NEURAL INFORMATION PROCESSING SYSTEMS

Adaptive Kernel Design for Bayesian Optimization Is a Piece of CAKE with LLMs

Richard Cornelius Suwandi, Feng Yin, Juntao Wang, Renjie Li,
Tsung-Hui Chang, Sergios Theodoridis

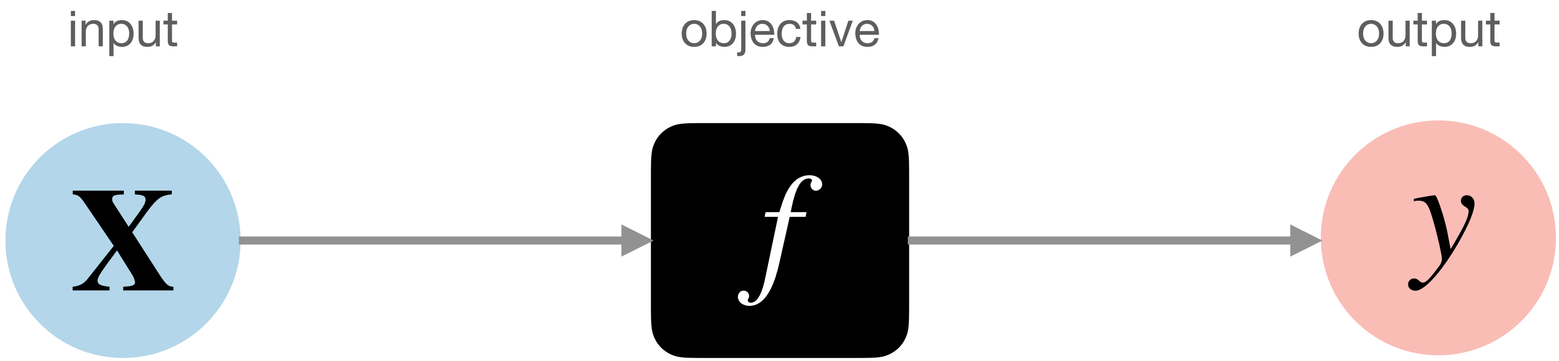
香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

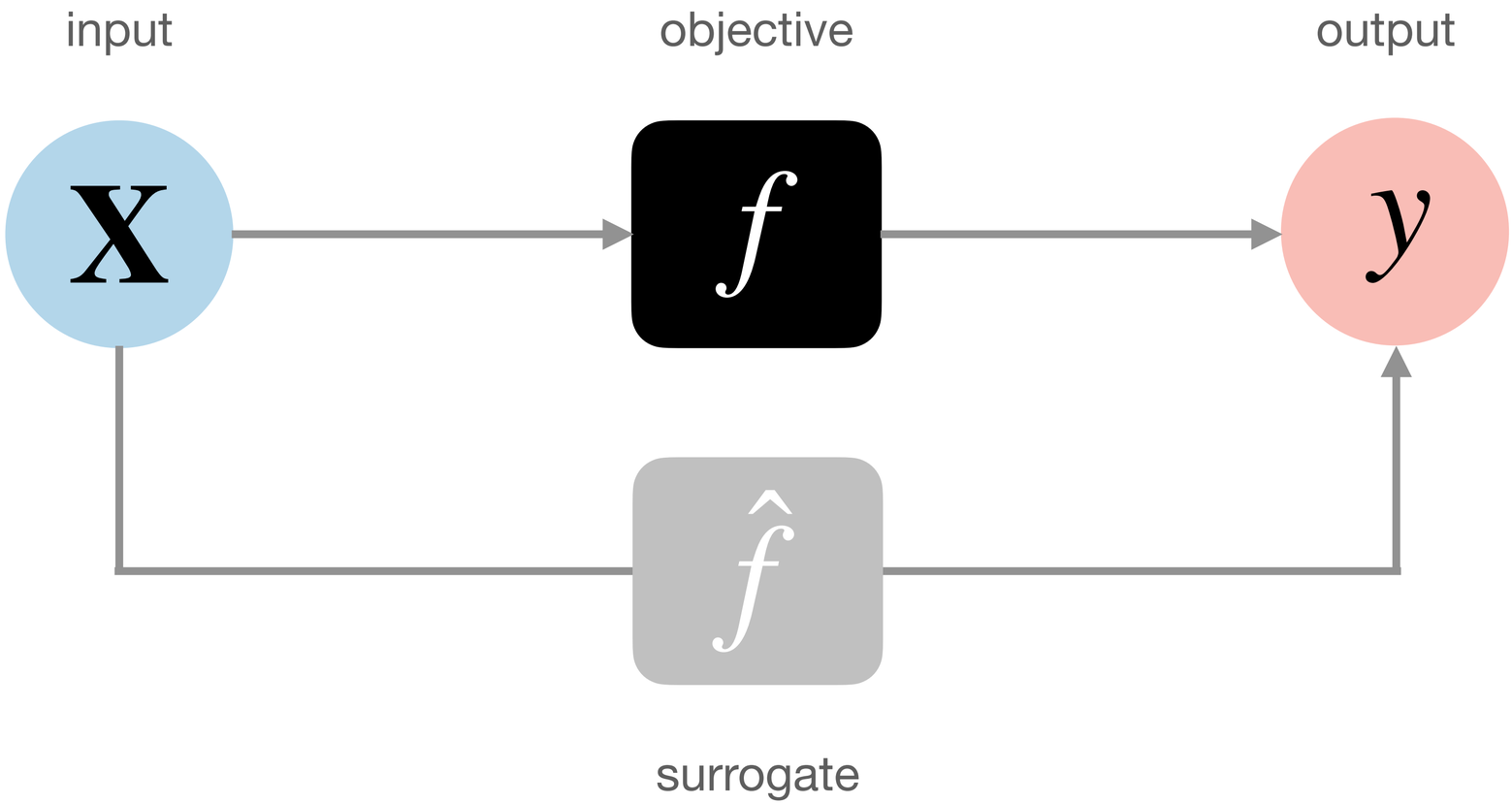
Black-box optimization

Consider a **black-box** objective that is **expensive-to-evaluate**



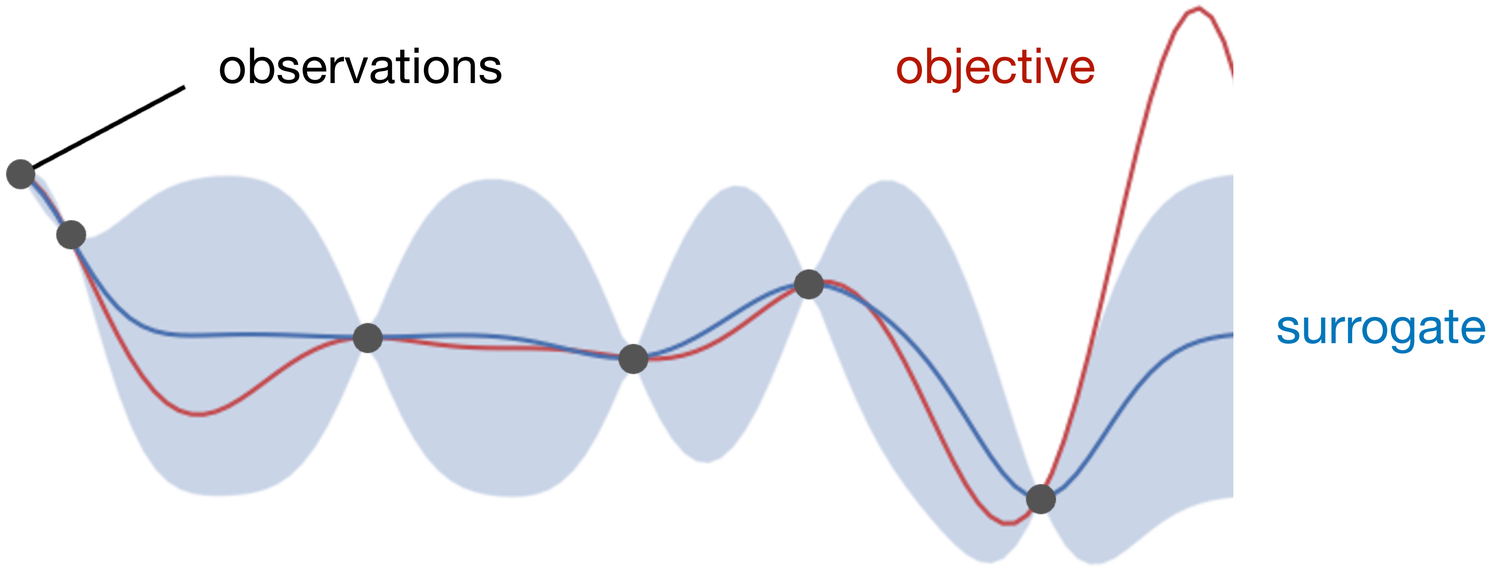
Black-box optimization

Estimate the objective with a **surrogate**



Bayesian optimization

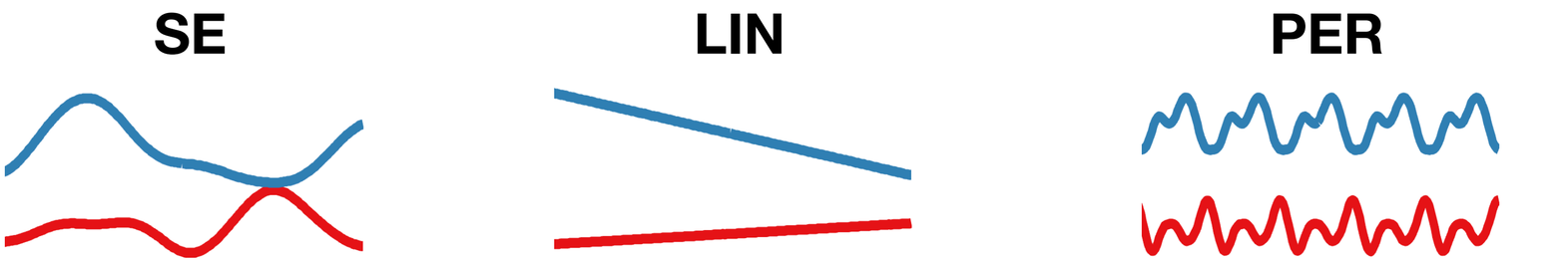
Use a **Gaussian process (GP)** surrogate: $\hat{f} \sim \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$



The behavior of a GP *heavily* depends on the **kernel choice**

Kernel design problem

- Different kernel encodes *different* assumptions:



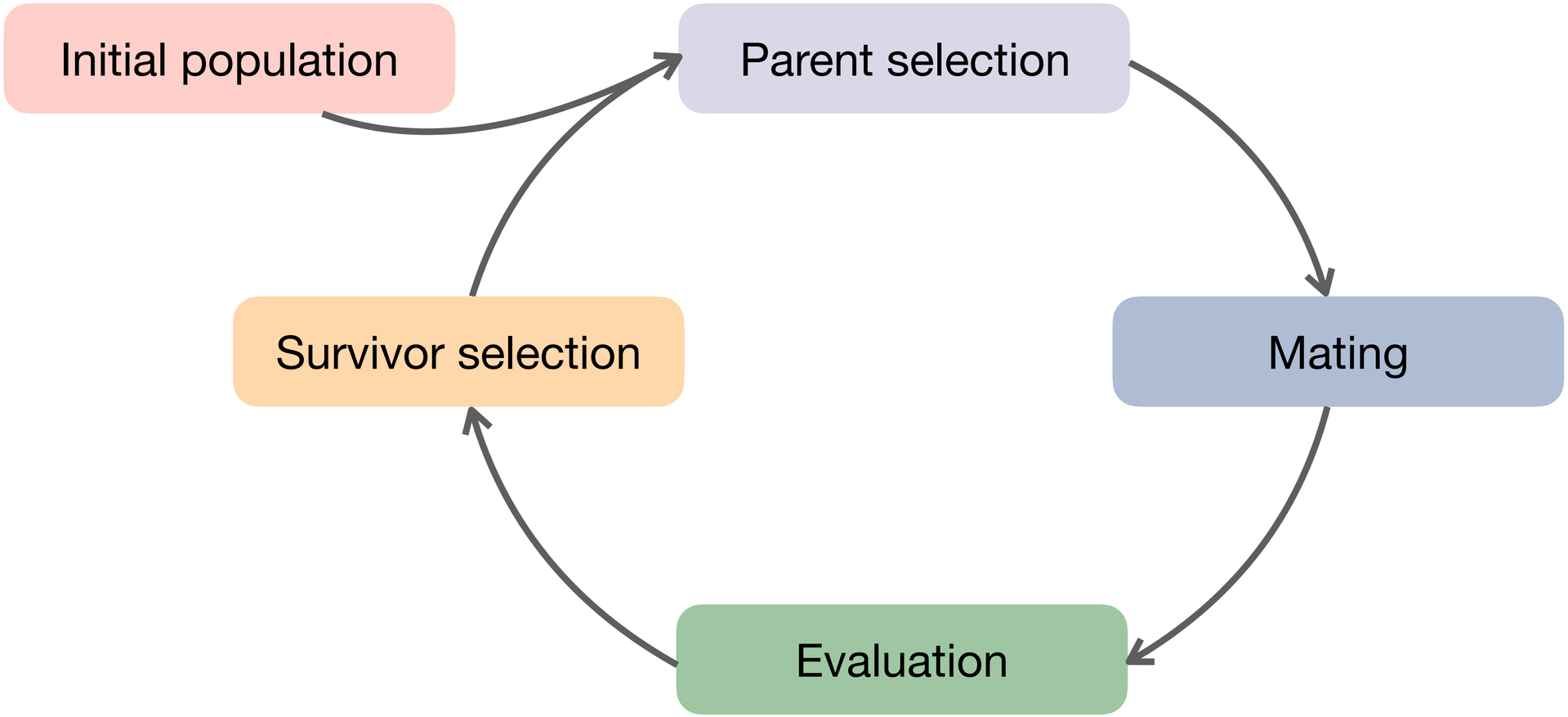
local variation | linear trend | repeating structure

- A poor kernel choice can **hinder convergence**

Which kernel should we use given a problem?

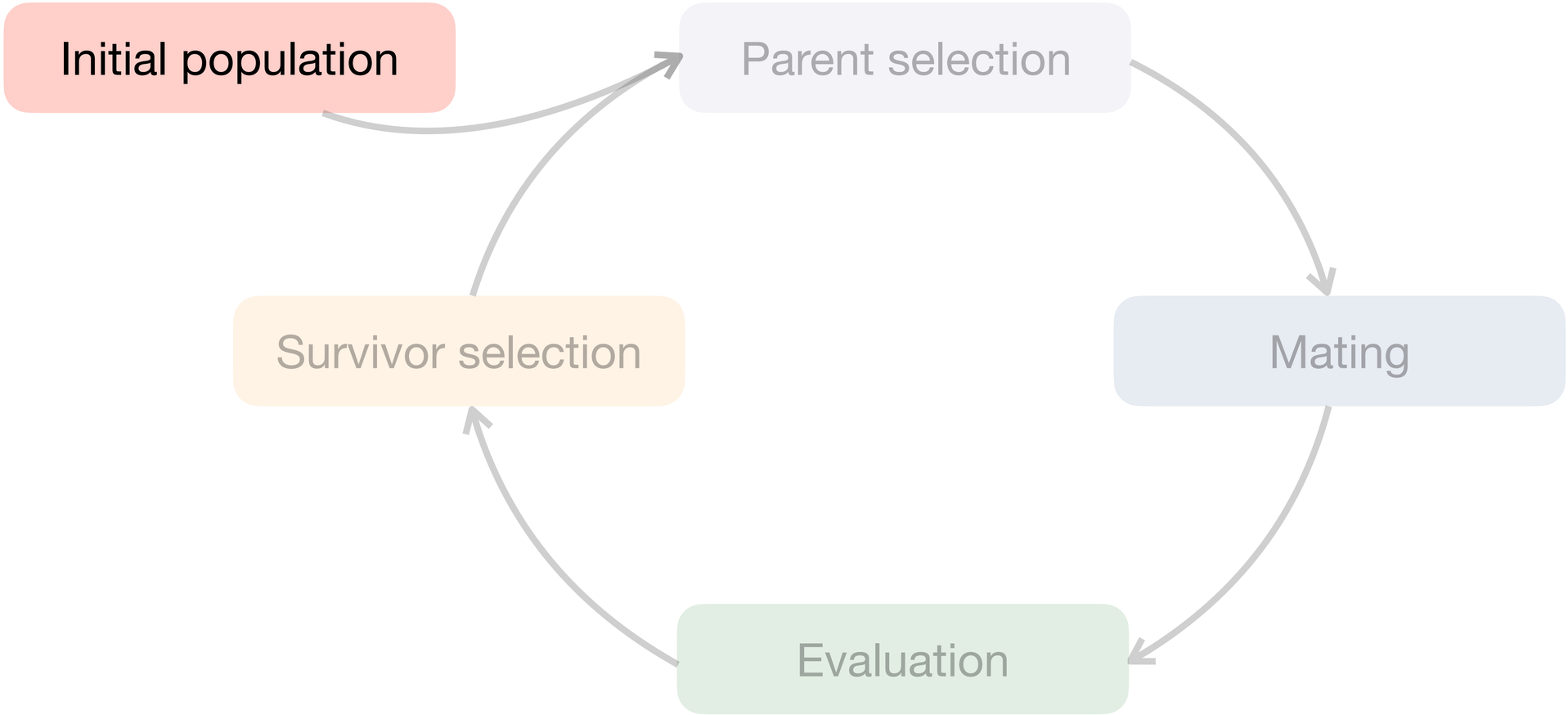
Context-Aware Kernel Evolution (CAKE)

Reframe the kernel design problem as an **evolutionary process**



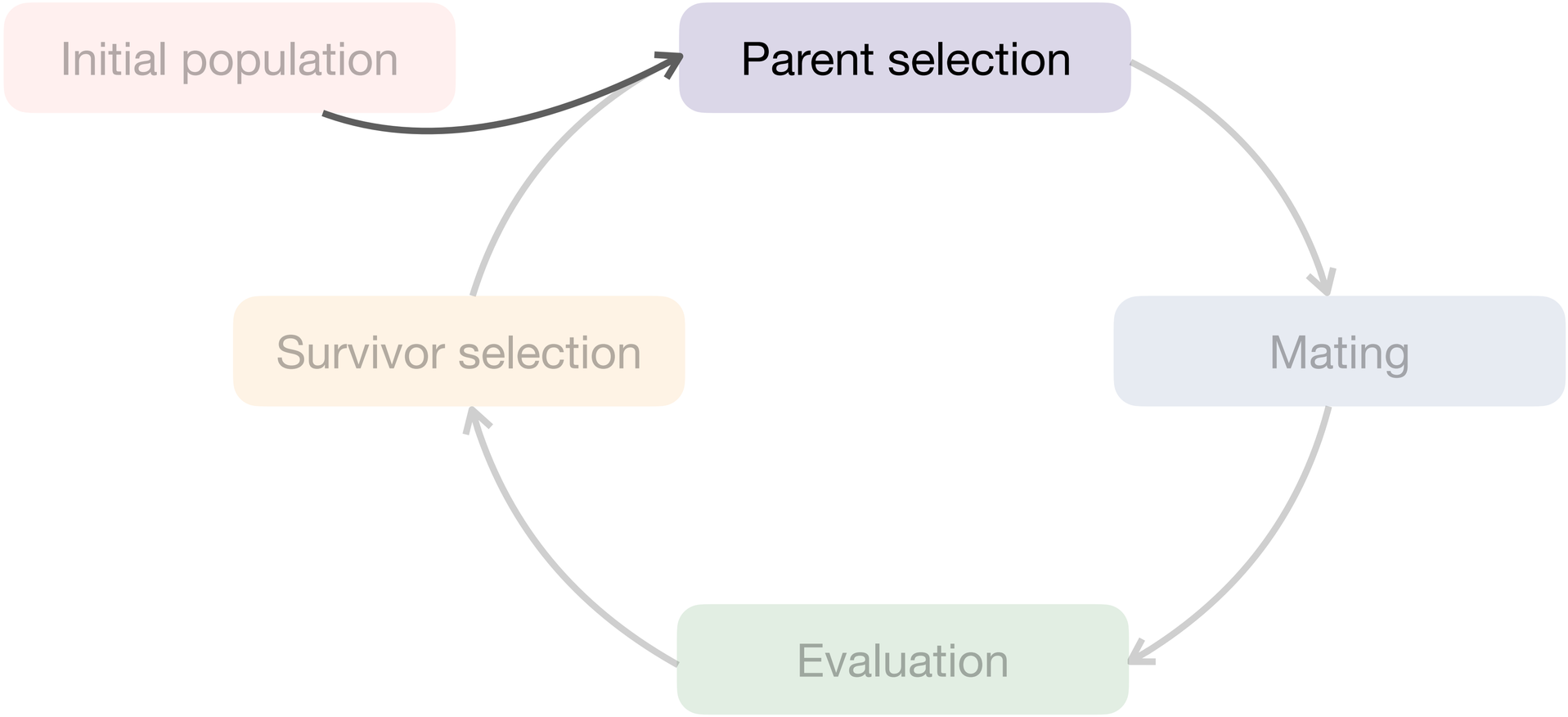
Context-Aware Kernel Evolution (CAKE)

Start with a set of **base kernels** $\mathbb{K} = \{\text{SE, LIN, PER, RQ}\}$



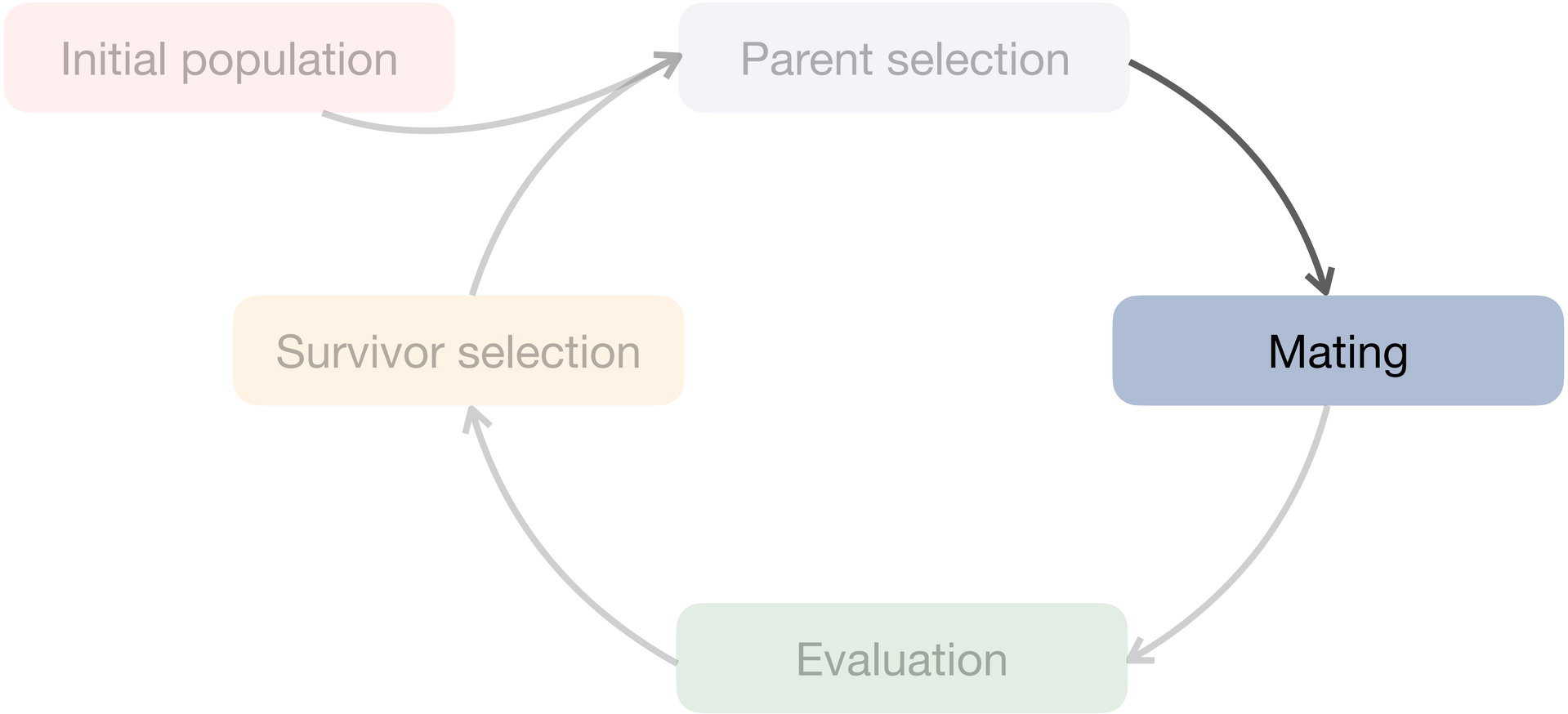
Context-Aware Kernel Evolution (CAKE)

Sample **parent kernels** k_1, k_2 from the population $\mathbb{K}$



Context-Aware Kernel Evolution (CAKE)

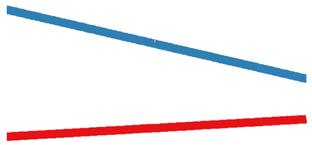
Generate **new kernels** using LLM as genetic operator



Mating

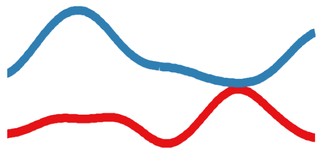
Perform **crossover** to generate new offspring kernel

LIN



linear trend

SE

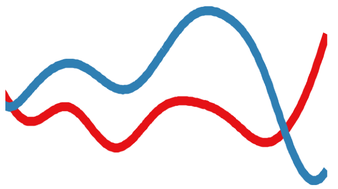


local variation

LLM

"Combine the parent kernels to improve model fit..."

LIN × SE

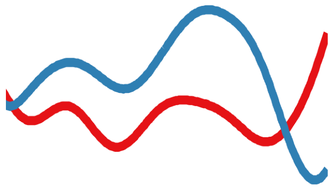


variation with trend

Mating

Perform **mutation** to refine the kernel

LIN × SE

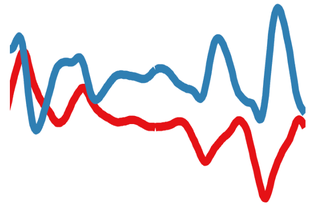


variation with trend

LLM

"Modify the given kernel to improve model fit..."

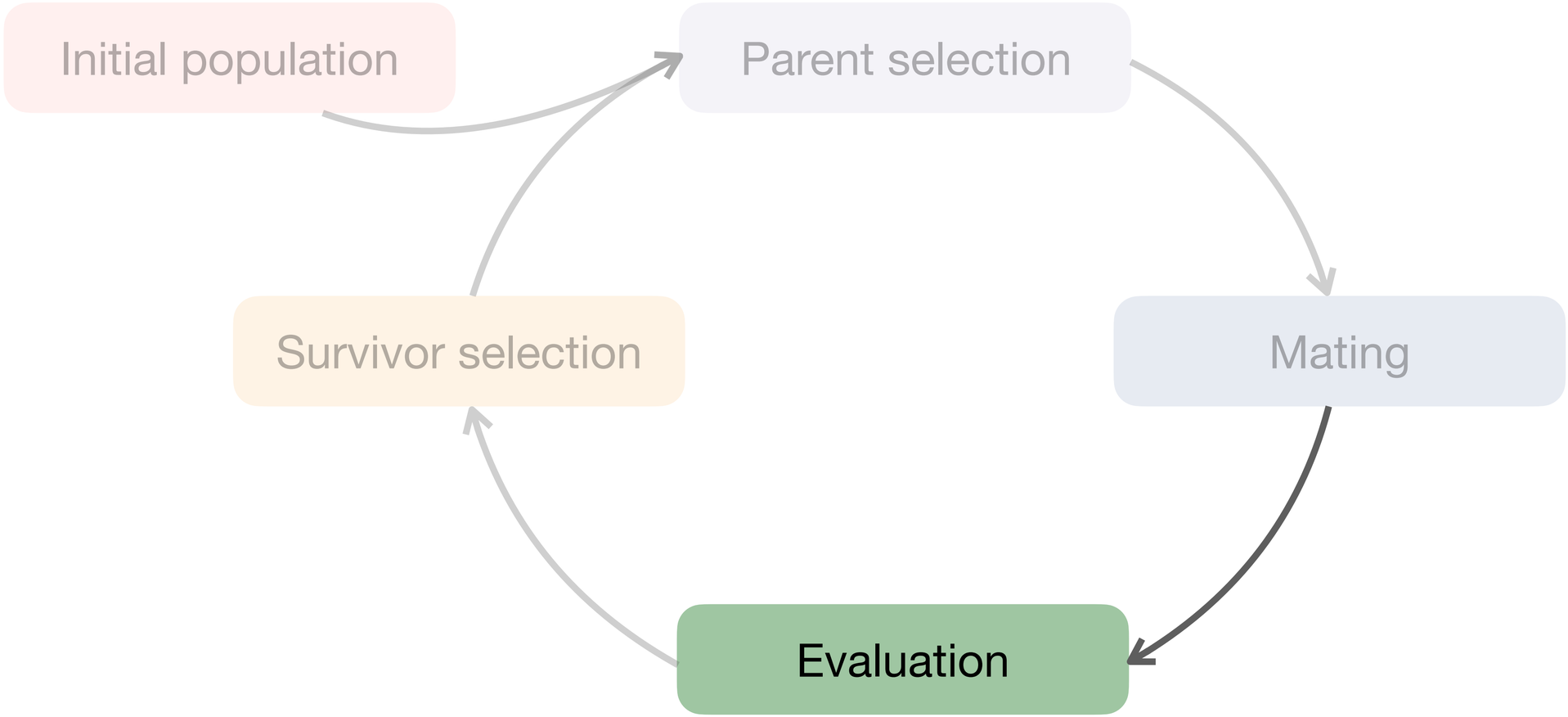
LIN × PER



growing amplitude

Context-Aware Kernel Evolution (CAKE)

Measure the **fitness scores** of each kernel in the population



Evaluation

Measure fitness using the normalized **Bayesian information criterion (BIC)**:

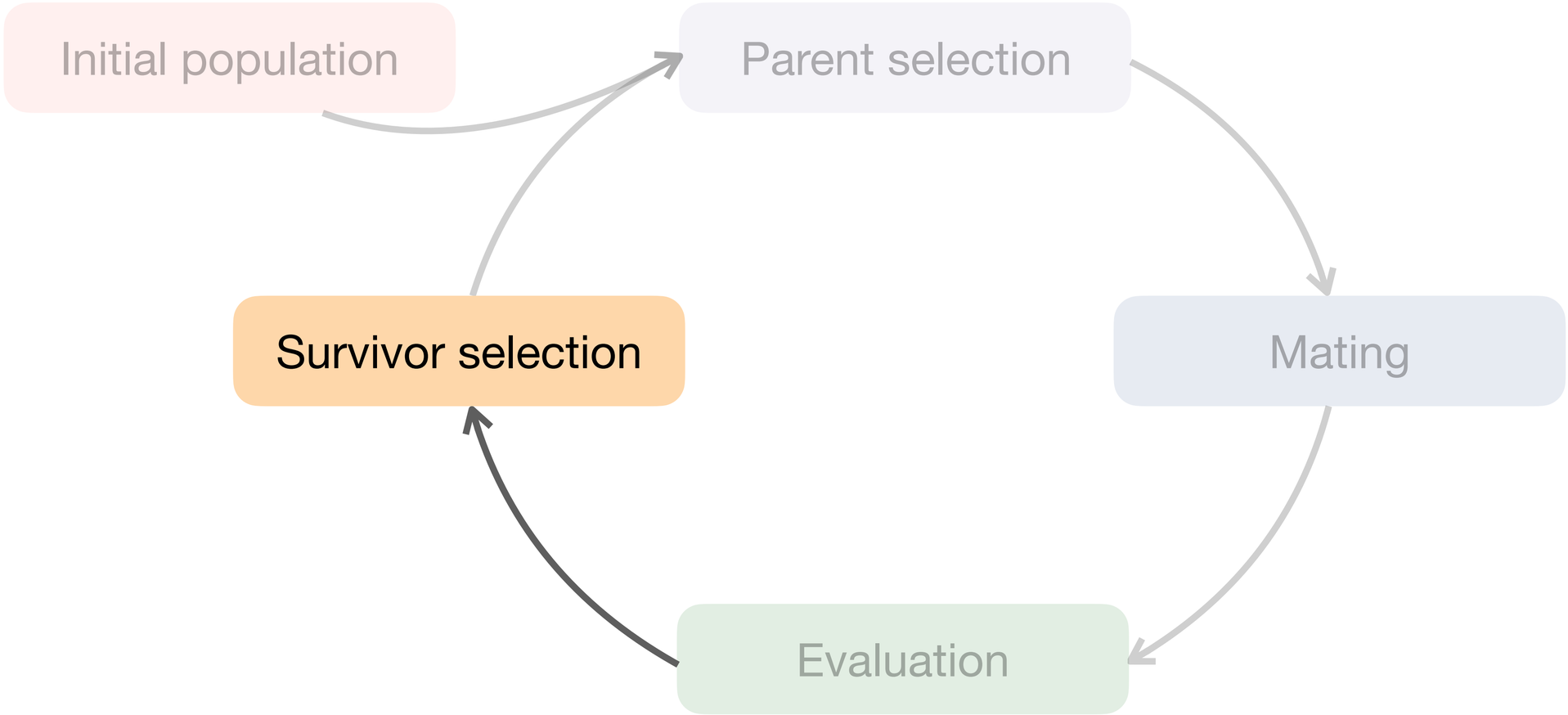
$$\text{fitness}_k = \frac{\exp(-\text{BIC}_k)}{\sum_{k'} \exp(-\text{BIC}_{k'})}$$

where

$$\text{BIC}_k = \underbrace{-2 \log L_k}_{\text{model fit}} + \underbrace{p_k \log n}_{\text{model complexity}}$$

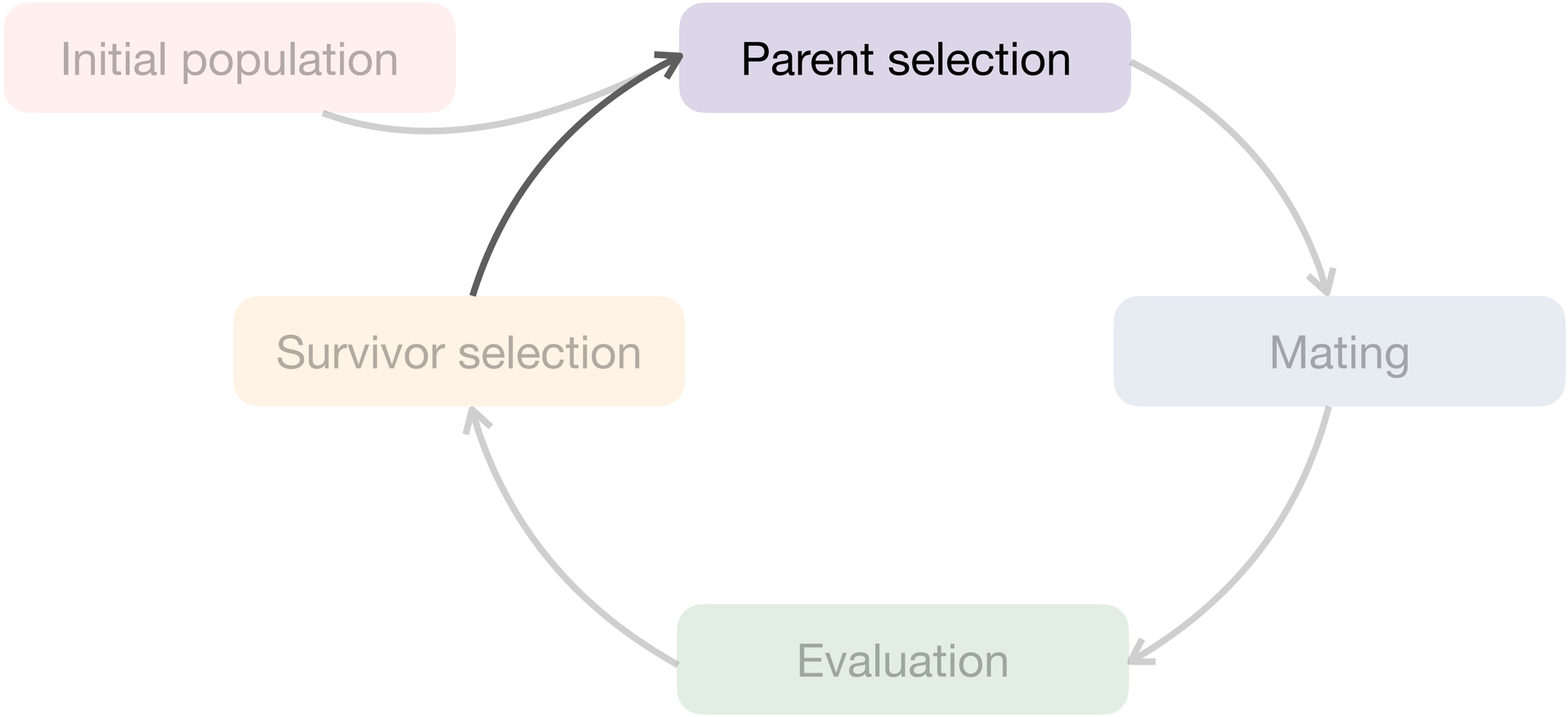
Context-Aware Kernel Evolution (CAKE)

Select the **top performing kernels** for the next generation

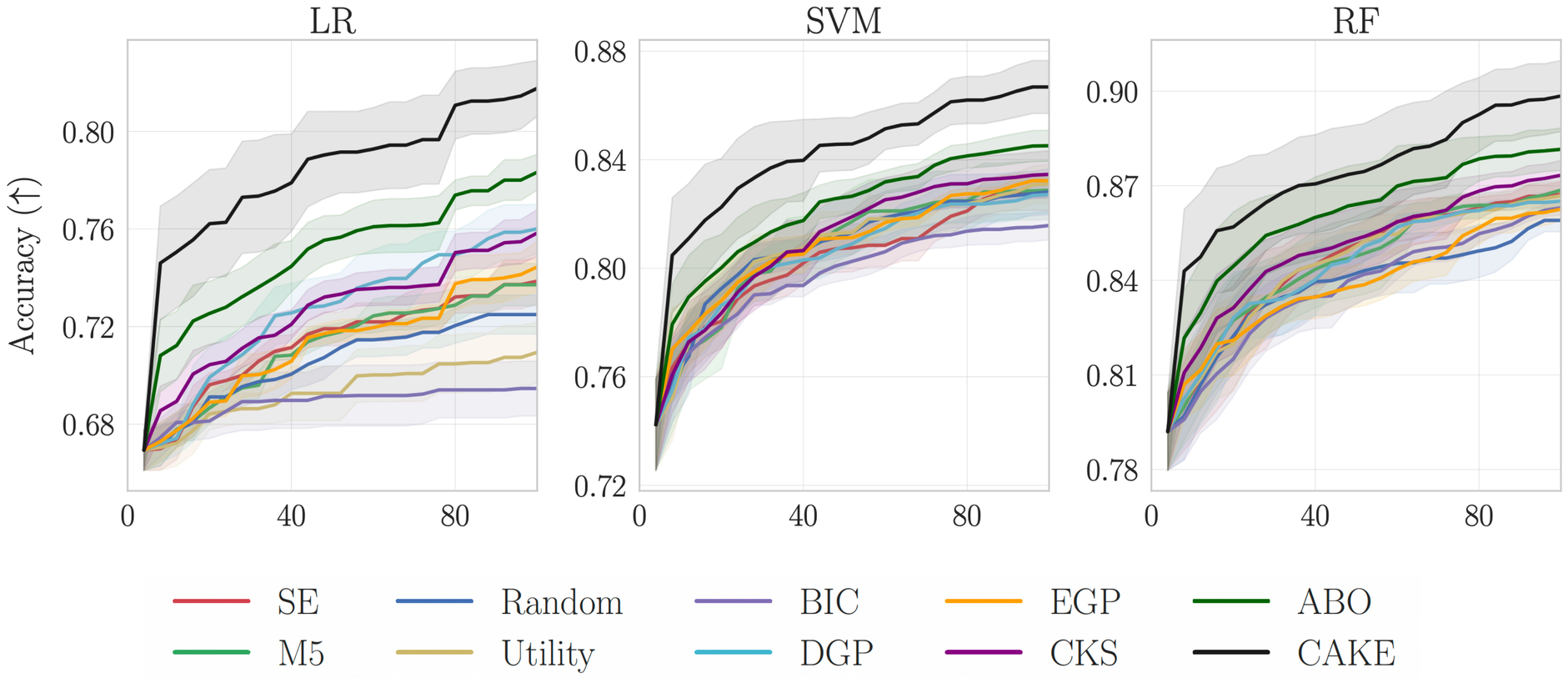


Context-Aware Kernel Evolution (CAKE)

Repeat until budget is exhausted

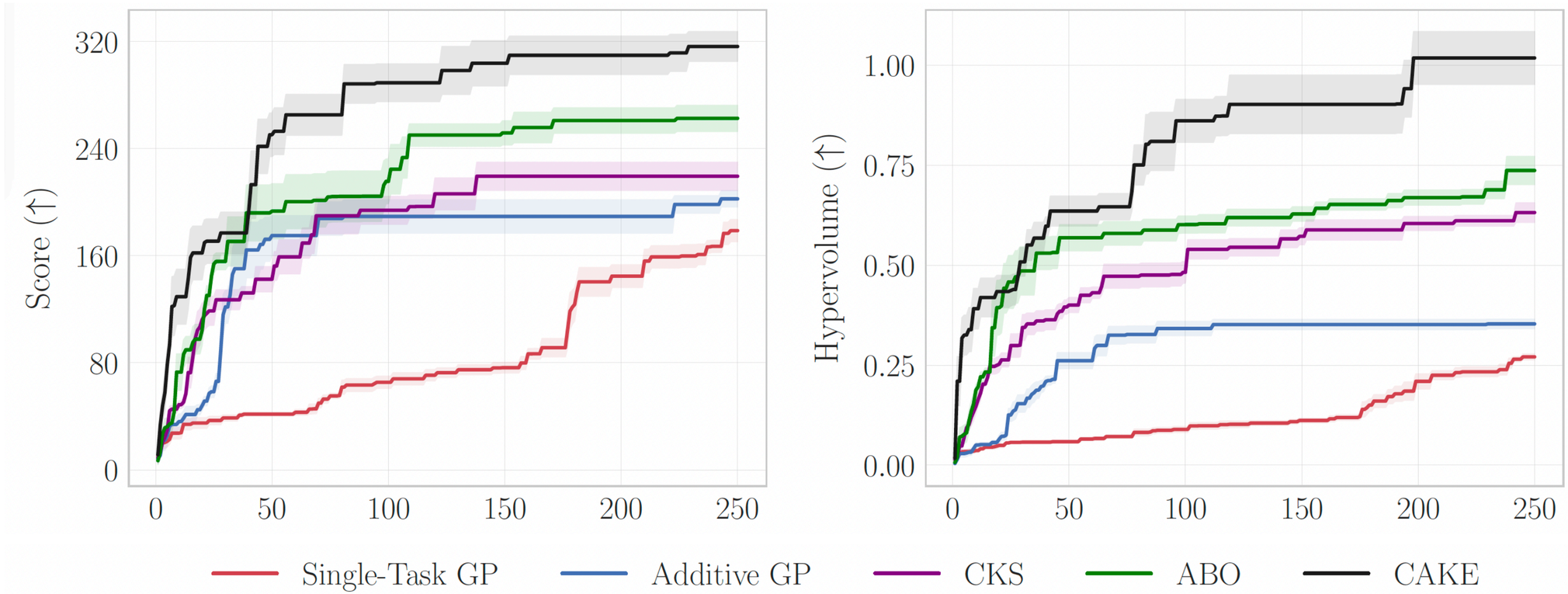


Hyperparameter optimization



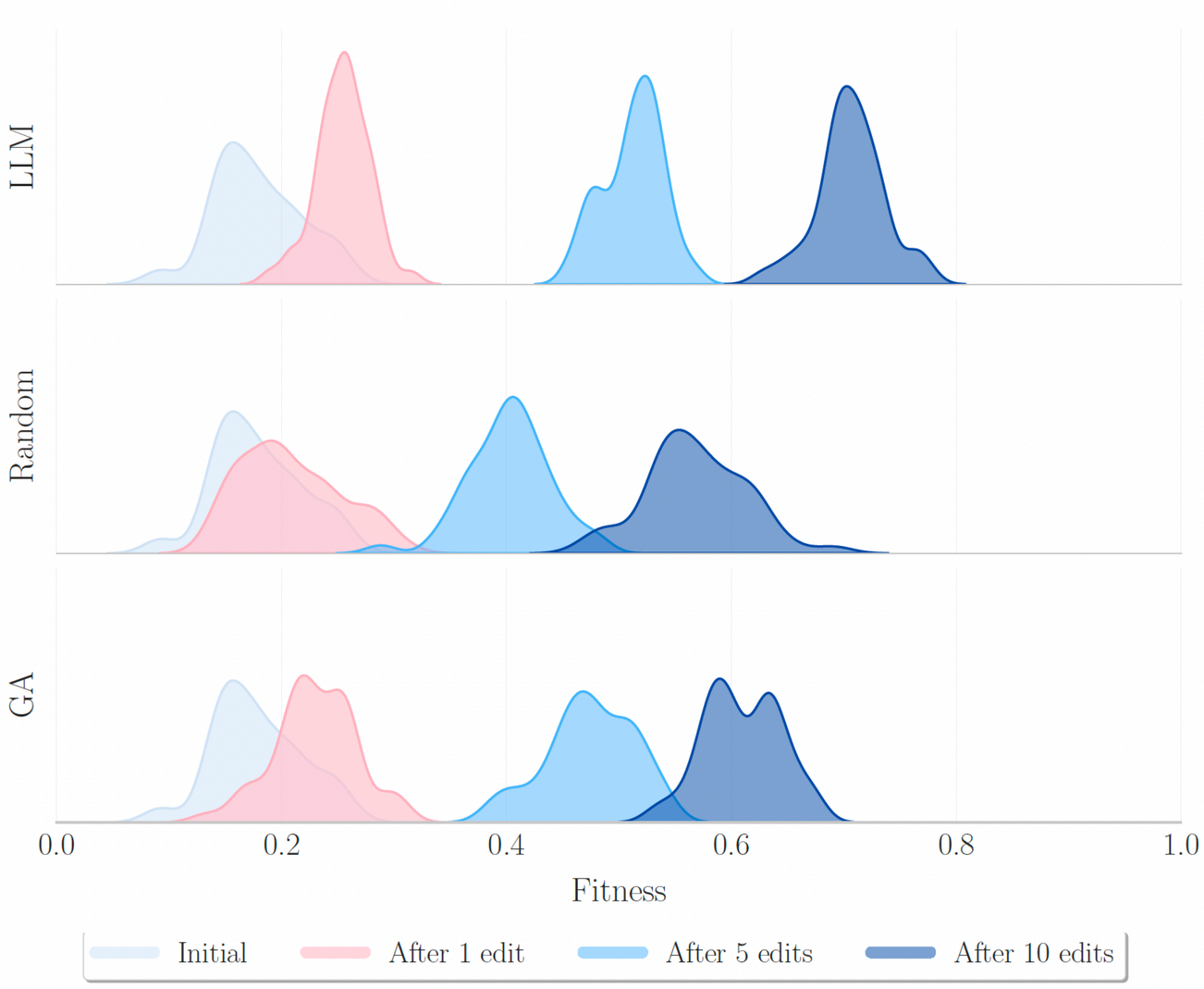
- Consistently achieved **highest test accuracy** on average
- Showed 67.5% of total improvement **within just 25% of the budget**

Photonic chip design



- Effectively **balanced trade-offs** between objectives
- Demonstrated **10× speedup** in finding high-quality solutions

Analysis



- After just 1 LLM edit, the distribution immediately shifts toward **higher values**
- LLM edits achieve **faster convergence** and **higher fitness** over time

Thanks for listening!

https://arxiv.org/abs/2509.17998

https://github.com/richardcsuwandi/cake

richardsuwandi@link.cuhk.edu.cn