# Data Fusion of Multi-Fidelity Systems via Latent Variable Gaussian Process for Active Learning

**Yi-Ping Chen*, Liwei Wang, Yigitcan Comlek, Wei Chen**

*Department of Mechanical Engineering, Northwestern University*

*Integrated DEsign Automation Laboratory (IDEAL)*

2nd MMLDE-CSET, Sep. 25, 2023, El Paso, TX

IDEAL
Integrated DEsign
Automation Laboratory

Northwestern University

# Outline

1. Introduction
2. LVGP & MF-LVGP
3. Multi-Fidelity Adaptive Sampling
4. Case Studies
5. Closure

# Outline

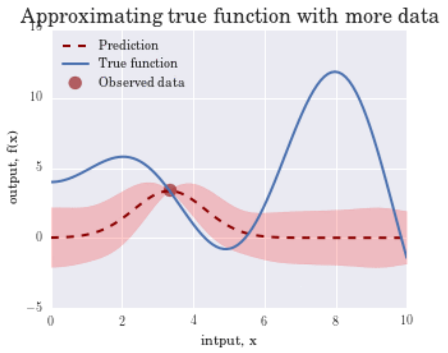
1. Introduction
2. LVGP & MF-LVGP
3. Multi-Fidelity Adaptive Sampling
4. Case Studies
5. Closure

# Why do We Need Multi-Fidelity Adaptive Sampling?

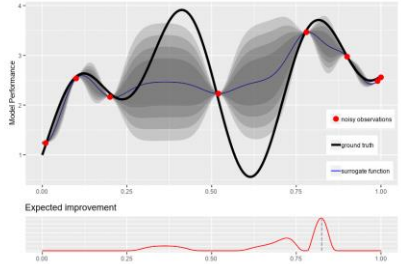
- Simulation cost of complex engineering system is expensive
- Building a surrogate model is a solution, while collecting sufficient data is a challenging task

## Adaptive Sampling

### Global Fitting (GF)




### Bayesian Optimization (BO)




## Multi-Fidelity (MF) Data Fusion

(LF)

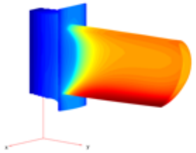
**Fidelity Spectrum**

(HF)

Low-fidelity models ⟷ High-fidelity models

- **Coarse** physical resolutions

- **Fast** runtime



- **Fine** physical resolutions

- **Slow** runtime

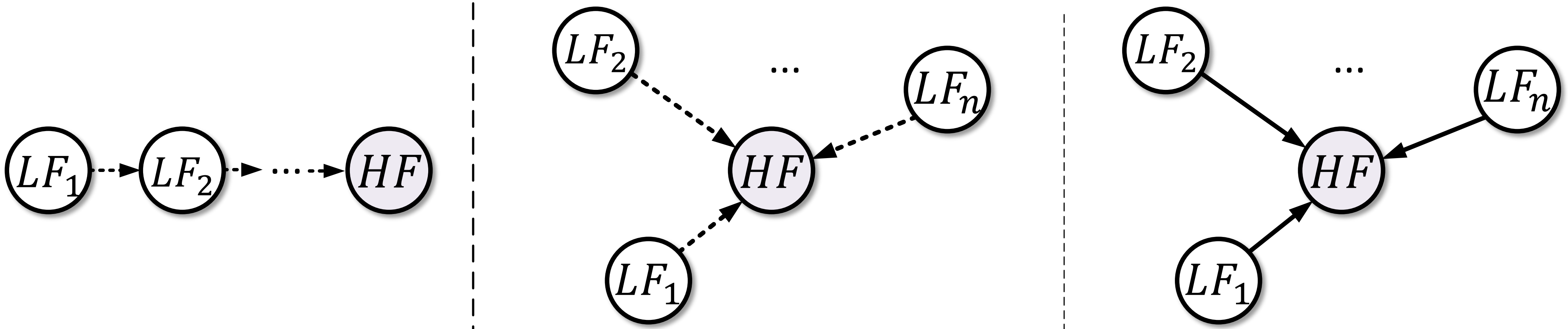


https://gist.github.com/ilanman/312d0489763b9c19164a, https://towardsdatascience.com/bayesian-hyperparameter-optimization-17dc5834112d

# Categorizing MF Methods via Architecture

$$HF = HF|LF_n| \dots |LF_2|LF_1$$

$$HF = HF|(LF_n, \dots, LF_2, LF_1)$$




(a) Hierarchical Architecture

(b) Non-hierarchical architecture with implicit use of correlation

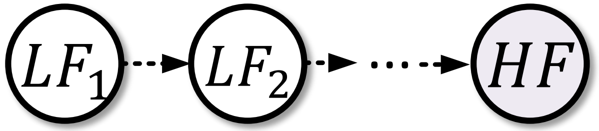
(c) Non-hierarchical architecture with explicit use of correlation

- Ranking fidelity models may not be practical
- Uncertainty propagation, not desirable for adaptive sampling

- **Implicit:** The correlation is only used for *response/uncertainty* prediction via *existing* data
- **Explicit:** The correlation is to quantify the *benefit* of the *future* samples as the sampling criteria

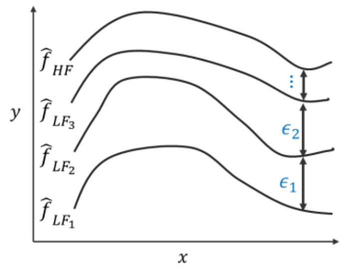
# MF Methods with Hierarchical Architecture

$$HF = HF|LF_n| \dots |LF_2|LF_1$$



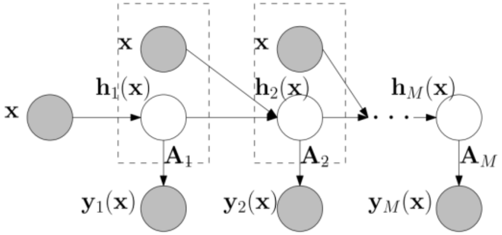

- The surrogate of the higher-fidelity models are conditioning on the information of the lower-fidelity models.
- Fidelity levels should be clearly ranked.
- Uncertainty propagates while the MF model is trained sequentially.

## Kennedy & O'Hagan



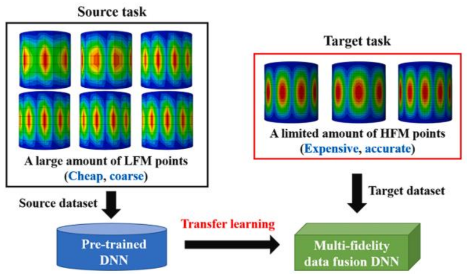

Kennedy,, and O'Hagan, A.. *Biometrika* 87.1 (2000)

## Deep Neural Network



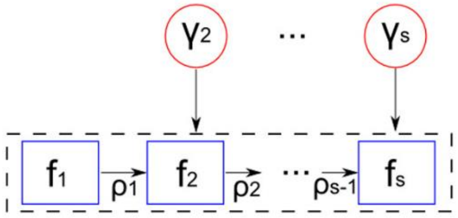

Li, S., Kirby, R. M., & Zhe, S., *Deep Multi-Fidelity Active Learning of High-dimensional Outputs,* 2020.

## Transfer Learning



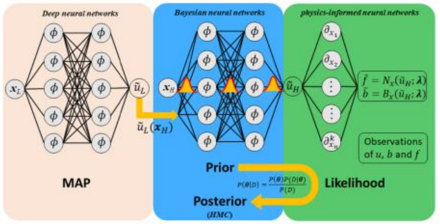

Li, Z., Zhang, S., Li, H., Tian, K., Cheng, Z., Chen, Y., & Wang, B. *Advanced Engineering Informatics*, 2022

## CoKriging



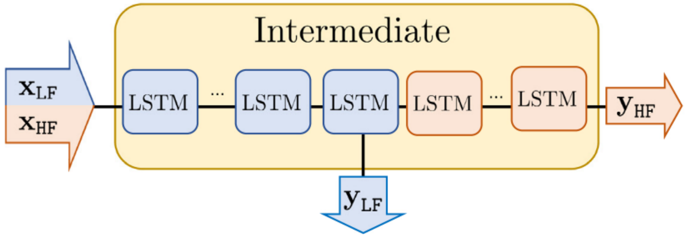

Le Gratiet, L., & Cannamela, C. *Technometrics*, *57*(3), 2015.

## Bayesian Neural Network




Meng, X., Babaee, H., & Karniadakis, G. E., Journal of Computational Physics, 2021.

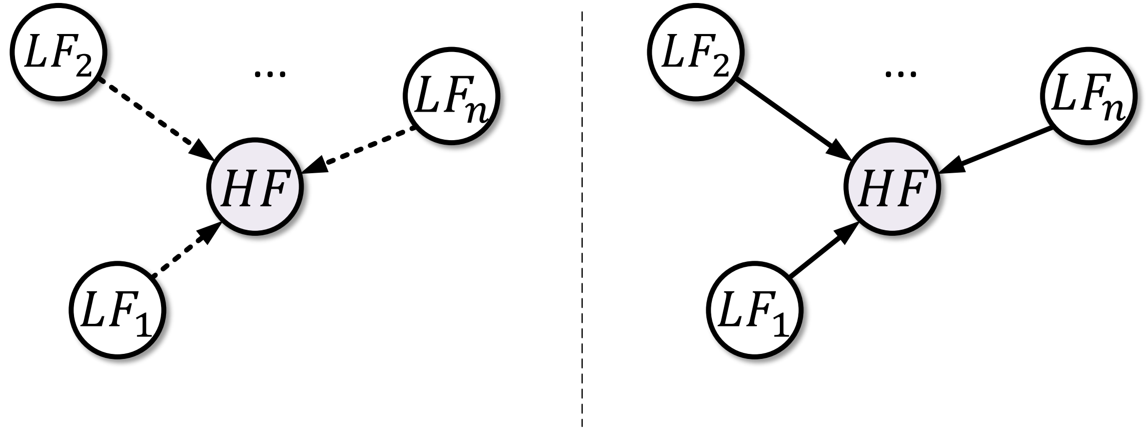
## LSTM




Conti, P., Guo, M., Manzoni, A., & Hesthaven, J. S., CMAME, 2022

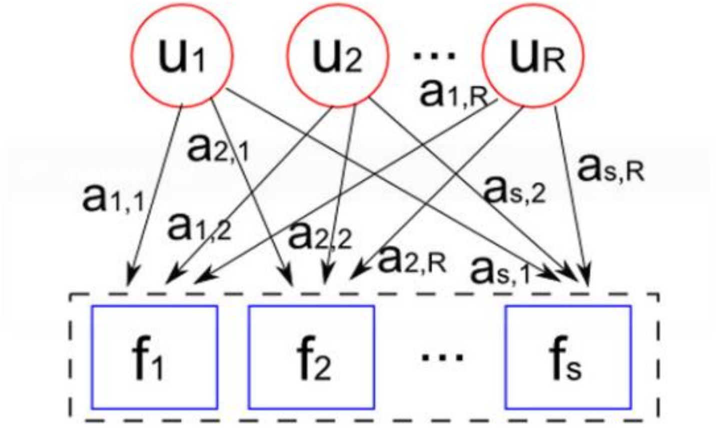
# Non-Hierarchical Architecture – MF Data Fusion

$$HF = HF|(LF_n, \dots, LF_2, LF_1)$$



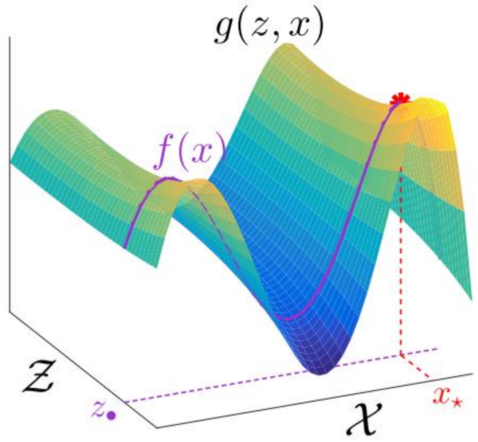

- HF surrogate conditions on all the LF samples simultaneously.
- Rank of fidelity may not be required.
- Latent methods provide better interpretability of correlation.

## Multi-Response Gaussian Process

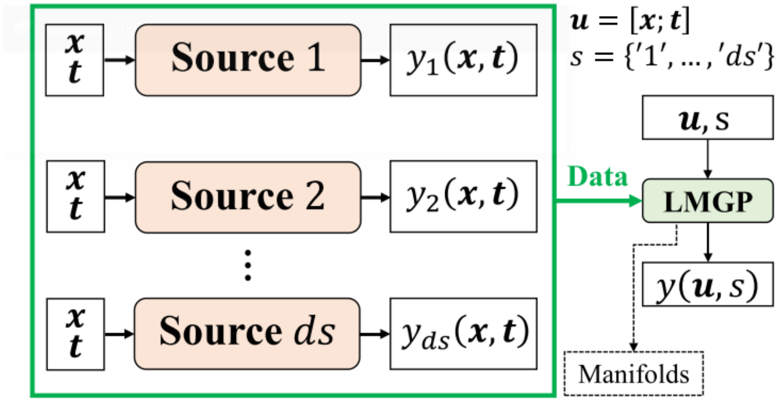



Brevault, L., Balesdent, M., & Hebbal, A., *Aerospace Science and Technology*, 2020

## Latent Representation



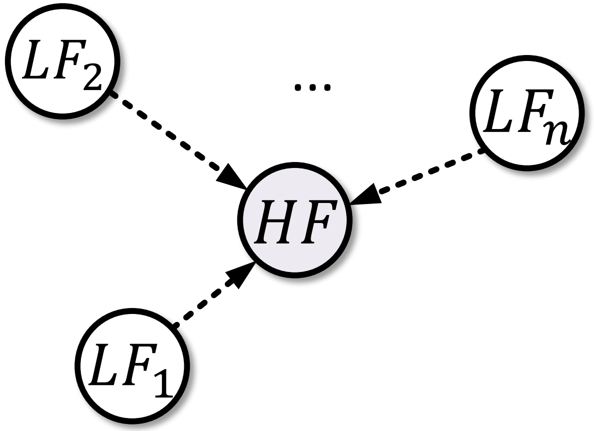

Kandasamy, K., Dasarathy, G., Schneider, J., & Póczos, B., 2017 *ICML.*




Zanjani Foumani, Z., Shishehbor, M., Yousefpour, A., & Bostanabad, R. CMAME, 2023.

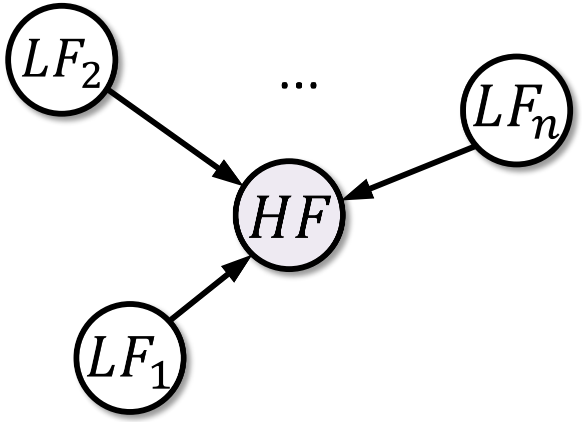
# Non-Hierarchical Architecture – Infill Sampling

## Implicit use of Correlation




- Assign separate acquisition function to different source, *neglect the interaction* between HF and LF models.
- The *contribution/benefit* of the infill LF samples to HF model is *unknown*.
- Less efficient resource allocation and robustness.

## Explicit use of Correlation




- Treat fidelity sources as the subregion of the same MF model and *include the interaction* between fidelities.
- *Quantify the benefit* before querying the infill sample.
- Allocate resource only for improving the performance of HF surrogate.

# Objectives of This Work

- Perform a non-hierarchical, interpretable MF data fusion methods.

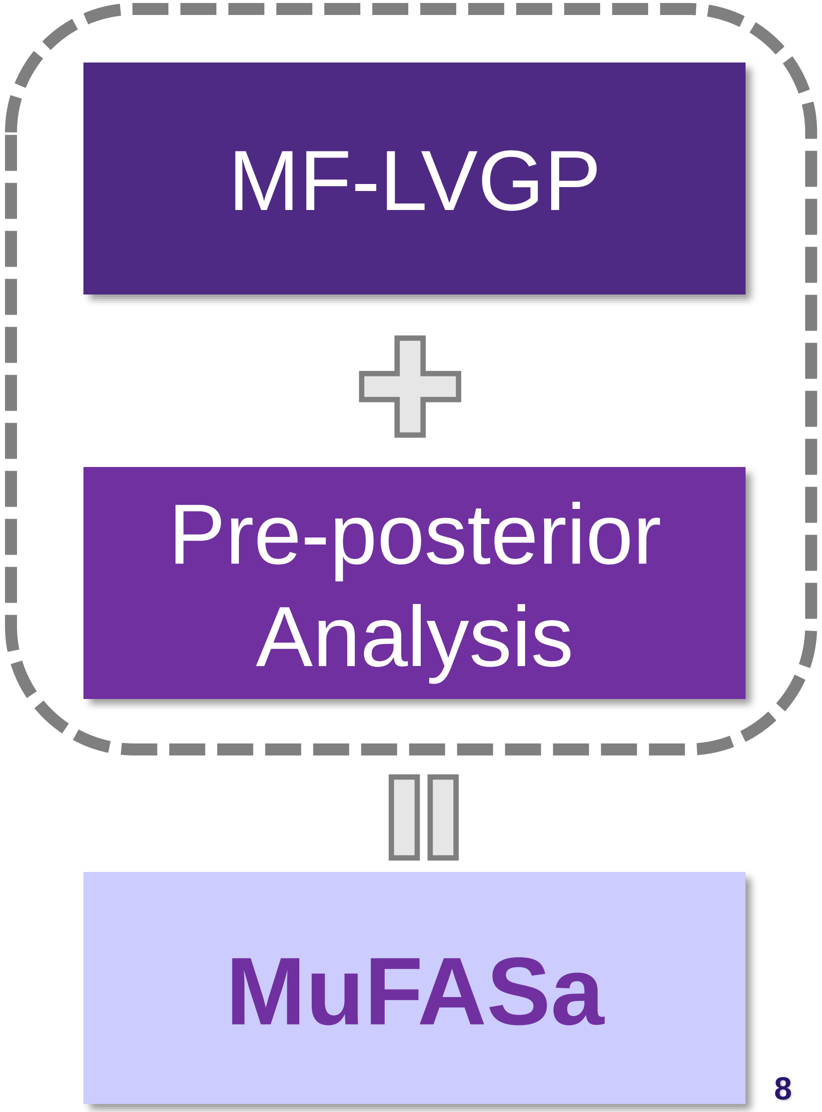
  *We perform MF data fusion with Latent Variable Gaussian Process (LVGP)*

- Quantify the future benefit of the candidate samples.

  *We implement the pre-posterior analysis on MF-LVGP to quantify the benefit*

- Propose an adaptive sampling framework that explicitly utilize the correlation.

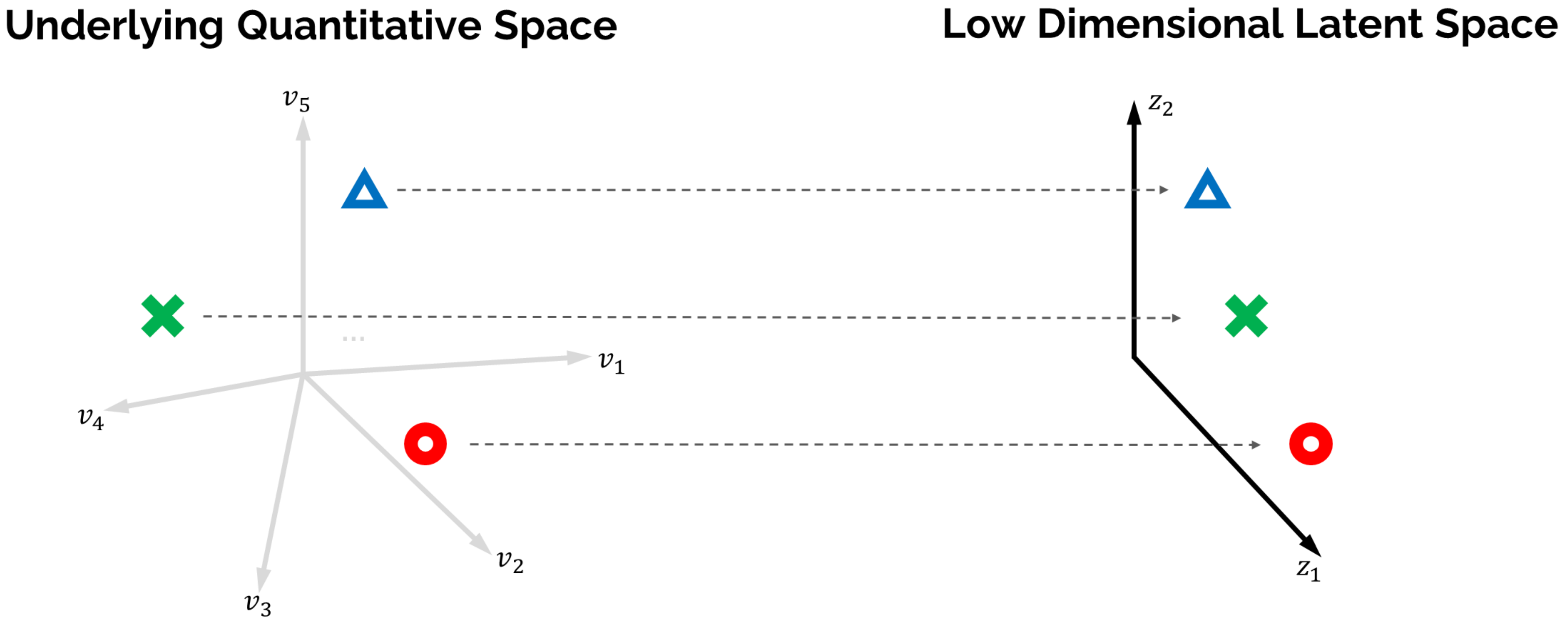
  *We proposed a flexible, benefit & cost-aware adaptive sampling framework*



# Outline

# Lightning Introduction of
# Latent Variable Gaussian Process
# (LVGP)

# Visualization of LVGP




**Qualitative Variables:**
It is known that every qualitative variable has an underlying, possibly high dimensional quantitative representation that influences the response.

**What LVGP does:**
Through physics-based dimension reduction, the numerical space can be reduced to lower latent spaces.


Zhang, Y., Tao, S., Chen, W., & Apley, D. W. (2020). A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors. *Technometrics*, *62*(3), 291–302.
Zhang, Y., Apley, D. W., & Chen, W. (2020). Bayesian Optimization for Materials Design with Mixed Quantitative and Qualitative Variables. *Scientific Reports*, *10*(1).

# Consider an example…

$$f(x_1, x_2) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10$$

Quantitative Variable

$$x_1 = [-5,10]$$

Qualitative Variable

$$x_2 = \{a, b, c, d, e\}$$

Training samples be like…

$$\mathbf{x}^1 = (3, a), \mathbf{x}^2 = (-1, b) \dots.$$

# *Training* with Ordinary Gaussian Process

$\{\mathbf{x}, \mathbf{y}\}$

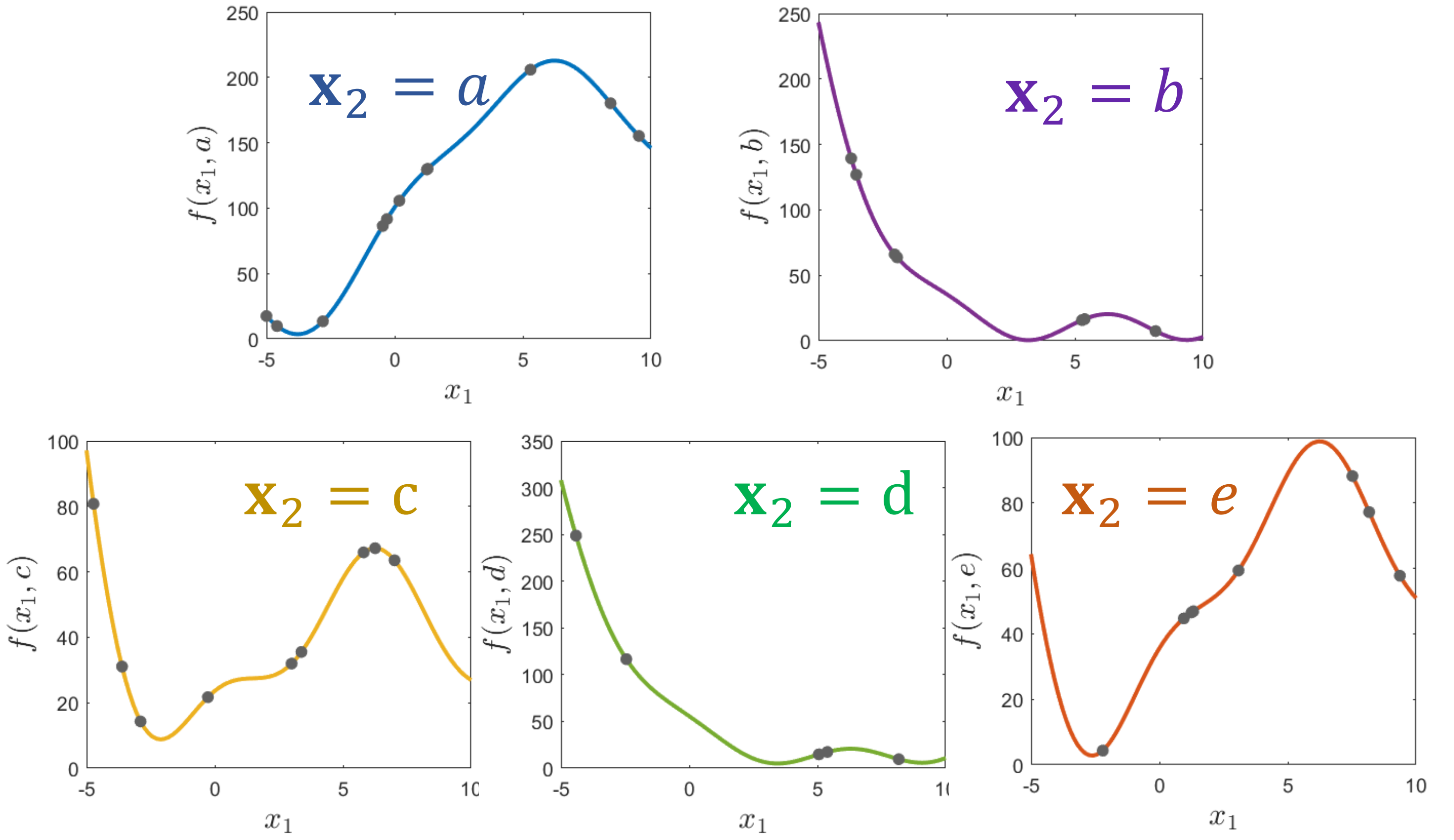
$GP(\mathbf{x}, \mathbf{y} | \mathbf{x}_2 = a)$

$GP(\mathbf{x}, \mathbf{y} | \mathbf{x}_2 = b)$

$GP(\mathbf{x}, \mathbf{y} | \mathbf{x}_2 = c)$

$GP(\mathbf{x}, \mathbf{y} | \mathbf{x}_2 = d)$

$GP(\mathbf{x}, \mathbf{y} | \mathbf{x}_2 = e)$



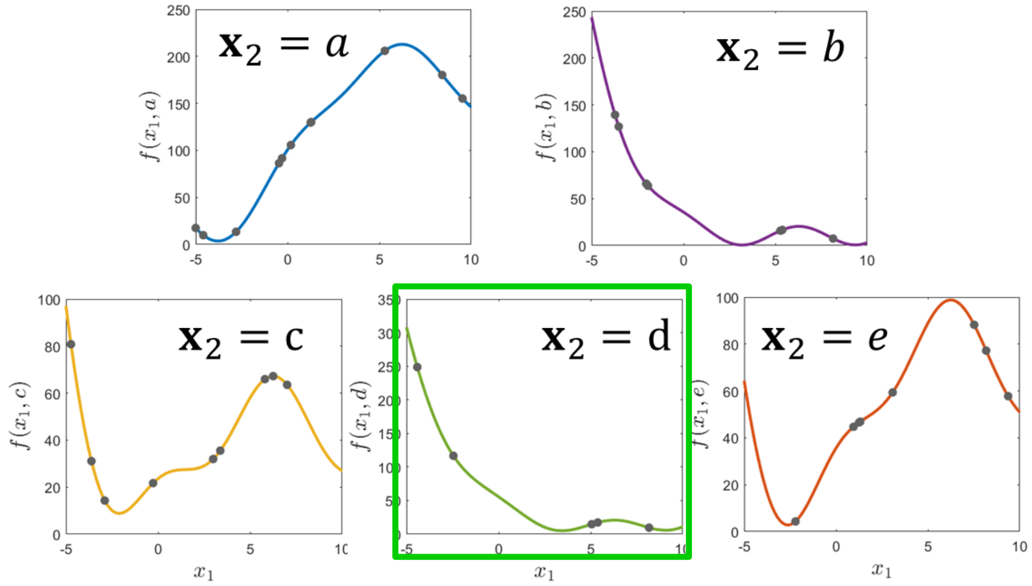

Training Samples

Fit separate GP model for each category

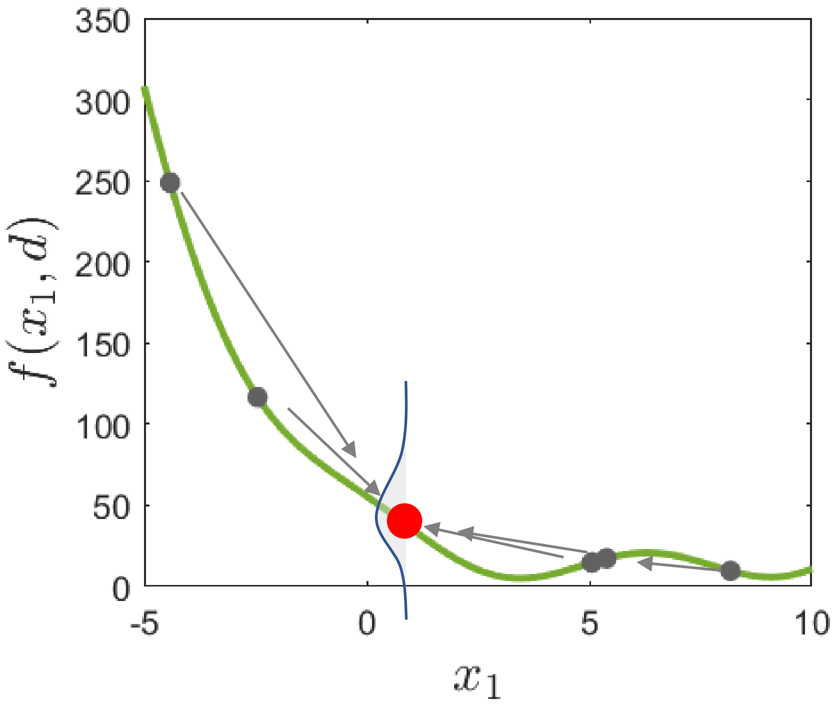
Response for each category

# *Predicting* with Ordinary Gaussian Process

$f(1, d)$ ⇨




⇨



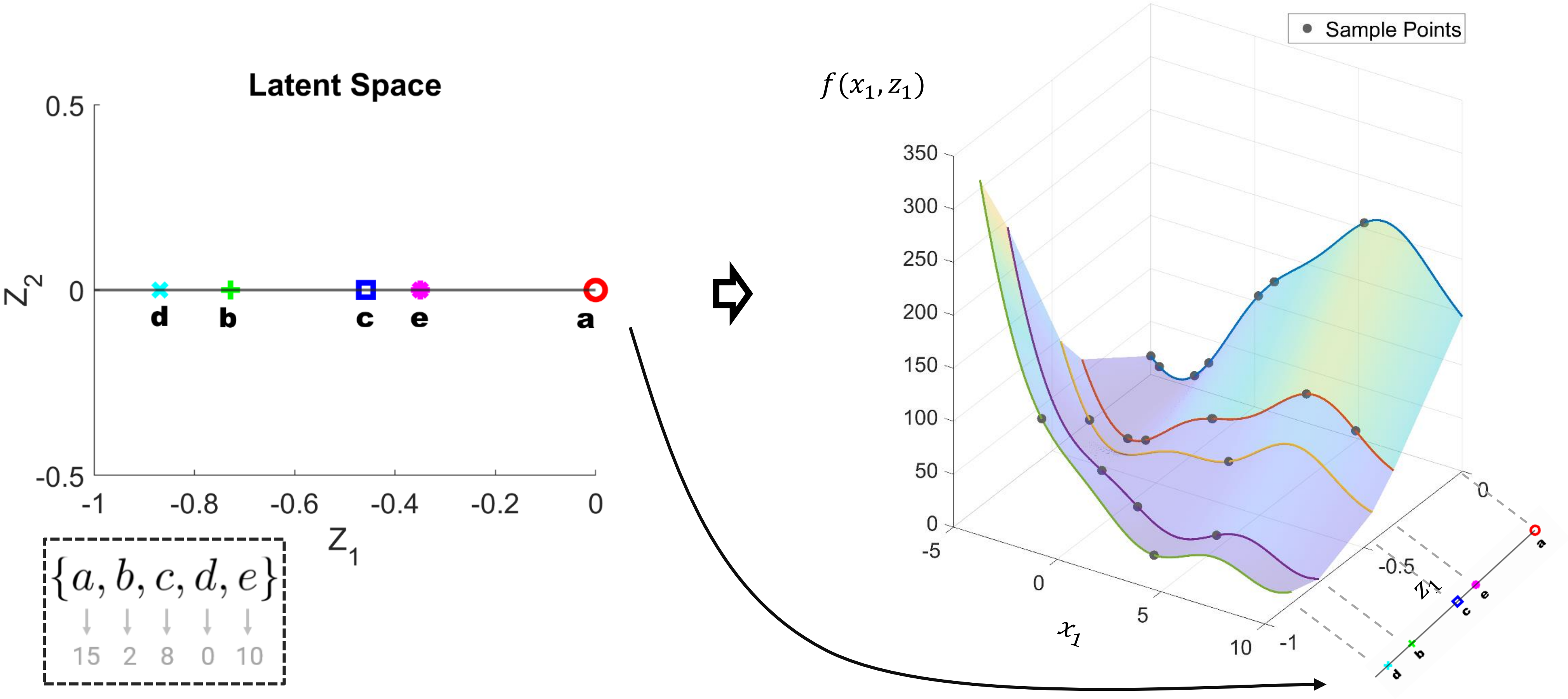

Predicting Target

Select corresponding response model

Response prediction is conditioned to the samples within the *same category*

# *Training* with LVGP

{x, y} ⇨



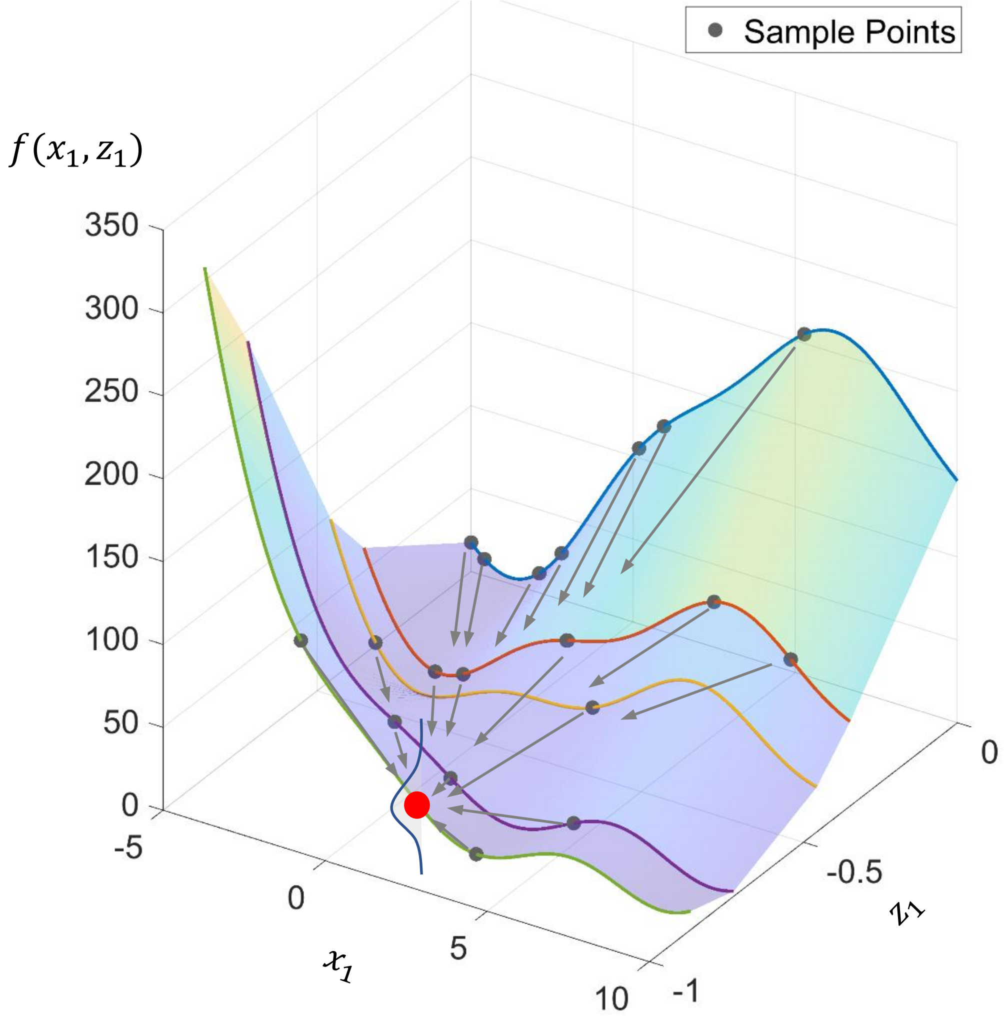

Training Samples

Map qualitative variables into latent space

Generate an output space with quantitative variables and latent variables

# *Predicting* with LVGP

$f(1, d)$ ⇨

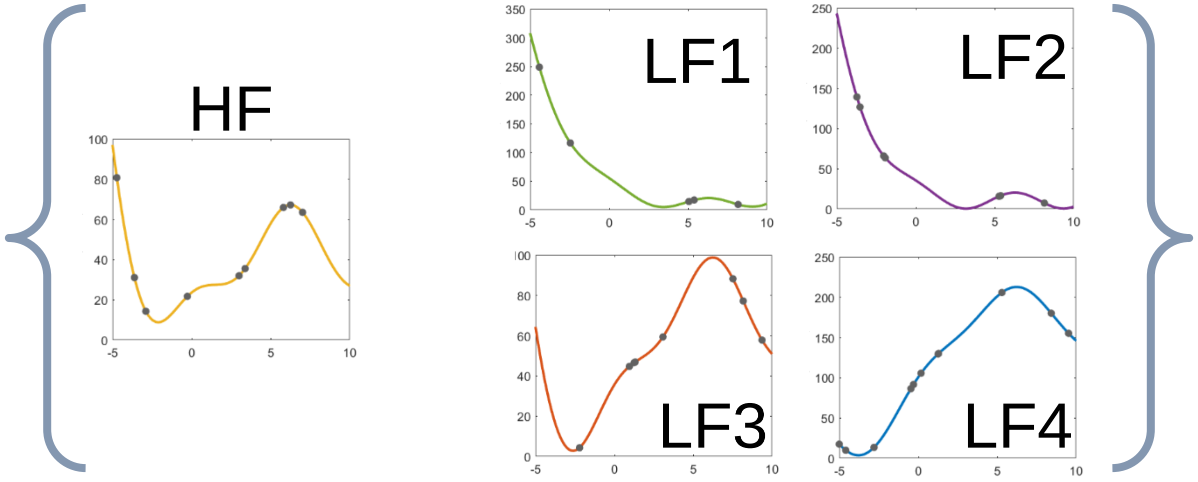


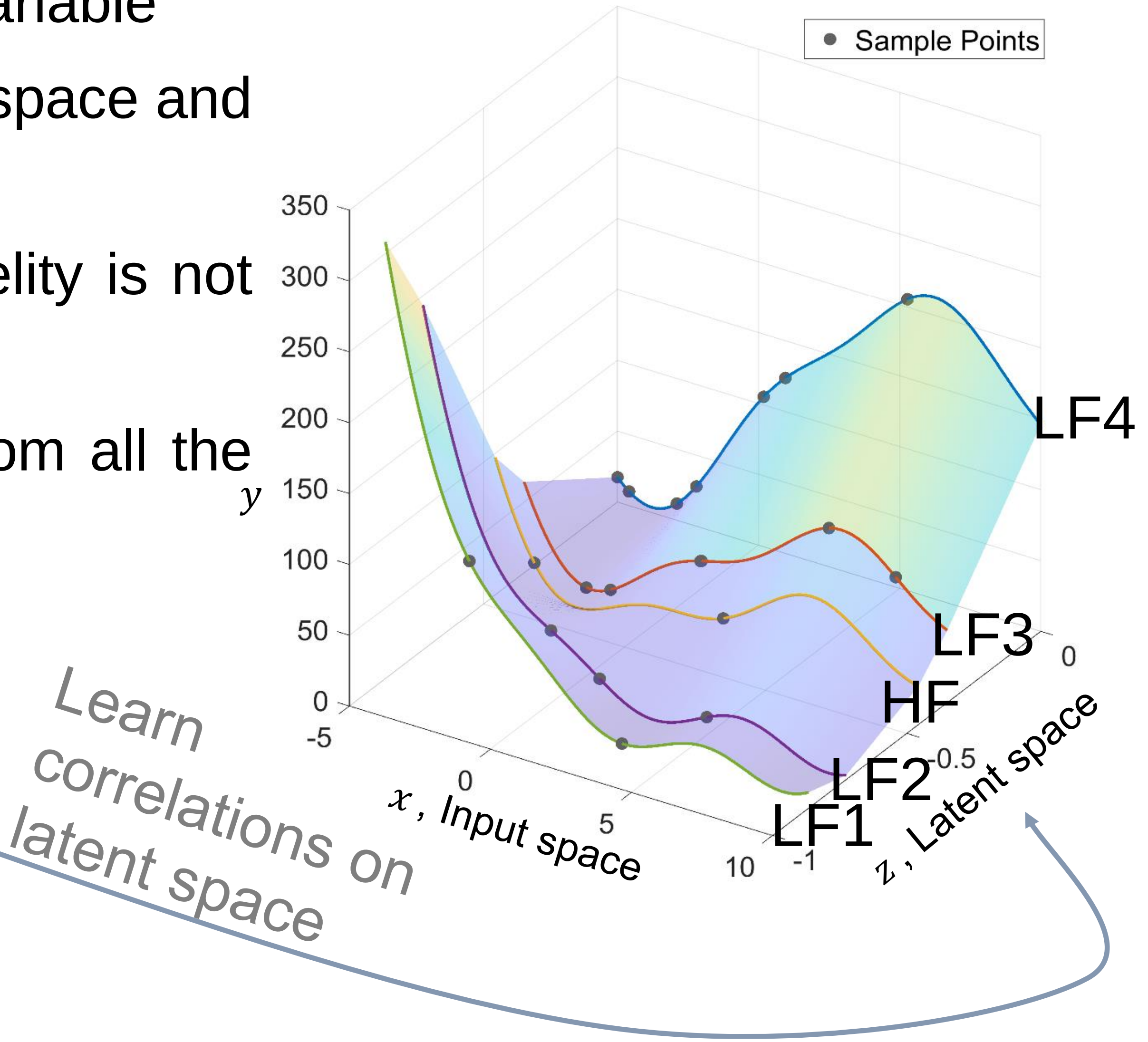

Predicting Target

The prediction is conditioned to all the training samples

# LVGP for Multi-fidelity Modeling (MF-LVGP)

- Assigning the fidelity level as a qualitative variable
- Mapping different fidelity models into latent space and capturing their correlations
- Prior knowledge about the hierarchy of fidelity is not required
- Predicting HF response with information from all the fidelity samples



# Outline

1 Introduction

2 LVGP & MF-LVGP

3 Multi-Fidelity Adaptive Sampling

4 Case Studies

5 Closure

# The Main Questions We Encountered

*Q: How to determine "Which Fidelity Model" and "Where to Sample"?*

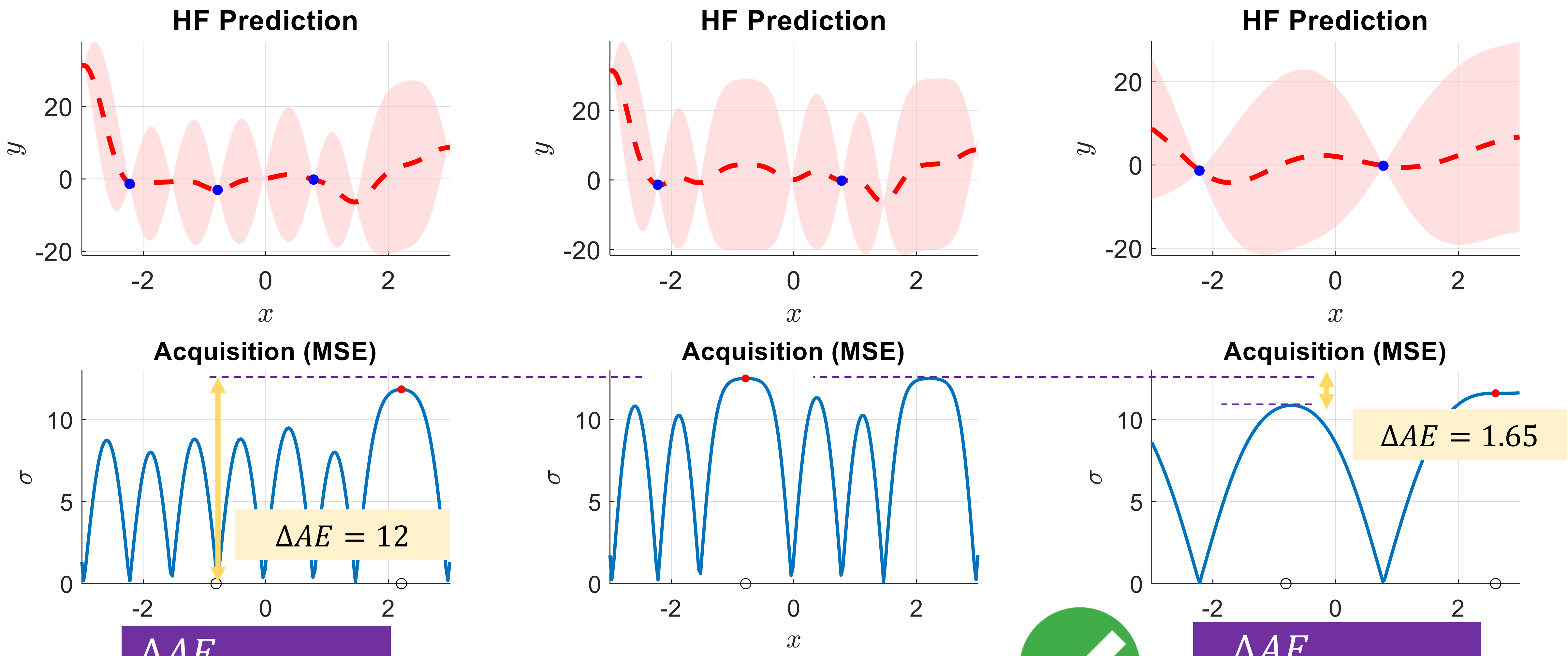
*A:* Infill the sample that provides the highest benefit-to-cost ratio.

*Q: What is the "benefit" and how do we quantify it?*

*A:* We define the benefit as the potential reduction of the acquisition.

# Reduction of Acquisition (MSE) with Infill Samples (GF)

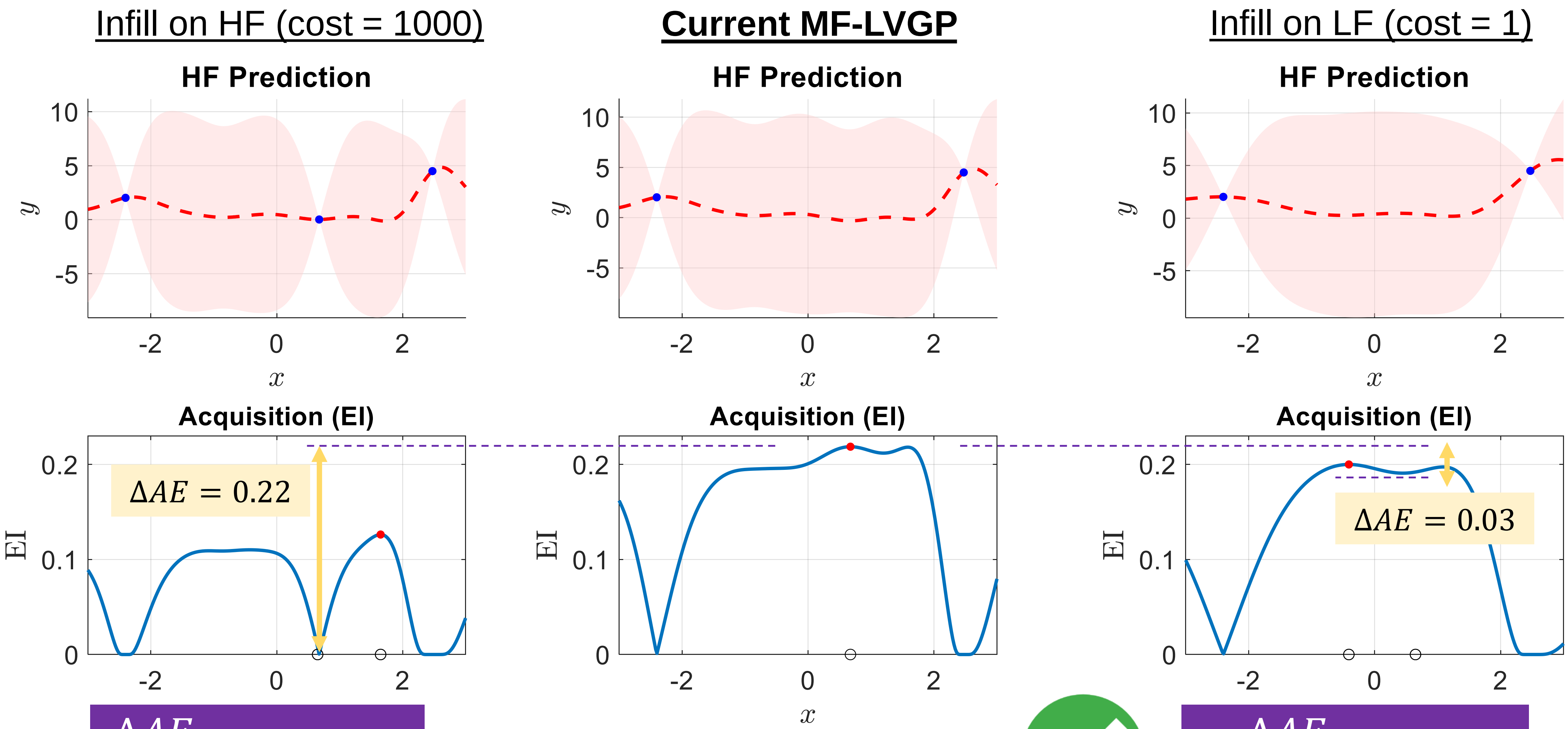
Infill on HF (cost = 1000) | **Current MF-LVGP** | Infill on LF (cost = 1)




$$\frac{\Delta AE}{cost} = 0.012$$

$$\frac{\Delta AE}{cost} = 1.65$$

# Reduction of Acquisition (EI) with Infill Samples (BO)

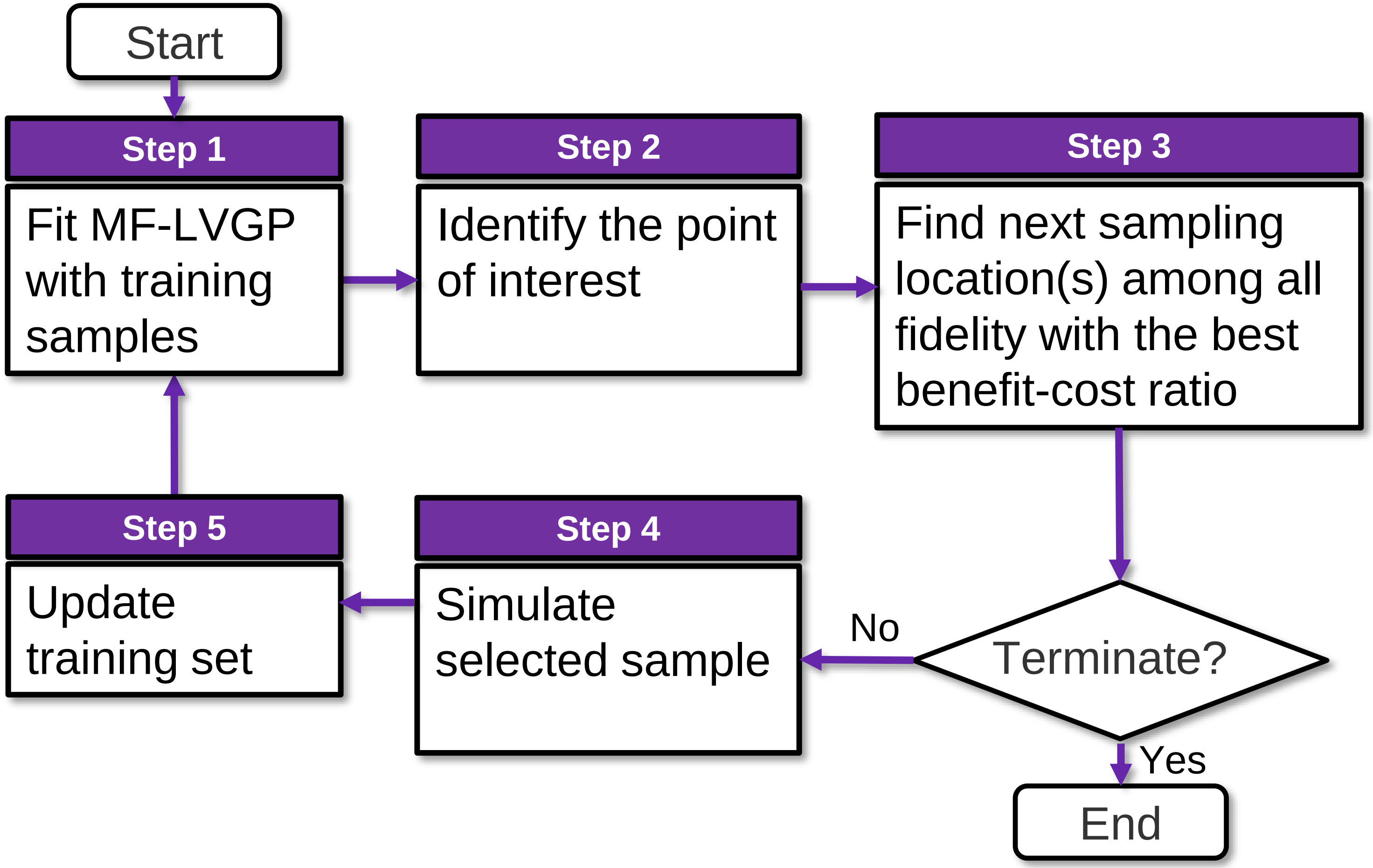


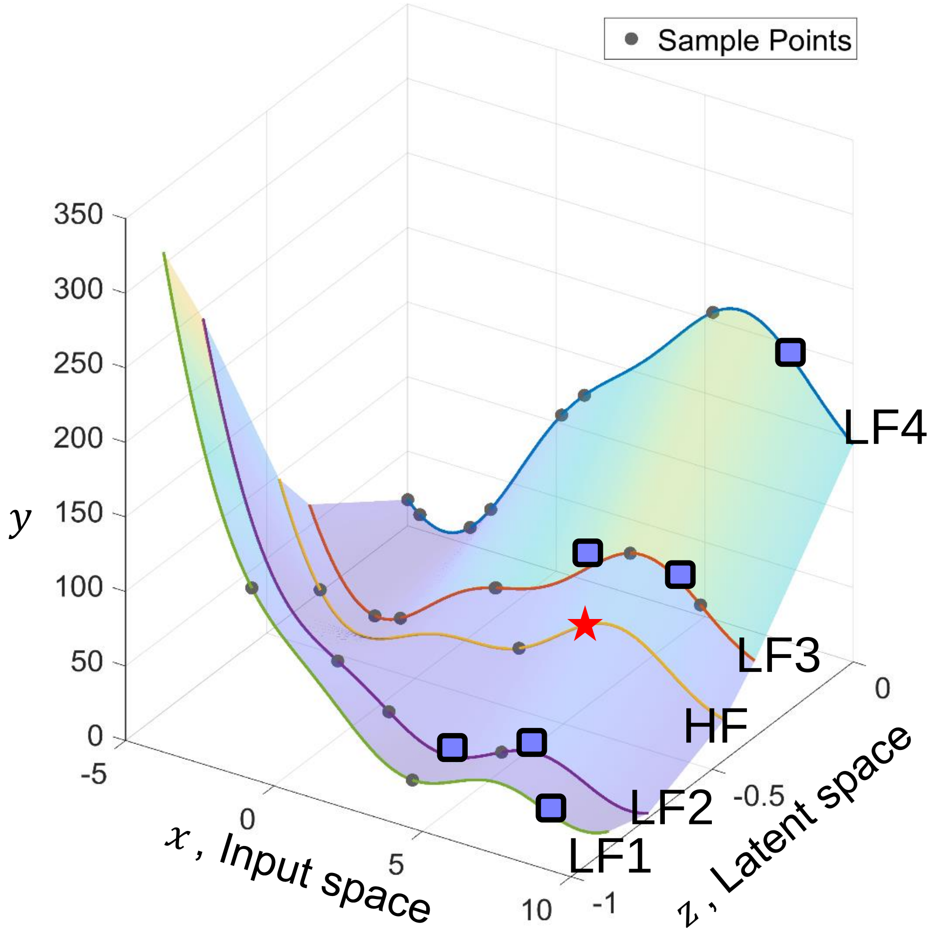

$$\frac{\Delta AE}{cost} = 0.00022$$

$$\frac{\Delta AE}{cost} = 0.03$$

# MF Adaptive Sampling Method: MuFASa

## Multi-Fidelity Adaptive Sampling (MuFASa)




★ Point of interest on HF subregion

▢ Potential sampling location on LF subregions

# Two-Stage Optimization

**First Stage**:
Determine the location of interest on the design space of HF model

$$x_{HF}^* = \arg\max(AF(x))$$

$AF(x)$ can be substituted by the common acquisition function for active learning.

**Global Fitting: Maximum MSE**

$$x_{HF}^* = \arg\max \hat{s}^2(x_{HF})$$

**Bayesian Optimization: EI**

$$EI(x) = (\hat{y}(x) - y^*)\Phi\left(\frac{\hat{y}(x) - y^*}{\hat{s}(x)}\right) + \hat{s}(x)\varphi\left(\frac{\hat{y}(x) - y^*}{\hat{s}(x)}\right)$$

**Second Stage**:
Identify the infill sample with the best benefit-cost ratio

$$x_{next} = \arg\max \frac{AF(x_i) - \widehat{AF}(x_i)}{cost(i)}$$

$\widehat{AF}(x) \rightarrow$ Pre-posterior analysis

**Global Fitting: Maximum MSE**

$$\Delta\widehat{AF}(x_{HF}^*, x_{next}) = \hat{s}^2(x_{HF}^*) - \hat{\hat{s}}^2(x_{HF}^*, x_{next})$$

**Bayesian Optimization: EI**

$$\Delta\widehat{AF}(x_{HF}^*, x_{next}) = EI^2(x_{HF}^*) - \widehat{EI}^2(x_{HF}^*, x_{next})$$

Ginsbourger, David, Rodolphe Le Riche, and Laurent Carraro. "Kriging is well-suited to parallelize optimization." *Computational intelligence in expensive optimization problems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. 131-162.

# Pre-posterior Analysis of LVGP

- candidate LF infill sample $x_{next}$.
- $let\ \hat{y} = LVGP(x_{next})$
- Update parameters $\widehat{\mathbf{R}}_{new} = \begin{bmatrix} \mathbf{R} & r(x_{next}, \boldsymbol{x})^T \\ r(x_{next}, \boldsymbol{x}) & r(x_{next}, x_{next}) \end{bmatrix}$, $\widehat{\mathbf{y}}_{new} = [\boldsymbol{y}, \hat{y}]^T$

### Parameters in pre-posterior LVGP

$$\hat{\hat{\mu}}_{new} = \left(\mathbf{1}^T(\widehat{\mathbf{R}}_{new})^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^T(\widehat{\mathbf{R}}_{new})^{-1}\,\widehat{\mathbf{y}}_{new}$$

$$\hat{\hat{\sigma}}^2{}_{new} = \frac{1}{n}\left(\widehat{\mathbf{y}}_{new} - \hat{\hat{\mu}}_{new}\mathbf{1}\right)^T(\widehat{\mathbf{R}}_{new})^{-1}\left(\widehat{\mathbf{y}}_{new} - \hat{\hat{\mu}}_{new}\mathbf{1}\right)$$

### Prediction of arbitrary input of pre-posterior LVGP

$$\boxed{\begin{matrix}\hat{\hat{y}}(x^*_{HF}) \\ \hat{\hat{s}}^2(x^*_{HF})\end{matrix}}\begin{matrix} = \hat{\hat{\mu}}_{new} + \boldsymbol{r}(x^*)(\widehat{\mathbf{R}}_{new})^{-1}(\widehat{\mathbf{y}}_{new} - \hat{\hat{\mu}}_{new}\mathbf{1}) \\ = \hat{\hat{\sigma}}^2{}_{new}\left(\boldsymbol{r}(x^*) - \boldsymbol{r}(x^*)\widehat{\mathbf{R}}_{new}{}^{-1}\boldsymbol{r}(x^*)^T\right)\end{matrix}$$

### Parameters in LVGP

$$\hat{\mu} = (\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1})^{-1}\mathbf{1}^T\mathbf{R}^{-1}\mathbf{y}$$

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \hat{\mu}\mathbf{1})^T\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}).$$

### Prediction of arbitrary input in LVGP

$$\hat{y}(\boldsymbol{x}^*) = \hat{\mu} + \boldsymbol{r}(\boldsymbol{x}^*)\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1})$$

$$\hat{s}^2(\boldsymbol{x}^*) = \hat{\sigma}^2(\boldsymbol{r}(\boldsymbol{x}^*) - \boldsymbol{r}(\boldsymbol{x}^*)\mathbf{R}^{-1}\boldsymbol{r}(\boldsymbol{x}^*)^T)$$

# Two-Stage Optimization

## First Stage:

Determine the location of interest on the design space of HF model

$$x_{HF}^{*} = \arg\max(AF(x))$$

$AF(x)$ can be substituted by the common acquisition function for active learning.

**Global Fitting: Maximum MSE**

$$x_{HF}^{*} = \arg\max \hat{s}^2(x_{HF})$$

**Bayesian Optimization: EI**

$$EI(x) = (\hat{y}(x) - y^*)\Phi\left(\frac{\hat{y}(x) - y^*}{\hat{s}(x)}\right) + \hat{s}(x)\varphi\left(\frac{\hat{y}(x) - y^*}{\hat{s}(x)}\right)$$

## Second Stage:

Identify the infill sample with the best benefit-cost ratio

$$x_{next} = \arg\max \frac{AF(x_i) - \widehat{AF}(x_i)}{cost(i)}$$

$\widehat{AF}(x) \rightarrow$ Pre-posterior analysis

**Global Fitting: Maximum MSE**

$$\Delta\widehat{AF}(x_{HF}^{*}, x_{next}) = \hat{s}^2(x_{HF}^{*}) - \hat{\hat{s}}^2(x_{HF}^{*}, x_{next})$$

**Bayesian Optimization: EI**

$$\Delta\widehat{AF}(x_{HF}^{*}, x_{next}) = EI^2(x_{HF}^{*}) - \widehat{EI}^2(x_{HF}^{*}, x_{next})$$

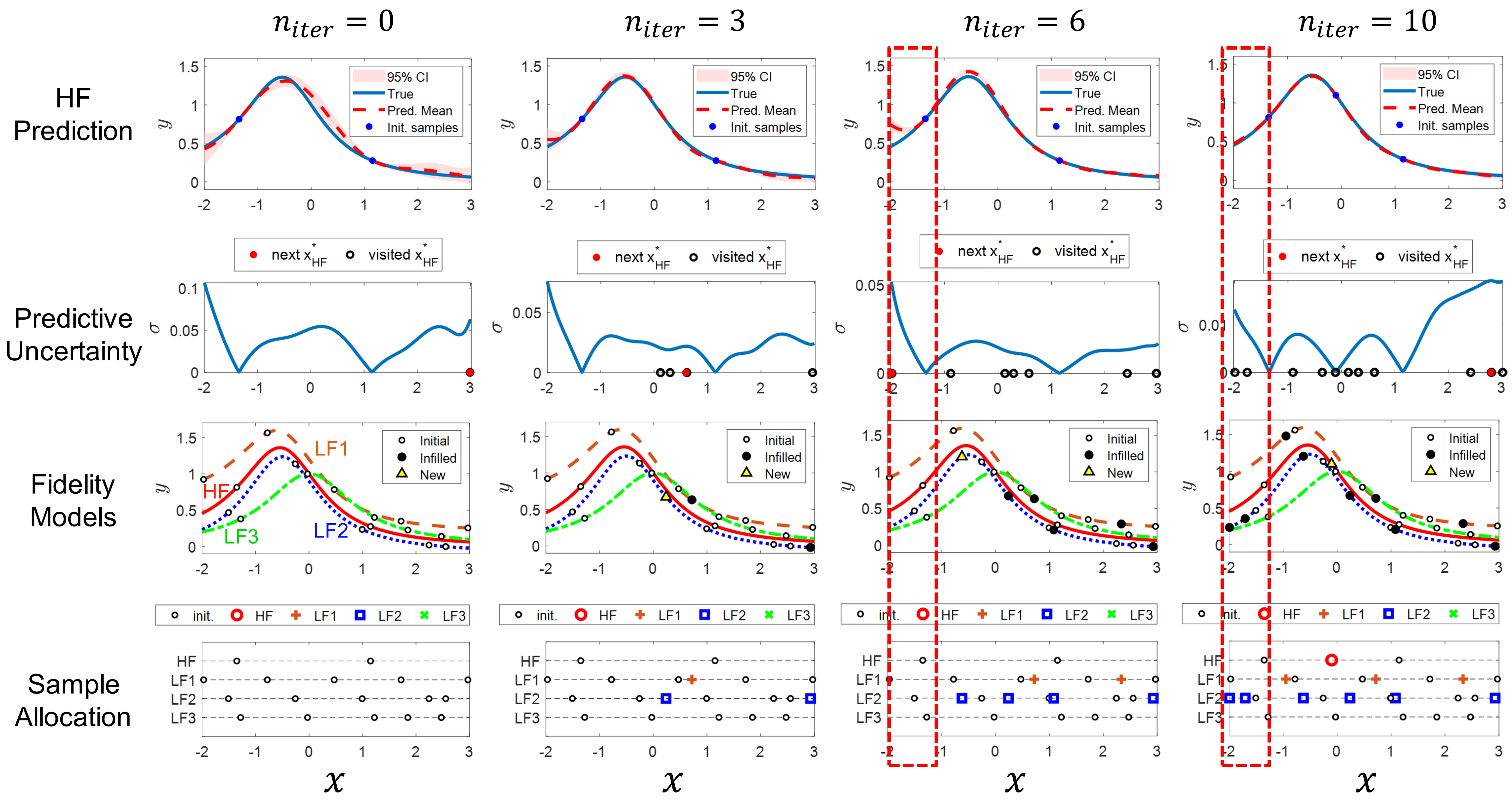
Ginsbourger, David, Rodolphe Le Riche, and Laurent Carraro. "Kriging is well-suited to parallelize optimization." *Computational intelligence in expensive optimization problems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. 131-162.

# Outline

# 1D Illustrative Example with Multiple Fidelities (GF)

| | Model $(-2 \le x \le 3)$ | Cost | Initi. Sample |
|---|---|---|---|
| HF | $y_{HF}(x) = \dfrac{1}{0.1x^3 + x^2 + x + 1}$ | 100 | 2 |
| LF1 | $y_{LF1}(x) = \dfrac{1}{0.2x^3 + x^2 + x + 1} + 0.2$ | 10 | 5 |
| LF2 | $y_{LF2}(x) = \dfrac{1}{x^2 + x + 2} - 0.1$ | 10 | 5 |
| LF3 | $y_{LF3}(x) = \dfrac{1}{x + 1}$ | 10 | 5 |

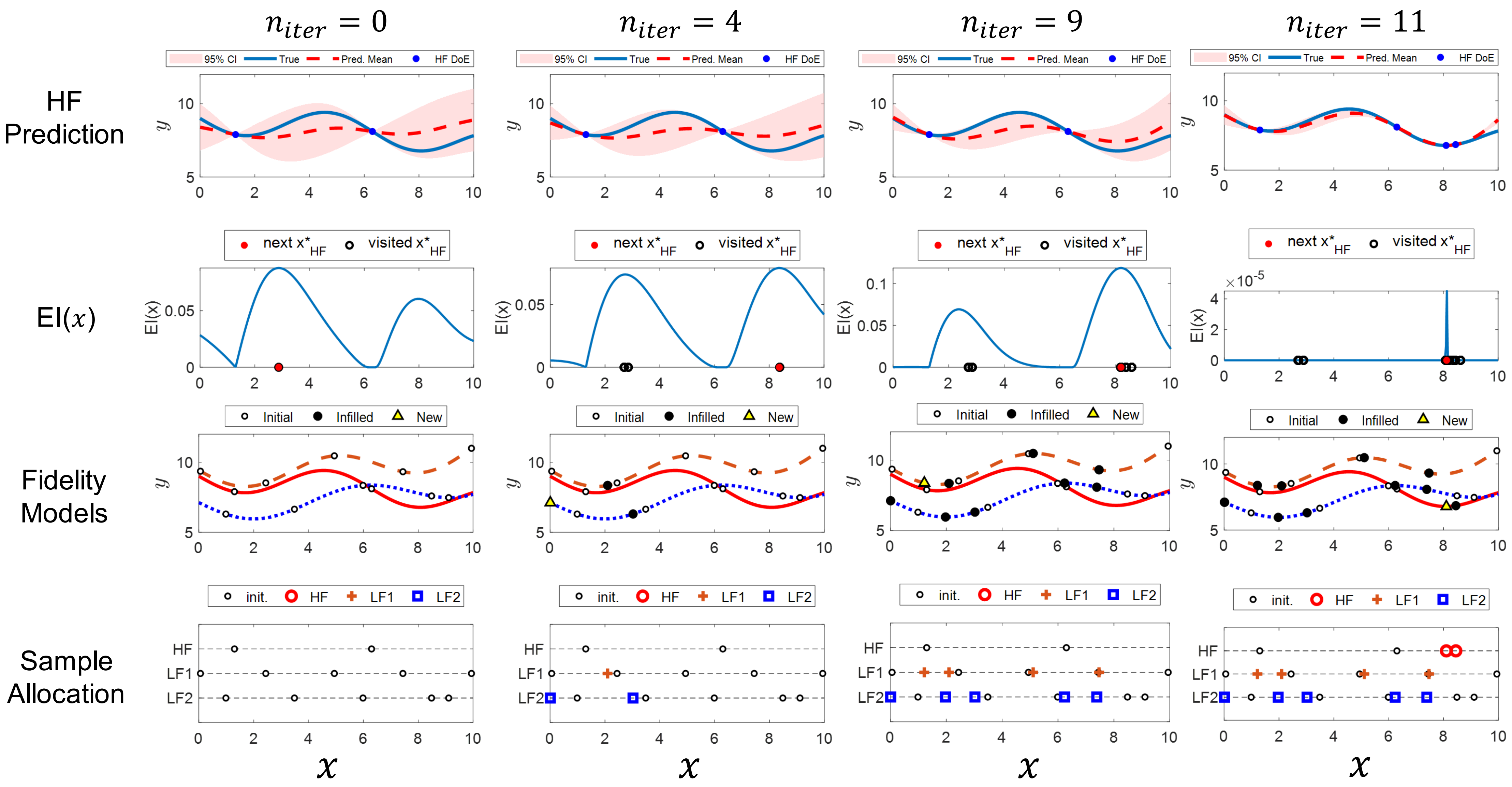
# 1D Illustrative Example with Multiple Fidelities (GF)



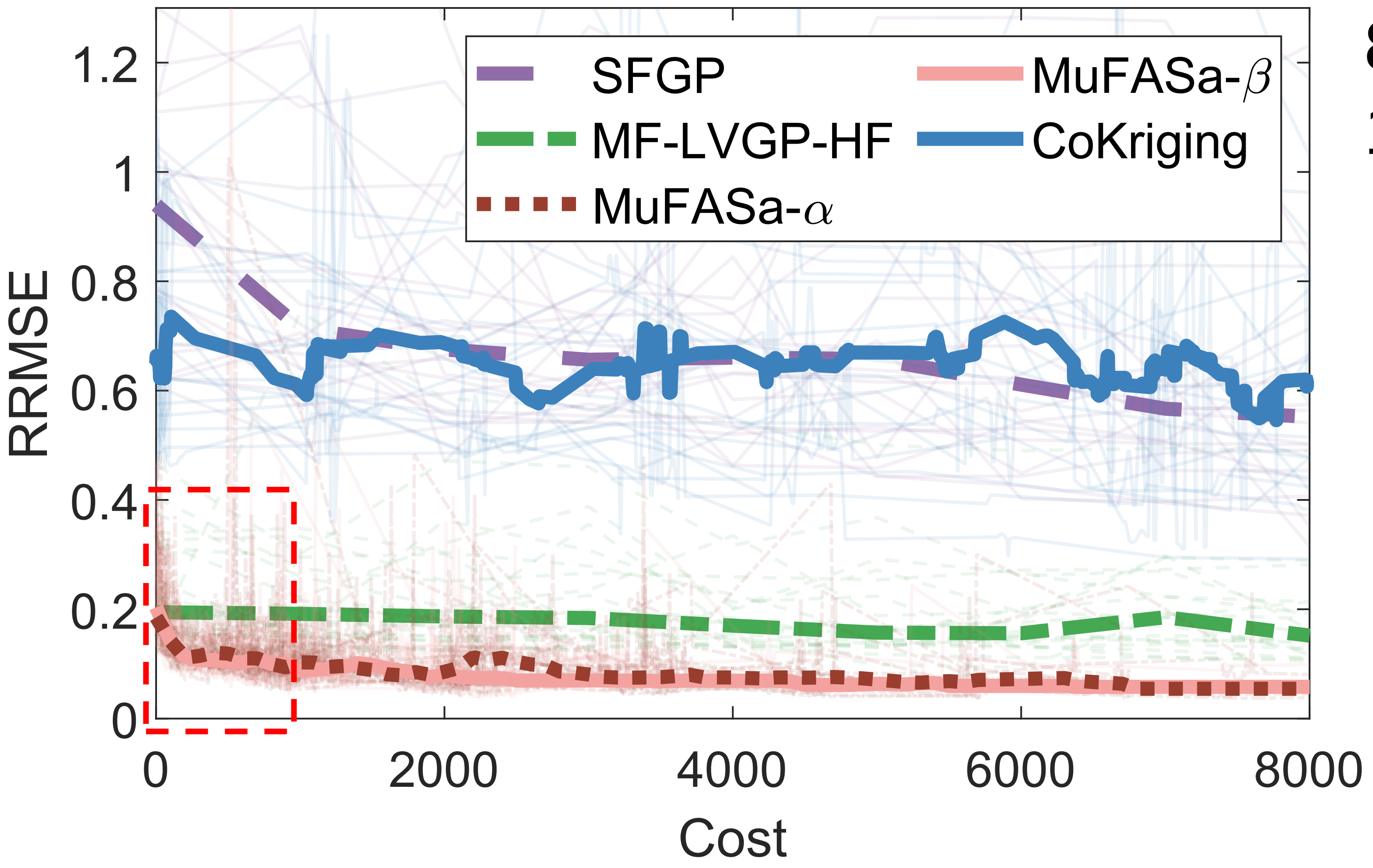
# 1D Illustrative Example with Multiple Fidelities (BO)

| | **Model** $(0 \le x \le 10)$ | **Cost** | **Initi. Sample** |
|---|---|---|---|
| HF | $y_{HF}(x) = -\sin x - \exp\left(\frac{x}{10}\right) + 10$ | 1000 | 2 |
| LF1 | $y_{LF1}(x) = -\sin(0.95x) - \exp\left(\frac{x}{50}\right) + 0.03(x-2)^2 + 10.3$ | 1 | 5 |
| LF2 | $y_{LF2}(x) = -\sin(0.8x) - \exp\left(\frac{x}{50}\right) + 0.03(x-2)^2 + 8$ | 1 | 5 |

# 1D Illustrative Example with Multiple Fidelities (BO)



# Convergence History for Borehole (GF)

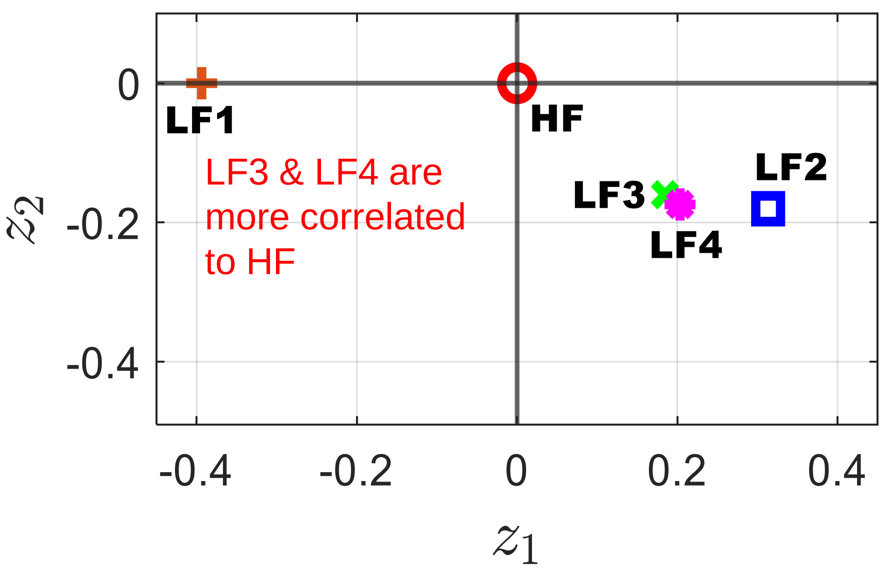



8 inputs,
1 HF + 4 LF

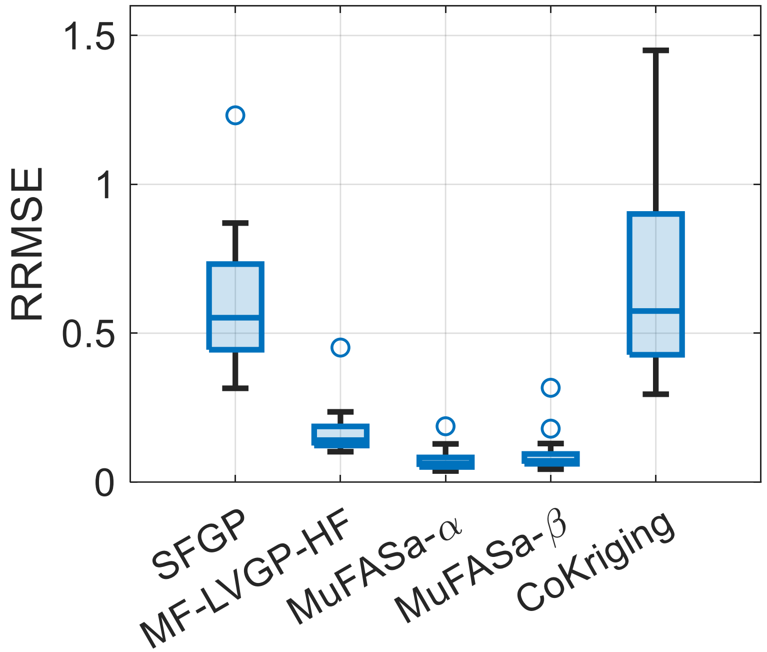
- MF-LVGP & MuFASa outperforms the CoKriging and SFGP in accuracy, even before adaptive sampling, RRMSE drops dramatically before the HF samples are infilled, mostly contributed by LF infill samples.
- RRMSE curves have several jumps due to the variation of the scaling factors via MLE.

# Replicate Tests for High-D Test Cases (GF)

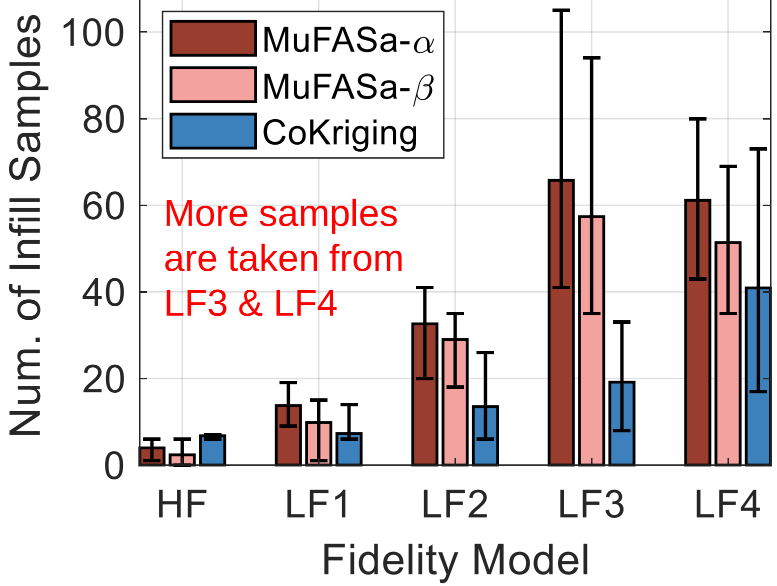
**Latent space of fidelity level**




**Termination RRMSE**

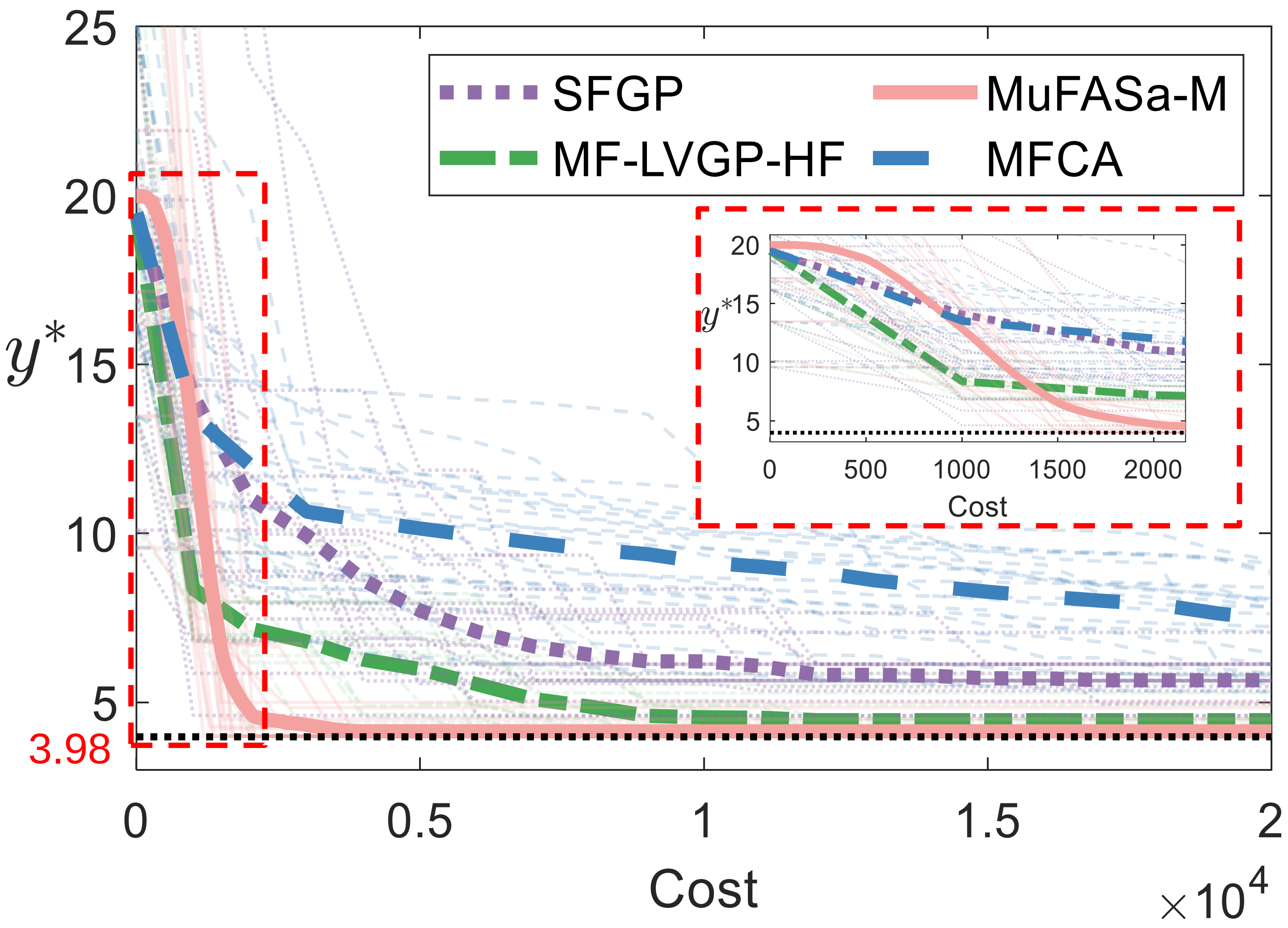



**Sample allocation**




| ID | HF | LF1 | LF2 | LF3 | LF4 |
|---|---|---|---|---|---|
| Cost | 1000 | 10 | 10 | 10 | 10 |
| Ini. DoE | 4 | 10 | 10 | 10 | 10 |

# Convergence History for Borehole (BO)

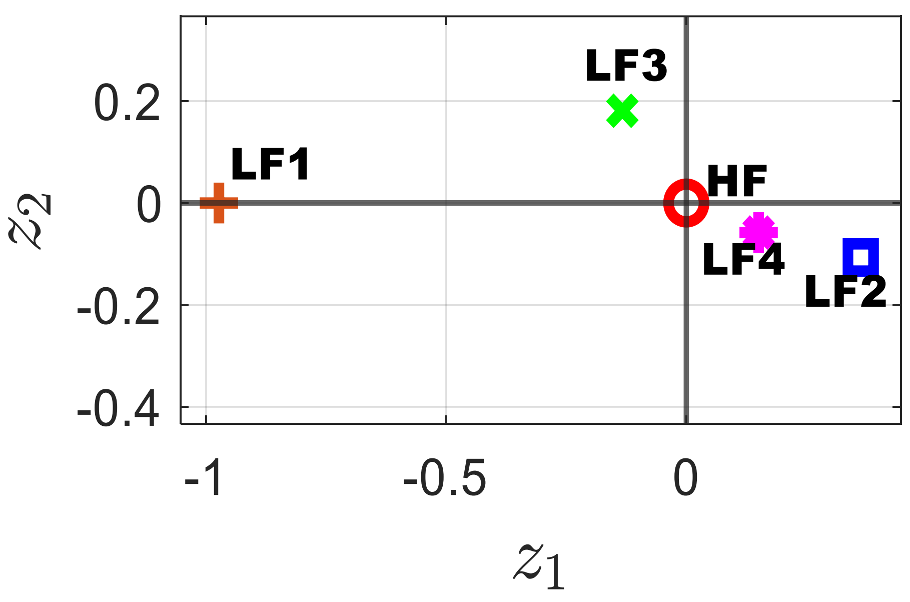



8 inputs,
1 HF + 4 LF

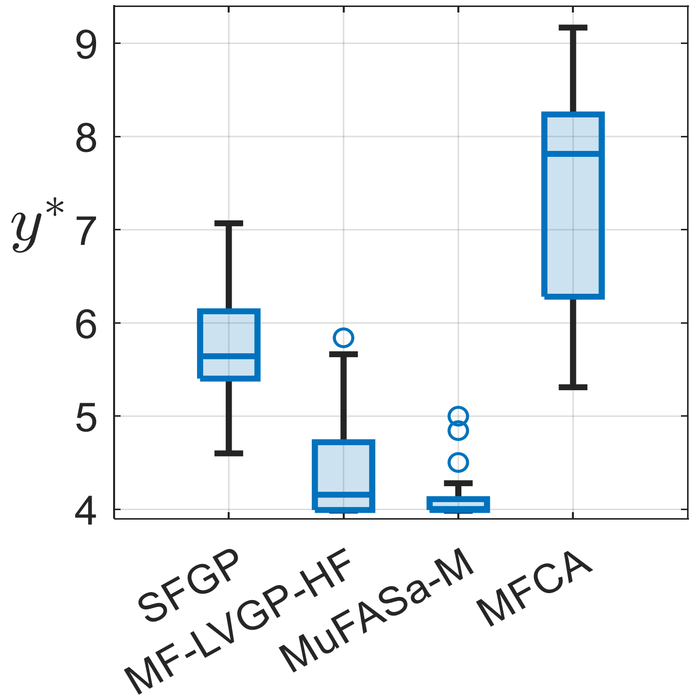
- MuFASa outperforms benchmark methods in both convergence rate and robustness.
- The optimal values do not drop immediately as the BO starts: more LF samples are taken to improve the model, then HF samples are infilled near the ground true optimum.

# Replicate Tests for Borehole (BO)

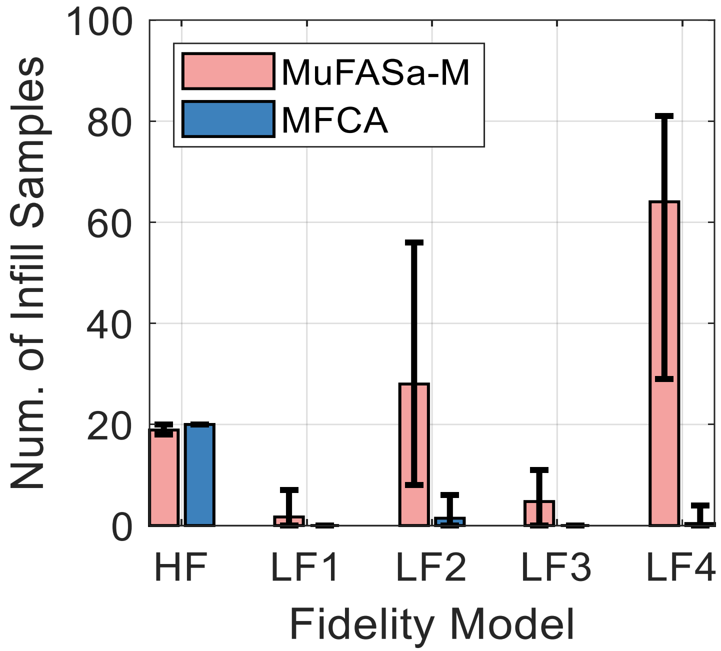
### Latent space of fidelity level




### Optimal Solution




### Sample Allocation




| ID | HF | LF1 | LF2 | LF3 | LF4 |
|---|---|---|---|---|---|
| Cost | 1000 | 100 | 10 | 100 | 10 |
| Ini. DoE | 5 | 5 | 25 | 5 | 25 |

# Outline

1. Introduction
2. LVGP & MF-LVGP
3. Multi-Fidelity Adaptive Sampling
4. Case Studies
5. Closure

# Closure

## Remarks

- **MF-LVGP:** non-hierarchical, interpretable MF data fusion.
- **Pre-posterior Analysis***:* **Quantify the future benefit** of the candidate samples.
- **MuFASa:** a benefit & cost-aware adaptive sampling framework.
- Address *GF* and *BO* with the same framework.
- Outperform benchmark methods with better convergence rate and robustness.

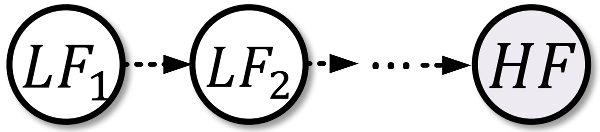
## Future Works

- *Scalable techniques* → accommodate big dataset.
- *Model-reduction techniques* → accommodate higher-dimensional cases.
- *greedy sampling/dynamic programming* → batch sampling, look multiple steps ahead.

# Thank you & Appreciate for the attentions!

# Categorize MF Methods via Architecture

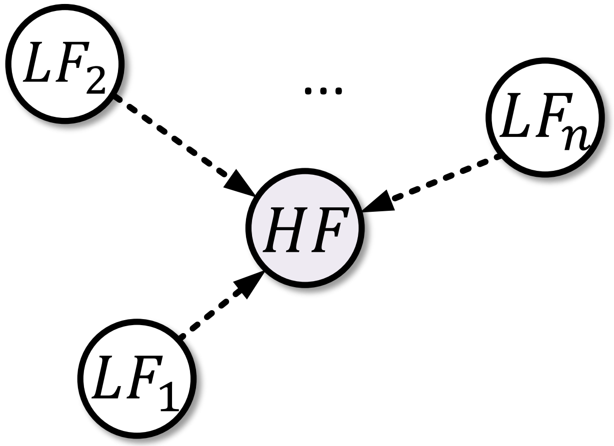
$$HF = HF|LF_n| \ldots |LF_2|LF_1$$




(a) Hierarchical Architecture

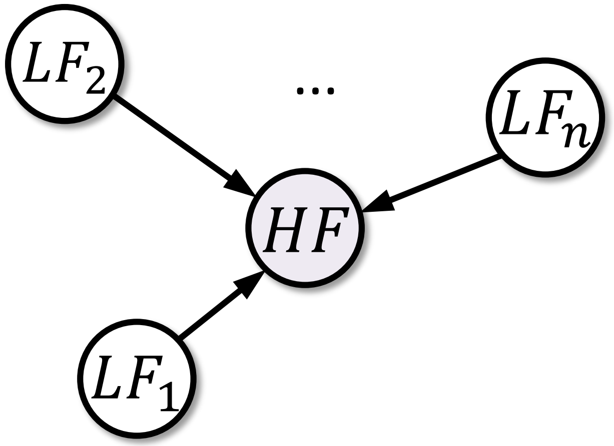
- Ranking fidelity models may not be practical
- Uncertainty propagation, not desirable for adaptive sampling

$$HF = HF|(LF_n, \ldots, LF_2, LF_1)$$




(b) Non-hierarchical architecture with implicit use of correlation

- The contribution of the LF samples are unknown.
- Treat LF and HF source as independent sources and neglect their interplay.




(c) Non-hierarchical architecture with explicit use of correlation

- Focus on improving HF surrogate only as the objective.
- Approximate the future benefit as the sampling criteria is the key.

# Borehole Function for Bayesian Optimization

| | Model | Cost | Ini. samples |
|---|---|---|---|
| HF | $y_{HF} = \dfrac{2\pi T_u(H_u - H_l)}{\ln\left(\frac{r}{rw}\right)\left(1 + \frac{2LT_u}{\ln\left(\frac{r}{rw}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$ | 1000 | 5 |
| LF 1 | $y_{LF1} = \dfrac{2\pi T_u(H_u - 0.8H_l)}{\ln\left(\frac{r}{rw}\right)\left(1 + \frac{1LT_u}{\ln\left(\frac{r}{rw}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$ | 100 | 5 |
| LF 2 | $y_{LF2} = \dfrac{2\pi T_u(H_u - H_l)}{\ln\left(\frac{r}{rw}\right)\left(1 + \frac{8LT_u}{\ln\left(\frac{r}{rw}\right) r_w^2 k_w} + 0.75\frac{T_u}{T_l}\right)}$ | 10 | 25 |
| LF 3 | $y_{LF3} = \dfrac{2\pi T_u(1.09H_u - H_l)}{\ln\left(\frac{4r}{rw}\right)\left(1 + \frac{3LT_u}{\ln\left(\frac{r}{rw}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$ | 100 | 5 |
| LF 4 | $y_{LF4} = \dfrac{2\pi T_u(1.05H_u - H_l)}{\ln\left(\frac{4r}{rw}\right)\left(1 + \frac{3LT_u}{\ln\left(\frac{r}{rw}\right) r_w^2 k_w} + \frac{T_u}{T_l}\right)}$ | 10 | 25 |

Input space:

$$100 \le T_u \le 1000$$
$$990 \le H_u \le 1110$$
$$700 \le H_l \le 820$$
$$100 \le r \le 10000$$
$$0.05 \le r_w \le 0.15$$
$$10 \le T_l \le 500$$
$$1000 \le L \le 2000$$
$$6000 \le K_w \le 12000$$

Ground Truth Optimum: 3.98

Zanjani Foumani, Z., Shishehbor, M., Yousefpour, A., & Bostanabad, R. (2023). Multi-fidelity cost-aware Bayesian optimization. *Computer Methods in Applied Mechanics and Engineering*, *407*, 115937.