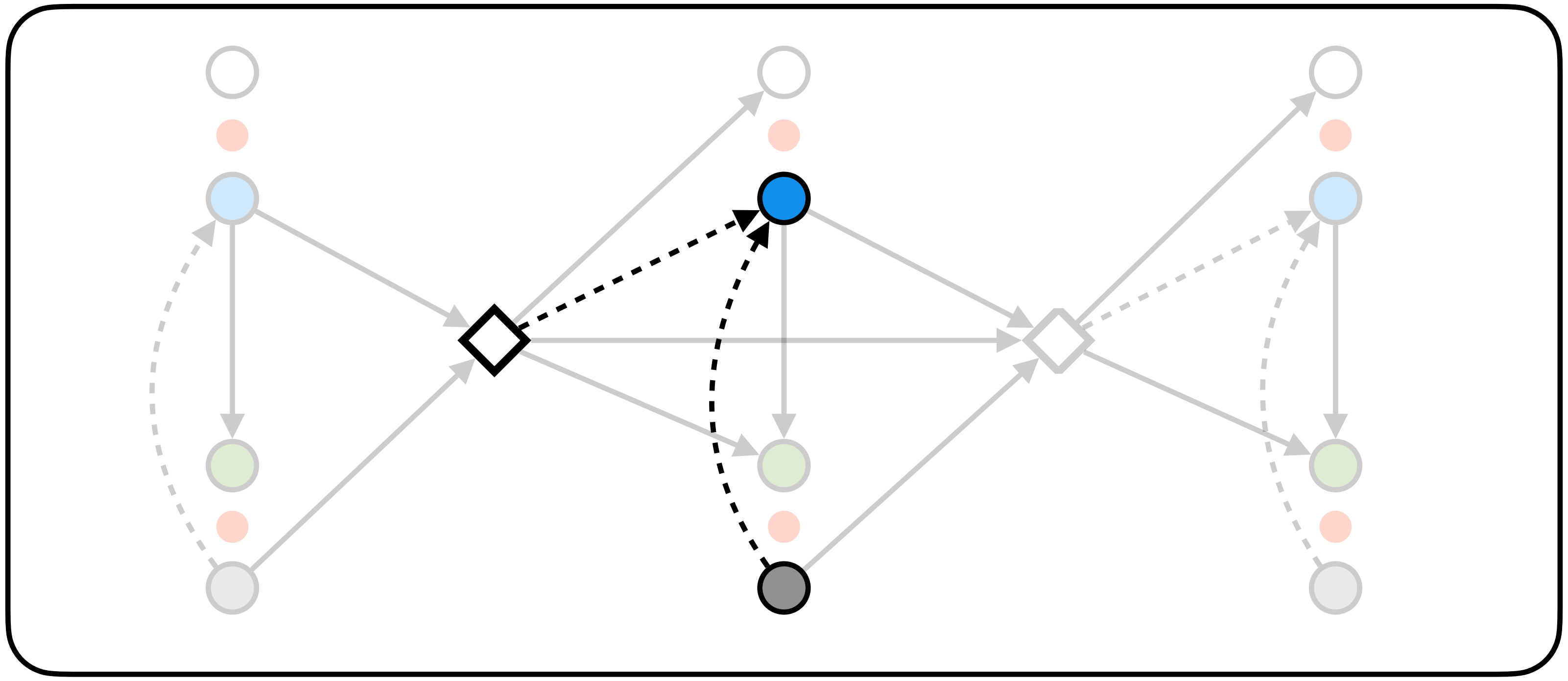
filtering inference $$q(\mathbf{z}_{\leq T}|\mathbf{x}_{\leq T}) = \prod_{t=1}^{T} q(\mathbf{z}_t|\mathbf{x}_{\leq t}, \mathbf{z}_{<t})$$

VRNN inference model form $$= \prod_{t=1}^{T} q(\mathbf{z}_t|\mathbf{x}_t, \mathbf{h}_{t-1})$$

Chung *et al.*, 2015

VRNN

inference model



approximate posterior:

$$q(\mathbf{z}_t|\mathbf{x}_t, \mathbf{h}_{t-1}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z},t}, \mathrm{diag}(\boldsymbol{\sigma}^2_{\mathbf{z},t}))$$

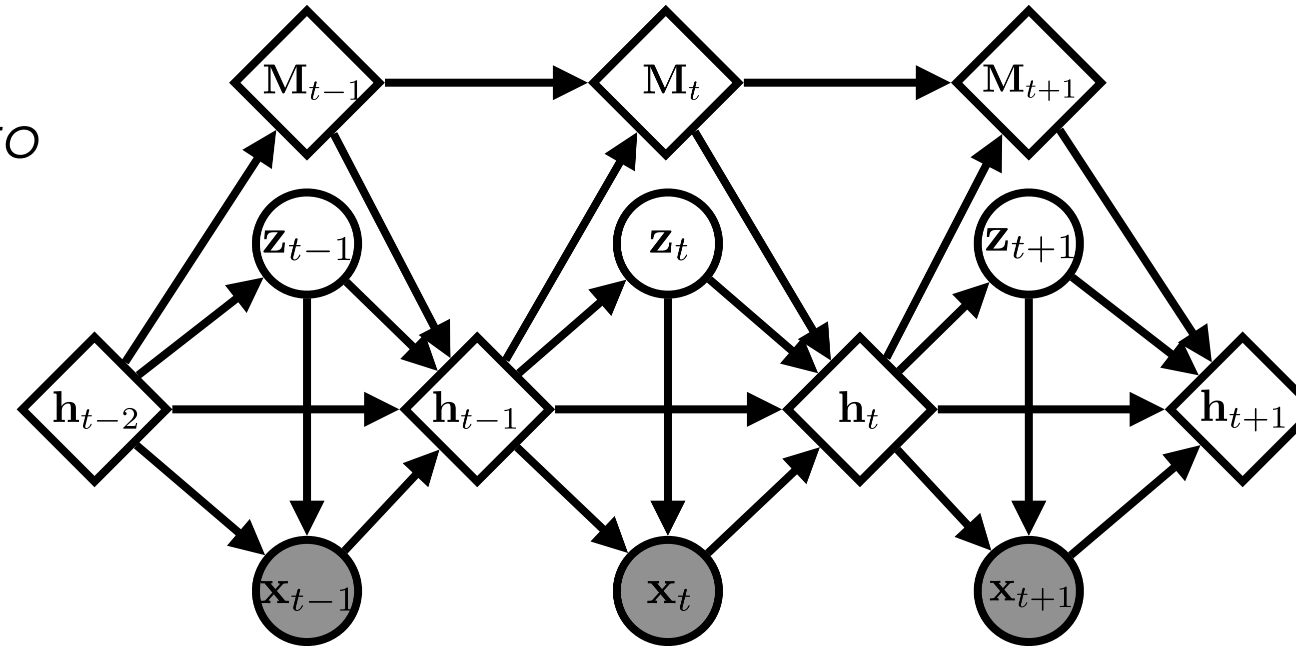
where $[\boldsymbol{\mu}_{\mathbf{z},t}, \boldsymbol{\sigma}_{\mathbf{z},t})] = \varphi_{\text{enc}}(\varphi_{\mathbf{x}}(\mathbf{x}_t), \mathbf{h}_{t-1})$

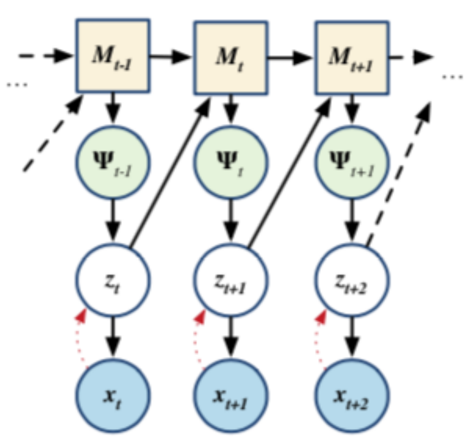
φ are fully-connected networks

Chung *et al.*, 2015

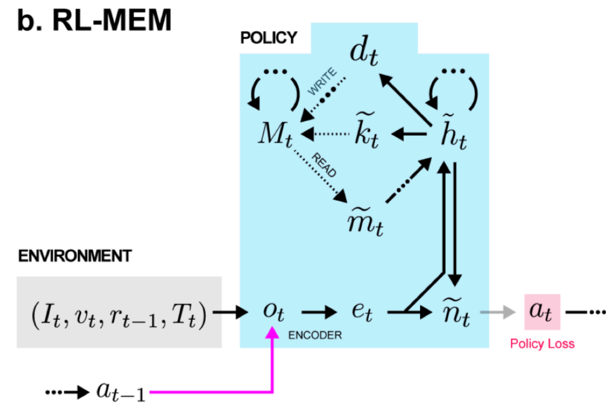
MEMORY

use a specialized memory module to model longer-term dependencies

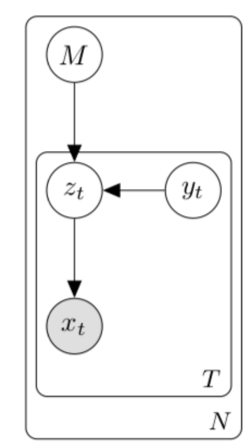




GTMM
Gemici *et al.*, 2017



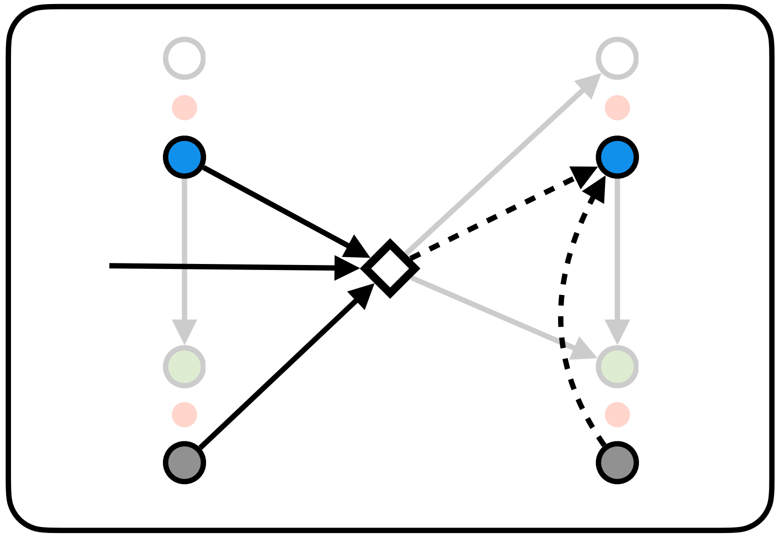
MERLIN
Wayne *et al.*, 2018



Kanerva Machine
Wu *et al.*, 2018

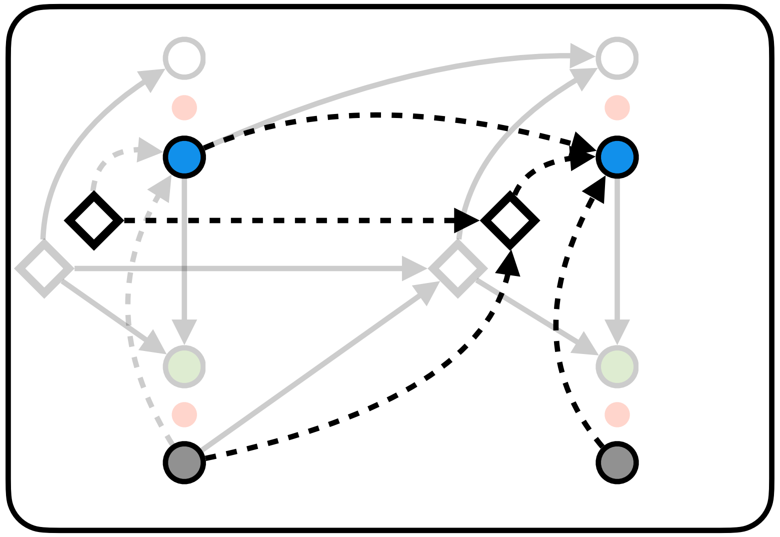
● ● ●

FILTERING INFERENCE MODELS



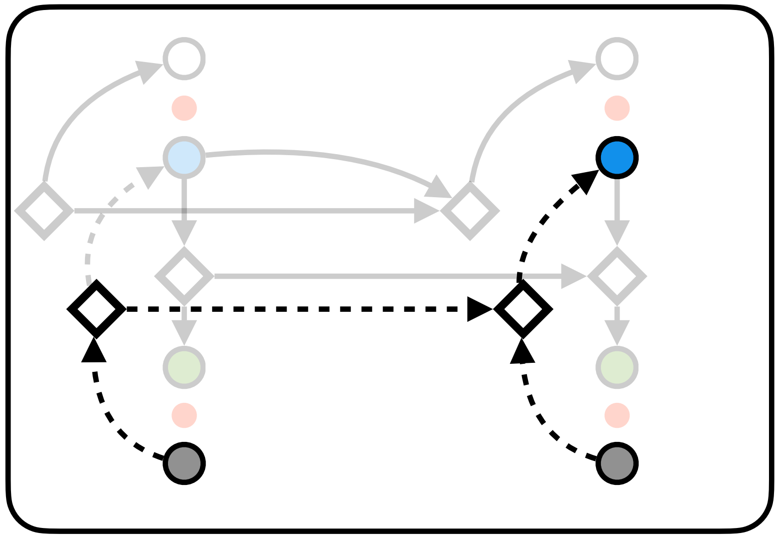
VRNN

Chung et al., 2015



SRNN

Fraccaro et al., 2016

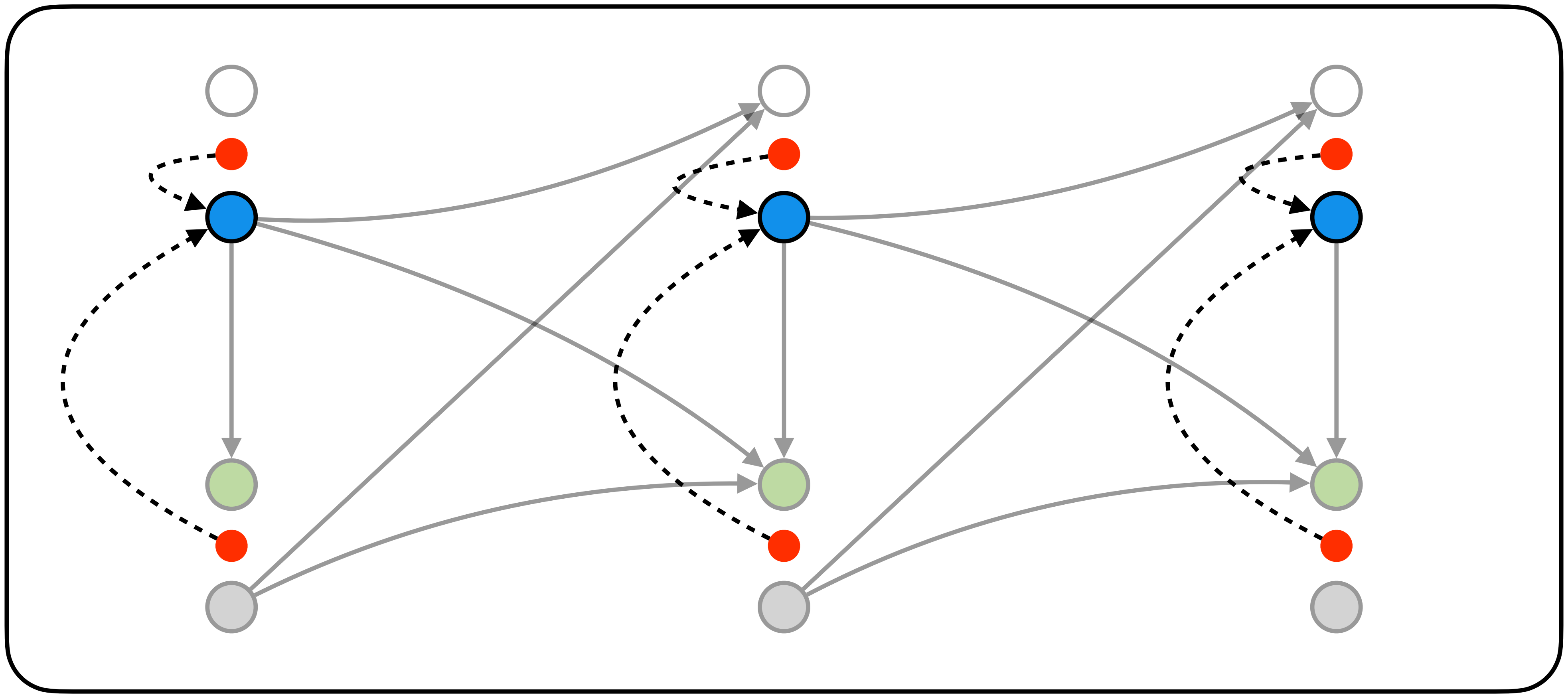


SVG

Denton & Fergus, 2018

custom-designed inference models

AMORTIZED VARIATIONAL FILTERING

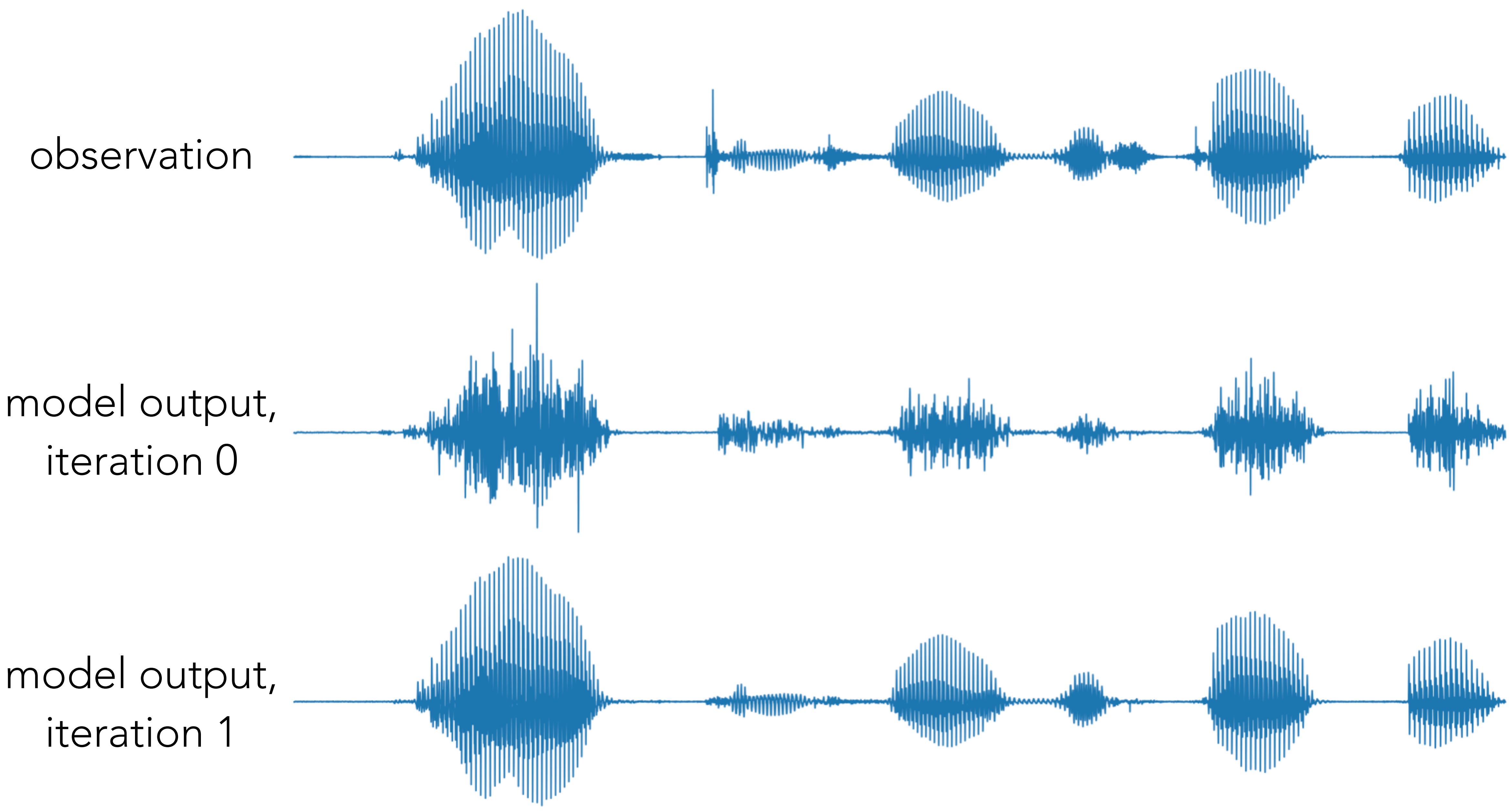


perform iterative amortized inference at each time step

Marino *et al.*, 2018b

INFERENCE IMPROVEMENT

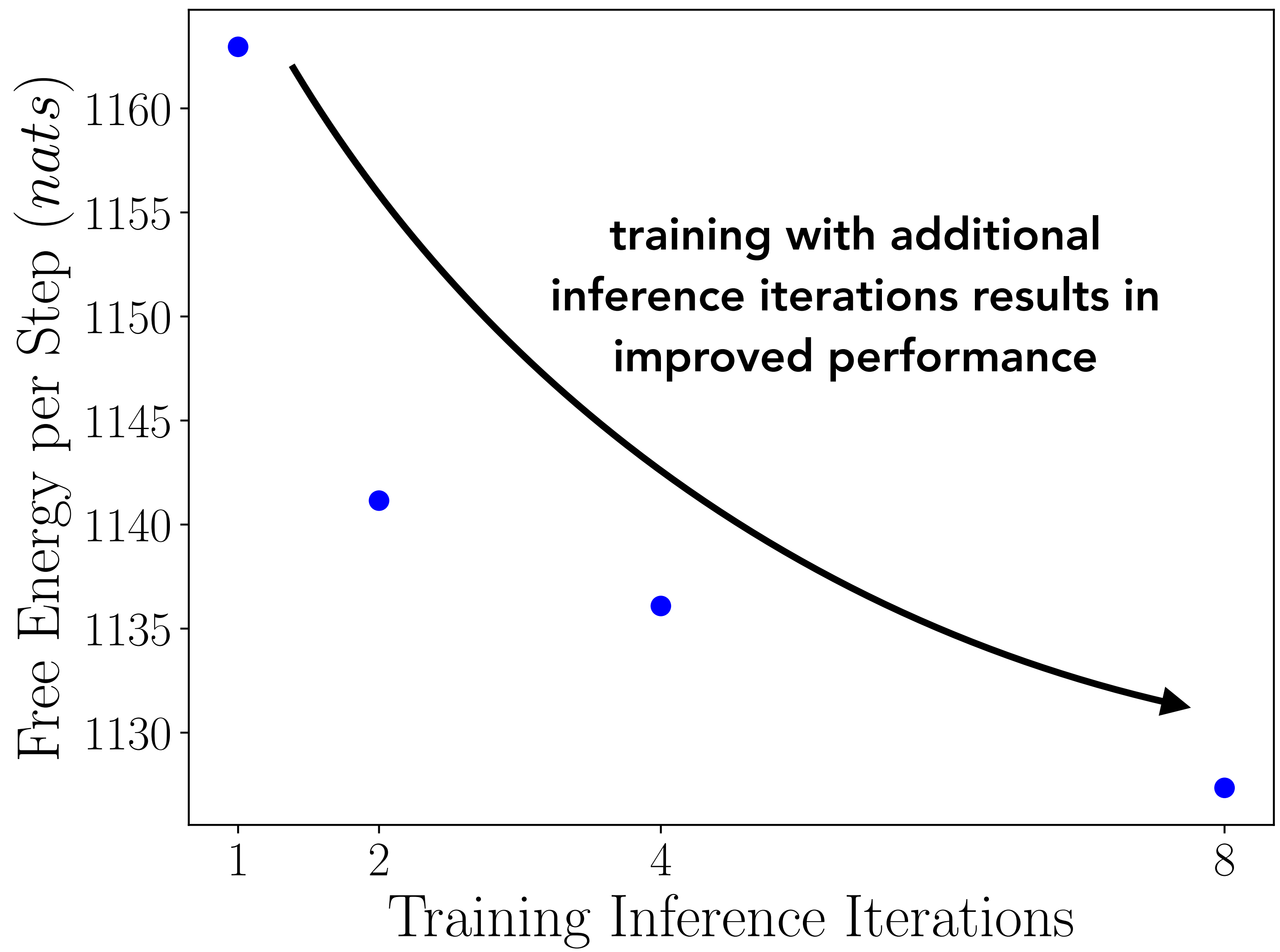
TIMIT audio waveforms



Marino *et al.*, 2018b

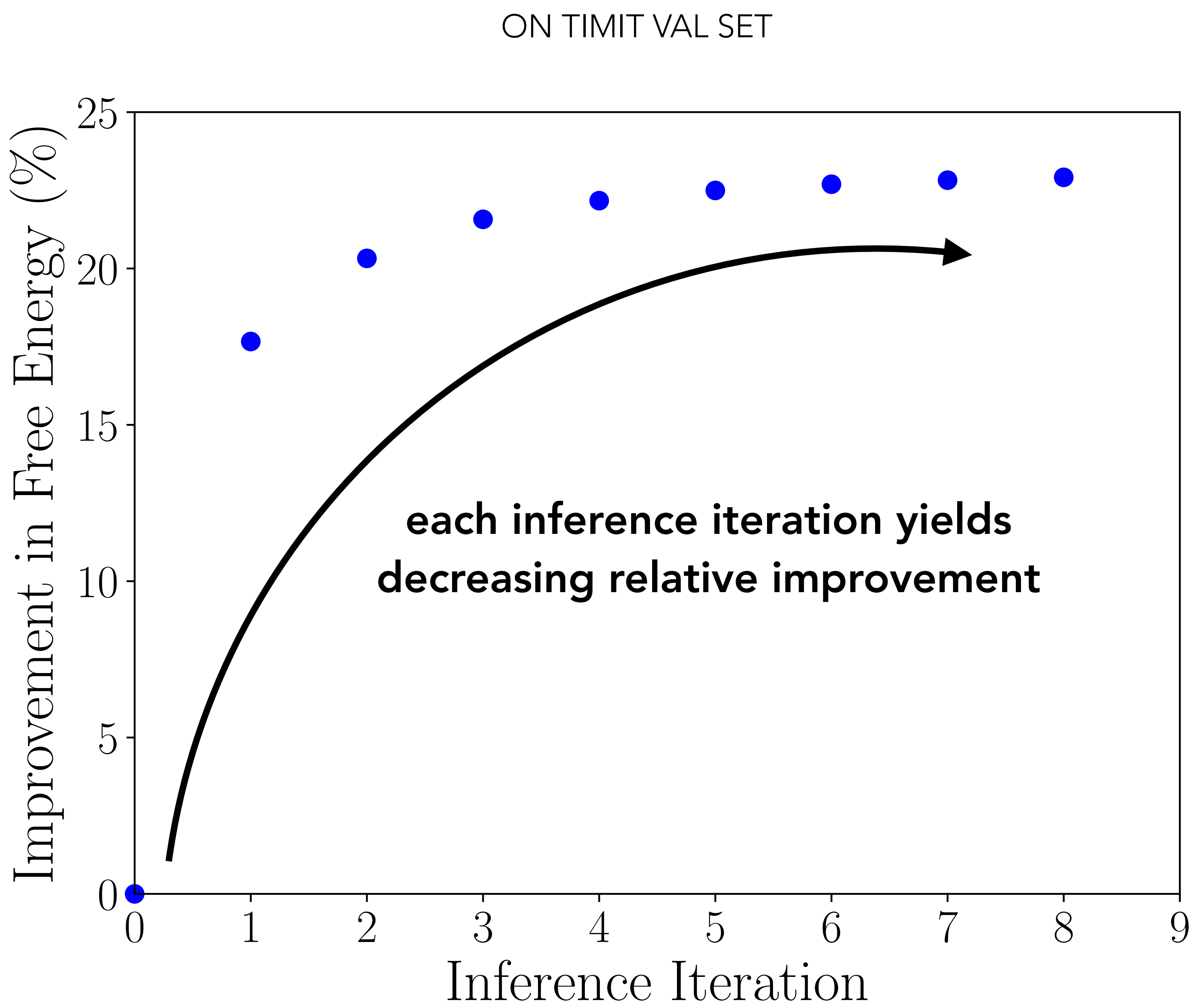
INFERENCE ITERATIONS

ON TIMIT VAL SET

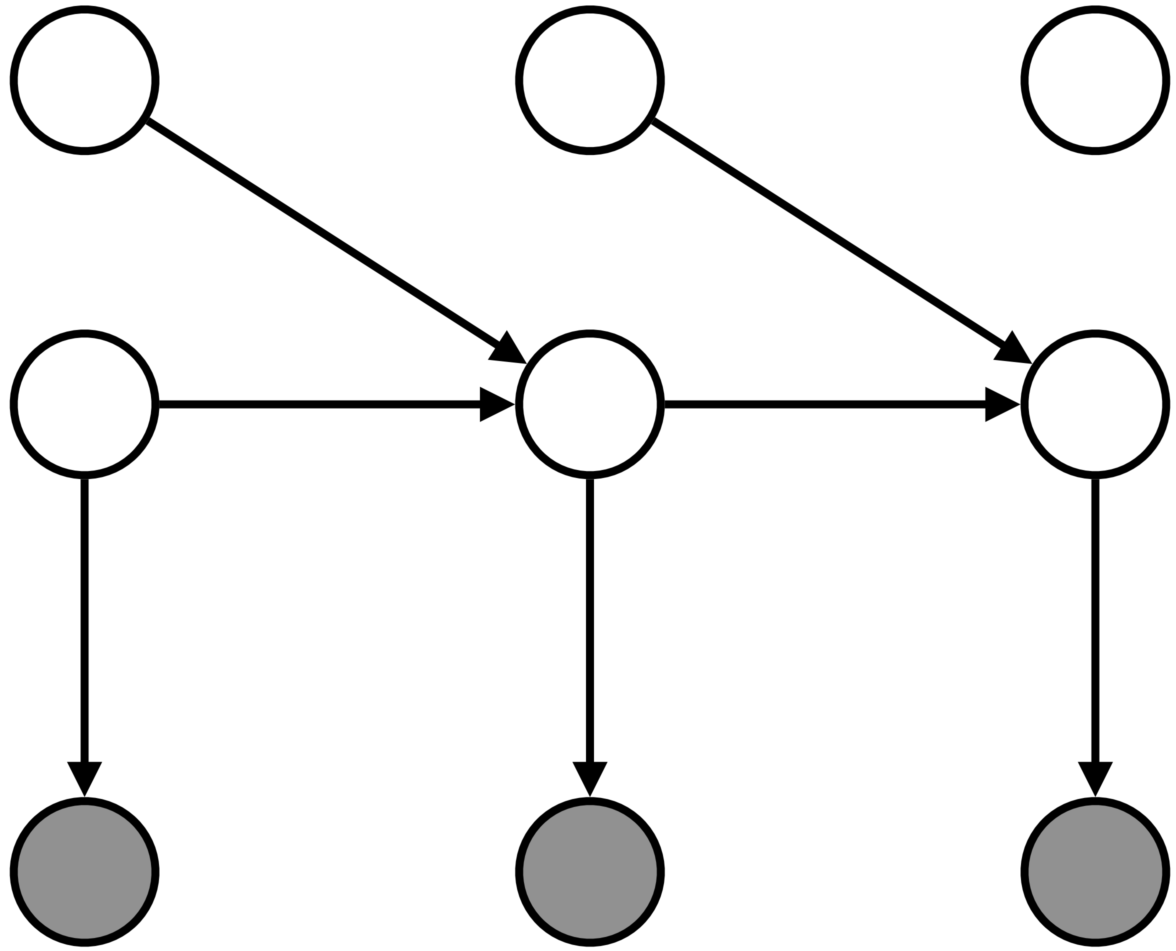


Marino *et al.*, 2018b

INFERENCE ITERATIONS



Marino *et al.*, 2018b



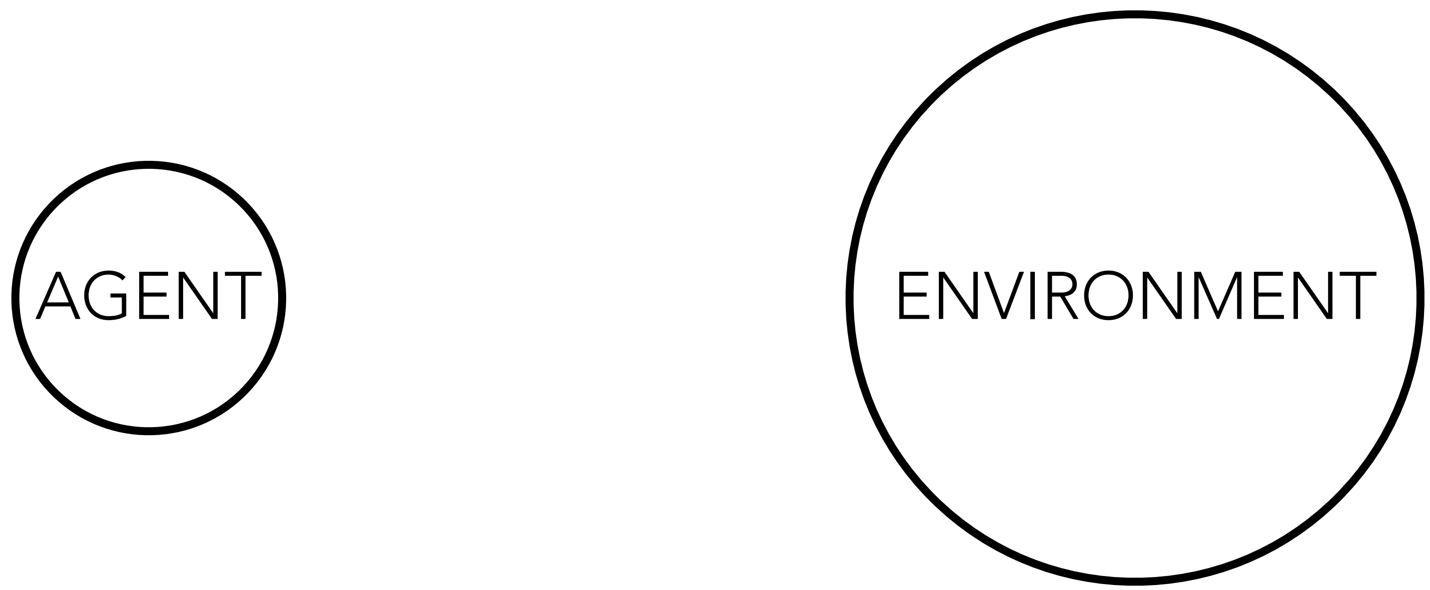
MODEL-BASED REINFORCEMENT LEARNING

REINFORCEMENT LEARNING

sequential decision making by maximizing expected future reward

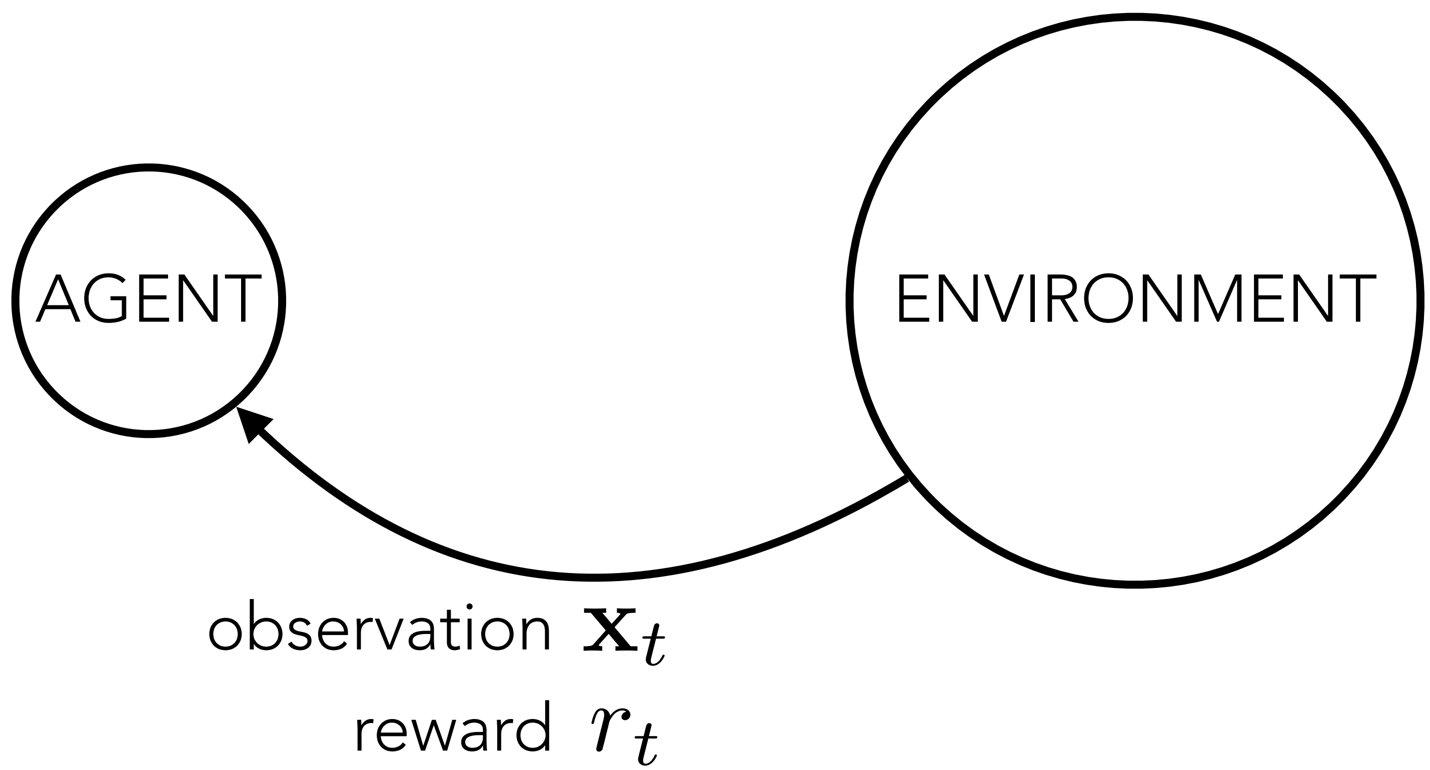
REINFORCEMENT LEARNING

partially-observable Markov decision process (POMDP)



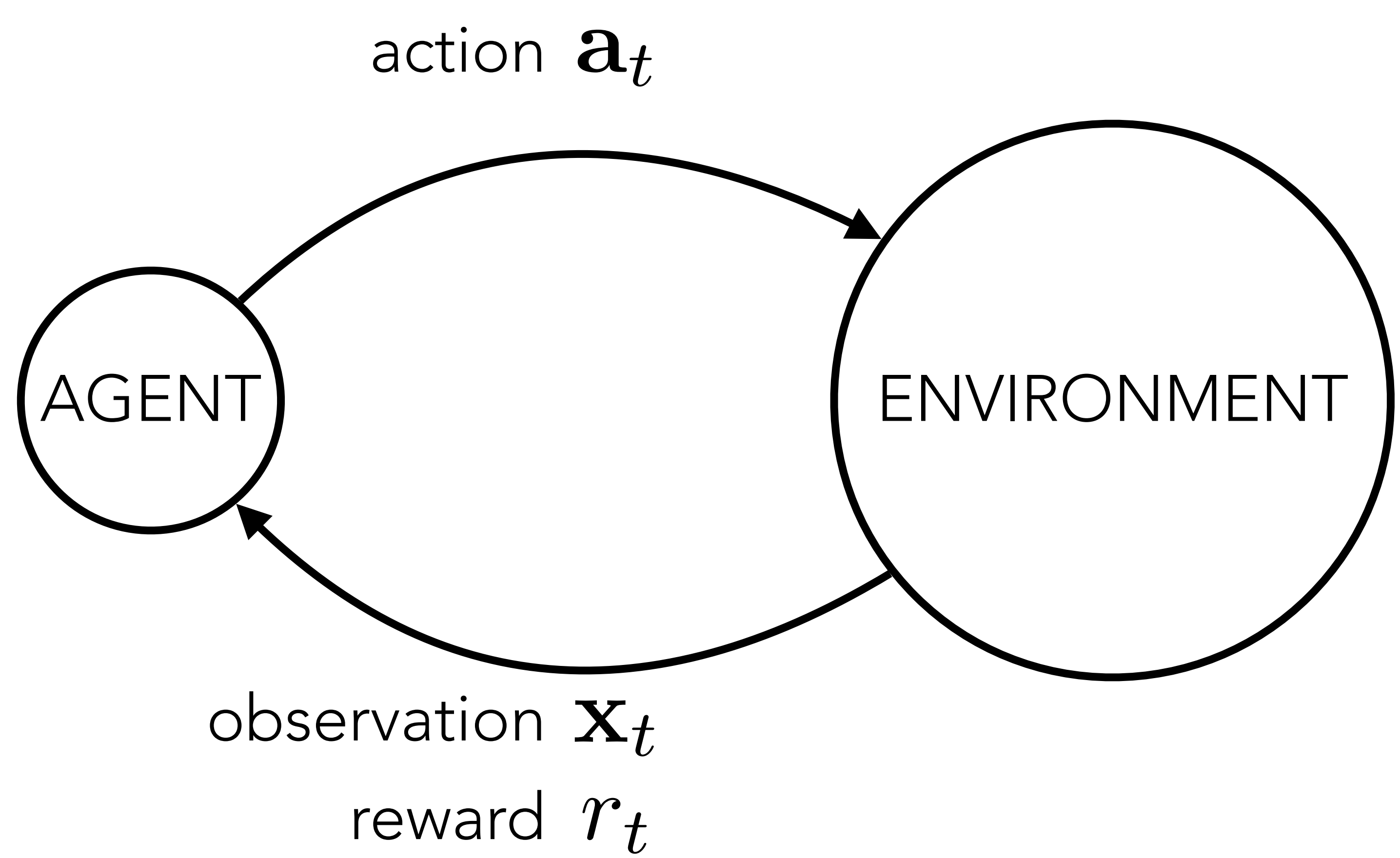
REINFORCEMENT LEARNING

partially-observable Markov decision process (POMDP)



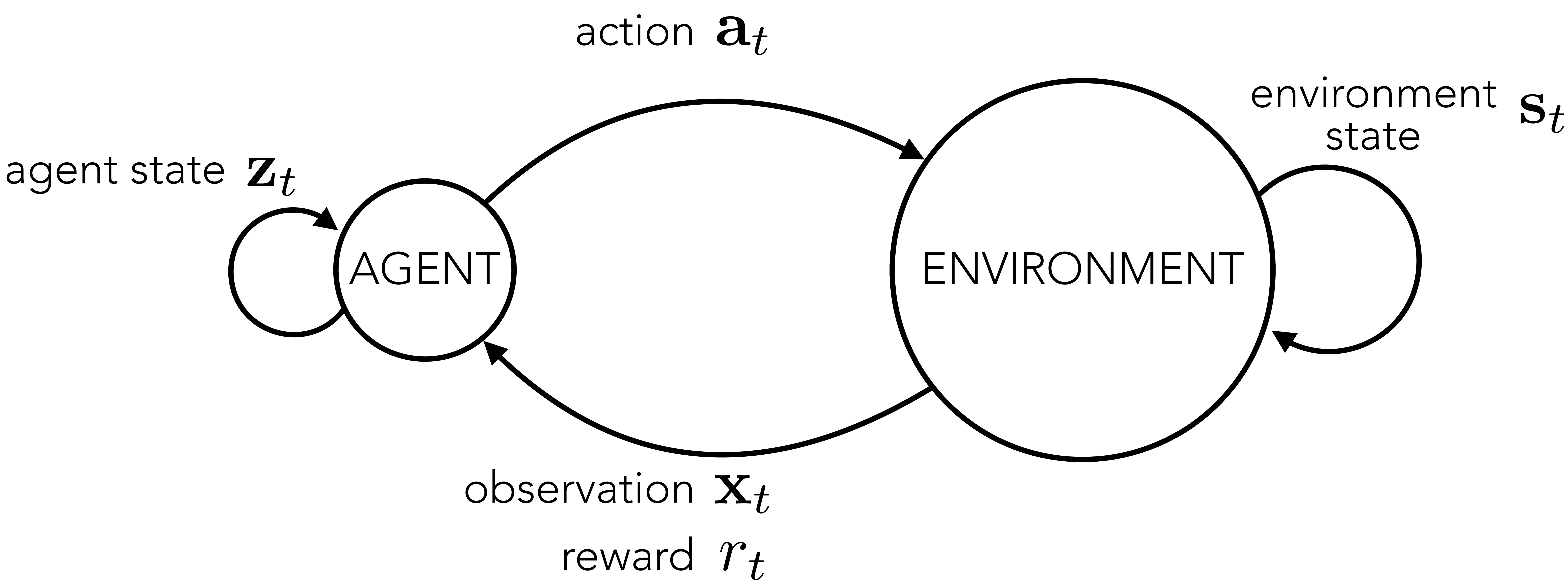
REINFORCEMENT LEARNING

partially-observable Markov decision process (POMDP)



REINFORCEMENT LEARNING

partially-observable Markov decision process (POMDP)



REINFORCEMENT LEARNING

a *policy* is a probability distribution over actions: $\mathbf{a} \sim \pi(\mathbf{a}|\cdot)$

RL objective:

maximize the expected sum of rewards (*return*)

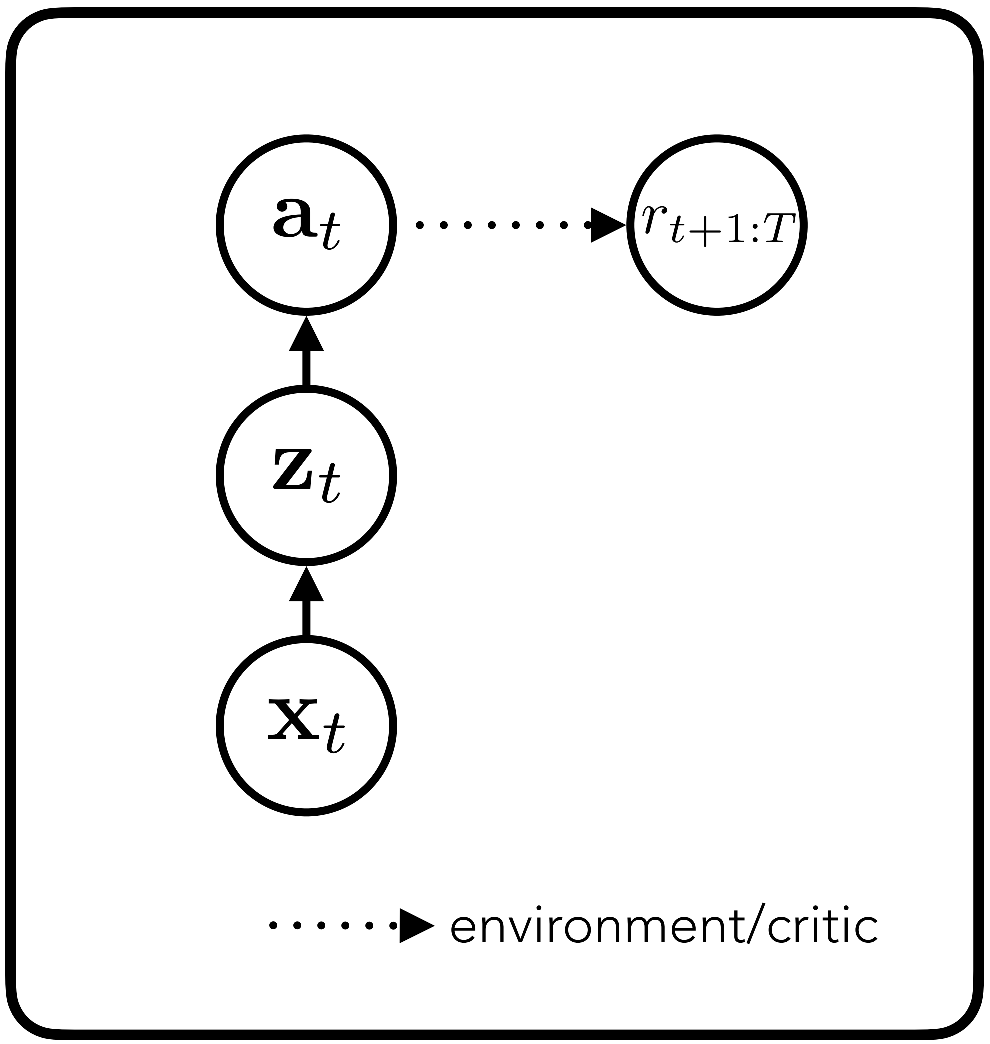
$$\pi(\mathbf{a}|\cdot) \leftarrow \arg\max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=1}^{T} r_t \right]$$

REINFORCEMENT LEARNING

approaches to policy optimization

model-free

direct mapping to actions

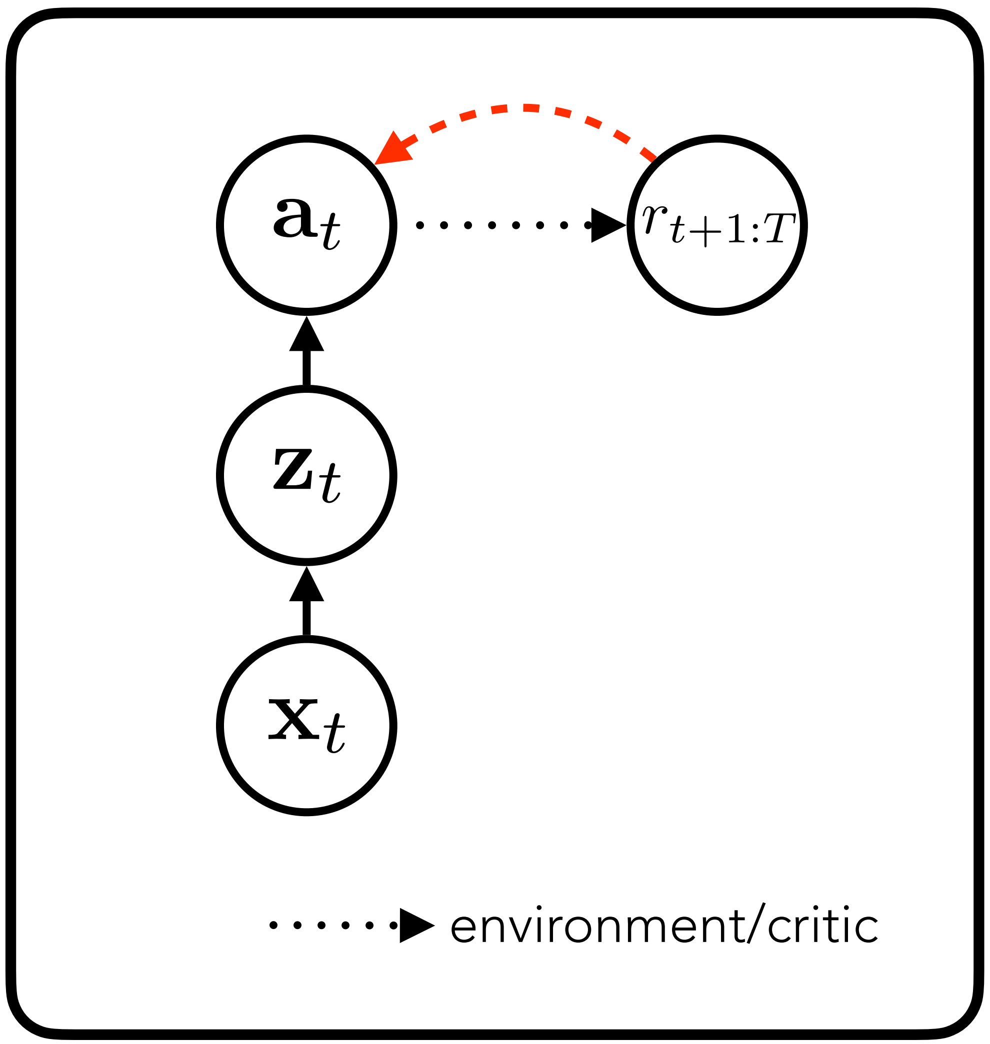


REINFORCEMENT LEARNING

approaches to policy optimization

model-free

direct mapping to actions

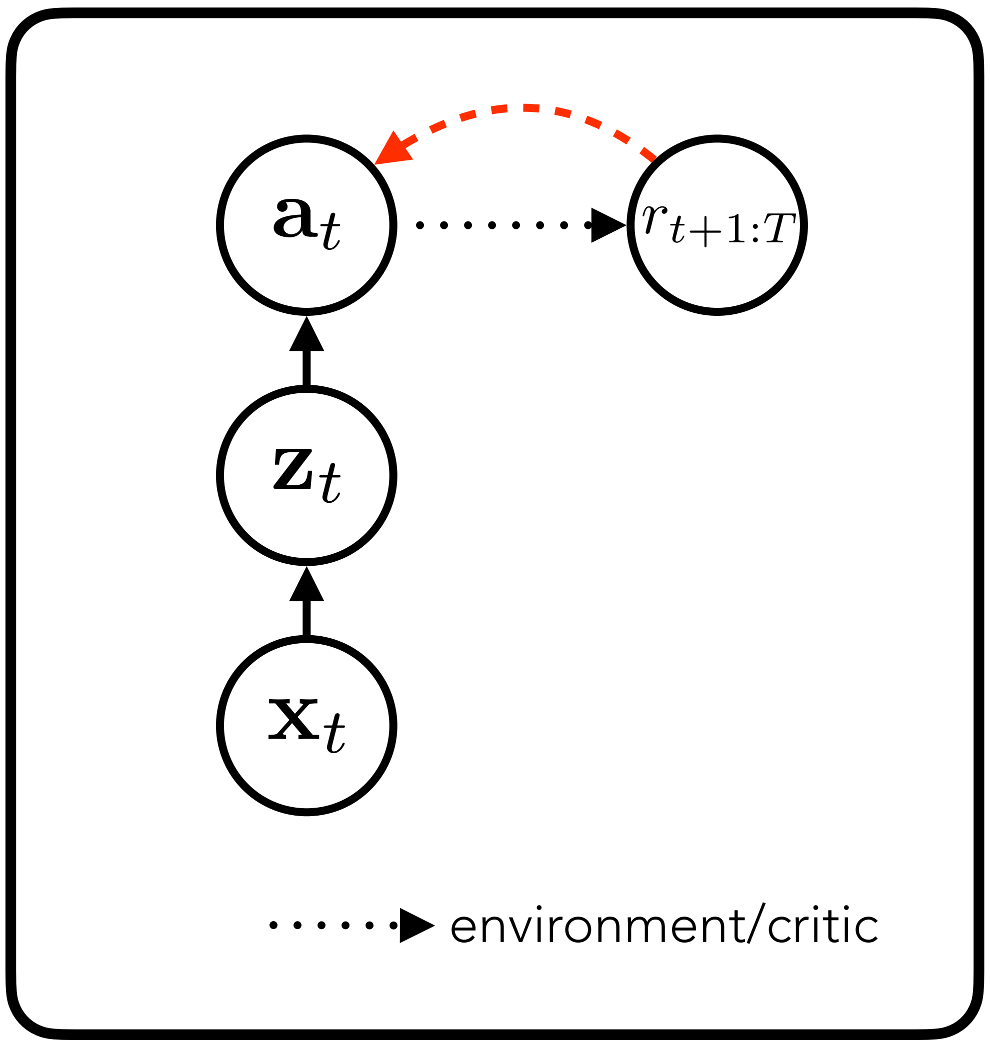


REINFORCEMENT LEARNING

approaches to policy optimization

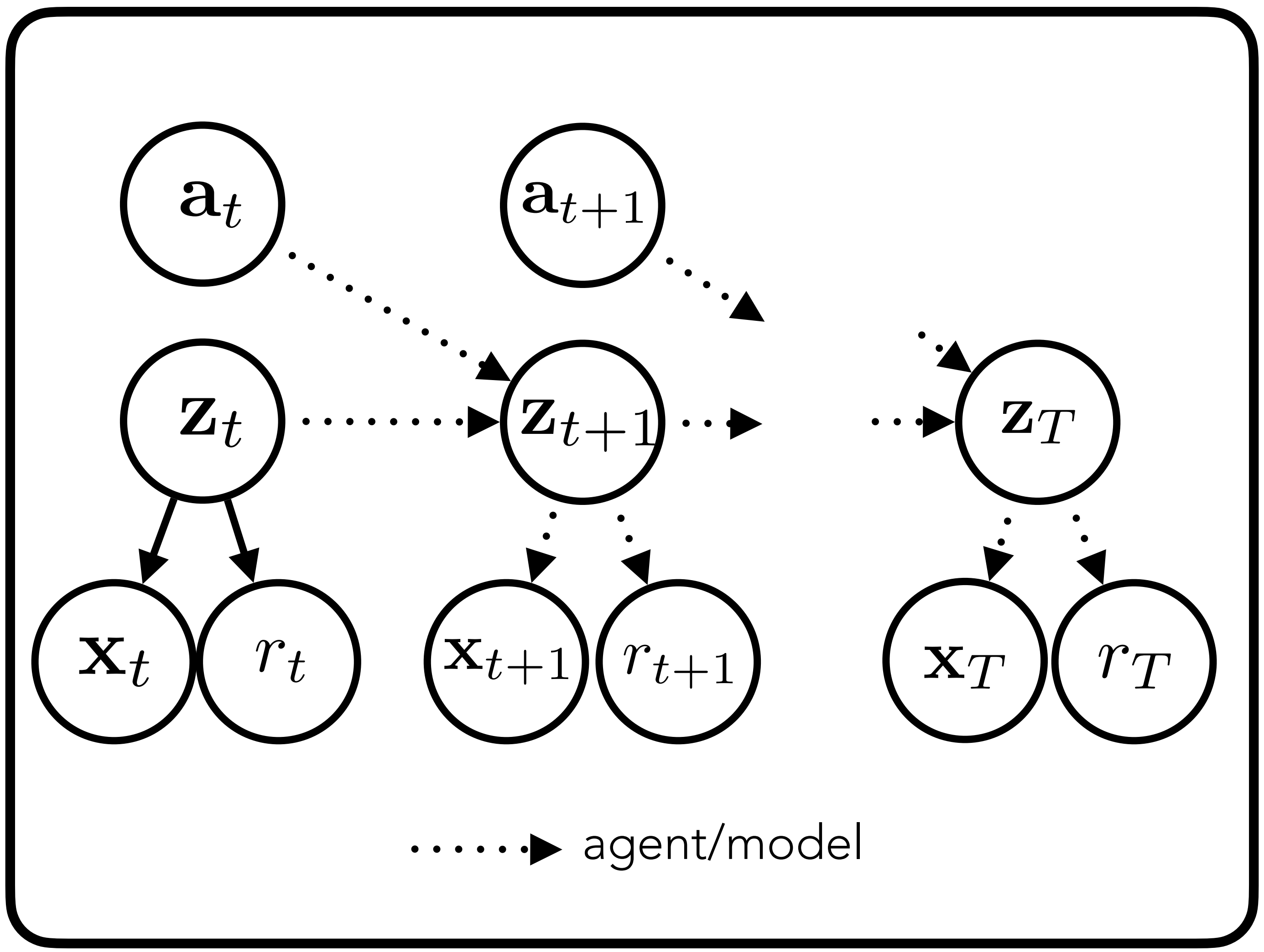
model-free

direct mapping to actions



model-based

unroll model to evaluate actions

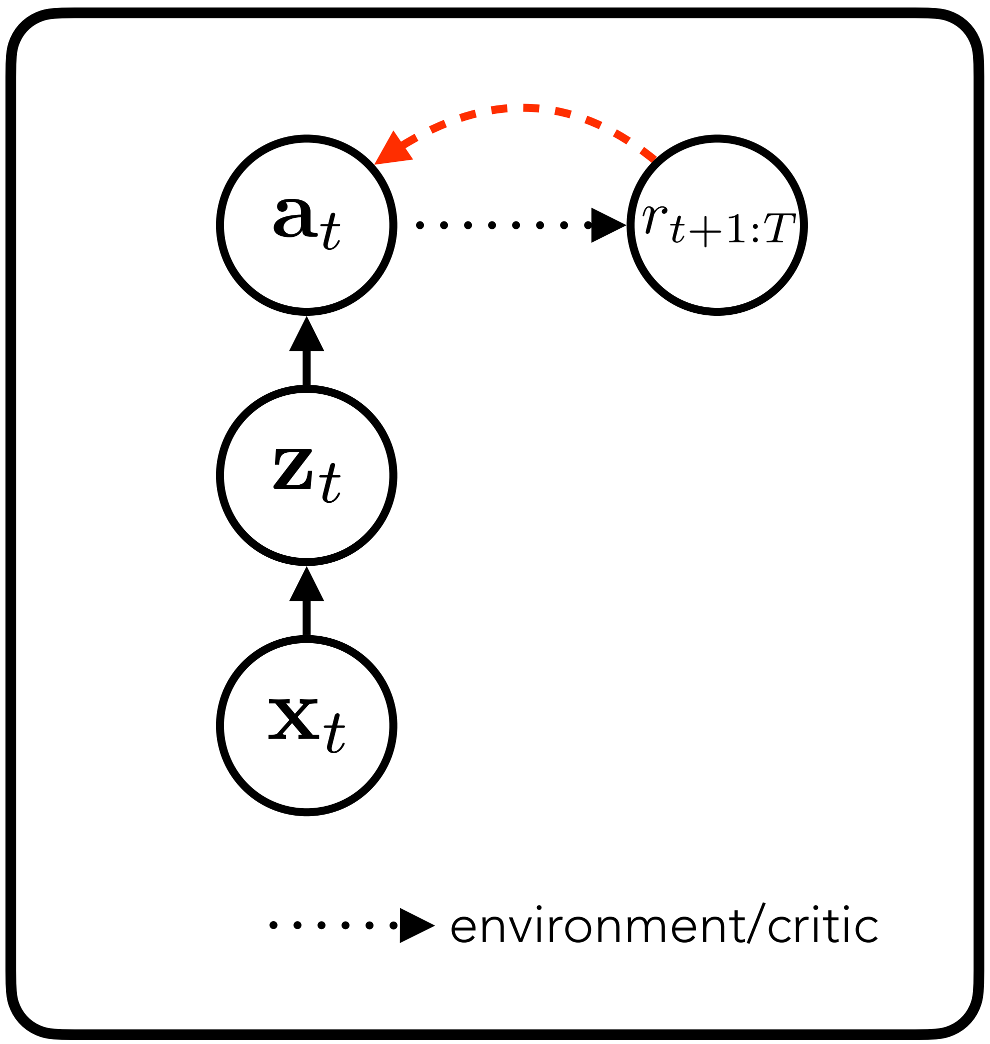


REINFORCEMENT LEARNING

approaches to policy optimization

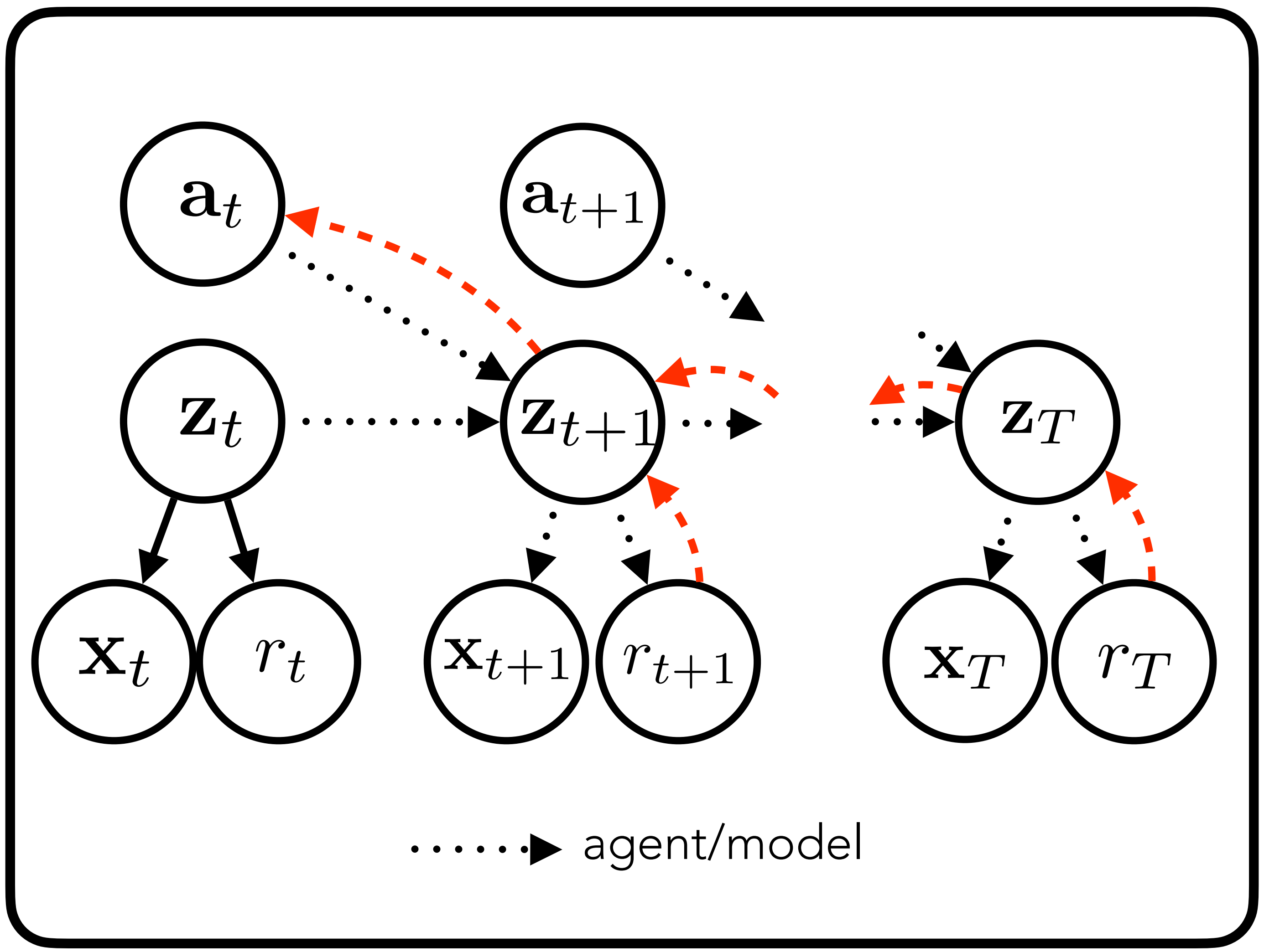
model-free

direct mapping to actions



model-based

unroll model to evaluate actions



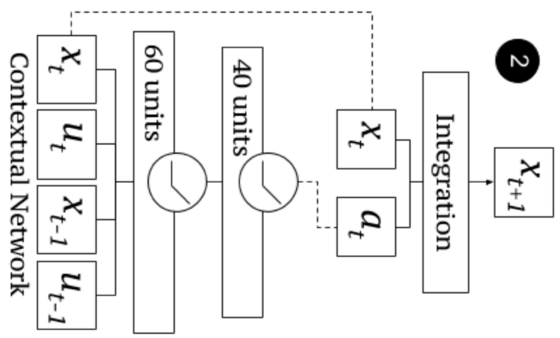
REINFORCEMENT LEARNING

approaches to policy optimization

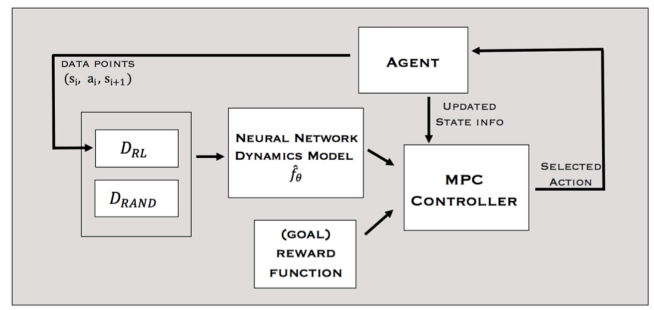
model-free	model-based
direct mapping to actions	*unroll model to evaluate actions*
+ **easy/fast to act**	**+** **often shorter to train**
— **often longer to train**	**—** **difficult/slow to act**
	also, **only as good as the model**

RECENT APPROACHES TO MODEL-BASED RL

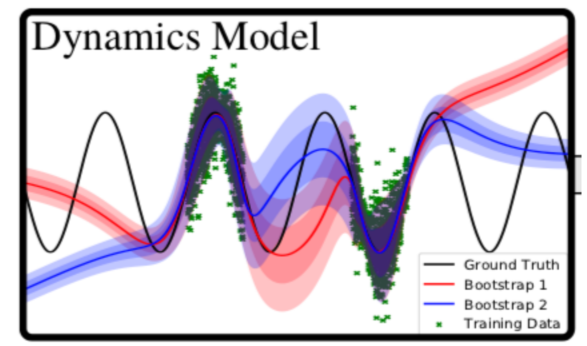
without latent variables (fully-observed):



Fu *et al.*, 2016



Nagabandi *et al.*, 2018

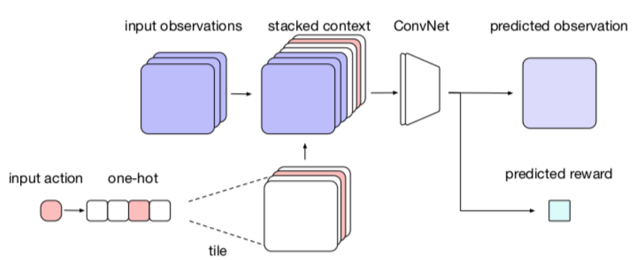


PETS

Chua *et al.*, 2018

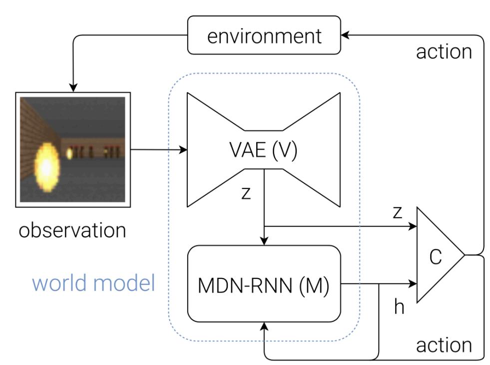
• • •

with latent variables (partially observed):



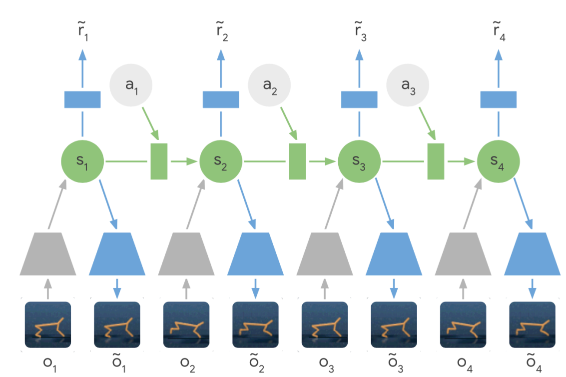
I2A

Weber *et al.*, 2017



World Models

Ha & Schmidhuber, 2018



Planning From Pixels

Hafner *et al.*, 2019

• • •

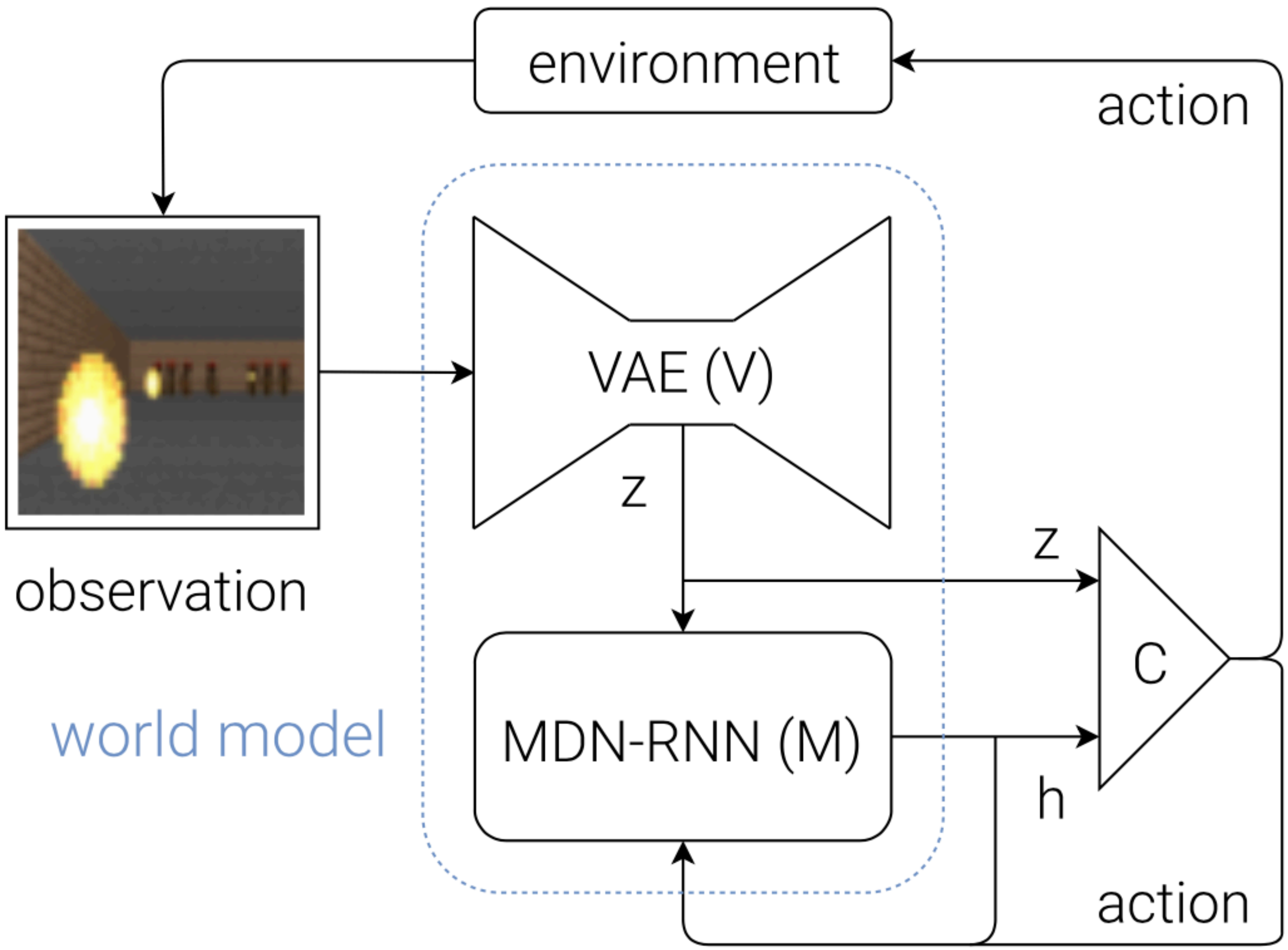
WORLD MODELS

- learn a generative model/compressed representation of environment from pixel observations
- use the model as a simulator to learn actions

Ha & Schmidhuber, 2018

WORLD MODELS

the model:

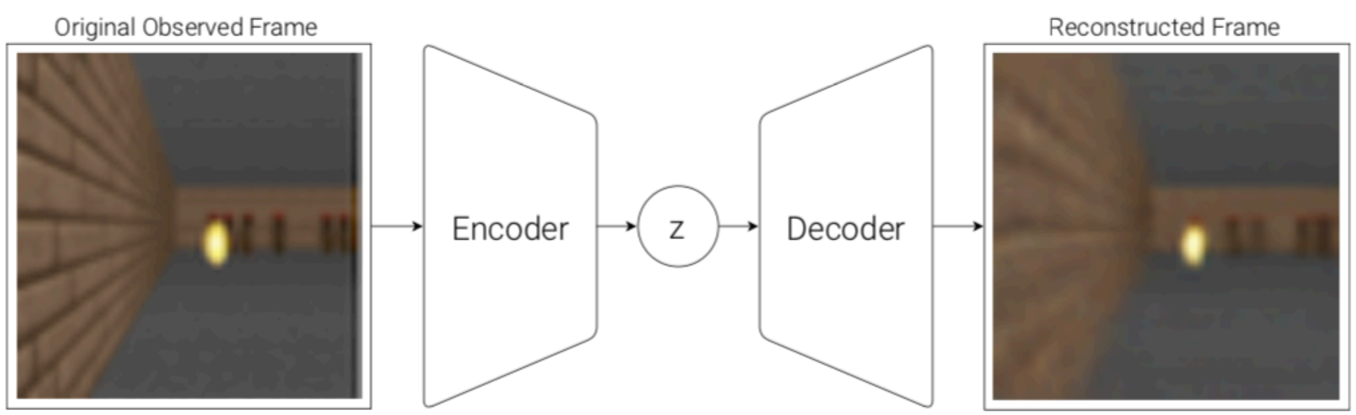


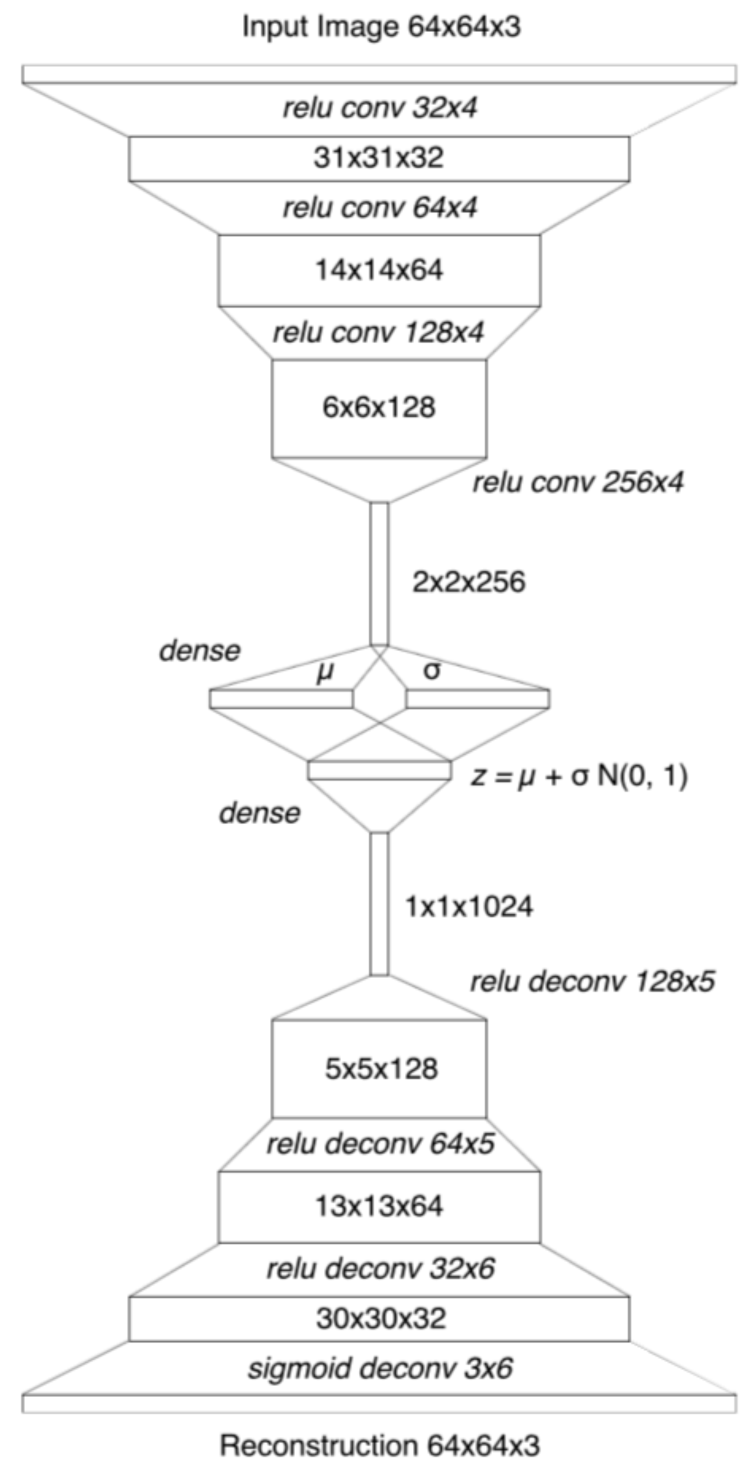
Ha & Schmidhuber, 2018

WORLD MODELS

the model (vision):

compress the observations



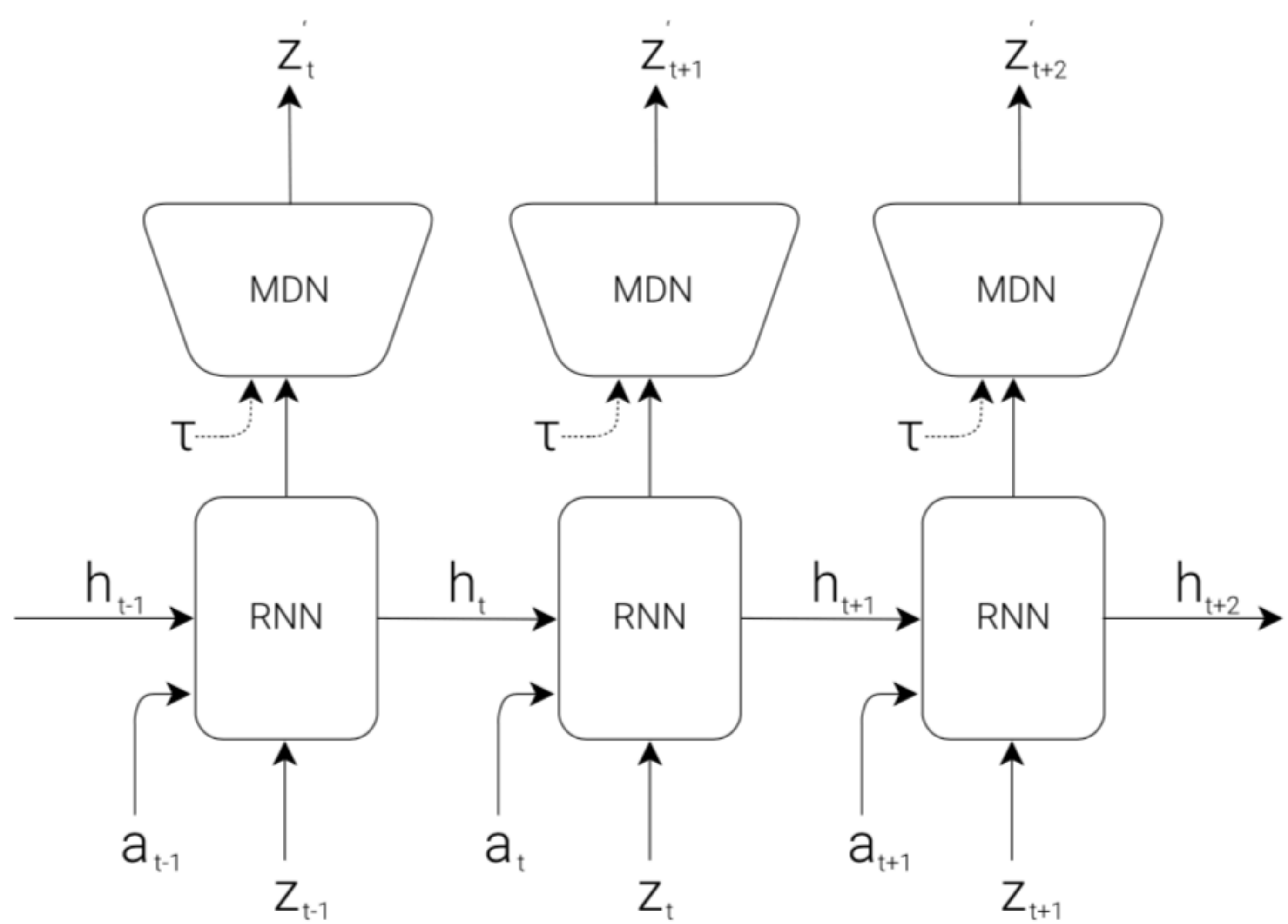


Ha & Schmidhuber, 2018

WORLD MODELS

the model (dynamics):

learn the dynamics of compressed state representations



Ha & Schmidhuber, 2018

WORLD MODELS

`CarRacing-v0`

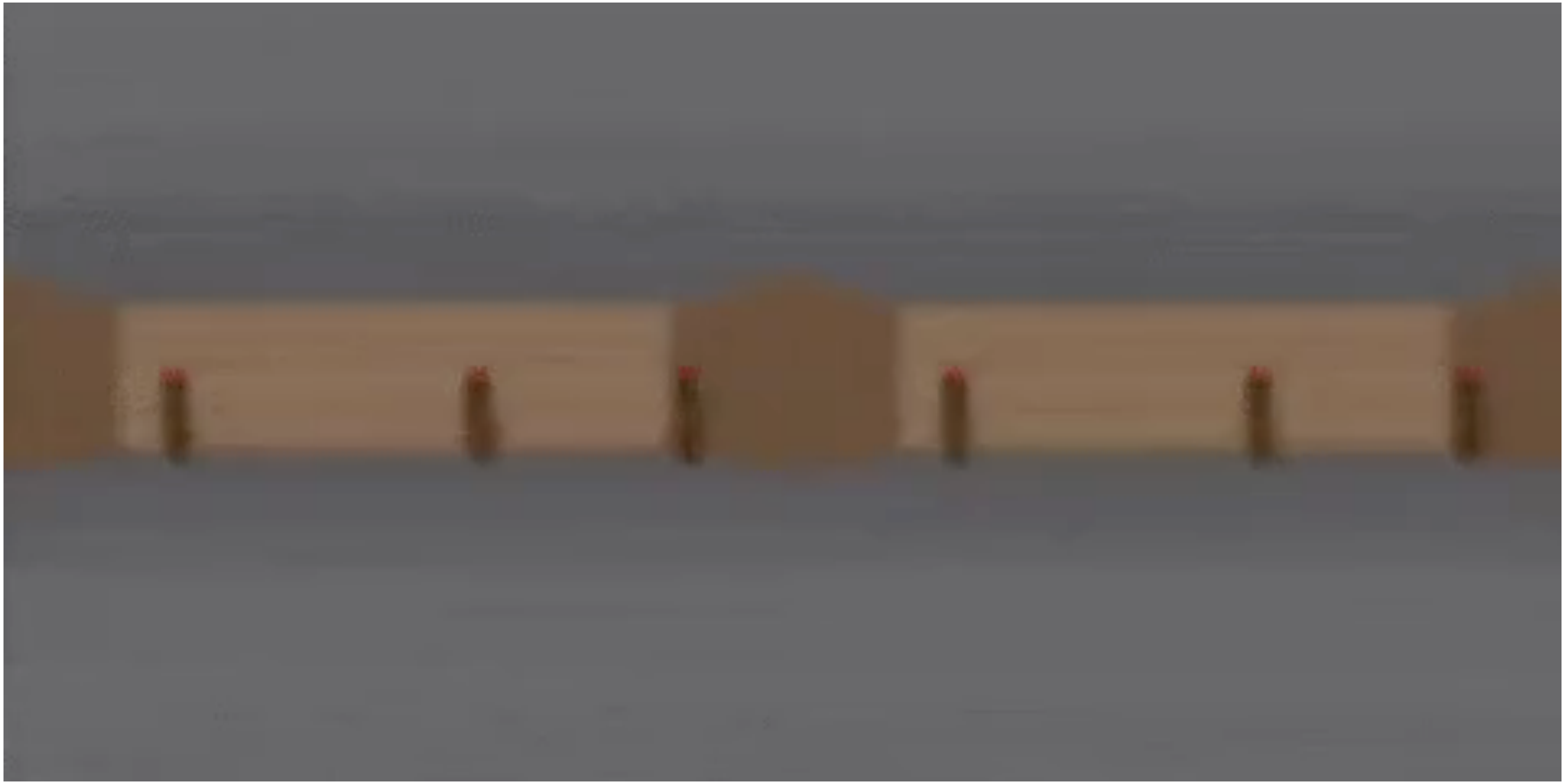
observations

reconstructions

Ha & Schmidhuber, 2018

WORLD MODELS

`VizDoomTakeCover`

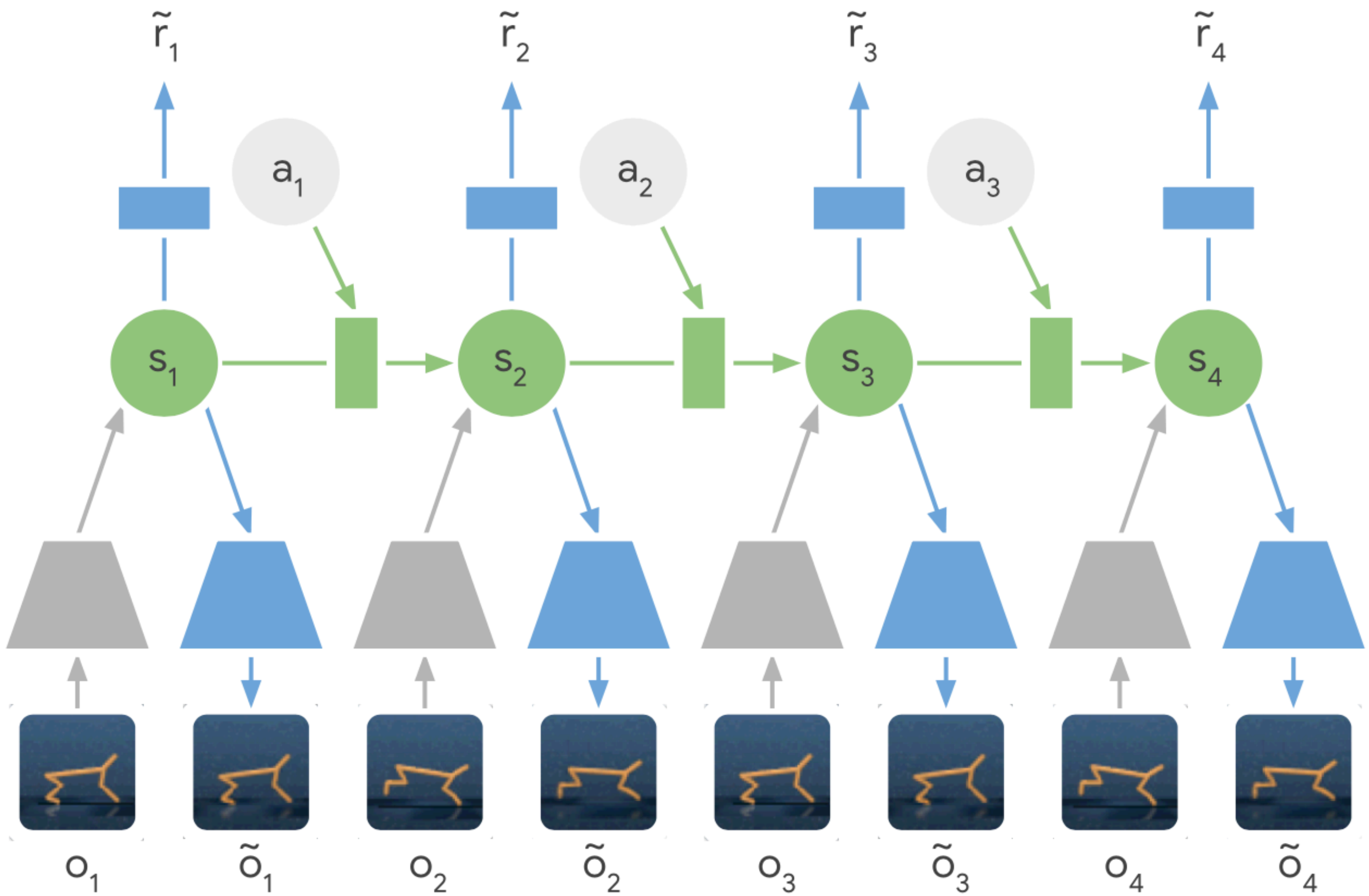


observations reconstructions

Ha & Schmidhuber, 2018

PLANNING FROM PIXELS

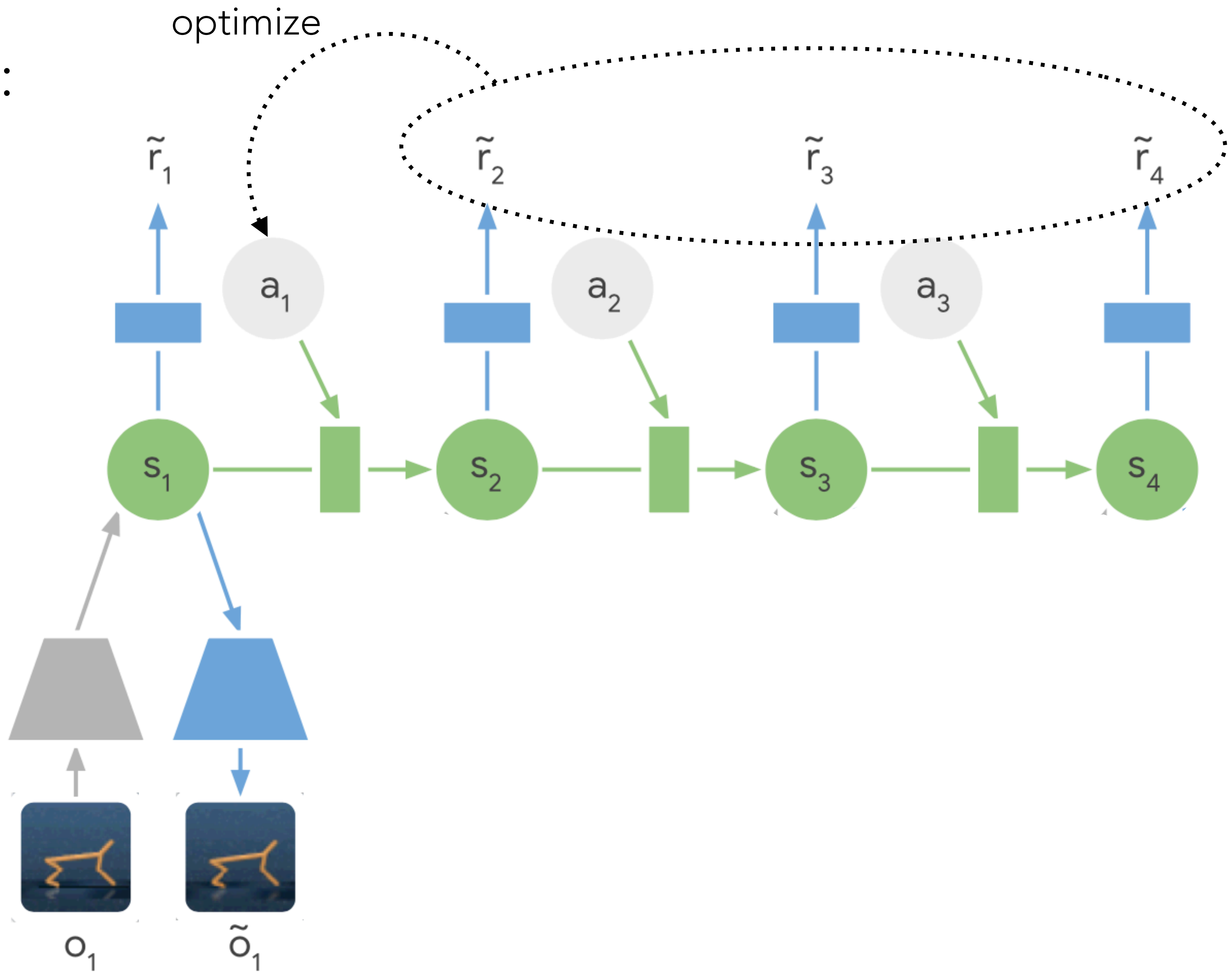
the model:



Hafner *et al.*, 2019

PLANNING FROM PIXELS

planning:

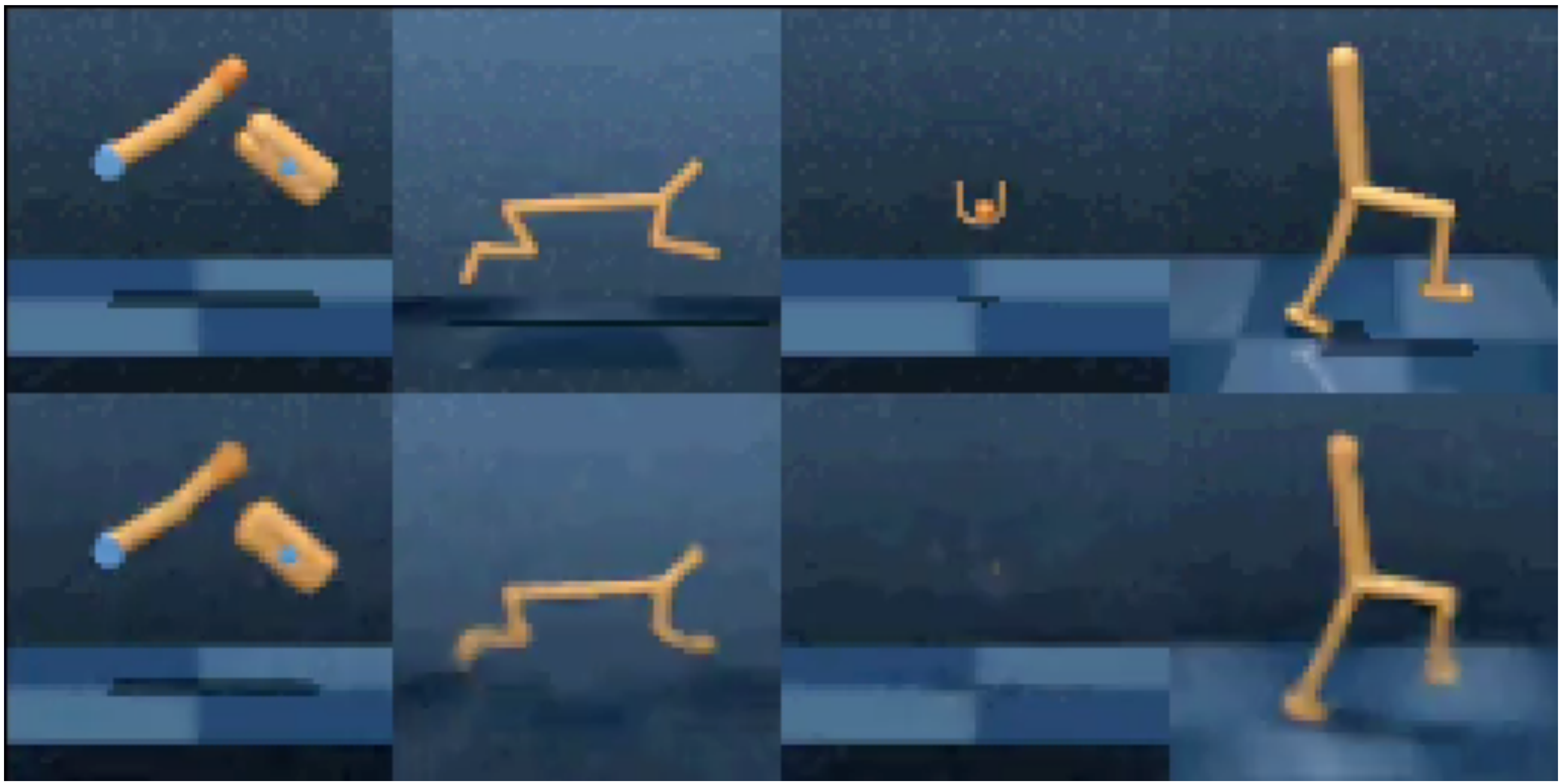


Hafner *et al.*, 2019

PLANNING FROM PIXELS

observations

predictions

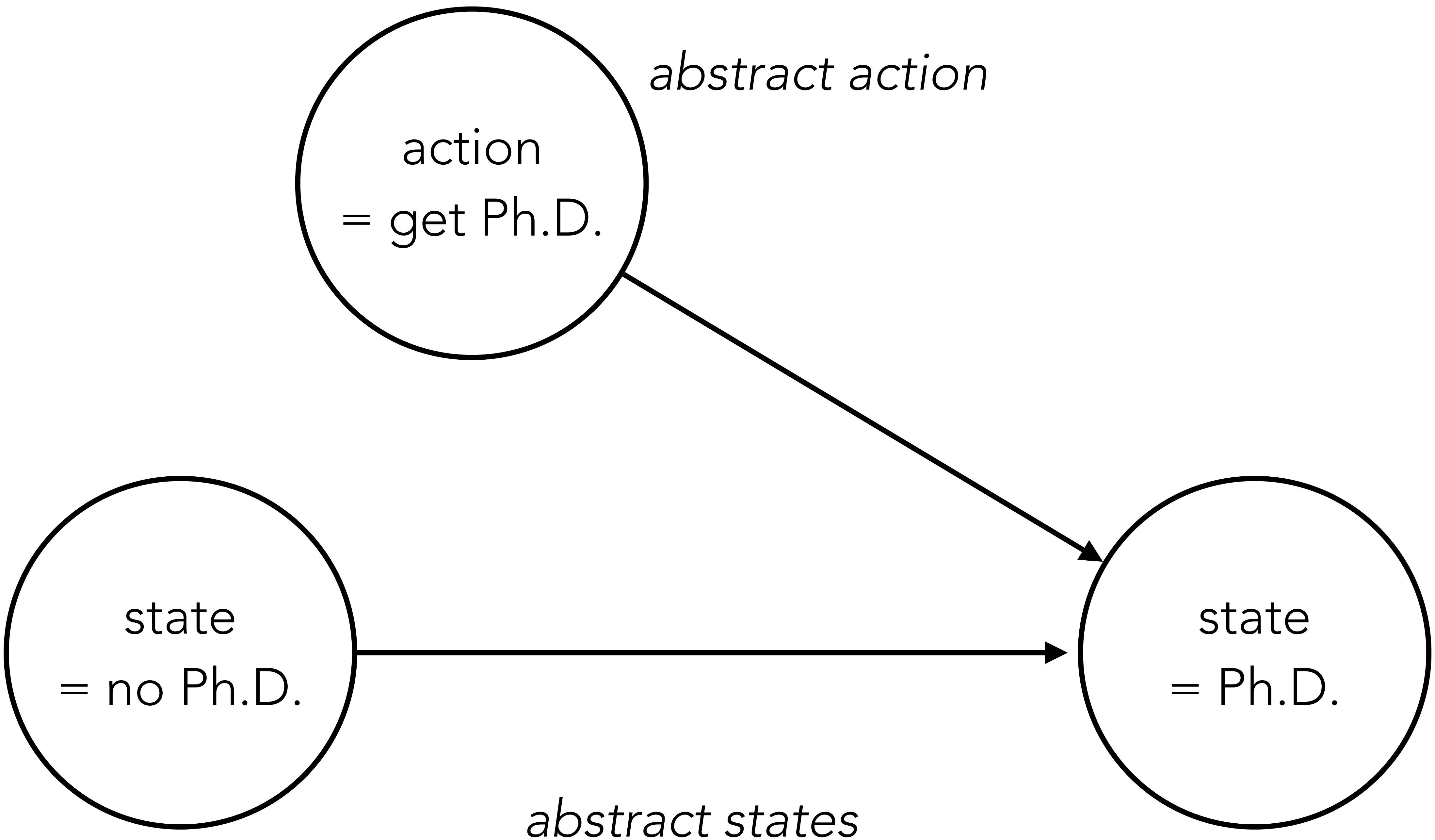


Hafner *et al.*, 2019

OPEN RESEARCH AREAS IN MODEL-BASED RL

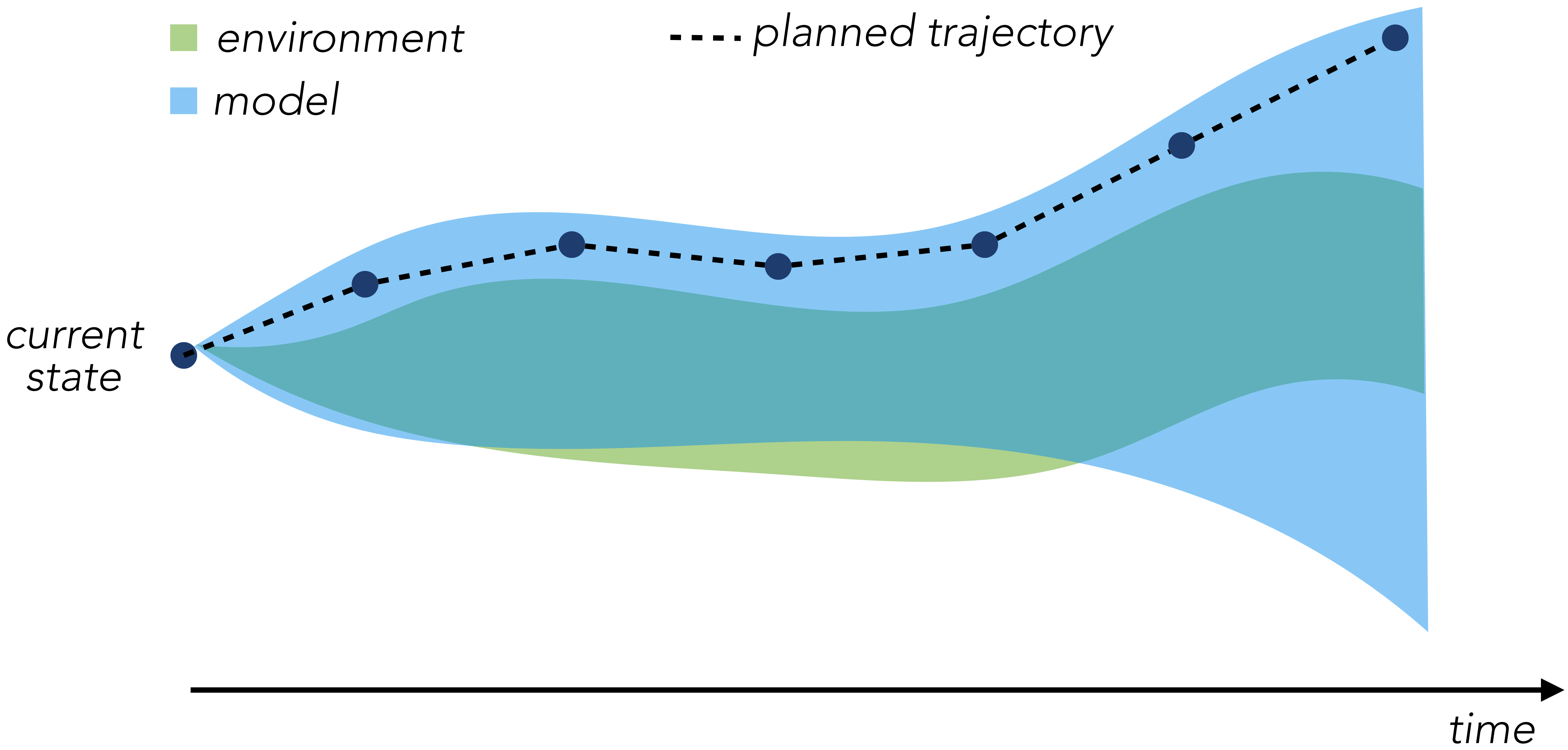
TEMPORAL ABSTRACTION

hierarchy of states and actions



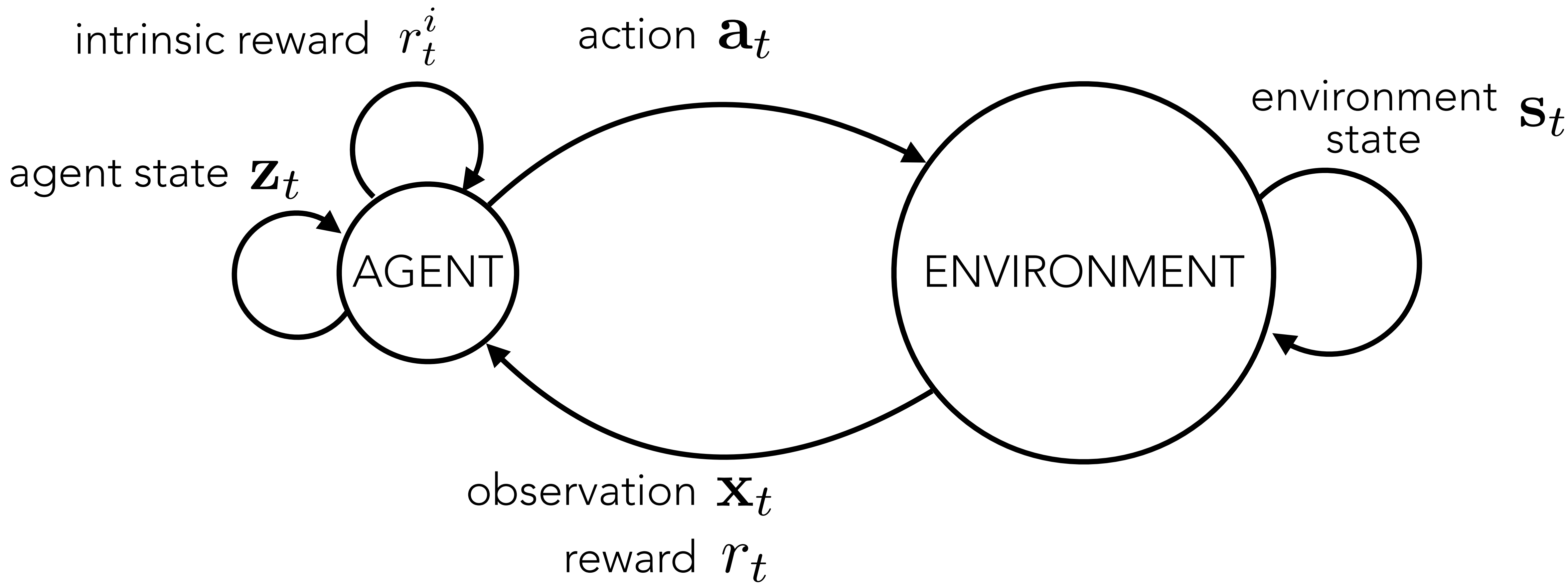
UNCERTAINTY ESTIMATION

distinguish between model uncertainty and environment stochasticity
prevent regions of exploitability in the model



INTRINSIC MOTIVATION

learning from intrinsic (non-environmental) rewards



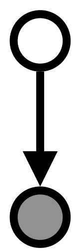
intrinsic reward signals:

surprise, empowerment, learning improvement, etc.

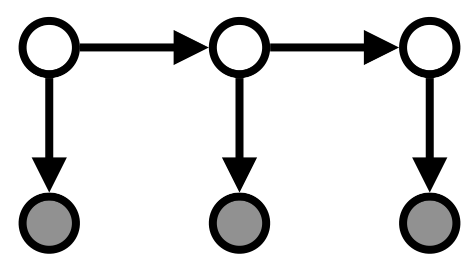
often helpful to have a model of the environment to estimate these quantities

OVERVIEW

LATENT VARIABLE
MODELS



DEEP SEQUENTIAL
LATENT VARIABLE MODELS



MODEL-BASED RL