Karlsruhe Institute of Technology

Autonome Lernende Roboter (ALR)
Prof. Gerhard Neumann

## Project Type

- [x] Master Thesis
- [ ] Bachelor Thesis
- [ ] Research Project

## Supervisors

Denis Blessing
denis.blessing@kit.edu
Henrik Schopmans
henrik.schopmans@kit.edu

## Difficulty

Algorithmic

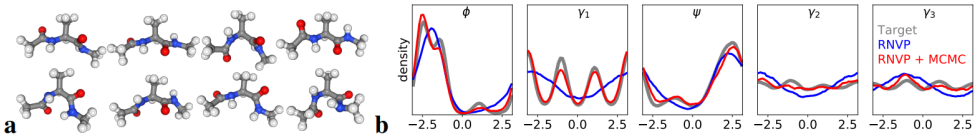
Math

Application

# Datasets for Molecular Geometry Generation from Energy Potentials using Variational Inference

## Description

Molecules play a vital role in a wide range of scientific disciplines, including drug discovery, materials science, and environmental research. The precise generation of molecular geometries is a pivotal task with extensive practical applications. Variational Inference (VI) presents an efficient approach to tackle this challenge by learning generative models. However, existing VI methods for molecular generation have primarily been tested on simple molecules, limiting their applicability to complex chemical systems [1, 2, 3].




This thesis is in cooperation with AIMat, a research group at KIT that is led by Prof. Friederich.

## Tasks

The tasks in this project will involve:

- Benchmark Dataset Creation. Curate a diverse dataset of molecules with varying complexities, properties, and structures.
- Benchmarking. Benchmark recent VI methods on the dataset, quantifying their effectiveness in generating molecules with desired properties.

## Prerequisites

We expect the student to possess the following prerequisites:

- Strong Programming Skills: Proficiency in Python programming is required to undertake this research successfully.
- Machine Learning Expertise: A solid foundation in both theoretical and applied Machine Learning concepts is helpful for this research project.
- Interest in Interdisciplinary Topics: This research project involves interdisciplinary aspects, so a genuine interest in exploring and collaborating on topics beyond Machine Learning is highly valued.

Please note that while we expect proficiency in the aforementioned areas, we do not require prior background knowledge in chemistry for this research.

## References

[1] Laurence Illing Midgley, Vincent Stimper, Gregor NC Simm, Bernhard Schölkopf, and José Miguel Hernández-Lobato. Flow annealed importance sampling bootstrap. *arXiv preprint arXiv:2208.01893*, 2022.

[2] Hao Wu, Jonas Köhler, and Frank Noé. Stochastic normalizing flows. *Advances in Neural Information Processing Systems*, 33:5933–5944, 2020.

[3] Qinsheng Zhang and Yongxin Chen. Path integral sampler: a stochastic control approach for sampling. *arXiv preprint arXiv:2111.15141*, 2021.