# Divide and Conquer : Local Gaussian Processes to design Covalent Organic Frameworks for Methane Deliverable Capacity

Nikhil Thota, Maitreyee Sharma Priyadarshini, Gigi Wang, Jarett Ren

March 28, 2024

In this project, we explore the use of local Gaussian Process models to accelerate materials discovery when the search spaces are very large. We evaluate the performance of the framework on a covalent organic framework (COF) dataset that consists of 69,840 2D and 3D COFs. [1] This dataset replicates some real-world scenarios wherein the search space to explore is very large. In this test, we used an initial training dataset of 680 points, *i.e.,* 10% of the total search space. These COF structures are designed for methane storage and our optimization target here is the deliverable capacity (v STP/v) of the COF structure.

Gaussian Process (GP) has been a popular choice of surrogate model in Bayesian Optimization due to its flexibility and uncertainty quantification. However, training a Gaussian Process involves several matrix inversions, which can dramatically scale up the computational cost as more data is obtained via Bayesian Optimization. Gaussian Process has a runtime complexity of $O(n^3)$, where $n$ is the number of training samples. Given its poor scalability, the application of GPs to high-dimensional problems with several thousand observations remains challenging. In this project, we aim to reduce the computational cost of GP-based Bayesian Optimization by breaking a global GP model into several local GP models. These local GP models will run in parallel, accelerating the optimization problem. We have also designed a new acquisition function to aggregate predictions from local GPs and select the next points to explore with a tunable parameter to adjust exploration vs. exploitation. We hypothesize that our method will significantly accelerate the runtime of Bayesian Optimization, and enable us to explore more points in the COFs dataset which cannot be done with a standard global GP model.

## References

[1] Rocío Mercado, Rueih-Sheng Fu, Aliaksandr V Yakutovich, Leopold Talirz, Maciej Haranczyk, and Berend Smit. In Silico Design of 2D and 3D Covalent Organic Frameworks for Methane Storage Applications. *Chemistry of Materials*, 30(15):5069–5086, 2018.

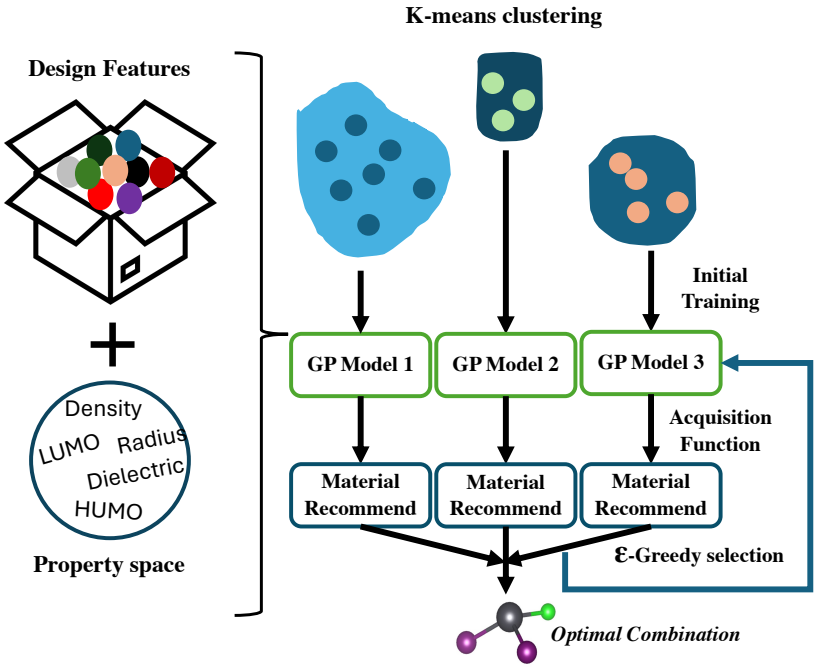


Figure 1: Workflow of our divide and conquer strategy