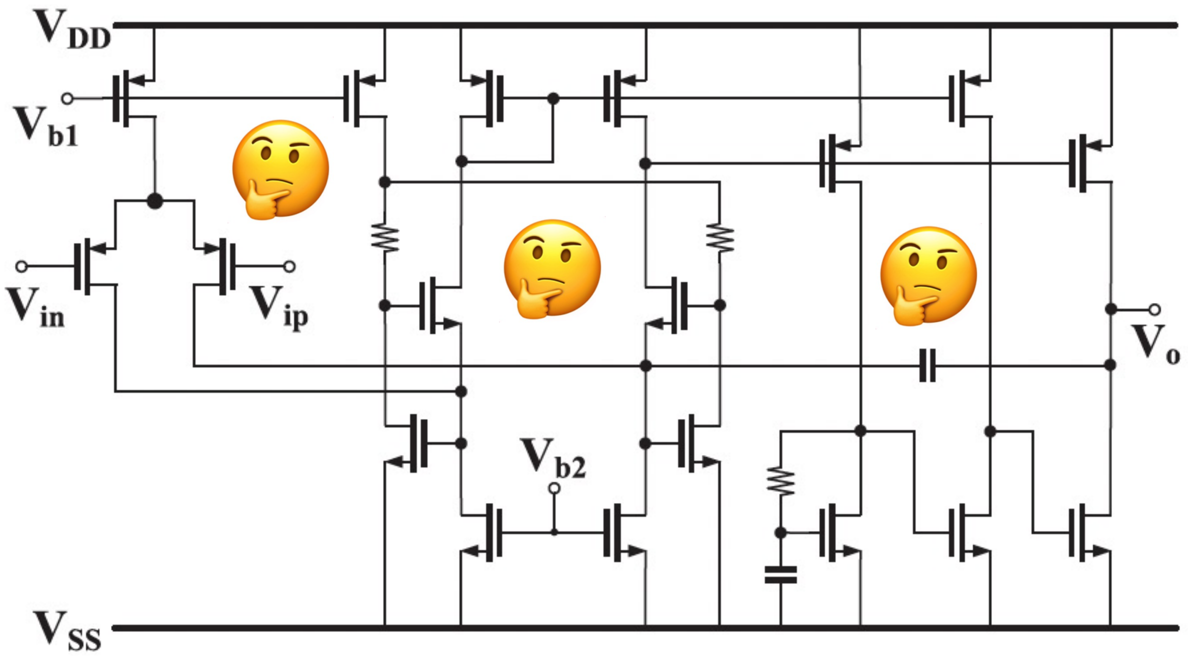
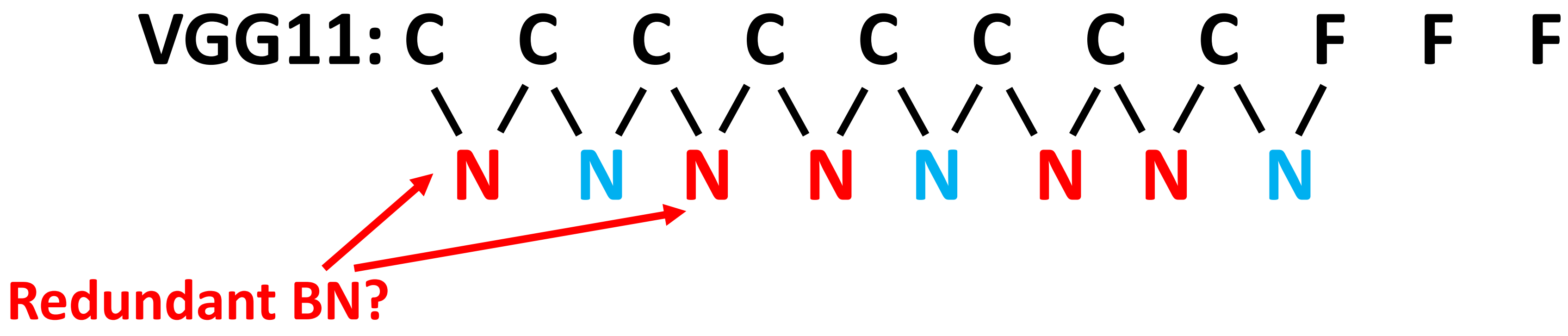
Analog design



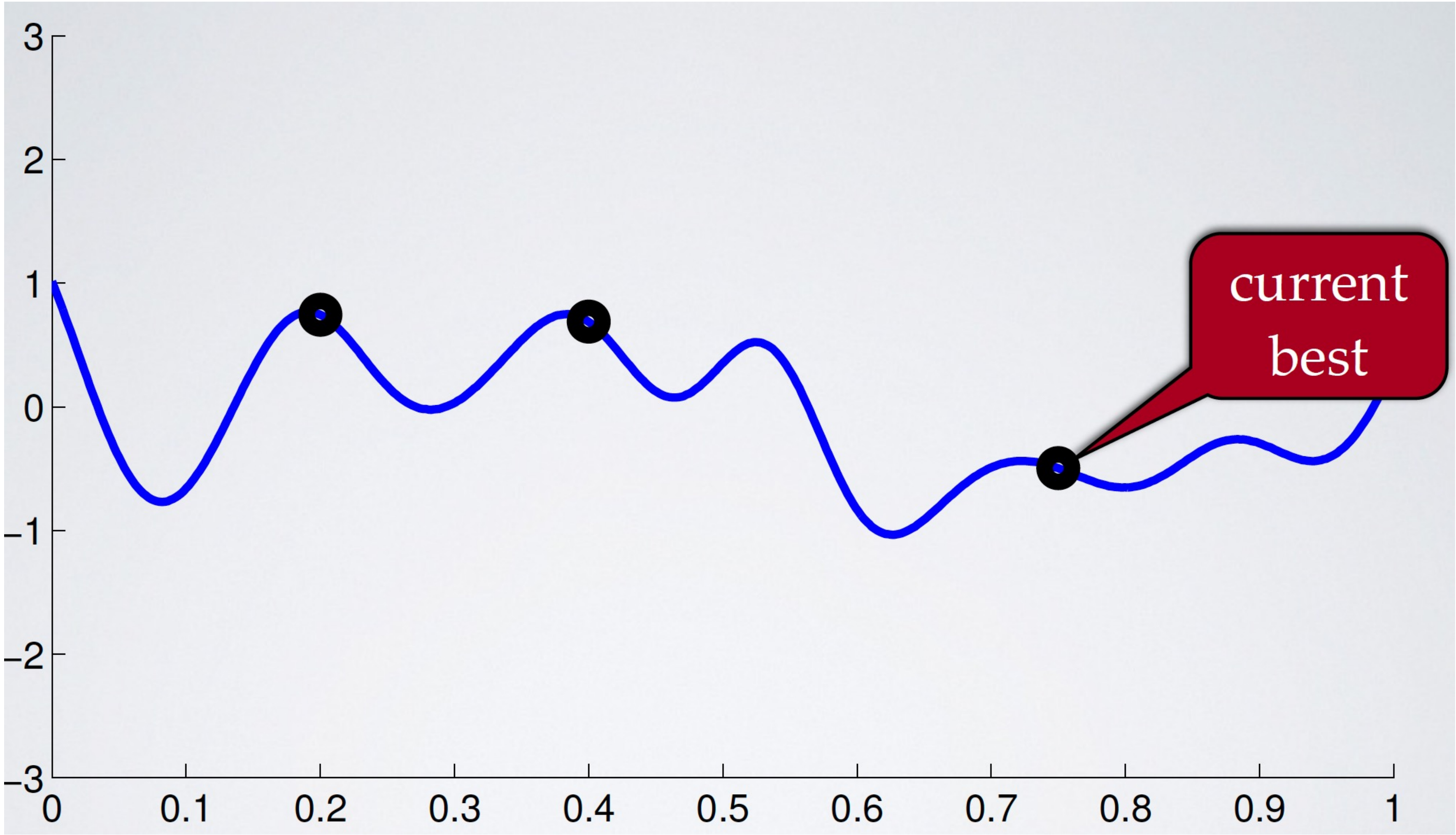
- Vout = f(Vin, Size1, R1,…)
- Given some requirements (e.g., gain, SNR)
- Design the circuit that meets requirements
- Use BO to find circuit parameters

ML for Analog IMC



- Not all BN layers are necessary
- Remove some BN layers to accelerate analog computation
- Use BO to find the BN layers

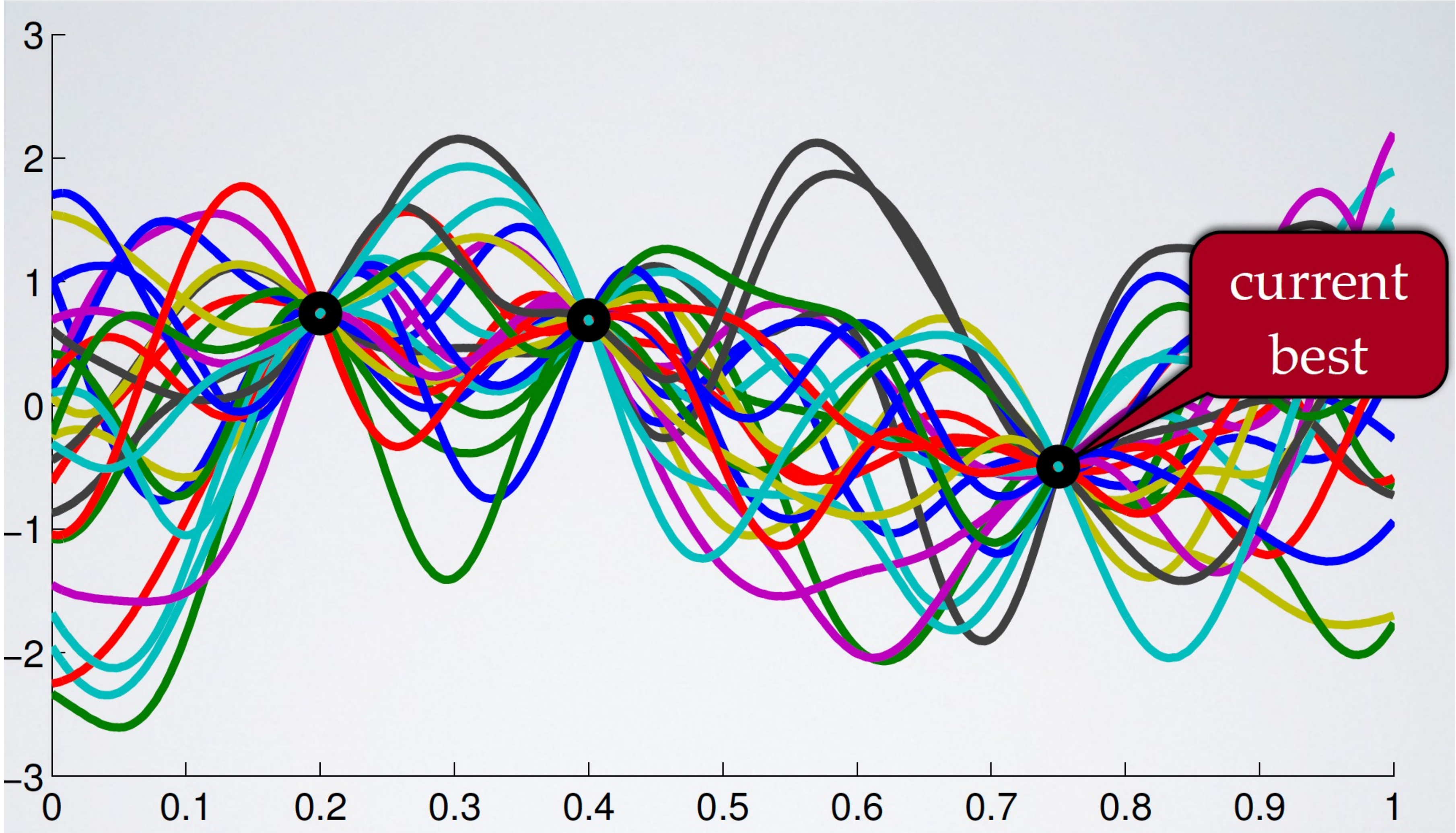
Bayesian optimization



- **Trying to approximate the original function using few data points**

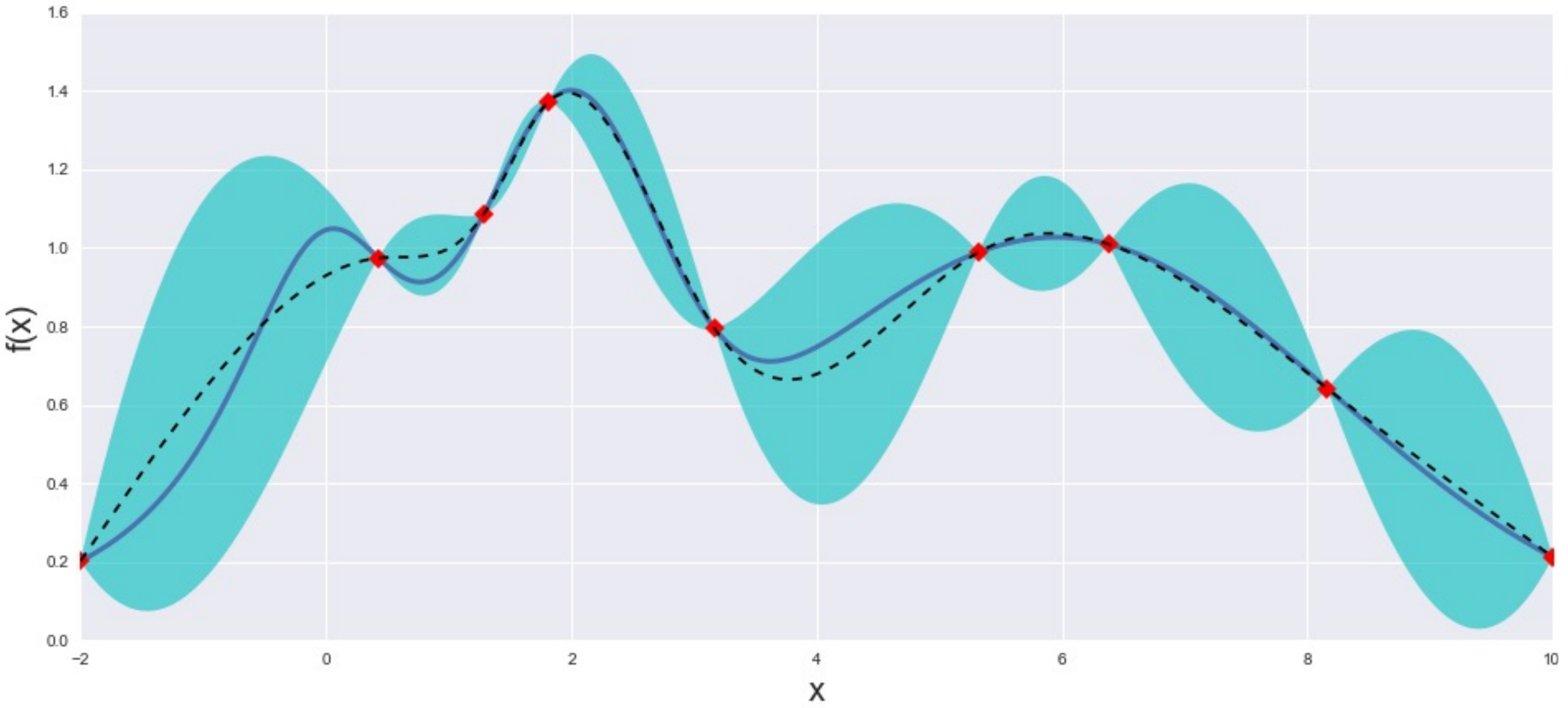
https://www.youtube.com/watch?v=M-NTkxfd7-8
https://www.youtube.com/watch?v=ECNU4WIuhSE

Bayesian optimization



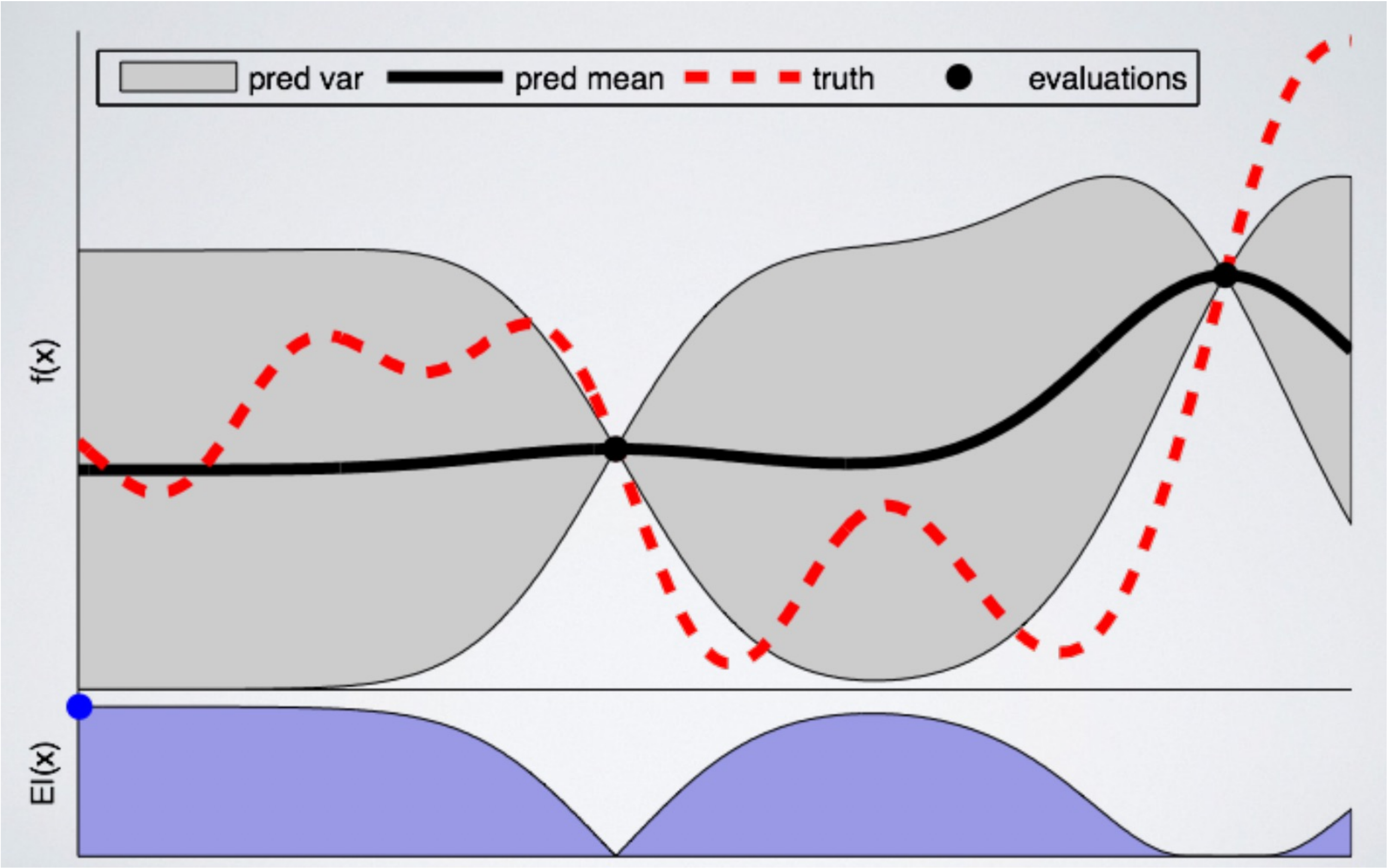
- $F(t) = a_1 t + a_2 t^2 + \cdots$
- **What are the values of a1, a2, etc?**

Uncertain regions using BO

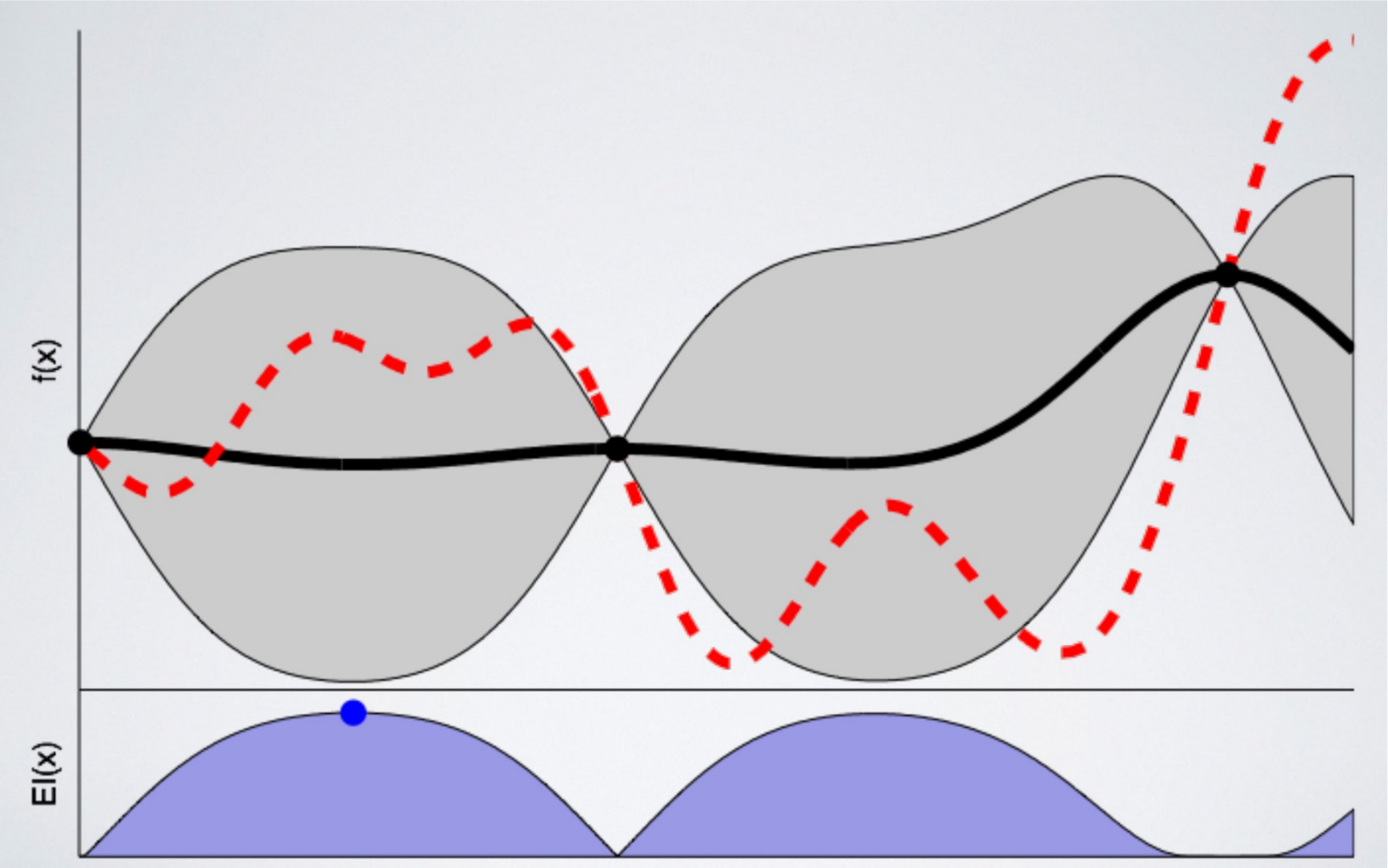


- We’re certain about the points we’ve already visited
- Uncertainty at every other points

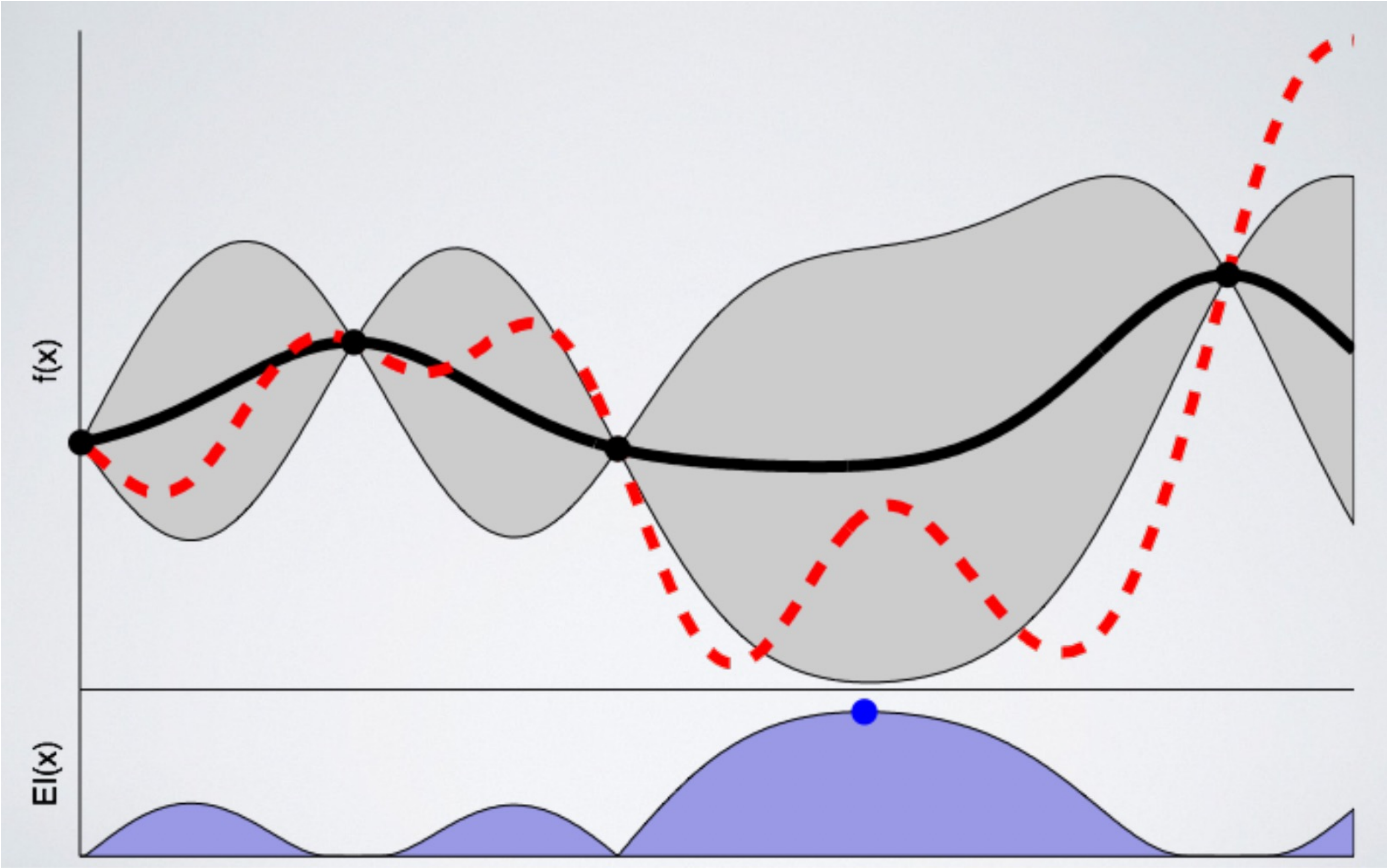
BO example (1)



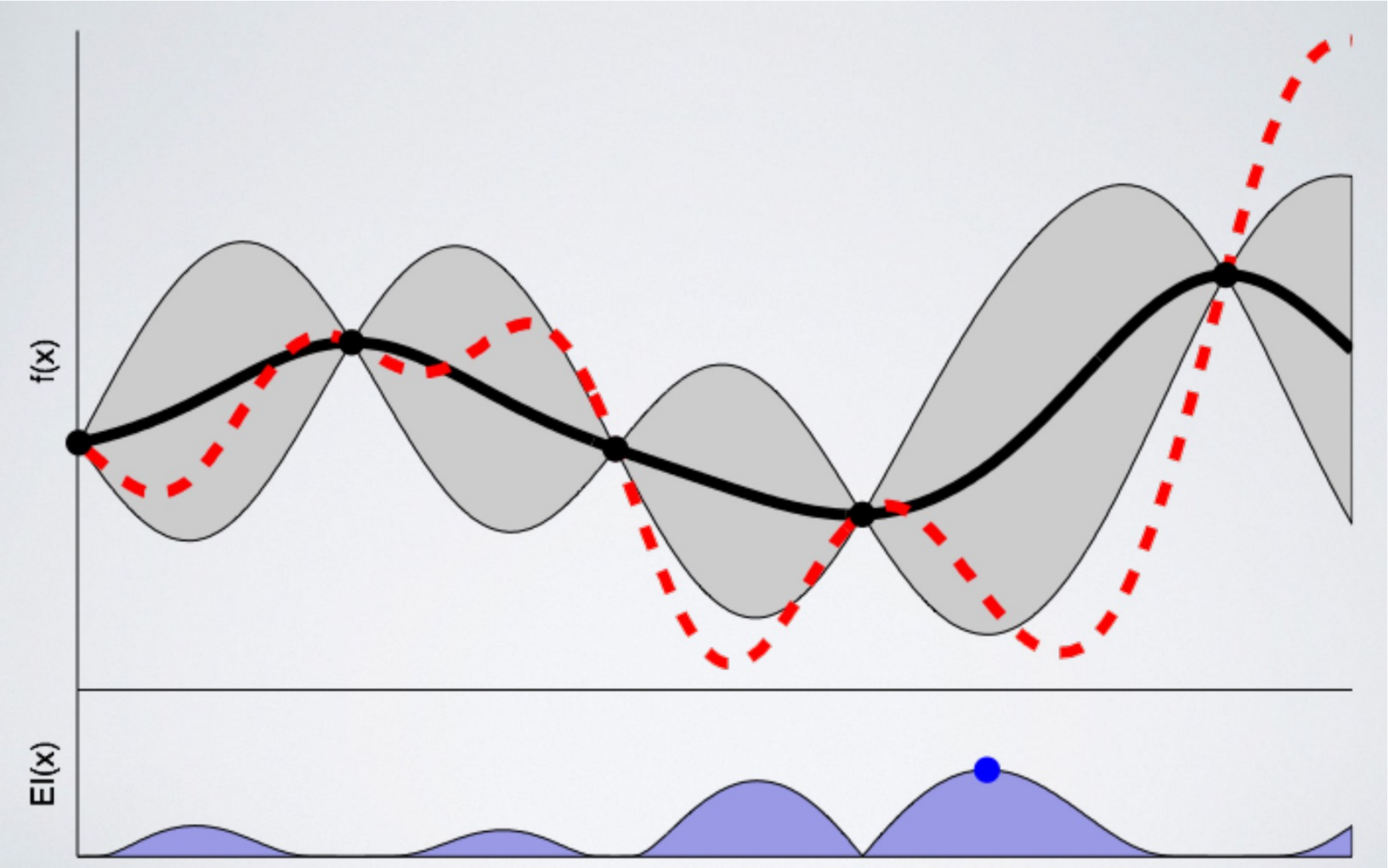
BO example (2)



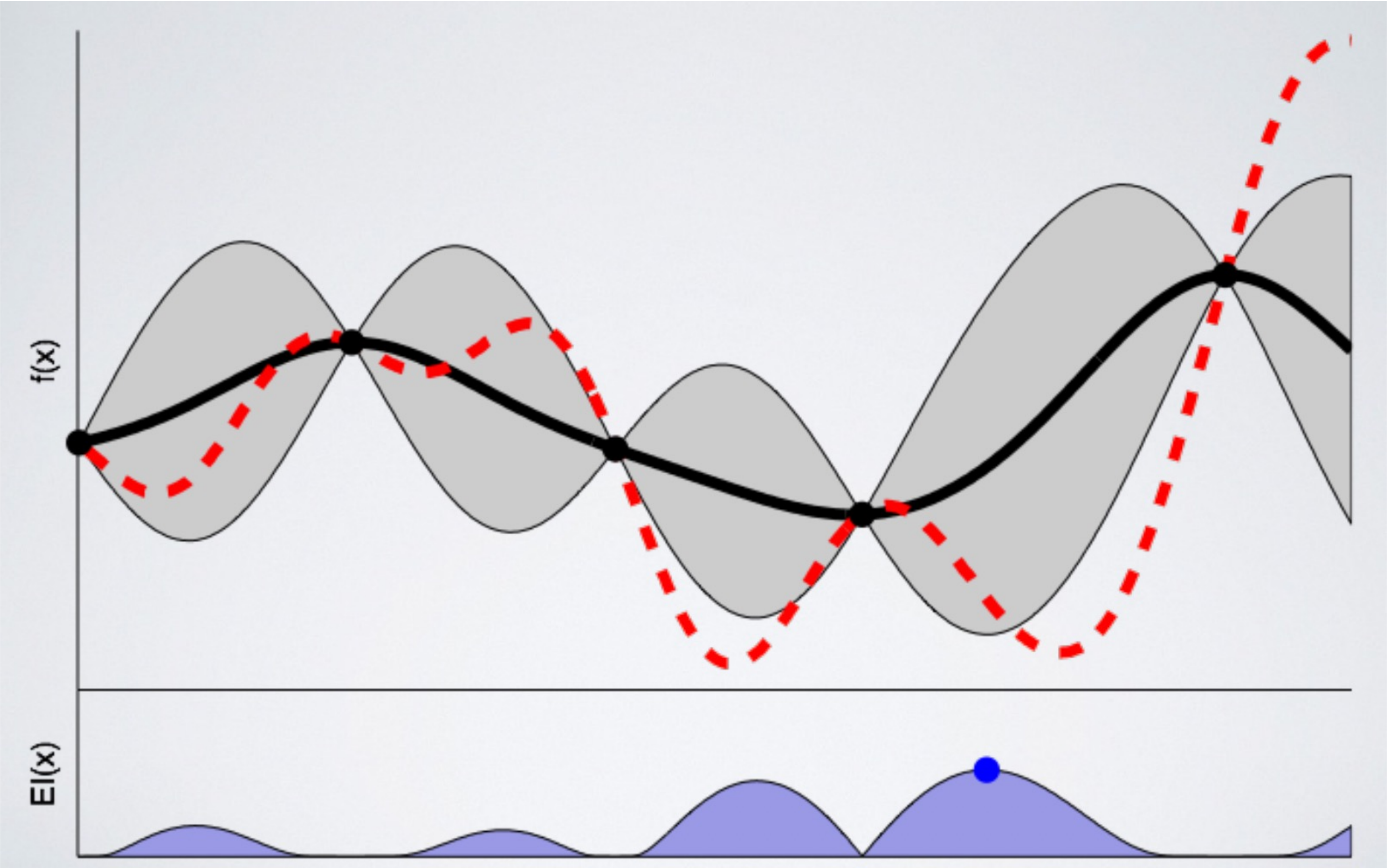
BO example (3)



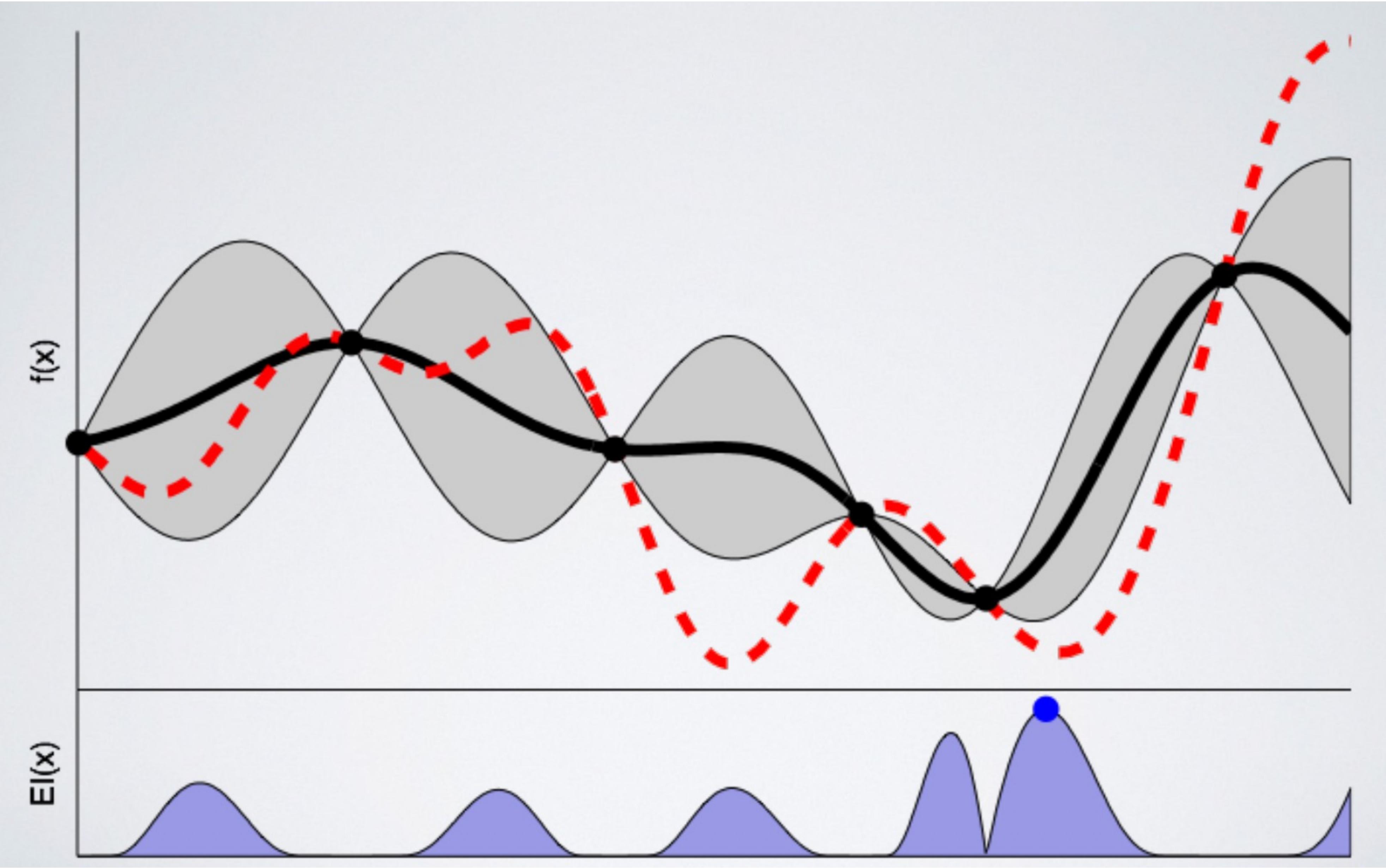
BO example (4)



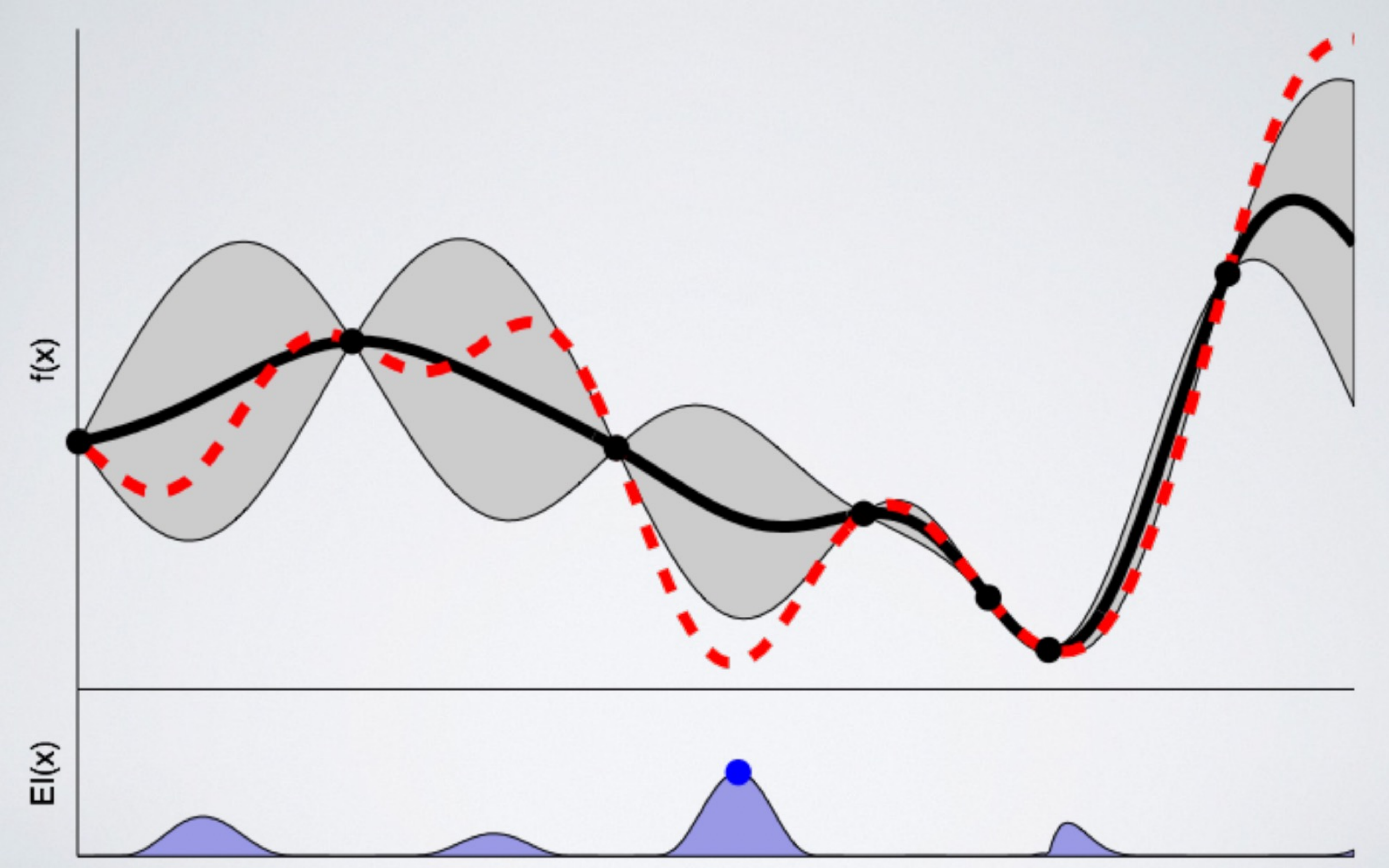
BO example (5)



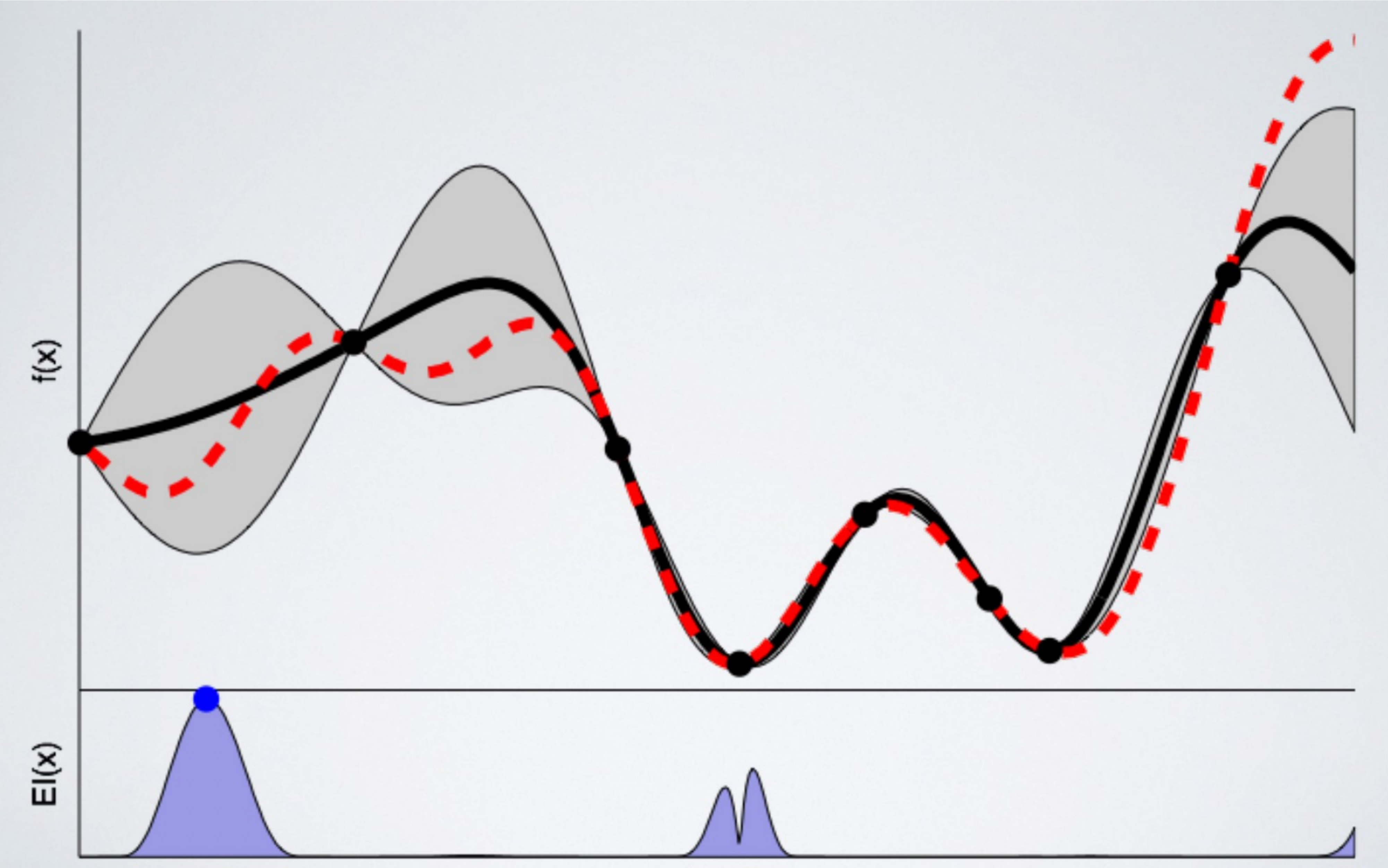
BO example (6)



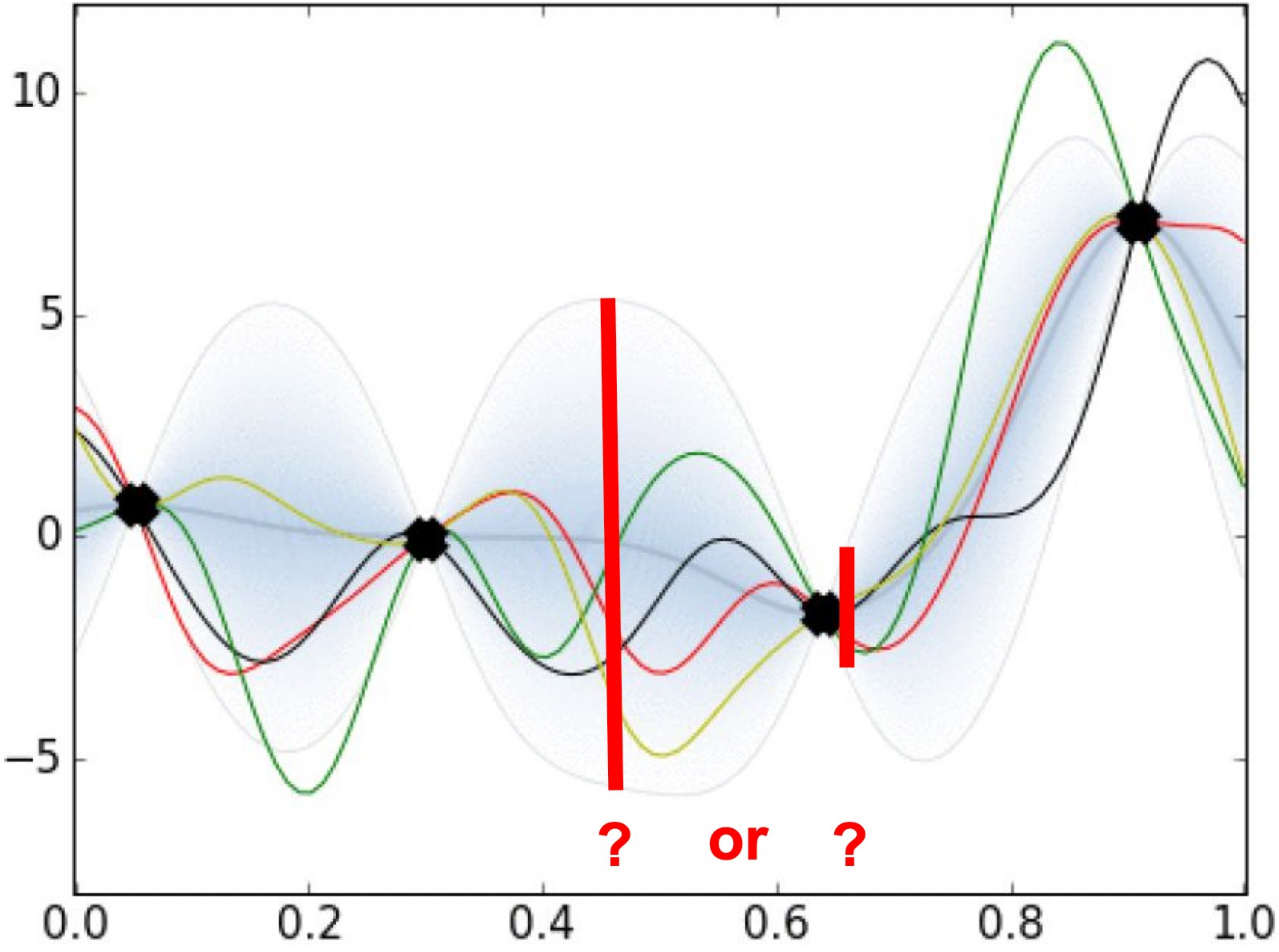
BO example (7)



BO example (8)

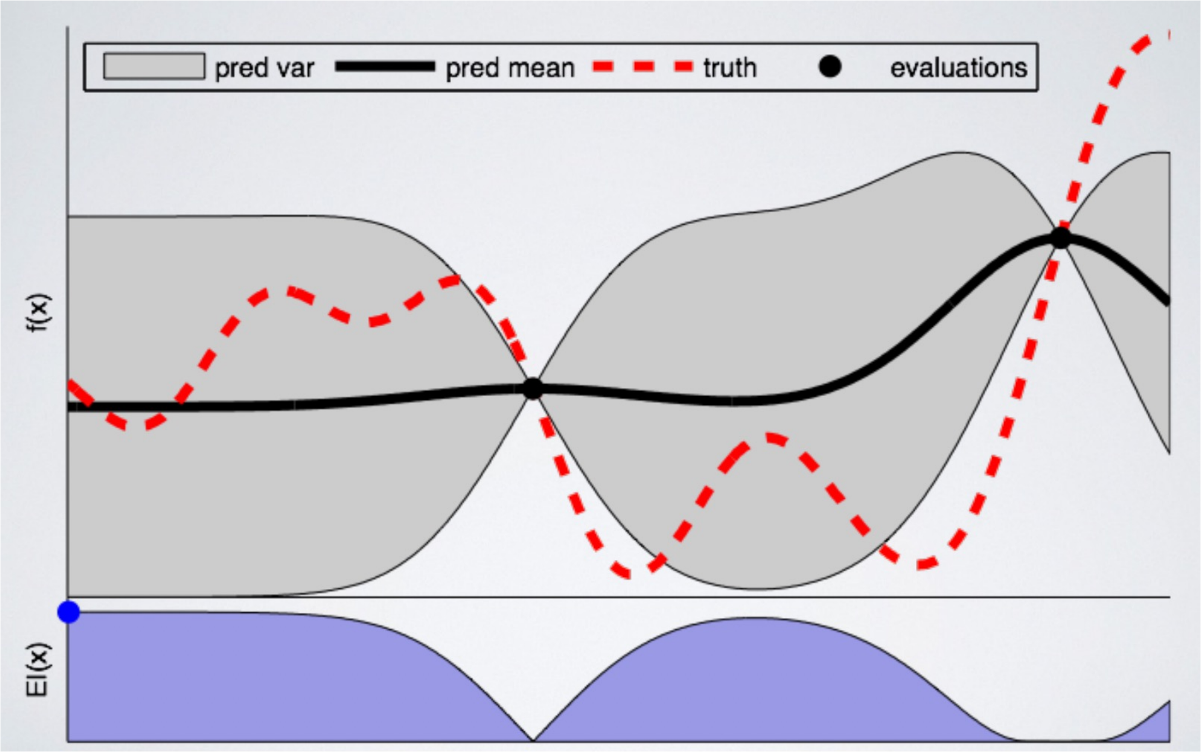


Choosing the next point

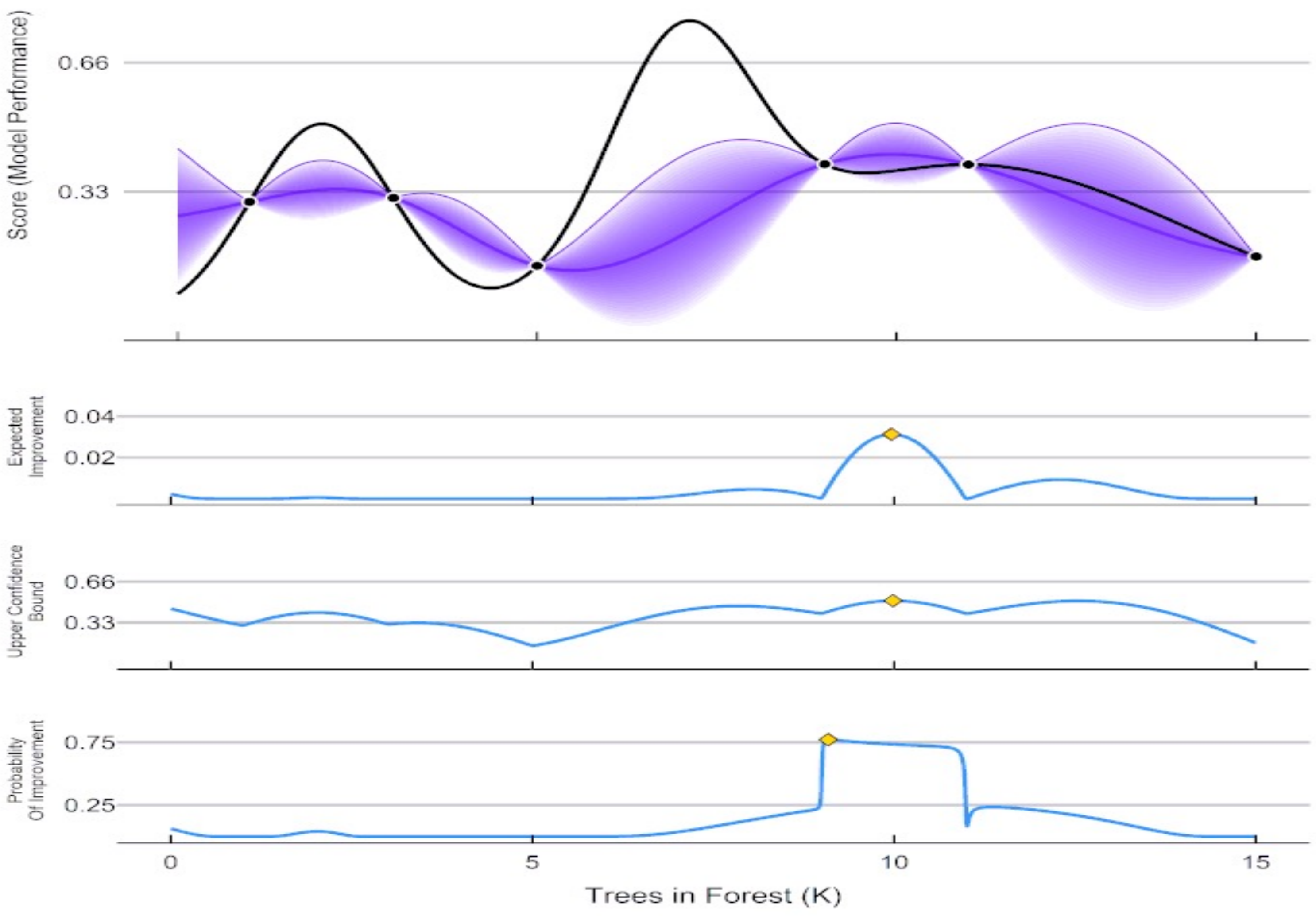


Acquisition functions

- How to choose the next point to evaluate?
  - Acquisition function
- Probability of improvement
- Expected improvement
- Upper confidence bound
- Lower confidence bound
- $\boldsymbol{LCB(x) = f(x) - k * \sigma(f(x))}$



ParBayesianOptimization in Action (Round 1)



BO pseudo code

Algorithm 1 Basic pseudo-code for Bayesian optimization

Place a Gaussian process prior on f

Observe f at n_0 points according to an initial space-filling experimental design. Set $n = n_0$.

while $n \leq N$ **do**

  Update the posterior probability distribution on f using all available data

  Let x_n be a maximizer of the acquisition function over x, where the acquisition function is computed using the current posterior distribution.

  Observe $y_n = f(x_n)$.

  Increment n

end while

Return a solution: either the point evaluated with the largest $f(x)$, or the point with the largest posterior mean.

When do we use BO

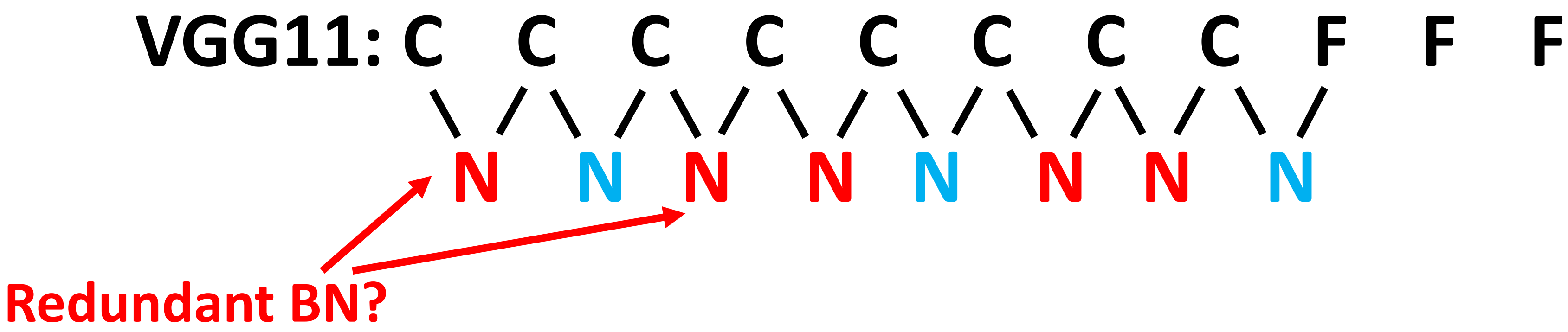
- The input x is in R^d for a value of d that is not too large. Typically, $d \leq 20$
- The objective function f is continuous. This will typically be required to model f using Gaussian process.
- f is "expensive to evaluate" in the sense that the number of evaluations that may be performed is limited (BO is sample efficient).
  - Evaluation takes a substantial amount of time (typically hours),
  - Each evaluation bears a monetary cost (e.g., from purchasing cloud computing power, or buying laboratory materials),
  - Each evaluation bears opportunity cost (e.g., if evaluating f requires asking a human subject questions who will tolerate only a limited number)

- **Applications:**
  - **Automated ML**
  - **Drug discovery**
  - **Design optimization, etc.**

Challenges of using BO

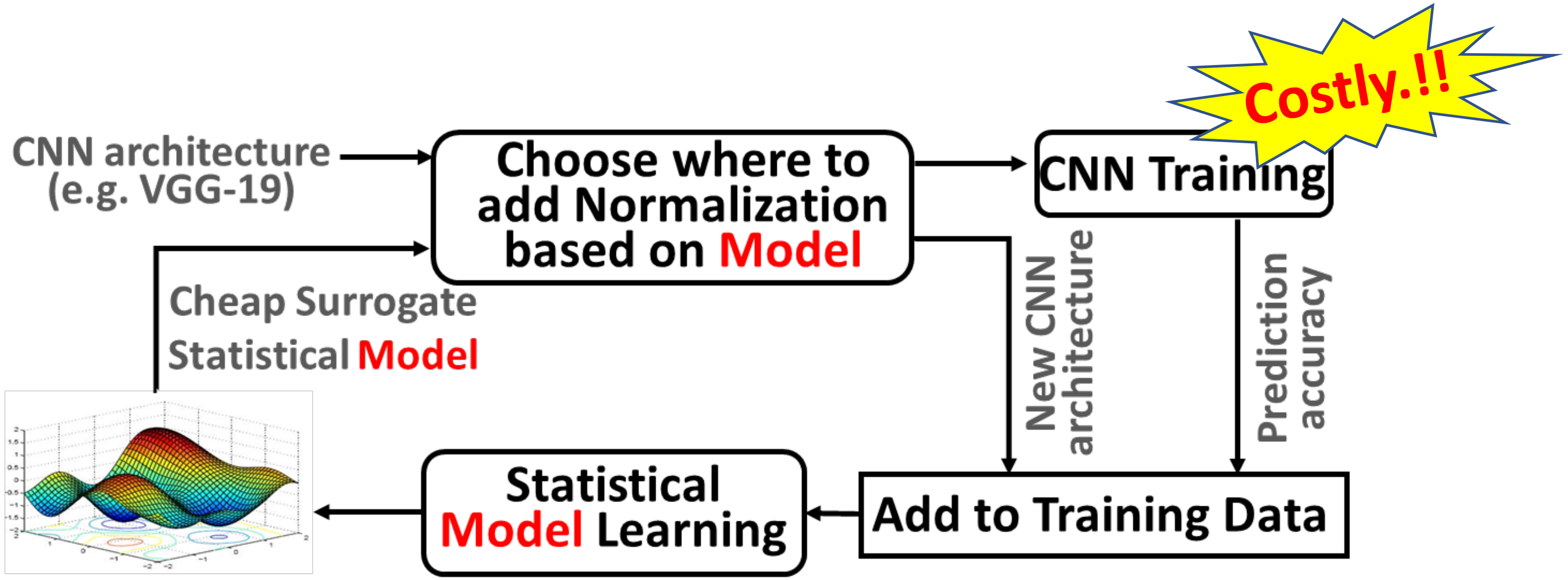
- **Not good if input has many dimensions (>20)**
- **Too slow when multiple dimensions involved.**
  - **Time complexity: O(n^4)**
- **Needs good starting points**
- **Multi-objective, /discrete variants not straightforward**

Our problem here



- Not all BN layers are necessary
- Remove some BN layers to accelerate analog computation
- Use BO to find the BN layers

BO to find the BN locations



- Prepare initializing data (N CNN configuration and accuracy pairs)
- Use surrogate model and acquisition function to choose next configuration
- Evaluate new configuration
- Improve the surrogate function
- Repeat the steps until MAX iterations