# Accurate and adaptive neural recognition in dynamic environments

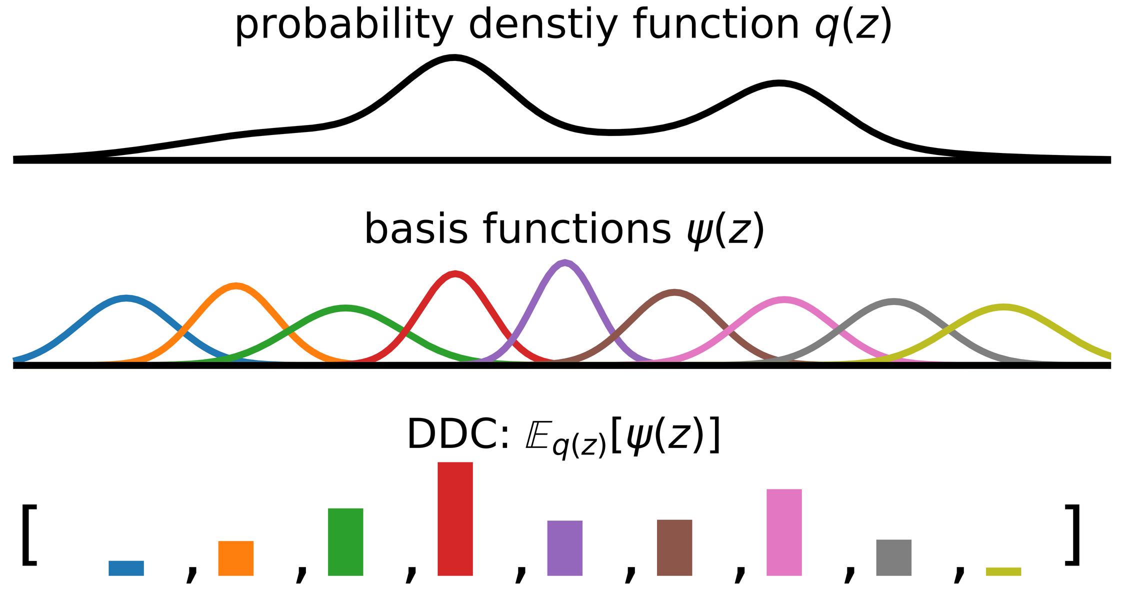
Li Kevin Wenliang, Eszter Vértes, Maneesh Sahani
Gatsby Computational Neuroscience Unit, University College London, {kevinli, eszter, maneesh}@gatsby.ucl.ac.uk

Poster III-36

## Distributed distributional code (DDC)

### Representation

DDC encode $q(\mathbf{z})$ with rate $\mathbf{r} := \mathbb{E}_{q(\mathbf{z})}[\psi(\mathbf{z})]$




(cf. Zemel, Dayan, and Pouget 1998; Sahani and Dayan 2003)

- Gaussian: $\psi(z) = [z, z^2]$ (mean and variance)
- Bernoulli: $\psi(z) = [z]$ (rate)
- exponential family: generic $\psi(z)$ (mean parameter)

**Accurate: no explicit distributional assumptions**

### Recognition

- Expectation approximation:

$$g(\mathbf{z}) \approx \alpha \cdot \psi(\mathbf{z}) \implies \mathbb{E}_{q(\mathbf{z})}[g(\mathbf{z})] \approx \alpha \cdot \mathbf{r}$$

- Bayes rule $q(\mathbf{z}|\mathbf{x}) = p(\mathbf{z}, \mathbf{x})/p(\mathbf{x})$

$$\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\psi(\mathbf{z})] \approx \mathbf{W}\phi(\mathbf{x}),$$
$$\mathbf{W} = \arg\min_{W'} \mathbb{E}_{p(\mathbf{z},\mathbf{x})} \| W'\phi(\mathbf{x}) - \psi(\mathbf{z})\|^2$$

**Biological: learning to infer by delta rule**

### Adaptation

- Adjust $\theta$ in internal model $p(\mathbf{z}, \mathbf{x}; \theta)$
- Gradients are expectations: $\mathbb{E}_q[\nabla_\theta p(\mathbf{z}, \mathbf{x}; \theta)]$

---

- The internal model is an Exp-Fam state-space model

$$p(\mathbf{z}_t|\mathbf{z}_{t\text{-}1}) = \exp\left[g_z(\mathbf{z}_{t\text{-}1}, \theta_z) \cdot T_z(\mathbf{z}_t) - \Phi_z(\mathbf{z}_{t\text{-}1}, \theta_z)\right]$$
$$p(\mathbf{x}_t|\mathbf{z}_t) = \exp\left[g_x(\mathbf{z}_t, \theta_x) \cdot T_x(\mathbf{x}_t) - \Phi_x(\mathbf{z}_t, \theta_x)\right]$$

- A DDC defines an **exponential family** of beliefs by maximum entropy.

$$q(\mathbf{z}|\rho) = \arg\max H(q) \text{ s.t. } \mathbb{E}_q[\psi; 1] = [\mathbf{r}; 1] \propto \exp(\rho \cdot \psi(\mathbf{z}))$$

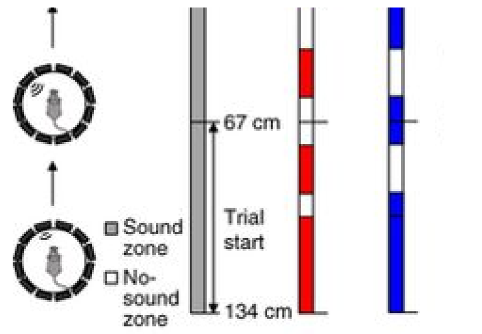
- Closed-form learning $\mathbf{W} = \mathbb{E}_p[\psi\phi^\intercal]\mathbb{E}_p[\phi\phi^\intercal]^{-1}$
- For models of $\mathbf{z}_1 \to \mathbf{z}_2 \to \mathbf{x}$, see (Vértes and Sahani 2018).

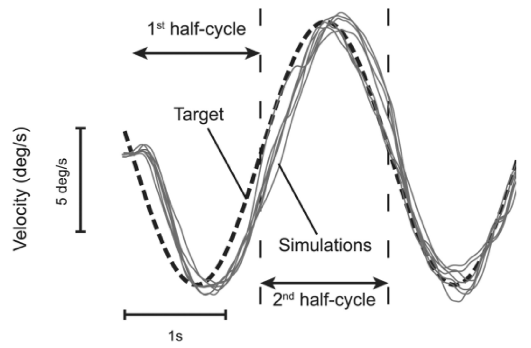
## Introduction

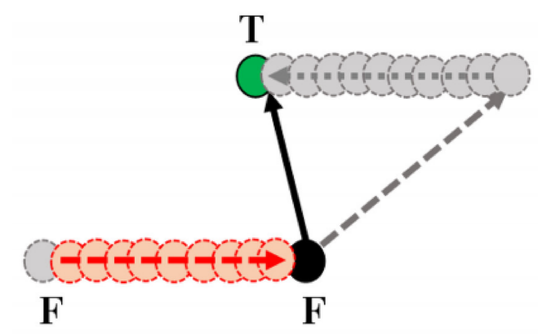
Survival often depends on the ability to accurately track and learn the behaviour of a latent feature in the world.
We propose an **accurate**, **adaptive** and **bio-plausible** method for latent feature recognition in dynamic environments.

**Examples:**

- localization (Funamizu, Kuhn, and Doya 2016)
- smooth pursuit (Xivry et al. 2013)
- spatial updating (Mohsenzadeh, Dash, and Crawford 2016)
- state representation in POMDP (**Poster III-65**)




**Previously proposed neural mechanisms**

- sampling (Kutschireiter et al. 2015; Legenstein and Maass 2014)
- probabilistic population codes (Makin, Dichter, and Sabes 2015; Sokoloski 2017)
- network attractor (Deneve, Duhamel, and Pouget 2007; Wilson and Finkel 2009)

## Recognition

### Compute posterior $\mathbf{r}_t = \mathbb{E}_{q(\mathbf{z}_t|\mathbf{x}_{1:t})}[\psi(\mathbf{z}_t)]$

- Assume a filtering function $\mathbf{r}_t = h_{\mathbf{W}_t}(\mathbf{r}_{t-1}, \mathbf{x}_t)$

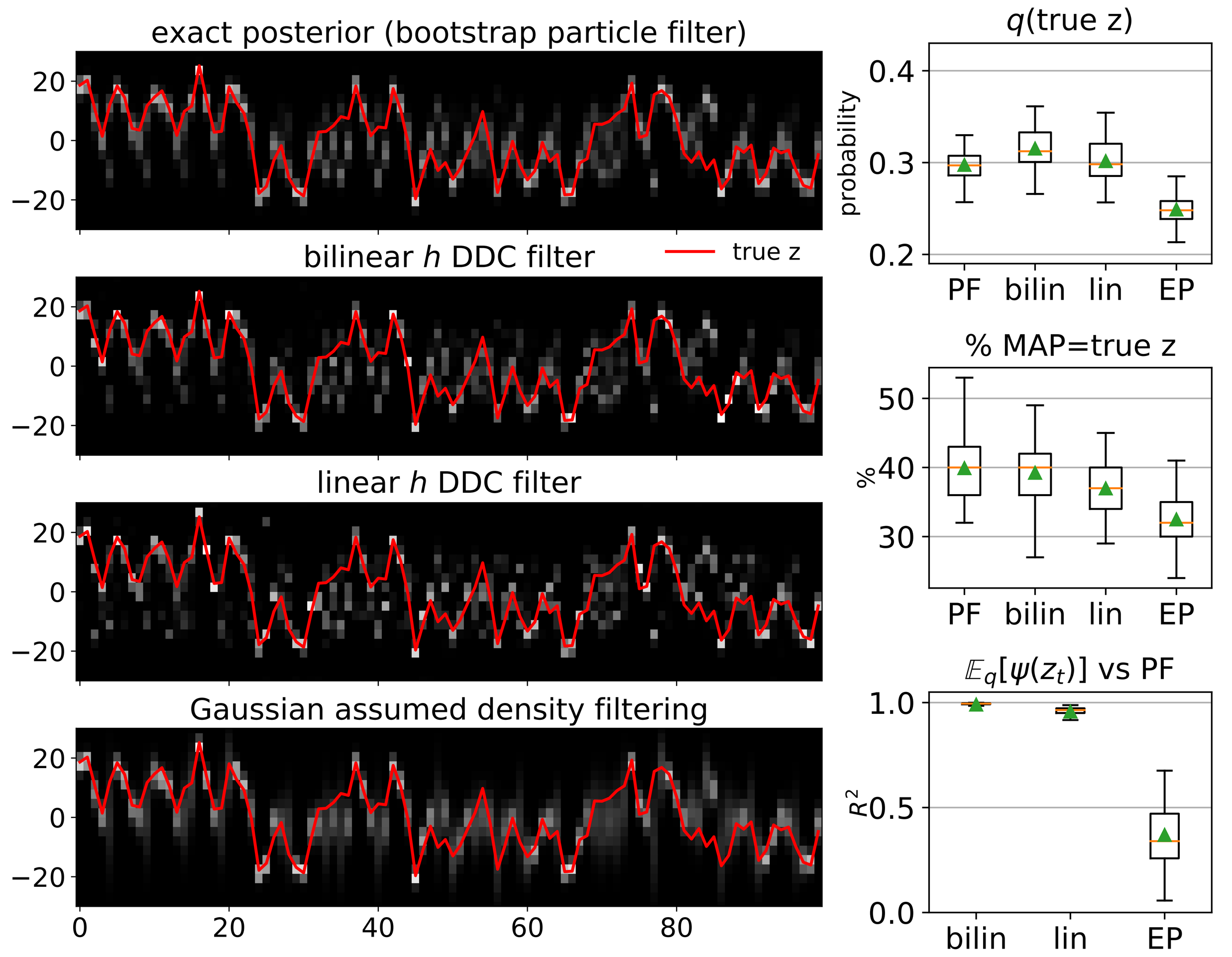
— linear: $\mathbf{r}_t = \mathbf{W}_t \cdot [\mathbf{r}_{t-1}; \phi_t]$

— bilinear: $\mathbf{r}_t = \mathbf{W}_t \cdot (\mathbf{r}_{t-1}\phi_t^\intercal) \Leftrightarrow r_{t,i} = W_{t,ijk} r_{t-1,j}\phi_{t,k}$

- Simulate $\mathbf{z}_{t\text{-}1}, \mathbf{r}_{t-1}, \mathbf{z}_t$, and $\mathbf{x}_t$ from internal model
- Least square learning:

$$\mathbf{W}_t = \arg\min_W \|h_W(\mathbf{r}_{t-1}, \mathbf{x}_t) - \mathbf{z}_t\|^2$$

### Results: multimodal posteriors




---

Formally, $\mathbf{r}_t = \mathbb{E}_{q(\mathbf{z}_t,\mathbf{x}_t|\mathbf{x}_{1:t\text{-}1})}[\psi_t\phi_t^\intercal]\mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_{1:t\text{-}1})}[\phi_t\phi_t^\intercal]^{-1}$

$$\mathbb{E}_{q(\mathbf{z}_t,\mathbf{x}_t|\mathbf{x}_{1:t\text{-}1})}[\psi_t\phi_t^\intercal] = \mathbb{E}_{q(\mathbf{z}_{t\text{-}1}|\mathbf{x}_{1:t\text{-}1})}\mathbb{E}_{p(\mathbf{z}_t,\mathbf{x}_t|\mathbf{z}_{t\text{-}1})}[\psi_t\phi_t^\intercal] = \mathbb{E}_q\mathbb{E}_p\left[\text{vec}(\psi_t\phi_t^\intercal)\phi_{t-1}^\intercal\right]\mathbb{E}_p[\phi\phi^\intercal]^{-1}\psi(\mathbf{z}_{t\text{-}1})$$
$$= \mathbb{E}_p\left[\text{vec}(\psi_t\phi_t^\intercal)\phi_{t-1}^\intercal\right]\mathbb{E}_p[\phi\phi^\intercal]^{-1}\mathbf{r}_{t-1}$$

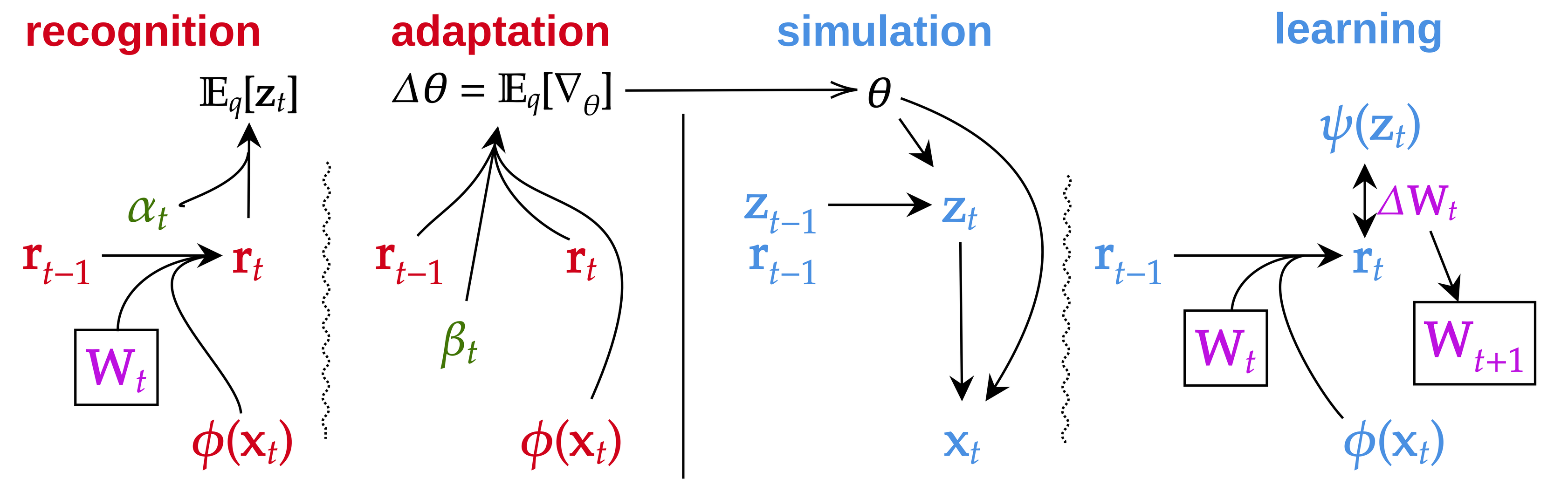
## Algorithm

An internal model of the observations

$$\mathbf{z}_t = f(\mathbf{z}_{t\text{-}1}; \theta) + \zeta_z(\theta)$$
$$\mathbf{x}_t = g(\mathbf{z}_t; \theta) + \zeta_x(\theta)$$

Infer distribution of latent $\mathbf{z}_t$ given signal history $\mathbf{x}_{1:t} := \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t\}$ by **wake-sleep**

- **recognition**: filter signal $\mathbf{x}_t$ using $\mathbf{W}_t$ giving $\mathbf{r}_t := \mathbb{E}_{q(\mathbf{z}_t|\mathbf{x}_{1:t})}[\psi(\mathbf{z}_t)]$
- **adaptation**: update $\theta$ using $\mathbf{r}_t$
- **simulation**: simulate data from internal model given $\theta$
- **learning**: update $\mathbf{W}_t$ using simulated data




## Summary

Using DDC,

- recognition is capable of tracking complex posteriors;
- posterior allows adaptation through expectation approximation;
- inference is trained using biological delta rule.

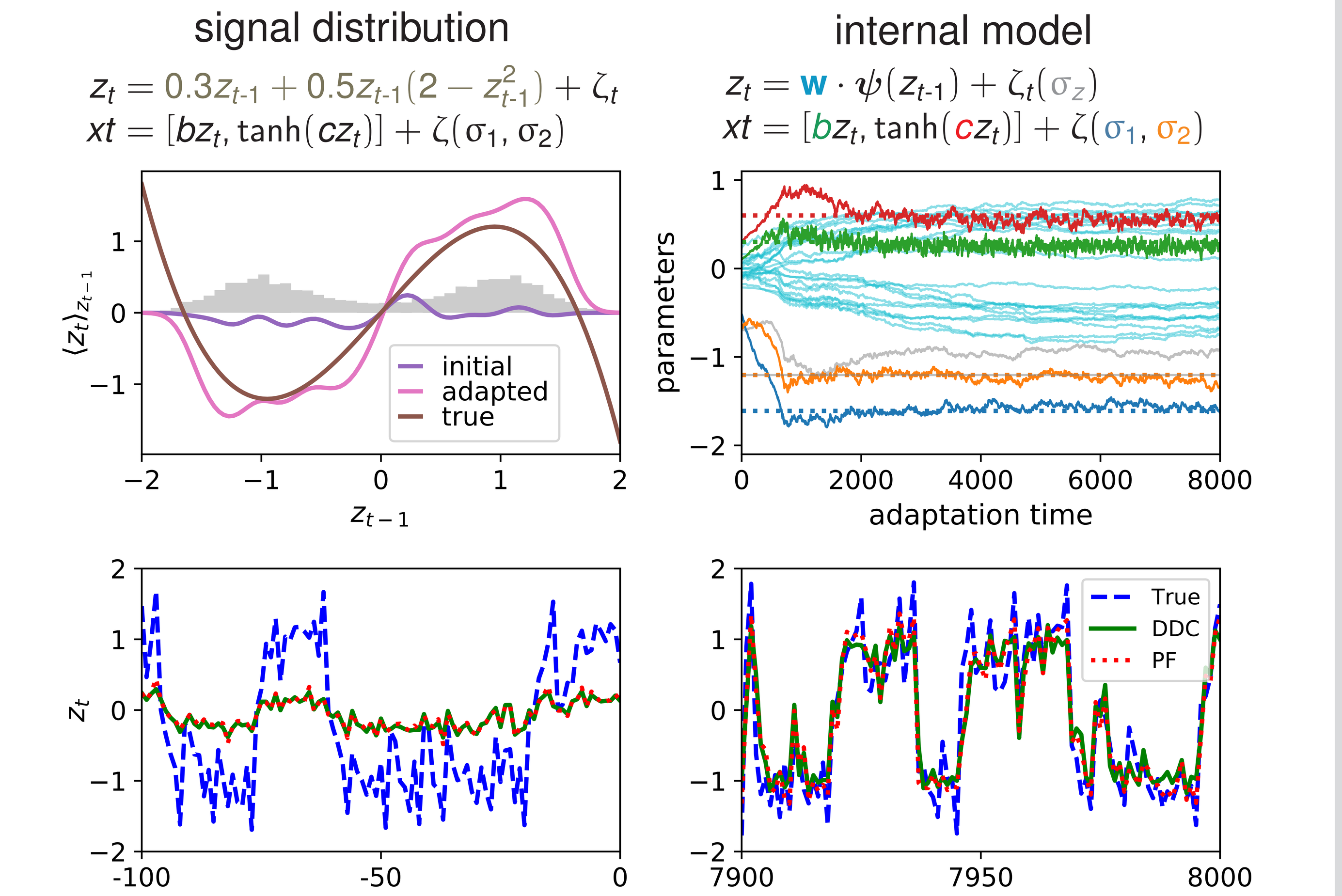
## Adaptation

### Update $\theta$ in internal model $p(\ldots; \theta)$

$$\Delta\theta \propto \mathbb{E}_{q(\mathbf{z}_t|\mathbf{x}_{1:t})}[\nabla_\theta \log p(\mathbf{z}_{t\text{-}1}, \mathbf{z}_t, \mathbf{x}_t; \theta)]$$

- approximated by DDC $\mathbf{r}$ and coefficients $\beta$
- $\beta$ assumed known by the internal model

### Results: nonlinear internal model




---

$$\mathbb{E}_q[\nabla_{\theta_x}\Phi_x(\mathbf{z}_t, \theta_x)] \approx \beta \cdot \mathbf{r}_t \quad \mathbb{E}_q[\nabla_{\theta_x} g_x(\mathbf{z}_t, \theta_x)] \approx \gamma \cdot \mathbf{r}_t \quad \mathbb{E}_q[\nabla_{\theta_z}\Phi_z(\mathbf{z}_{t\text{-}1}, \theta_z)] \approx \eta \cdot \mathbf{r}_t^{t-1}$$

The pairwise term: note that $p(\mathbf{z}_t, \mathbf{z}_{t\text{-}1}|\mathbf{x}_{1:t}) = p(\mathbf{z}_{t\text{-}1}|\mathbf{x}_{1:t})p(\mathbf{z}_t|\mathbf{z}_{t\text{-}1}, \mathbf{x}_t)$

$$\mathbb{E}_{\mathbf{z}_t,\mathbf{z}_{t\text{-}1}|\mathbf{x}_{1:t}}[\nabla_{\theta_z} g_z(\mathbf{z}_{t\text{-}1}, \theta_z) \cdot T_z(\mathbf{z}_t)] = \mathbb{E}_{\mathbf{z}_{t\text{-}1}|\mathbf{x}_{1:t}}[\mathbb{E}_{\mathbf{z}_t|\mathbf{z}_{t\text{-}1},\mathbf{x}_t}[\nabla_{\theta_z} g_z(\mathbf{z}_{t\text{-}1}, \theta_z) \cdot T_z(\mathbf{z}_t)]]$$
$$= \mathbb{E}_{\mathbf{z}_{t\text{-}1}|\mathbf{x}_{1:t}}[\mathbf{V} \cdot (\psi_{t\text{-}1}\phi_t^\intercal)] = \mathbf{V} \cdot (\mathbf{r}_t^{t-1}\phi_t^\intercal)$$