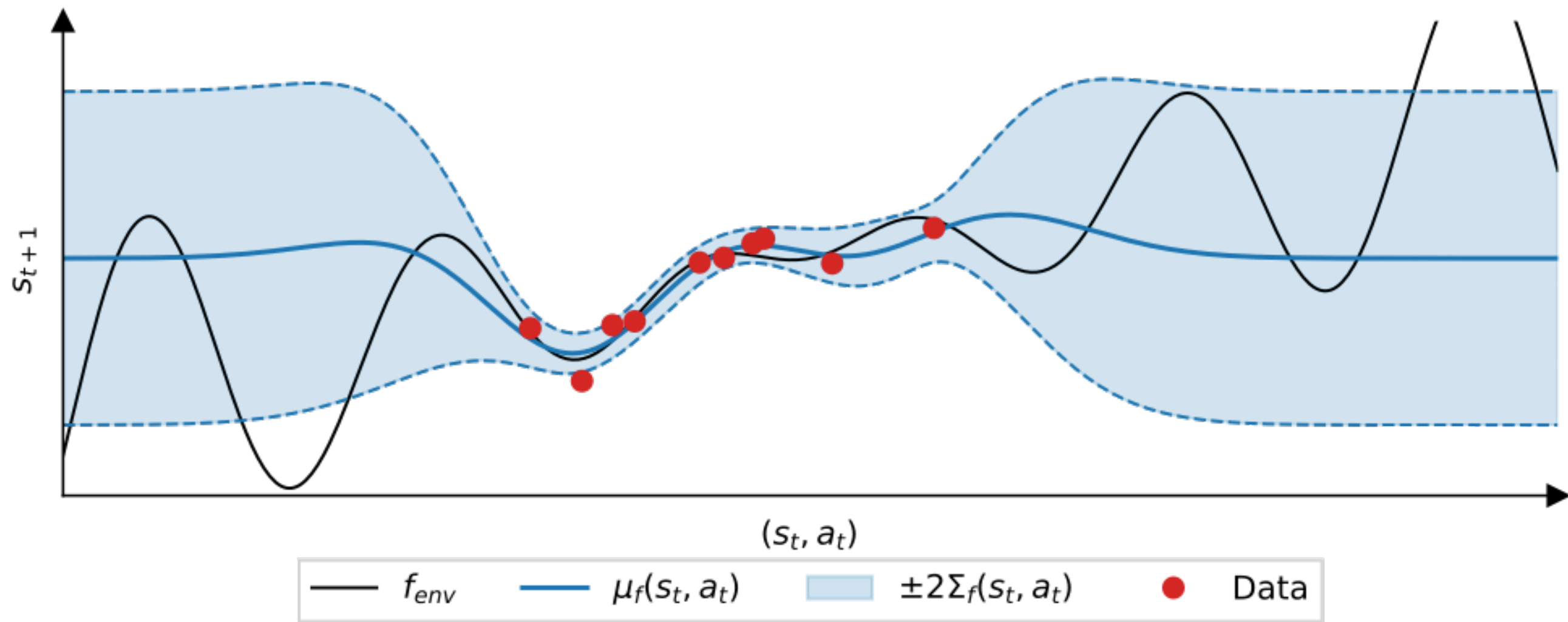# Exploration via Upper Confidence Bound

$$\pi^{UCB} = \arg\max_{\pi} \max_{f \in \mathcal{M}} J(\pi; f)$$

$$\mathcal{M} = \left\{ f \mid \|f(s,a) - \mu_f(s,a)\| \leq \beta \Sigma_f(s,a) \right\}$$

Legend: f_{env} — $\mu_f(s_t, a_t)$ — $\pm 2\Sigma_f(s_t, a_t)$ — Data

Axes: s_{t+1} (vertical), (s_t, a_t) (horizontal)

$$f(s, a) = \mu_f(s, a) - \beta \Sigma_f(s, a)$$

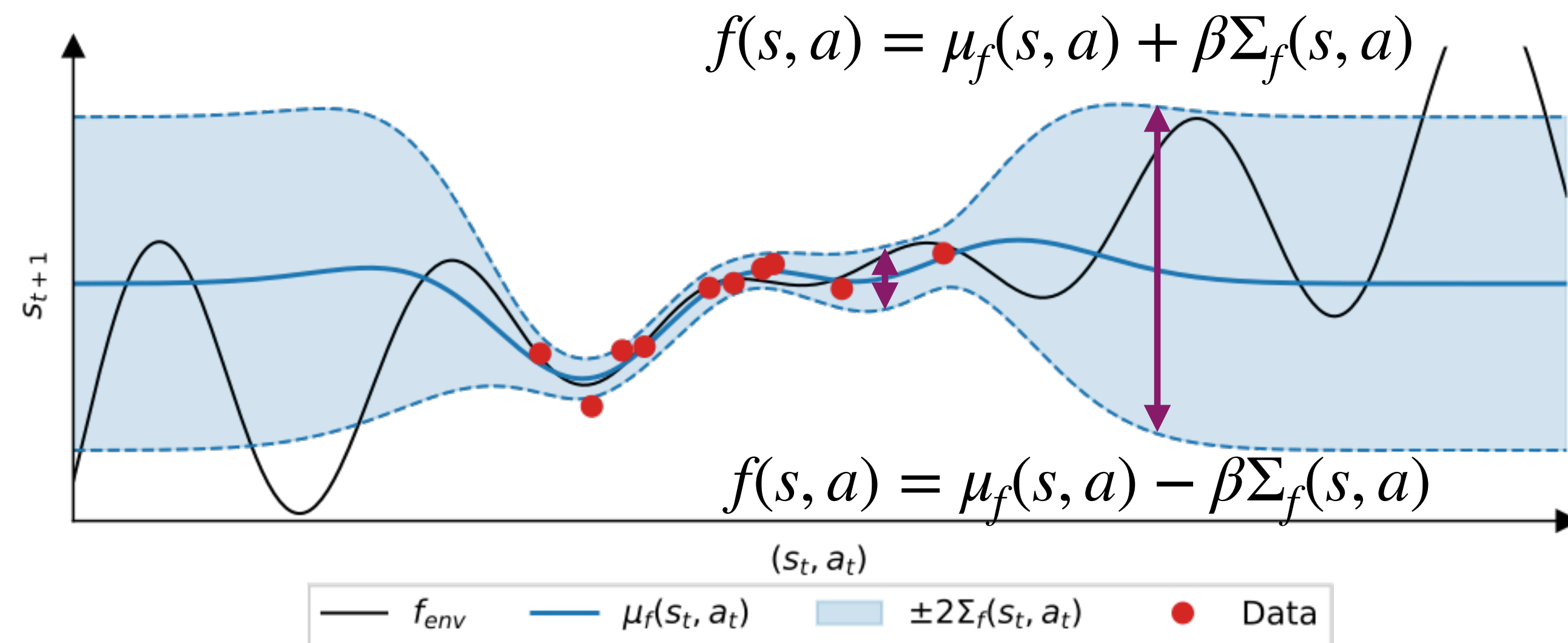$$f(s, a) = \mu_f(s, a) + \beta \Sigma_f(s, a)$$

Curi et al. (2020). Efficient Model-Based Reinforcement Learning through Optimistic Policy Search and Planning. NeurIPS.

Extra hyperparamter β

Exploration via Upper Confidence Bound

$$\pi^{UCB} = \arg \max_{\pi} \max_{f \in \mathcal{M}} J(\pi; f)$$

$$\mathcal{M} = \left\{ f \mid \|f(s,a) - \mu_f(s,a)\| \leq \beta \Sigma_f(s,a) \right\}$$

$$f(s,a) = \mu_f(s,a) + \beta \Sigma_f(s,a)$$

Extra hyperparamter β

$$f(s,a) = \mu_f(s,a) - \beta \Sigma_f(s,a)$$

Legend: — f_{env} — $\mu_f(s_t, a_t)$ ▢ $\pm 2\Sigma_f(s_t, a_t)$ ● Data

Axis labels: s_{t+1}, (s_t, a_t)

Curi et al. (2020). Efficient Model-Based Reinforcement Learning through Optimistic Policy Search and Planning. NeurIPS.

FCAI   **fcai.fi**

How to Quantify Uncertainty in Dynamics?

Scannell et al. (2024). Function-space Parameterisation of Neural Networks for Sequential Learning. ICLR.

FCAI

fcai.fi