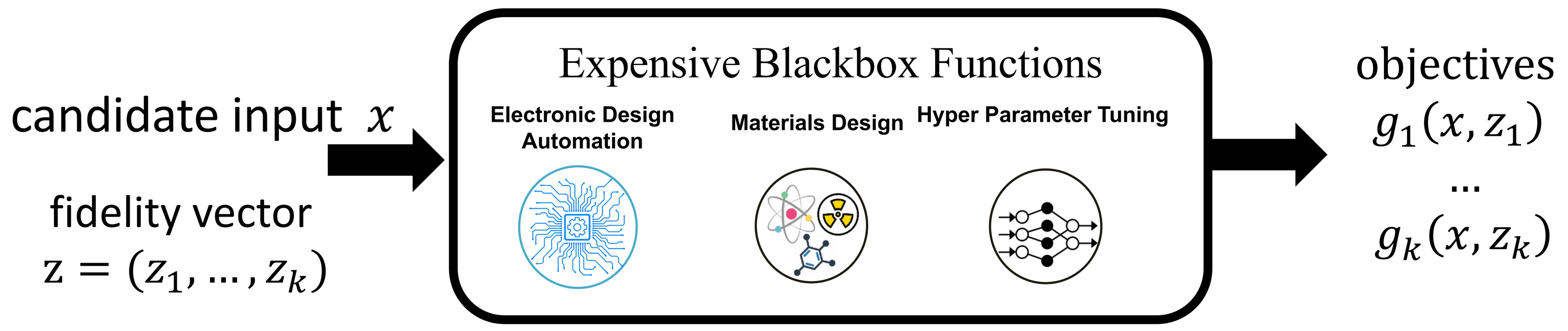
Information-Theoretic Multi-Objective Bayesian Optimization with Continuous Approximations

Washington State University

Syrine Belakaria, Aryan Deshwal, and Jana Doppa

Continuous-Fidelity Multi-Objective BO

❑ Bayesian optimization (BO) is a framework to **maximize expensive black-box functions** using the following elements:



➢ **Statistical models** as a prior for the functions: continuous fidelity Gaussian processes (CFGPs) can provide prediction $\mu(x,z)$ and uncertainty via variance $\sigma(x,z)$ for each fidelity

➢ **Acquisition function** to score the utility of evaluating input x and z

➢ **Optimization procedure** to select the best x and z for evaluation

Prior Work and Our Contributions

❑ **Single-fidelity multi-objective optimization:**

➢ *Scalarization*: relies on random scalars that can be sub-optimal

➢ *Hypervolume improvement*: not scalable for high-dimensional input spaces and large number of objective functions

➢ *Input space entropy-based acquisition function:* maximizes information gain about the optimal Pareto set X^*. Relies on approximating a very expensive and high-dimensional $(m.d)$ distribution over input space

Cannot handle multiple fidelities

Input dimension d

Requires approximation

$$\alpha(\mathbf{x}) = I(\{\mathbf{x},\mathbf{y}\}, \boxed{\mathcal{X}^*} \mid D) = H(\mathbf{y} \mid D, \mathbf{x}) - \boxed{\mathbb{E}_{\mathcal{X}^*}[H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{X}^*)]}$$

❑ **Application-specific multi-fidelity MO optimization**

➢ Single-fidelity methods with low-fidelity as initialization

➢ Strong assumptions that may not hold for general multi-fidelity settings

❑ **Multi-fidelity single-objective optimization**

➢ Acquisition functions to optimize single-objective in multi-fidelity setting has been extensively studied

Our Approach:

❑ **iMOCA algorithm** selects the candidate input $\boldsymbol{x}$ and fidelity vector $\boldsymbol{z}$ for evaluation that maximizes the information gain per unit cost about the optimal **Pareto front Y^***

➢ Relies on a computationally cheap and low-dimensional $m.k \ll m.d$ distribution, where k is the number of objectives

❑ **Key advantages of iMOCA's acquisition function approximations (AF)**

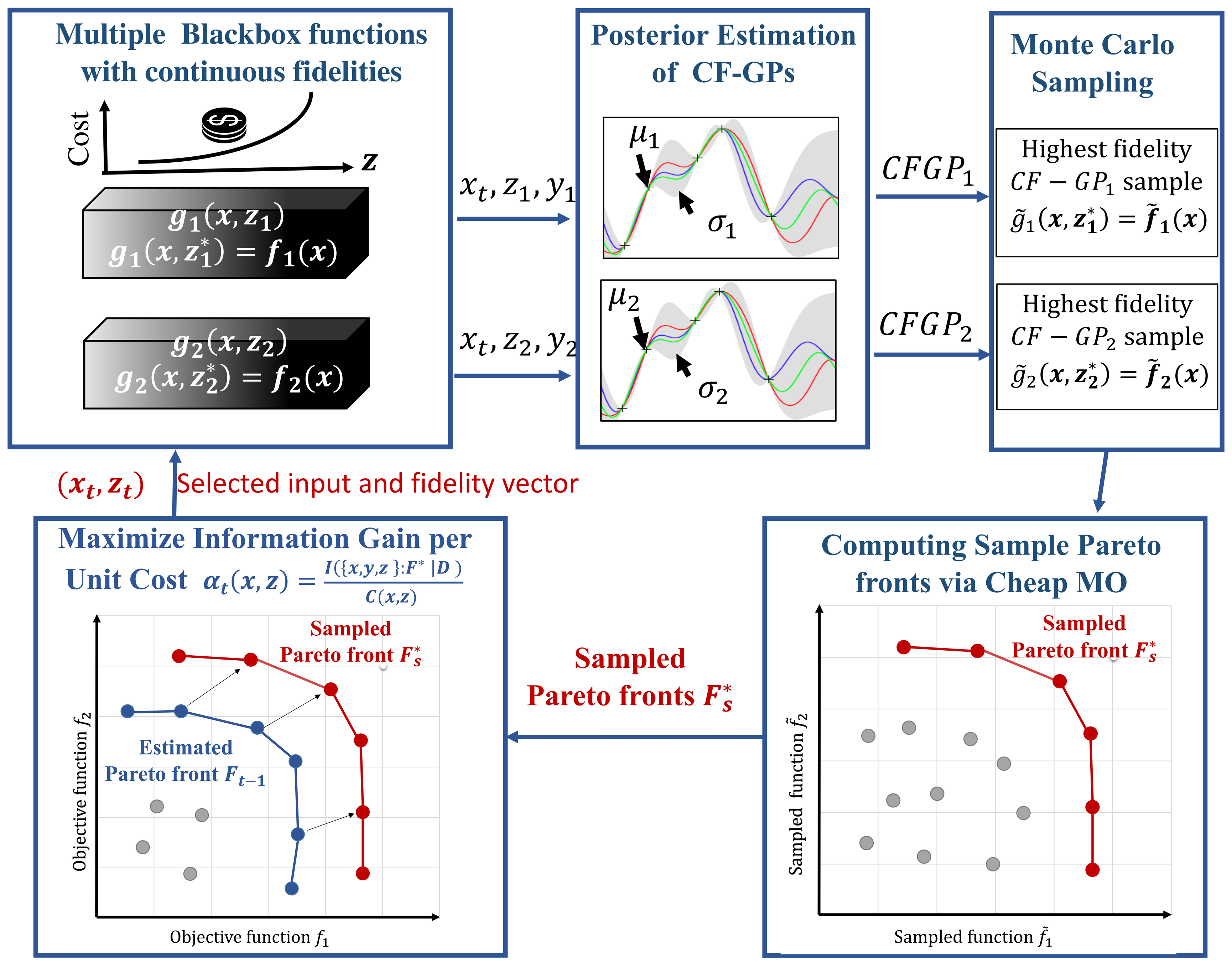
➢ Robust to the number of samples for AF computation

➢ Scalable for high-dimensions via output space entropy search

➢ Two tight approximations: iMOCA-T with closed-form expression and iMOCA-E with numerical integration over one-dimensional variable

➢ Comprehensive experiments to support all the claims

iMOCA Algorithm



❑ **Continuous-fidelity Output space entropy-based AF**

Can handle continuous fidelities

Output dimension $k << d$

Can have a closed form

$$\alpha_t(\mathbf{x},\mathbf{z}) = I(\{\mathbf{x},\mathbf{y},\mathbf{z}\}, \boxed{\mathcal{F}^*} D)/\mathcal{C}(\mathbf{x},\mathbf{z})$$

$$I(\{\mathbf{x},\mathbf{y},\mathbf{z}\}, \mathcal{F}^* | D) = H(\mathbf{y}|D,\mathbf{x},\mathbf{z}) \boxed{- \mathbb{E}_{\mathcal{F}^*}[H(\mathbf{y}|D,\mathbf{x},\mathbf{z},\mathcal{F}^*)]}$$

❑ **iMOCA's Acquisition Function**

➢ **iMOCA-T: Truncated Gaussian approximation**

Closed-form

$$\alpha_t(\mathbf{x},\mathbf{z},\mathcal{F}^*) = \frac{1}{\mathcal{C}(\mathbf{x},\mathbf{z})S}\sum_{j=1}^{K}\sum_{s=1}^{S}\frac{\gamma_s^{(g_j)}\phi(\gamma_s^{(g_j)})}{2\Phi(\gamma_s^{(g_j)})} - \ln(\Phi(\gamma_s^{(g_j)}))$$

➢ **iMOCA-E: Extended-skew Gaussian approximation**

Tight approximation

$$\alpha_t(\mathbf{x},\mathbf{z},\mathcal{F}^*) = \frac{1}{\mathcal{C}(\mathbf{x},\mathbf{z})S}\sum_{j=1}^{K}\sum_{s=1}^{S}\tau^2\frac{\gamma_s^{(f_j)}\phi(\gamma_s^{(f_j)})}{2\Phi(\gamma_s^{(f_j)})} - \ln(\Phi(\gamma_s^{(f_j)})) + \mathbb{E}_{u\sim\Gamma_{f_s^{j*}}}[\ln(\Phi(\frac{\gamma_s^{(f_j)} - \tau u}{\sqrt{1-\tau^2}}))]$$

where $\gamma_s^{(g_j)} = \frac{f_s^{j*} - \mu_{g_j}}{\sigma_{g_j}}$, $\boldsymbol{\phi}$ and $\boldsymbol{\Phi}$ are the p.d.f and c.d.f of a standard normal distribution respectively

❑ **Approach to Reduce Fidelity Search Space**

➢ **The continuity of the fidelity space results in infinite number of choices**

➢ **To select an informative and meaningful fidelity we reduce the search space of fidelities for each function as follow:**

$$\mathcal{Z}_t^{(j)}(\mathbf{x}) = \{\{z_j \in \mathcal{Z}_{\setminus\{z_j^*\}}, \sigma_{g_j}(\mathbf{x}, z_j) > \gamma(z_j), \xi(z_j) > \beta_t^{(j)}\|\xi\|_\infty\} \cup \{z_j^*\}\}$$

Where $\gamma(z_j) = \xi(z_j)(\frac{\mathcal{C}_j(\mathbf{x},z_j)}{\mathcal{C}_j(\mathbf{x},z_j^*)})^q$, $q = \frac{1}{p_j+d+2}$; $\xi(z_j) \approx \frac{\|z_j - z_j^*\|}{h_j}$

Experiments and Results

❑ **Evaluation metrics**

➢ *The hypervolume difference:* between optimal Pareto front (PF) Y^* and the best uncovered PF estimated by optimizing the posterior mean of the models at the highest fidelities

➢ R_2 Indicator: the average distance between two pareto fronts

➢ *Cost reduction factor:* gain in the cost compared to the best performing baseline

❑ **iMOCA vs. State-of-the-art**

➢ iMOCA consistently performs better than all baselines. Both the variants of iMOCA converge at a much lower cost.

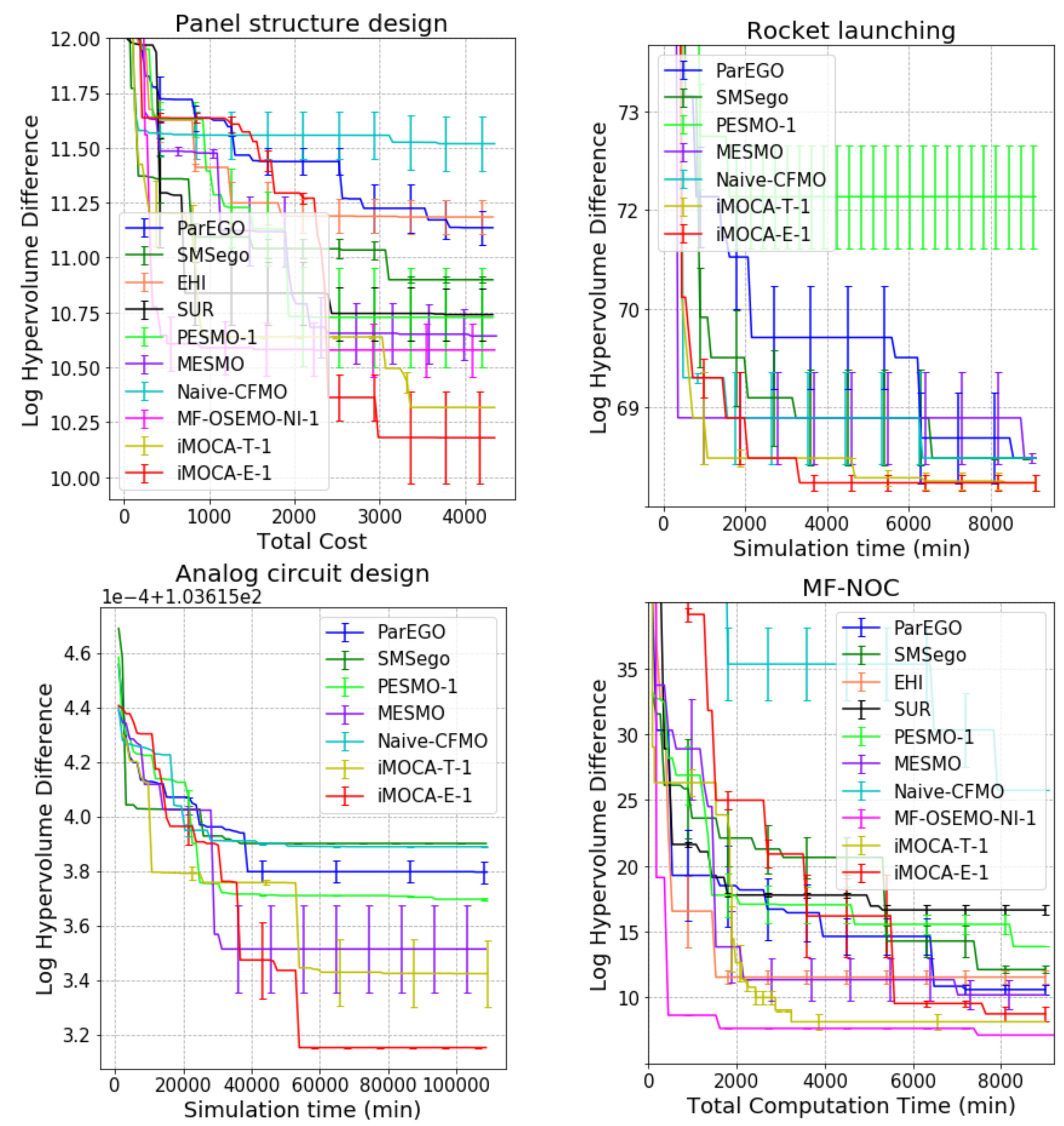
➢ Even with loose approximation, iMOCA-T provides consistently competitive results

❑ **Robustness of iMOCA**

➢ iMOCA is robust to the number of samples compared to PESMO: *maintains better performance consistently even with a single sample!*

❑ **Cost reduction factor**

➢ Although the metric gives advantage to baselines, the results in the table show a consistently high gain ranging from 52% to 85 %



Name	BC	ARS	Circuit	Rocket
$\mathcal{C}_B$	200	300	115000	9500
$\mathcal{C}$	30	100	55000	2000
$\mathcal{G}$	85%	66.6%	52.1%	78.9%

Table: *Best* convergence cost from all baselines $\mathcal{C}_B$, *Worst* convergence cost for iMOCA $\mathcal{C}$, and cost reduction factor $\mathcal{G}$.