

Industry-Academia Consortium for Co-creating Polymer Property Database

Aiko Takahashi, Yoshihiro Hayashi, and Ryo Yoshida

The Institute of Statistical Mathematics, Tokyo, Japan

Email: a-taka@ism.ac.jp

ID: PST08

Poster

Aug. 24, 16.00 - 17.30

Background and Objective. An interdisciplinary field of materials research called "Materials Informatics (MI)" is gaining significant attention. However, the development of polymeric material database to be utilized for data-driven research has been significantly delayed due to several reasons such as (1) the complexity of the materials space, (2) the high cost of data acquisition, and (3) the lack of incentive for researchers to disclose their laboratory data.

To overcome the hurdle of limited data resource for polymeric materials, we have developed a fully automated polymer property calculation system RadonPy [1,2] and a large computational database. The database is used as a dataset of machine learning, specifically Simulation-to-Real (Sim2Real) transfer learning: a source model pre-trained with the RadonPy database is fine-tuned to obtain a highly generalizable predictor for real-world systems using limited experimental data.

Currently, an industry-academia consortium has been formed to develop one of the world's largest open databases for polymer properties using RadonPy. The goal is to create a comprehensive database encompassing more than 100,000 molecular skeletons with a wide variety of thermophysical, mechanical, electrical and optical properties for various polymer systems.

Consortium Overview. The consortium currently consists of 241 members from 1 national research institute, 9 universities, and 36 companies, collaboratively producing and sharing data across organizational boundaries. Since 2021, our project has been supported by the "Fugaku" supercomputer program, providing massive amounts of computational resources. This enabled the calculation of various properties for approximately 110,000 amorphous polymers using RadonPy. Our database successfully revealed a polymer property world map. The monthly meetings and Slack communication facilitate collaboration and knowledge exchange, fostering the development of members who may not have expertise in MD simulation or high-performance computing.

Outlook. The consortium aims to release an open database of polymer properties encompassing over 100,000 molecular skeletons. Starting this fiscal year, it seeks to involve automated synthesis and experimental research groups to establish a foundation for data-driven polymer materials research.

References

1. Yoshihiro Hayashi, Junichiro Shiomi, Junko Morikawa, Ryo Yoshida, "RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics", *npj Comput. Mater.* **8**, 222 (2022).
2. Source code: <https://github.com/RadonPy/RadonPy>

Keywords: Polymers, industry-academia consortium, supercomputer, database, molecular dynamics, materials informatics.