BOFormer: Learning to Solve Multi-Objective Bayesian Optimization via Non-Markovian RL

Yu-Heng Hung

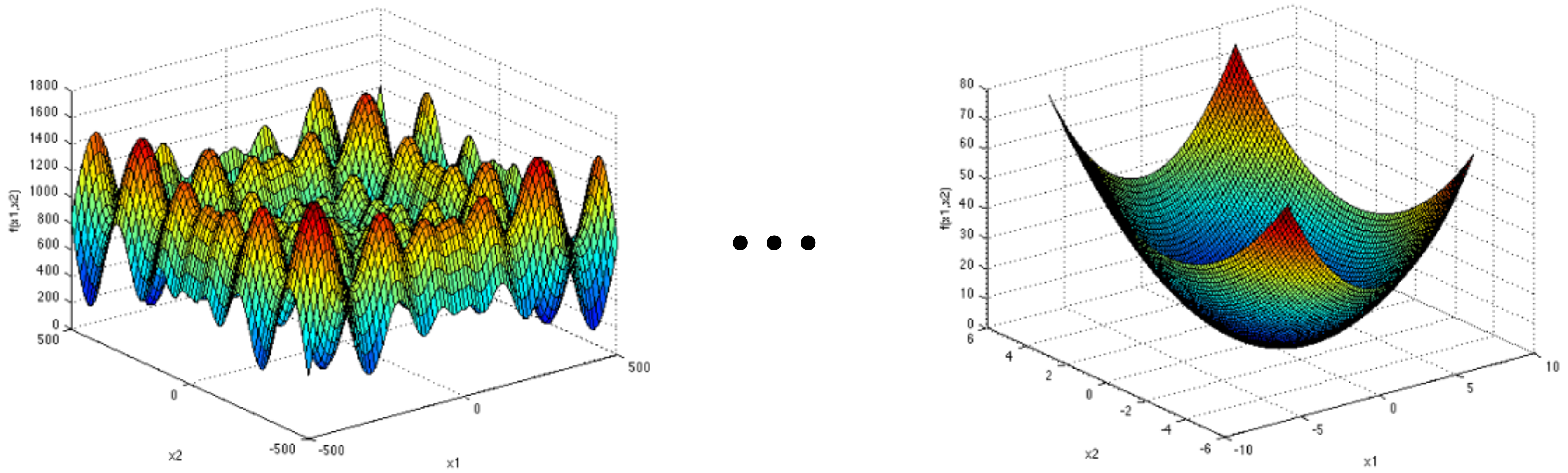
Department of Computer Science
National Yang Ming Chiao Tung University (NYCU)

Joint work with Kai-Jie Lin (NYCU), Yu-Heng Lin (NYCU), Ping-Chun Hsieh (NYCU), Cheng Sun (NVIDIA) and Chien-Yi Wang (NVIDIA)

Email: hungyh.cs08@nycu.edu.tw

Black-Box Optimization

Maximize $f_1(x), f_2(x), \cdots, f_K(x)$



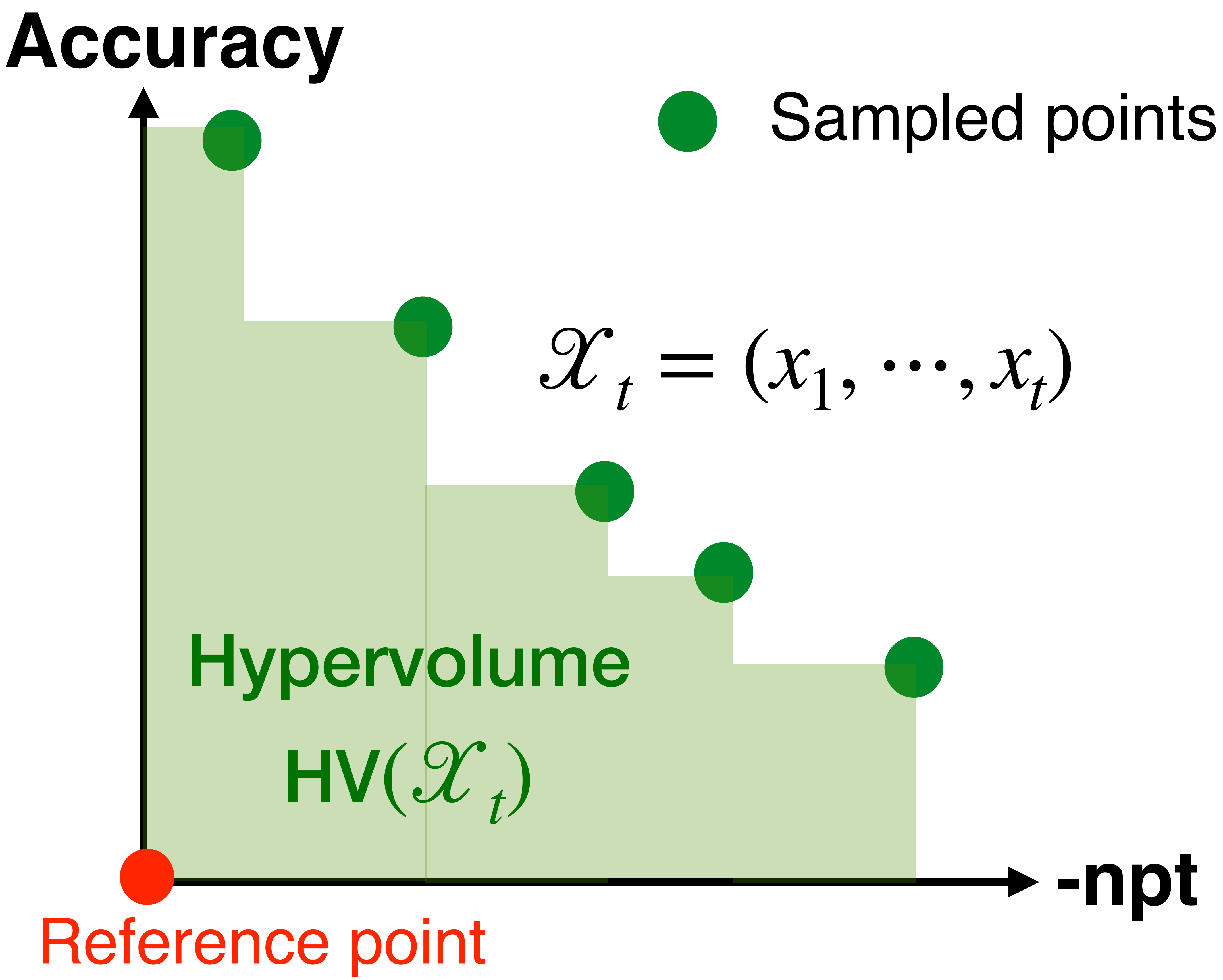
Next iterate x_{t+1}

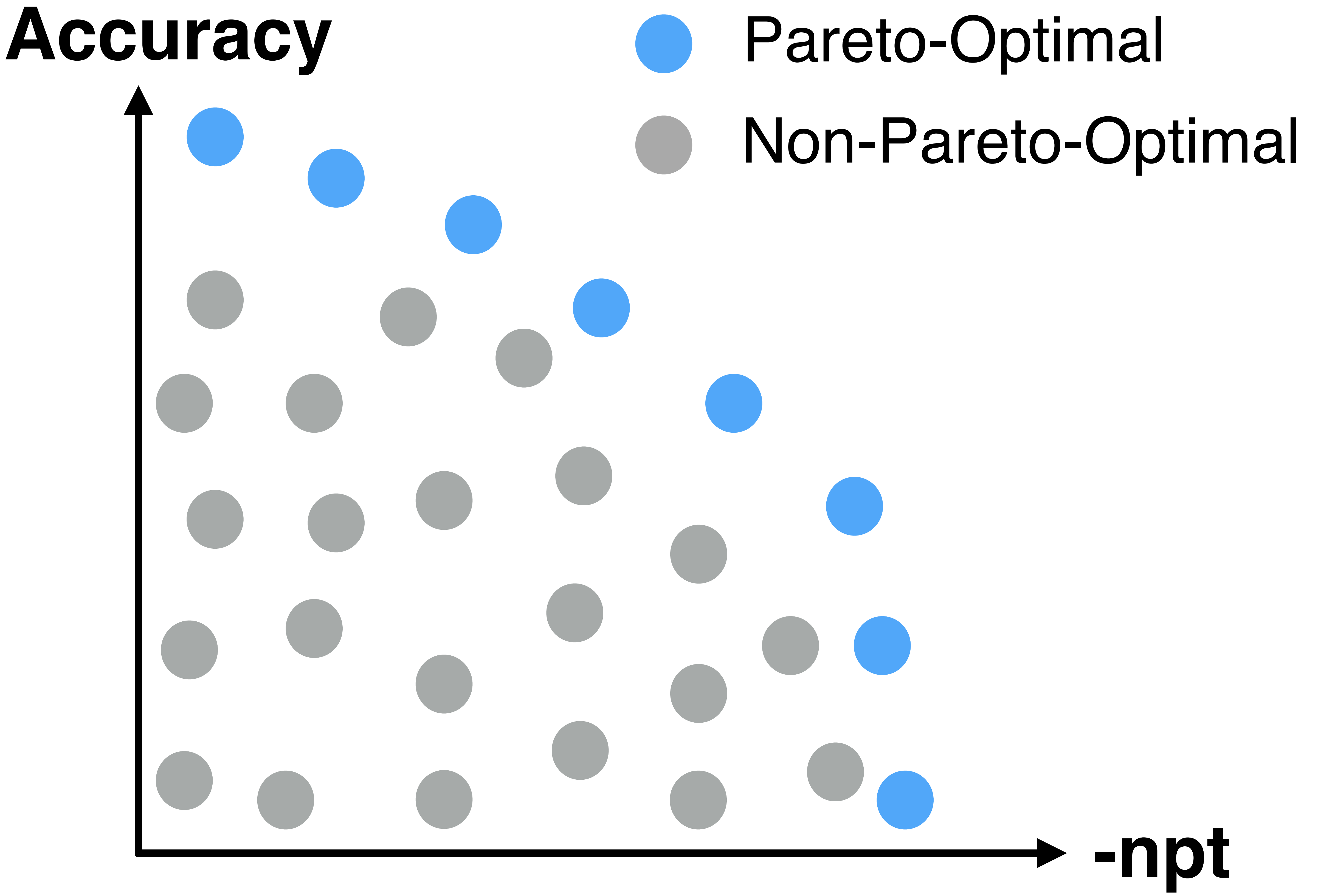
$f_1(x_t), f_2(x_t), \cdots, f_K(x_t)$

Sampling Strategy

- $f_1(x), f_2(x), \cdots, f_K(x)$ can be accessed only by sequential sampling
- No gradient information available
- Sampling is usually assumed expensive or time-consuming

Multi-Objective Optimization: Hypervolume





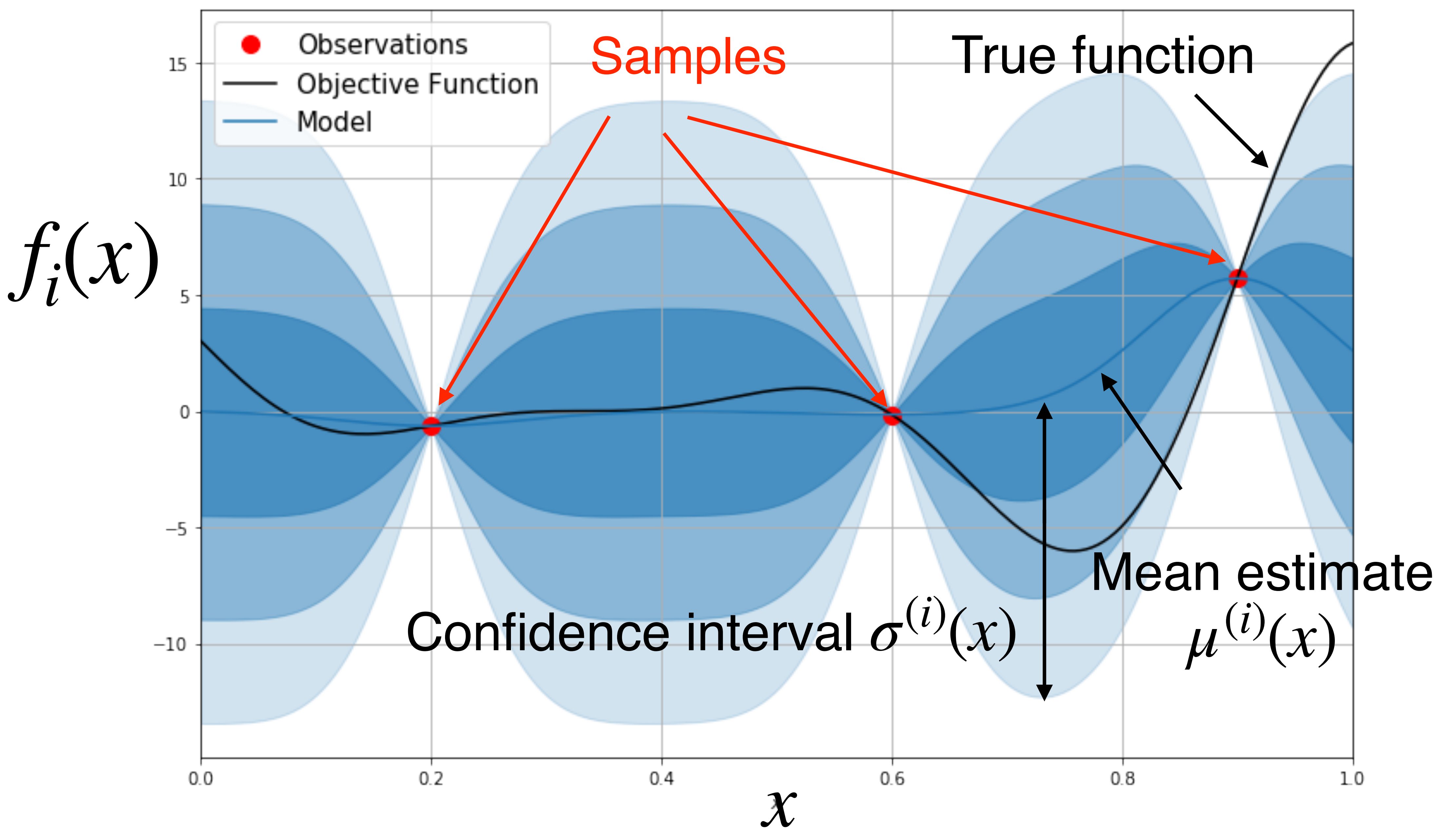
(Each point denotes the attained 2 objectives of a chosen capacitor value)

Find all Pareto-optimal points $\equiv$ Maximize hypervolume

Multi-Objective Bayesian Optimization (MOBO)

1. Gaussian process prior

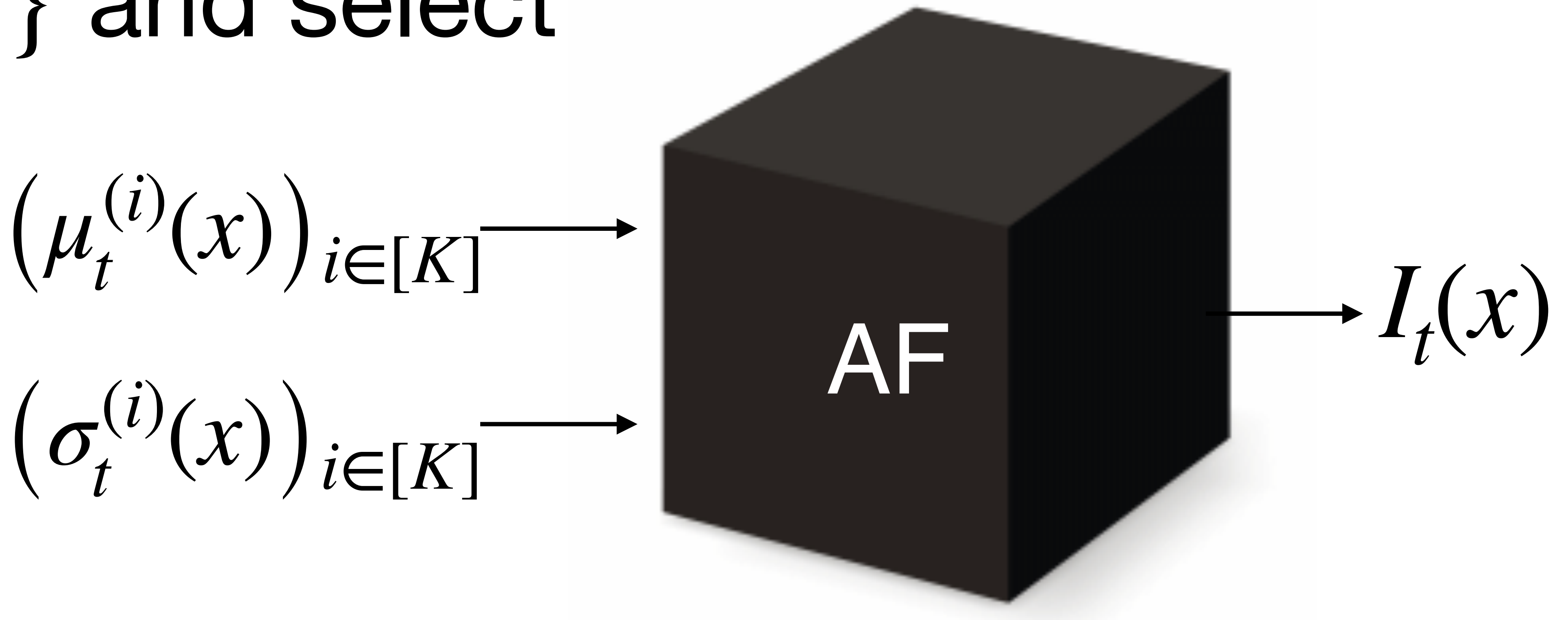
- The posterior enjoys a closed form by posterior mean and variance $(\mu_t^{(i)}(x), \sigma_t^{(i)}(x)^2)_{i\in[K]}$



2. Acquisition function

- Determines the next sample by an acquisition function (AF)
- At each t, AF computes $I_t(x)$ from $\{\mu_t^{(i)}(x), \sigma_t^{(i)}(x)^2\}$ and select

$$x_{t+1} = \arg\max_x I_t(x)$$

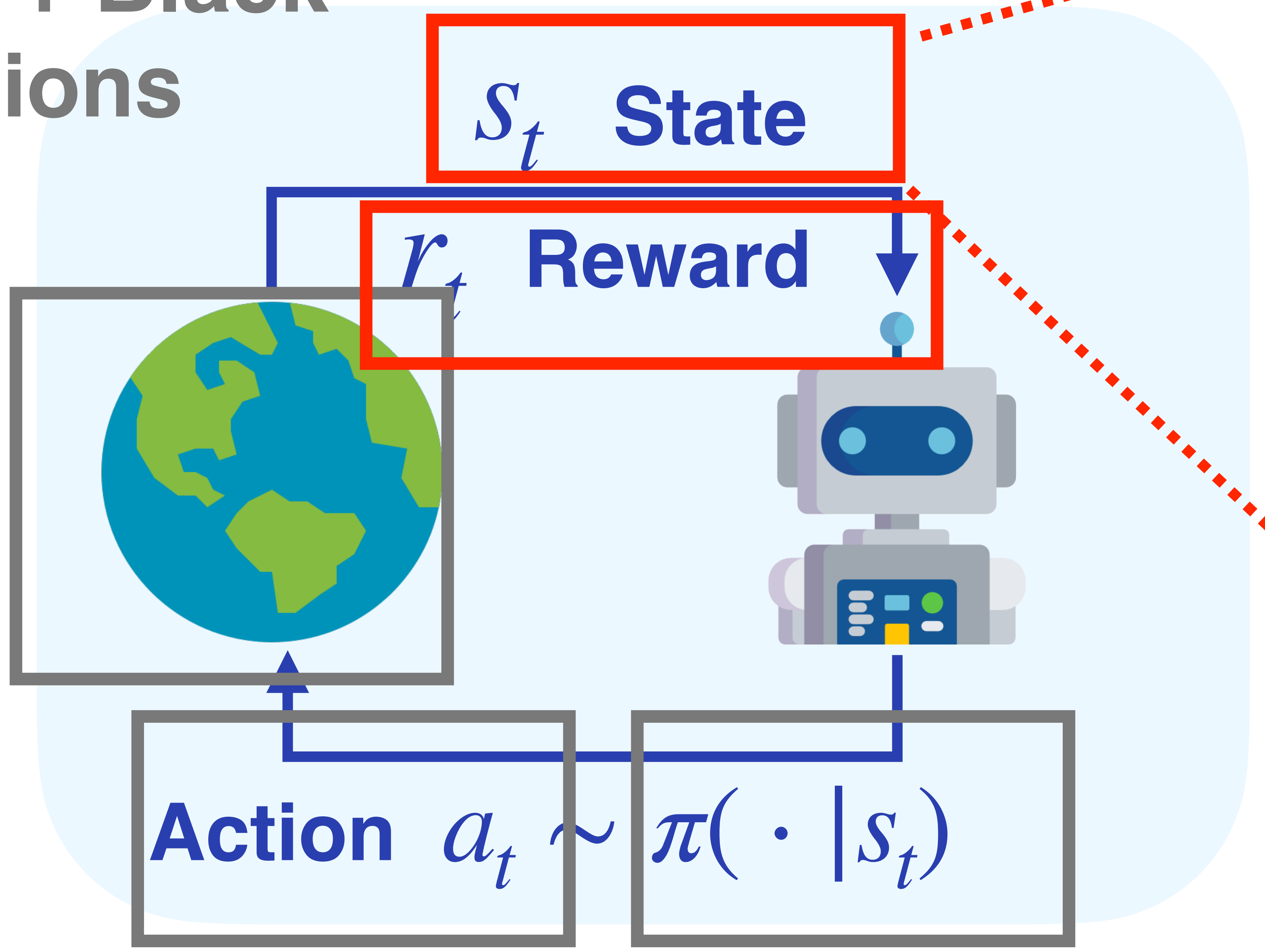


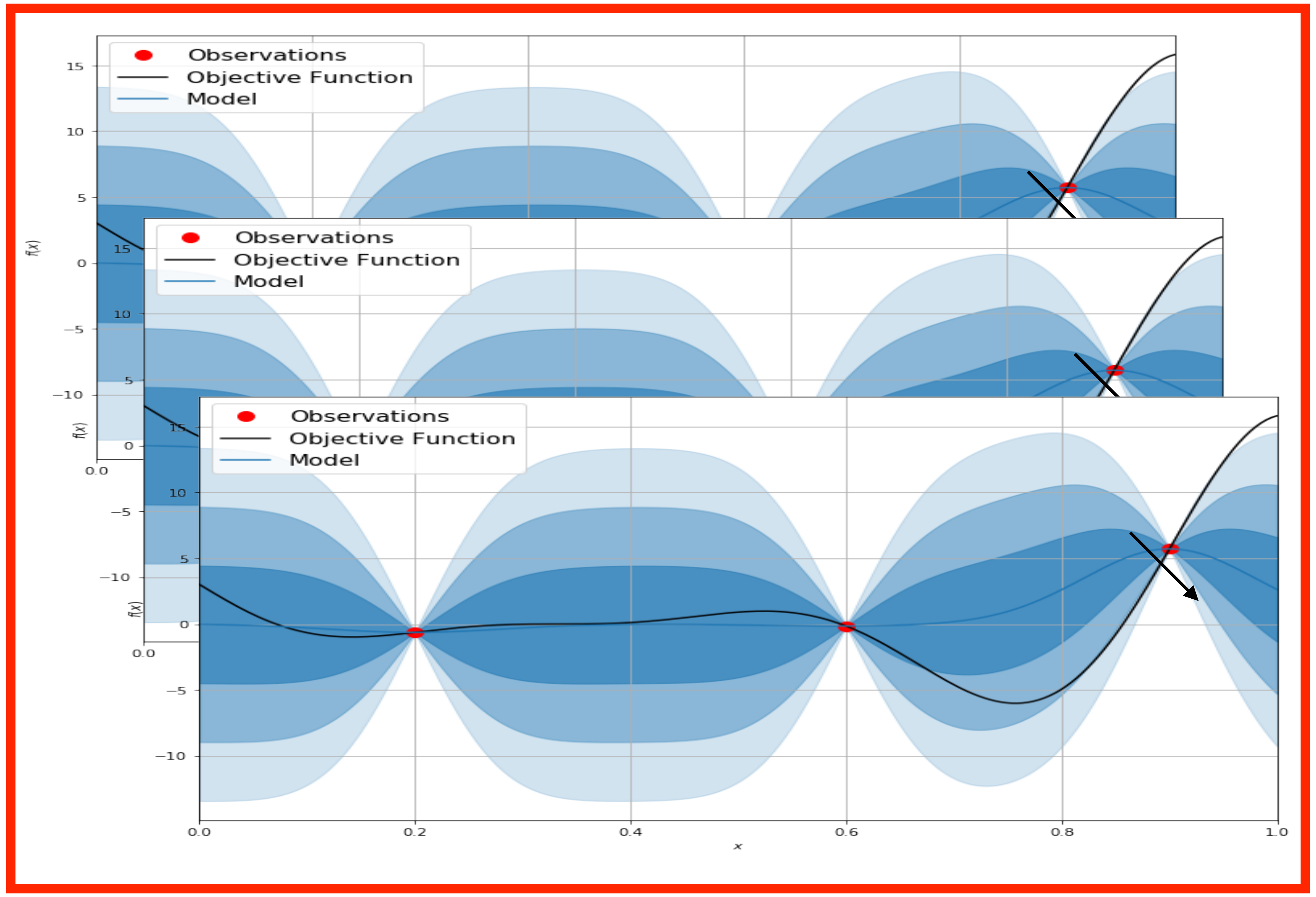
- Such AFs are domain-agnostic

MOBO can also be Interpreted as an RL Problem

2. State = Posterior of multiple functions

1. Env = GP + Black-box functions





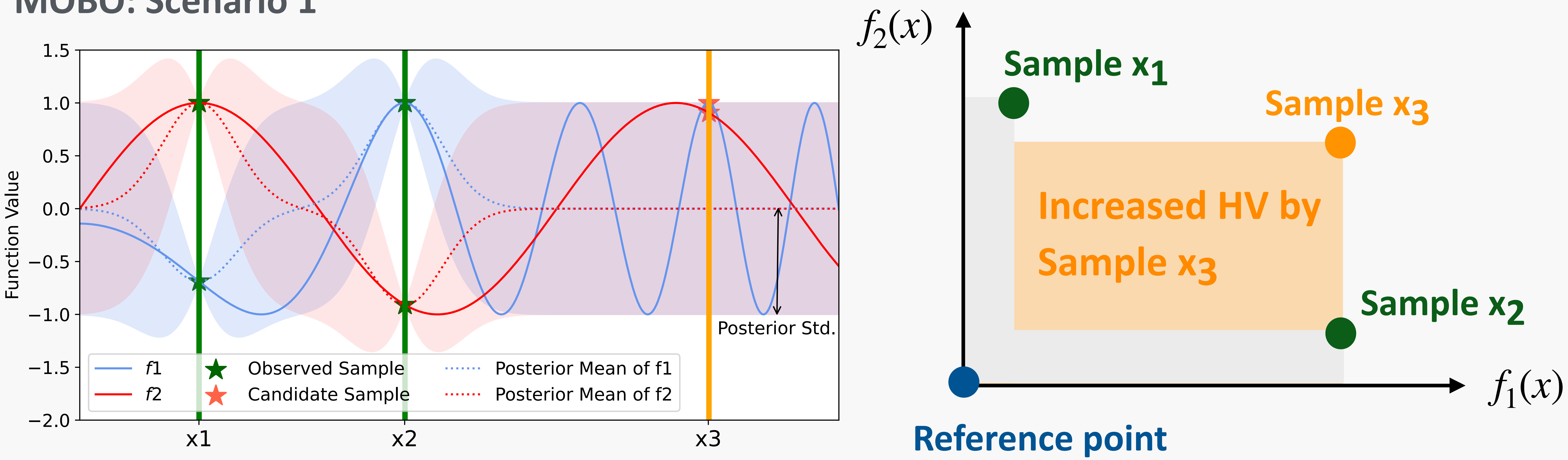
3. Action = next sample point

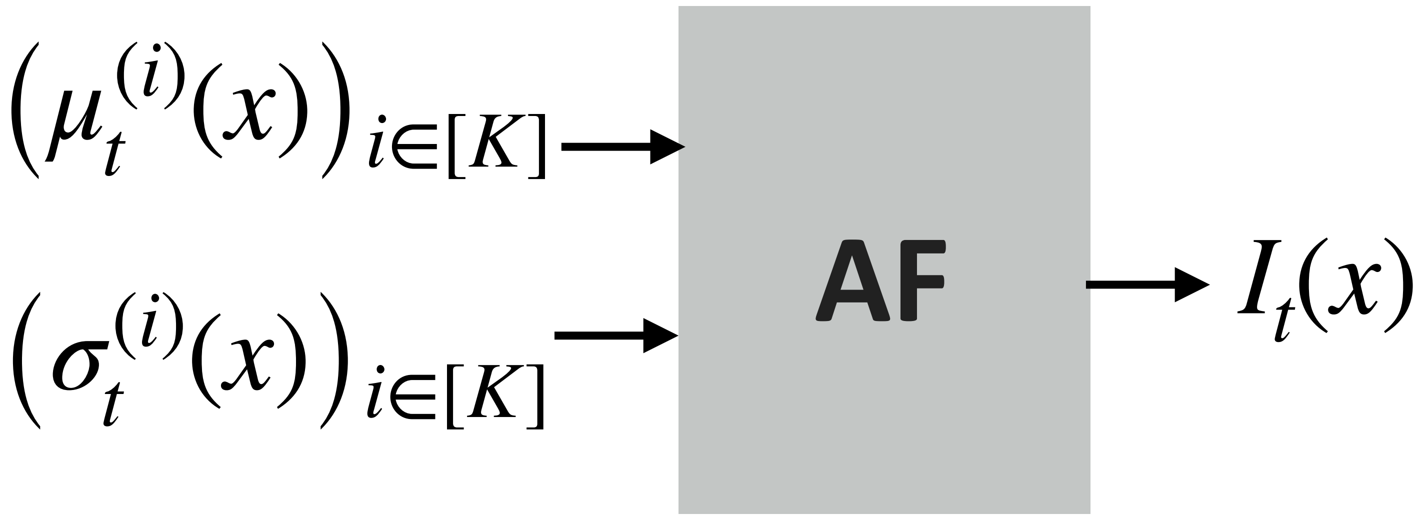
4. Policy determined by AF

5. Reward could be a function of “hypervolume”

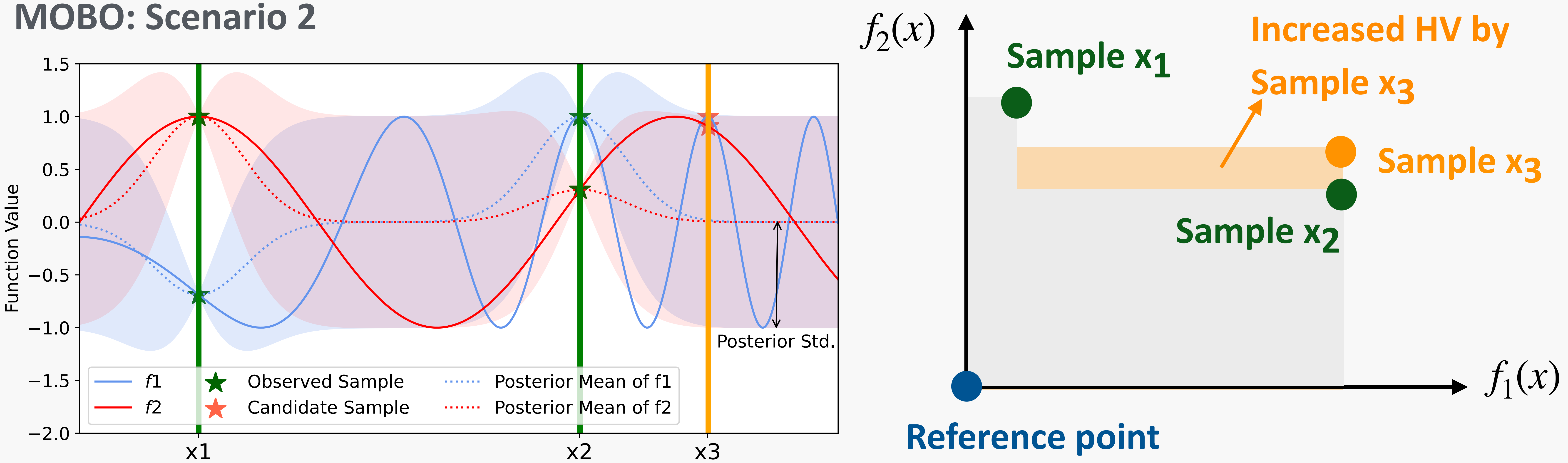
Hypervolume Identifiability Problem

MOBO: Scenario 1





MOBO: Scenario 2



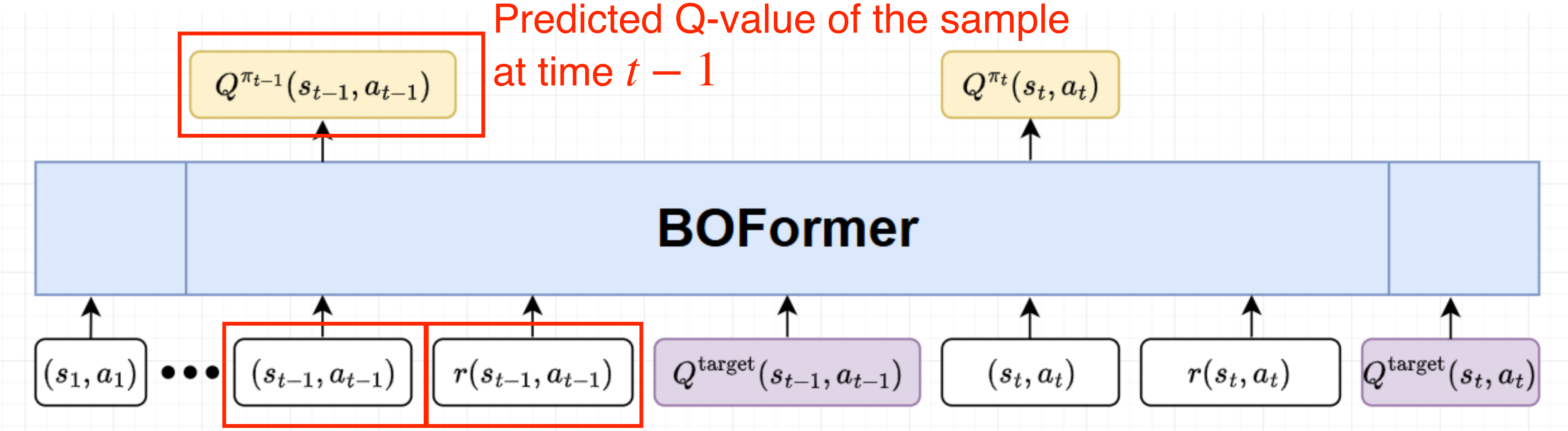
Upon sampling x_3:

- Same posterior for both scenarios
- Large difference in increased HV

BOFormer: Sequential Modeling + Non-Markovian RL

- Main idea #1: Use a Transformer to capture sampling history
- Main idea #2: Use Non-Markovian DQN to learn a Q-Function as AF

Predicted Q-value of the sample at time $t-1$

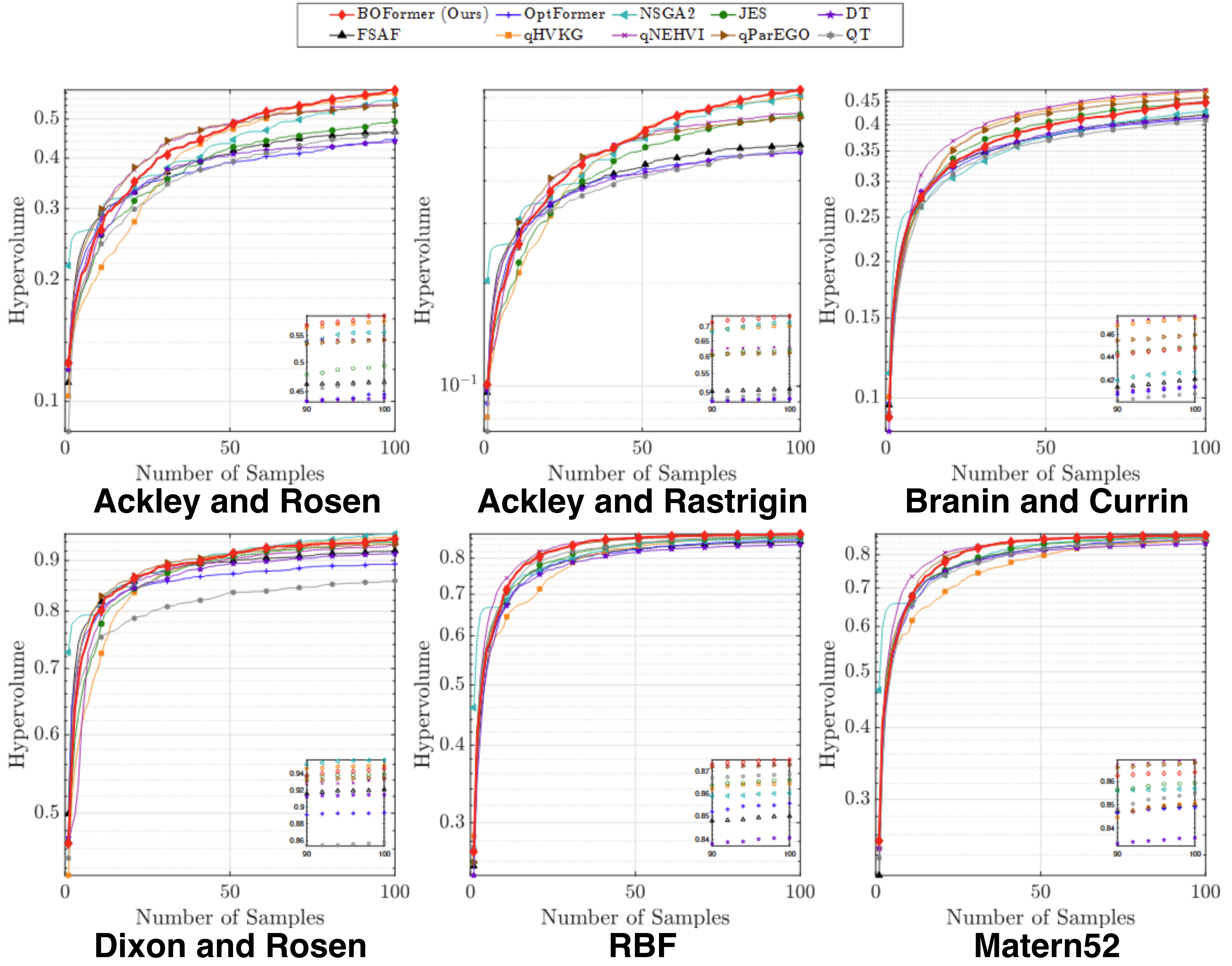


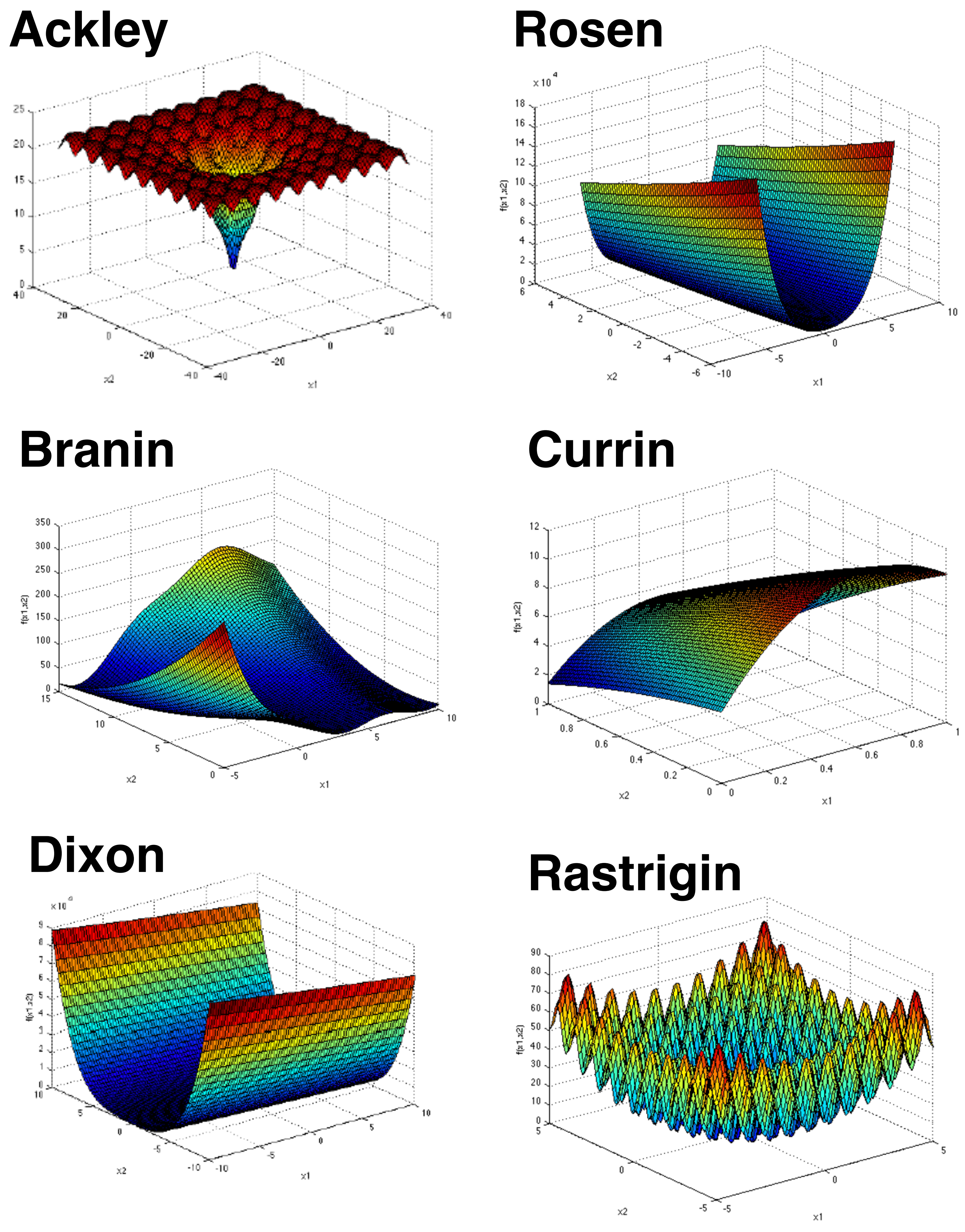
Joint state-action representation of the sample at time $t-1$

Immediate reward (increased HV) incurred by the sample at time $t-1$

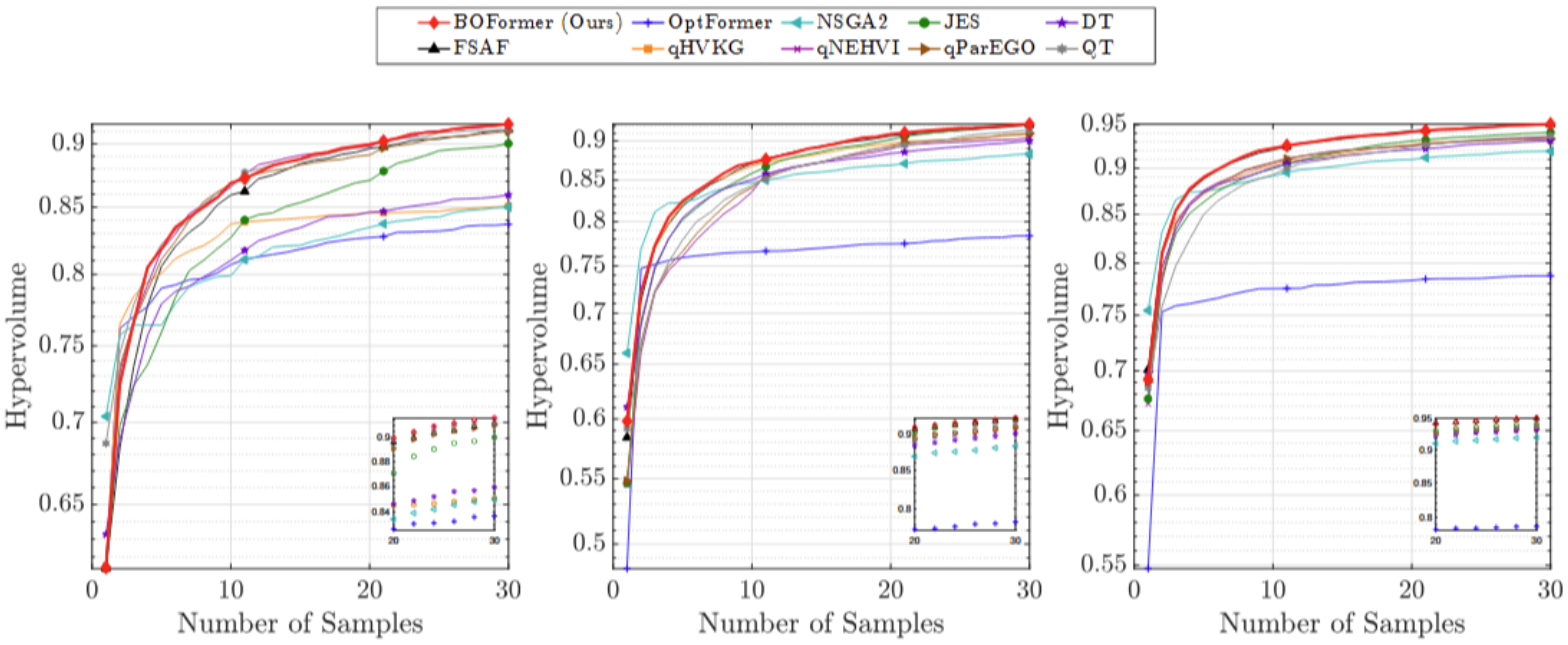
$$L(\theta) := \sum_{\tau \in B} \sum_{i=1}^{T-1} \left(Q_\theta \left(h_i, o_i(x_i) \right) - \left(r_i + \gamma \max_{x \in \mathbb{X}} Q_{\bar{\theta}}(h_{i+1}, o_{i+1}(x)) \right) \right)^2$$

Evaluation: BOFormer *on Synthetic Functions*





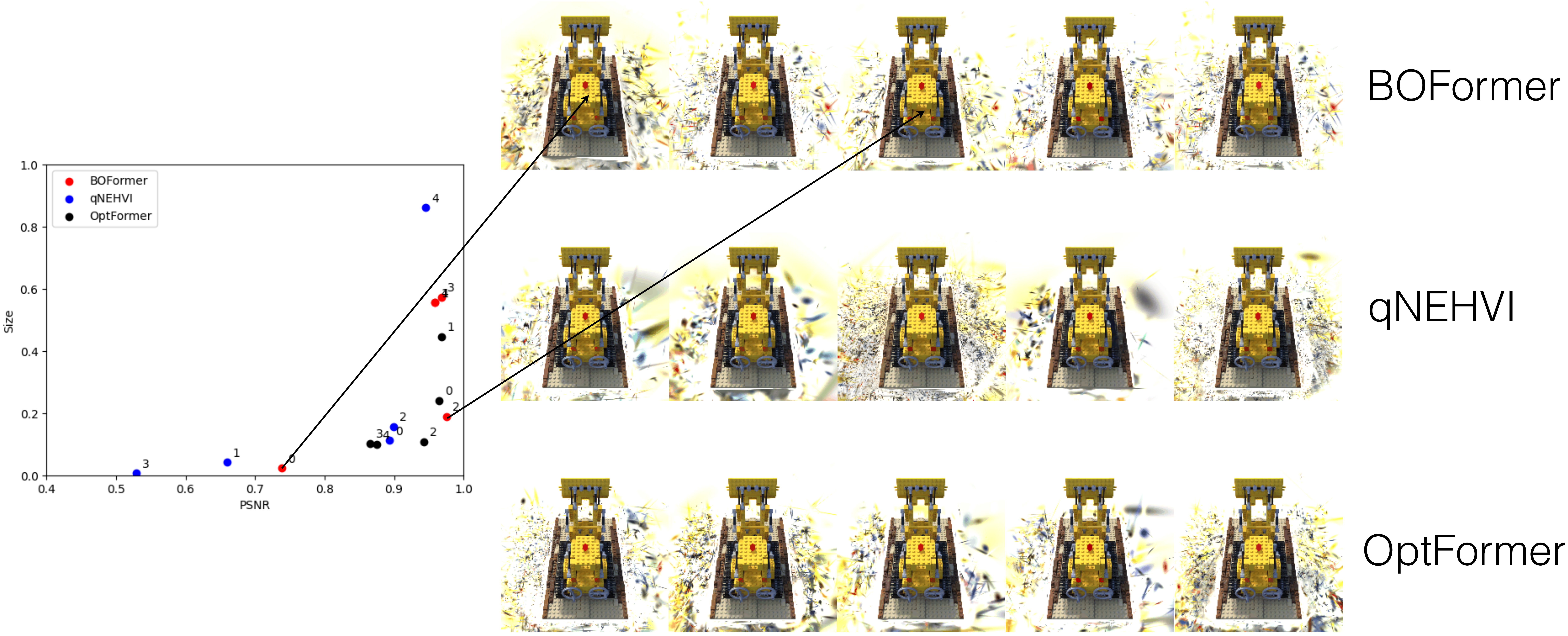
Evaluation: BOFormer *on HPO-3DGS*



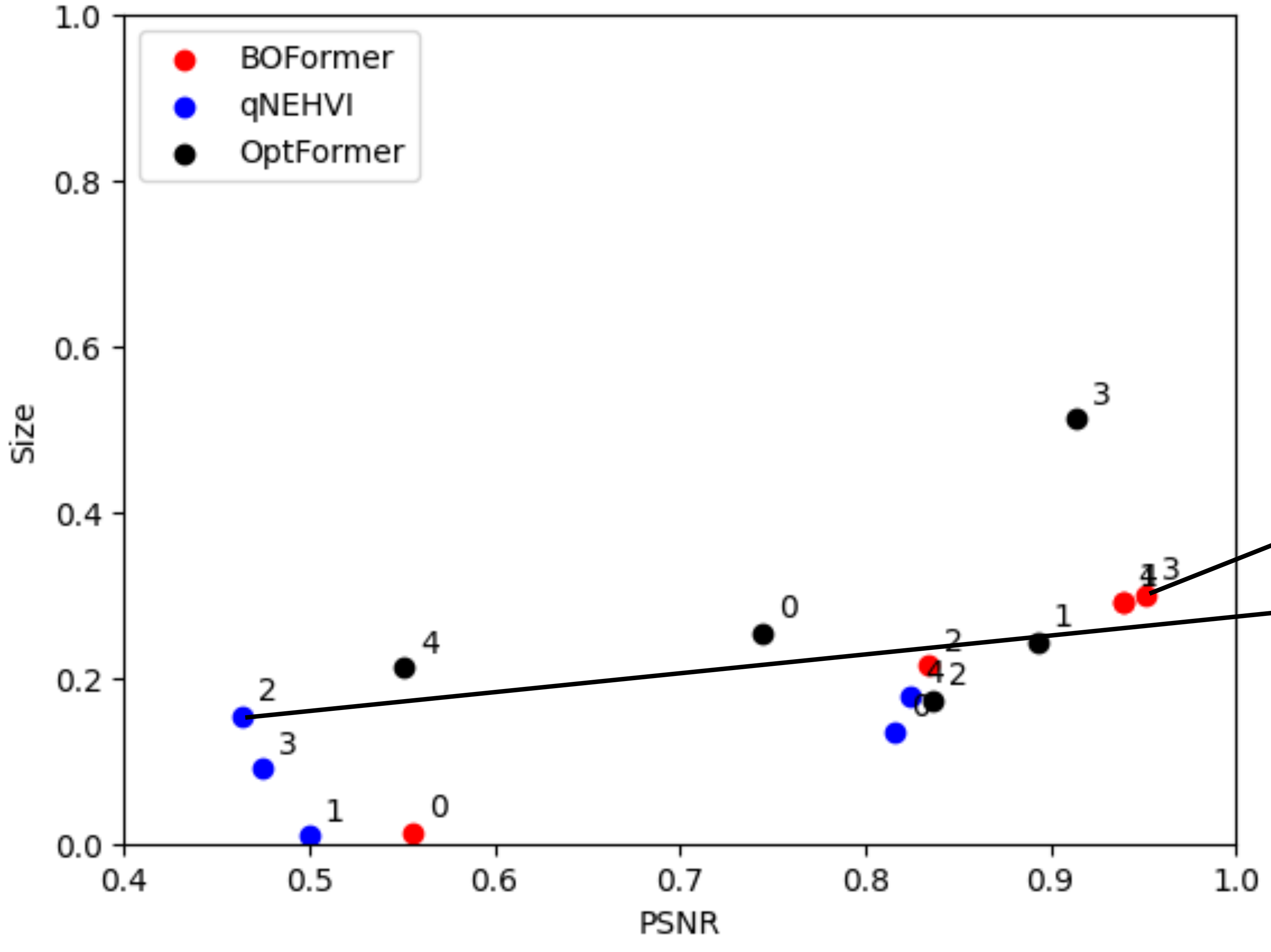
(a) Chairs (b) Materials (c) Lego

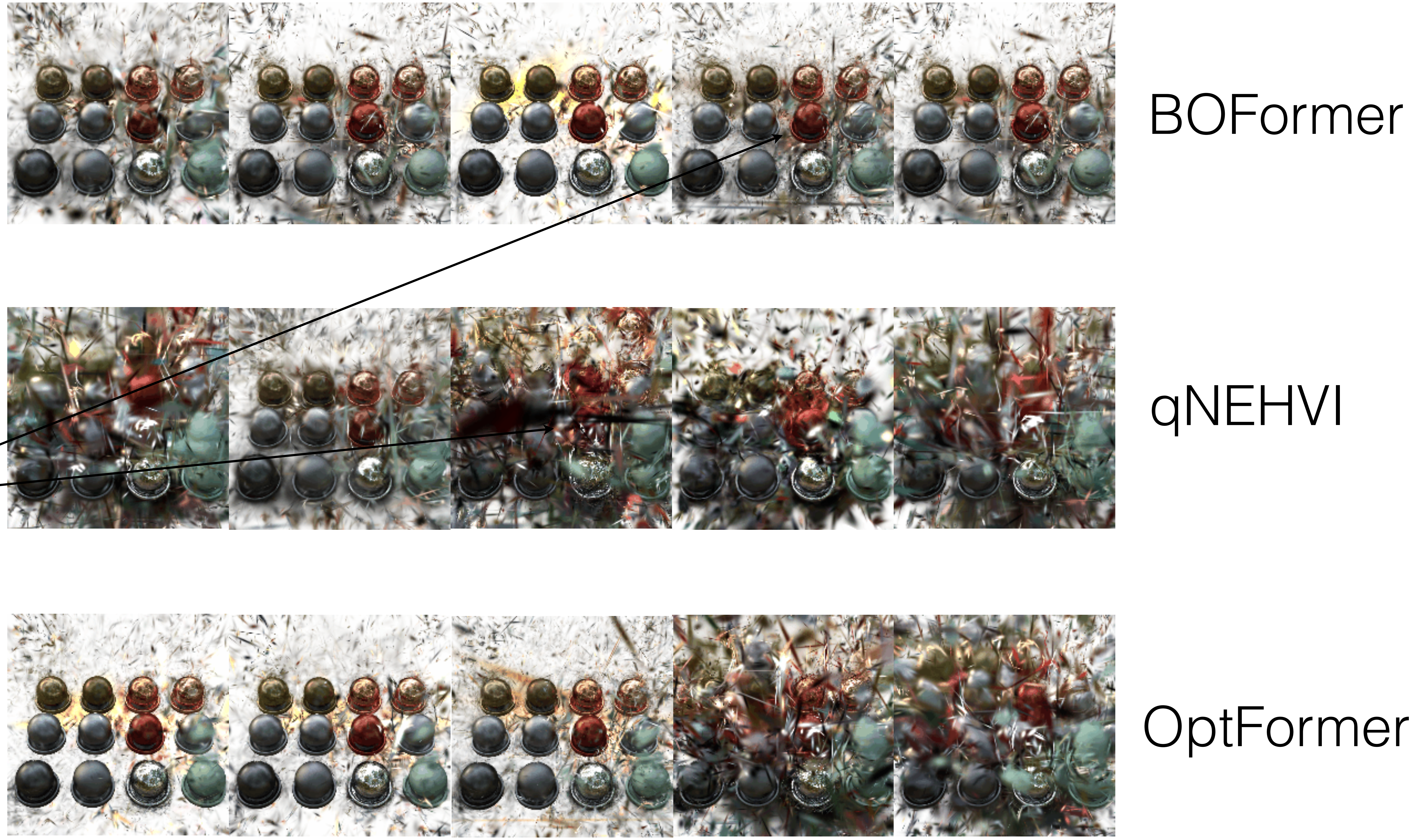
- BOFormer outperforms both the existing rule-based AFs and other Transformer-based RL benchmark methods.
- BOFormer can be trained on GP functions, which are easily to created, and achieve zero-shot transfer to other unseen testing functions.

Demo: BOFormer *on HPO-3DGS - Lego*



Demo: BOFormer *on HPO-3DGS - Materials*





Main Takeaway

- Learning to solve Multi-Objective Bayesian Optimization (MOBO) needs to handle the "*hypervolume identifiability problem*"

- Sequence modeling + Non-Markovian RL can nicely address this identifiability problem