Karlsruhe Institute of Technology

Institute for Applied Materials – Microstructure Modelling and Simulation

Bayesian optimization framework for data-driven materials design

G. Tosato[1], A. Koeppe[1], M. Selzer[1,2] and B. Nestler[1,2]

[1]Institute for Applied Materials – Microstructure Modelling and Simulation (IAM-MMS), Karlsruhe Institute of Technology (KIT), Straße am Forum 7, 76131 Karlsruhe, Germany
[2]Institute for Digital Materials Science (IDM), Hochschule Karlsruhe - University of Applied Sciences (HKA), Moltkestrasse 30, 76133 Karlsruhe, Germany

Introduction

Exploration of a large design space
In the context of materials design, microstructure parameters have a major influence on the resulting properties of the material. Multiple interlinked parameters can be tuned within a large design space.

Generation of machine learning training data
Generating enough data for ML while covering the parameter space sufficiently is challenging when dealing with expensive simulations and/or experiments.

Our strategy
Design of experiments methodology utilizing Bayesian Optimization:
- efficiently explore and exploit the search space,
- minimize number of required evaluations,
- adaptively select the next most informative evaluation to perform,
- learn from the already acquired dataset.

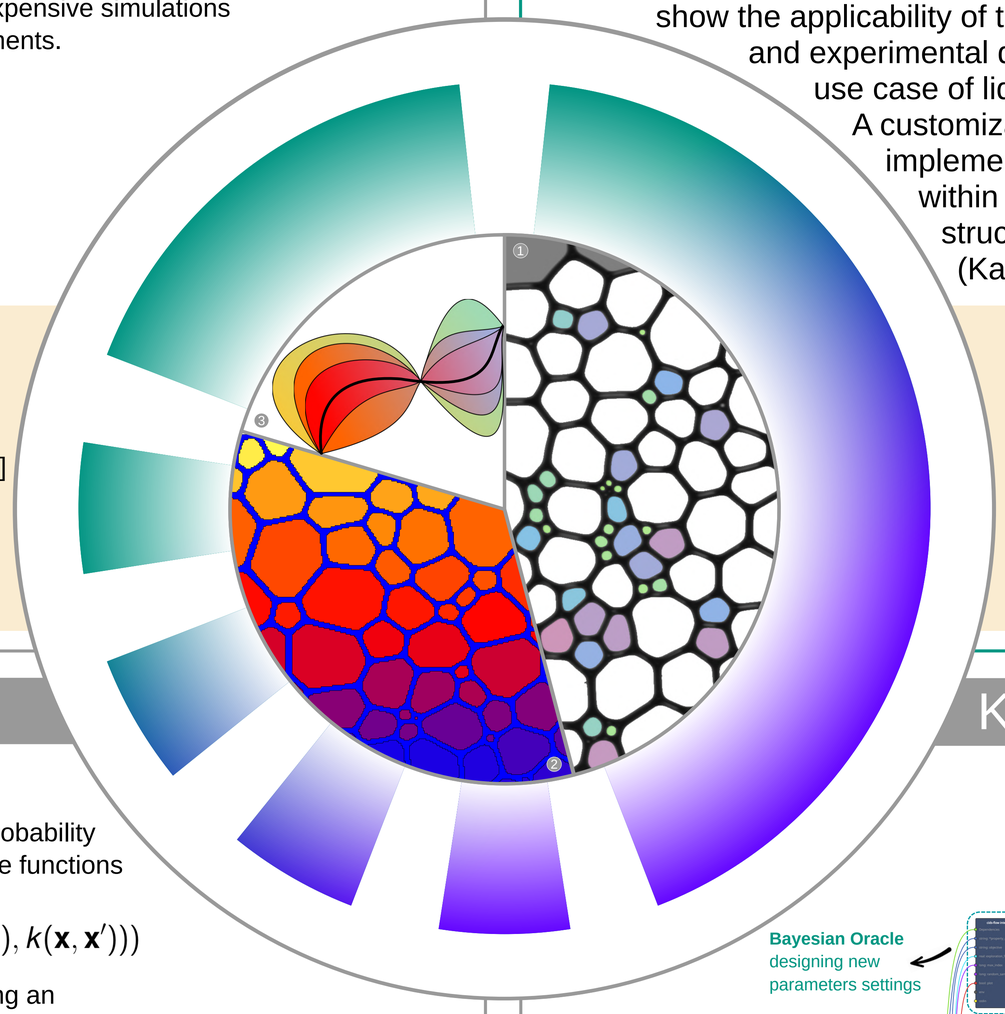
In-house developed workspace

Kadi4Mat
open-source platform for FAIR research data management [1,2]

KadiStudio
desktop application for modelling scientific processes as workflows [3]

KadiAI
interface between Kadi4Mat and machine learning tools

cids tools
Computational Intelligence and Data Science framework

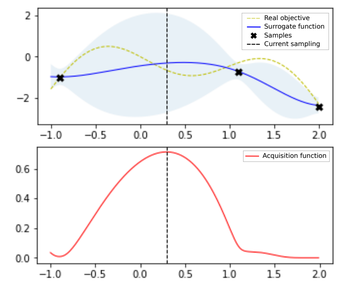


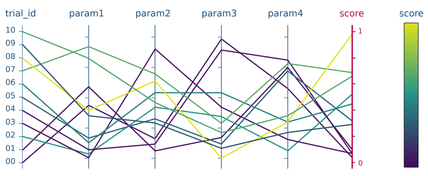
Main takeaways

In our work, we propose a methodology for generating small but information-rich datasets through the use of Bayesian Optimization and its memory of previously acquired data. To show the applicability of the method to both simulative and experimental datasets, we apply it to the use case of liquid foam optimization.
A customizable framework has been implemented as workflow nodes within the Karlsruhe Data Infrastructure for Materials Science (Kadi4Mat).

Bayesian Optimization (BO)[3]

Algorithm





Properties

- uses **Gaussian Process (GP):** probability distribution over cheaper surrogate functions of the objective

$$f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

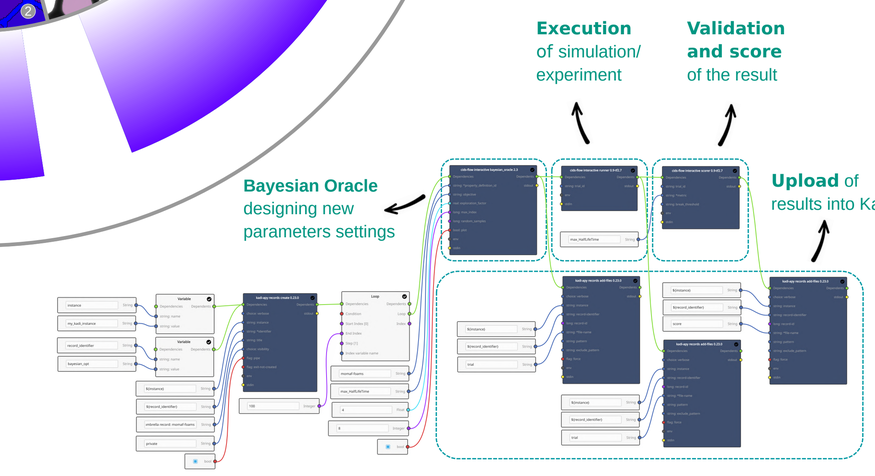
- selects next sample by maximizing an **acquisition function** used to:
  - score the results given by the surrogate
  - balance exploration and exploitation

$$u(x) = \mu(x) + \beta\sigma(x)$$

e.g. Upper Confidence Bound

- extension to multidimensional case and visualization available in Kadi4Mat.

Kadi-workflow

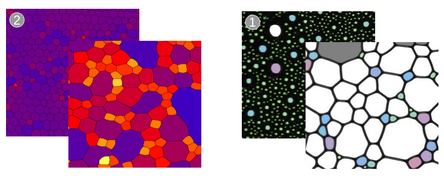


Example workflow implementation of the BO framework within KadiStudio.

Maximizing efficiency with fully-automated labs

Application:
Liquid Foams design

Optimization of foam stability is possible providing control of many interlinked parameters
⟶ large search space
⟶ partially unknown relationships between parameters



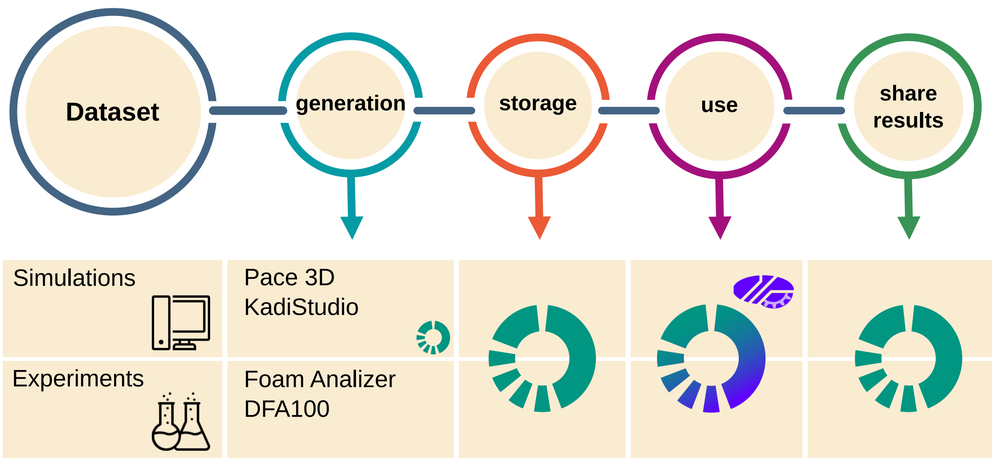
Simulation Experiment

Integrated environment combining simulative and experimental datasets

Hybrid model that combines and learns both from simulative and experimental data.

Opportunity towards accelleration of research process through automated laboratories.

Autonomous iteration of:
- machine learning based design
- execution (in the lab or digital simulation)
- validation
- uploading in Kadi4Mat



Acknowledgement and references.

[1] https://kadi4mat.iam-cms.kit.edu/
[2] Brandt, N., Griem, L., Herrmann, C., Schoof, E., Tosato, G., Zhao, Y., Zschumme, P. and Selzer, M., 2021. Kadi4Mat: A research data infrastructure for materials science. Data Science Journal, 20(1).
[3] Griem, L, Zschumme, P, Laqua, M, Brandt, N, Schoof, E, Altschuh, P and Selzer, M. 2022. KadiStudio: FAIR Modelling of Scientific Research Processes. Data Science Journal, 19: XX, pp. 1–17. DOI: https://doi.org/10.5334/dsj-2022-017
A special thanks to Jana Holland-Cunz for the design consultation.

MOMAF

This work is funded by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK-BW), in the project MoMaF--Science Data Center (grant number: 34-7547.222).
This work is funded by the BMBF and MWK-BW as part of the Excellence Strategy of the German Federal and State Governments in the project Kadi4X.
This work was [partly] carried out with the support of the Karlsruhe Nano Micro Facility (KNMFi, www.knmf.kit.edu), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT, www.kit.edu).

giovanna.tosato@kit.edu
www.iam.kit.edu/mms