DI

ML

BI

MACHINE LEARNING-ASSISTED HIGH-THROUGHPUT COMPUTATIONAL DESIGN OF SOLVENTS FOR LIQUID-EXFOLIATION

Allocation: Strategic/325 Knh
PI: Kimani Toussaint[1]

[1]University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Computational study and machine learning are two advanced tools of scientific discovery. In this project, the research team performed high-throughput computational studies to understand and obtain data on the liquid-phase exfoliation (LPE) process with various solvents. Then, the team analyzed the data using various machine learning algorithms to obtain optimal solvent composition, which facilitates the LPE process. Considering the nanoscale of phenomena occurring in the LPE process, the combination of computational physics and machine learning is one of the best methods to optimize the solvent to further improve the LPE process. This can then guide experimentalists in initial steps by reducing the number of expensive experimental attempts, especially since commonly used solvents in the LPE process are room temperature ionic liquids (RTILs) that have up to several million possible solvents.

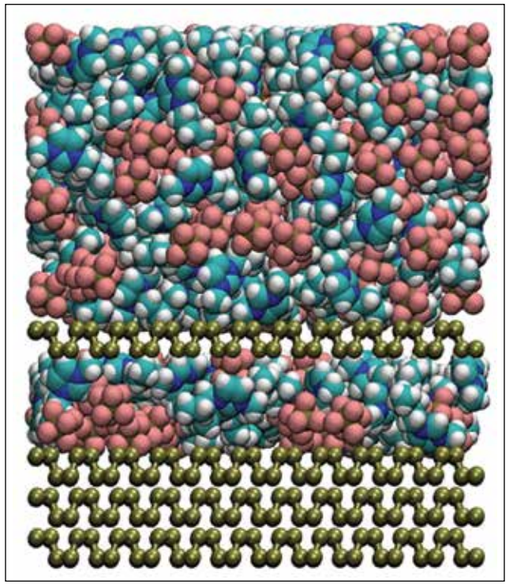

Figure 1: Schematic representation of the liquid-phase exfoliation process; solvent molecules move between layers of bulk materials, separating them from each other.

RESEARCH CHALLENGE

LPE is an advanced mass production method for 2D materials with applications in various technologies such as water desalination, energy storage, optoelectronic devices, DNA sequencing, and energy generation. Compared with other 2D material production methods such as mechanical exfoliation or chemical vapor deposition, LPE produces higher yields with lower costs. LPE is also a simple method that uses an external ultrasonic force on a bulk material immersed in a solvent [1,2]. The bottleneck of the LPE process is the design of a solvent that optimizes the yield, stability, and quality of produced 2D materials.

Because the LPE process occurs in the interface of a bulk material and a solvent, it has a nanoscale nature (Fig. 1). Therefore, knowledge of macroscopic properties of the solvent and bulk material does not allow experimentalists to select the optimal solvent. However, recent computational studies using molecular dynamics (MD) simulations have provided insight into LPE [3]. Even though these studies are crucial in understanding the key parameters of the LPE process, they do not contribute directly to the design and screening of all possible solvents. In this project, the research team combined high-throughput computational studies of various solvents with machine learning algorithms to relate interactions between 2D nanosheets with the composition of the solvents. Based on the applied machine learning algorithm, the team predicted the best solvent composition for the LPE process. The study, therefore, can significantly improve 2D material production, particularly since one of the objectives is to release a computational tool for users in industry and academia.

METHODS & CODES

The research team performed MD simulations using the GROningen MAchine for Chemical Simulations (GROMACS) package. GROMACS is an open-source classical MD code for simulation of various physiochemical and biological systems. The GROMACS package supports various potential forms used in the MD simulation of both solvents (RTILs) and 2D materials. The machine learning model development was performed using the TensorFlow and scikit-learn packages, both of which are accessible through Python (bwpy) on Blue Waters. Fig. 2 shows the workflow of the current study, in terms of both data generation and training as well as the inference and design phases.

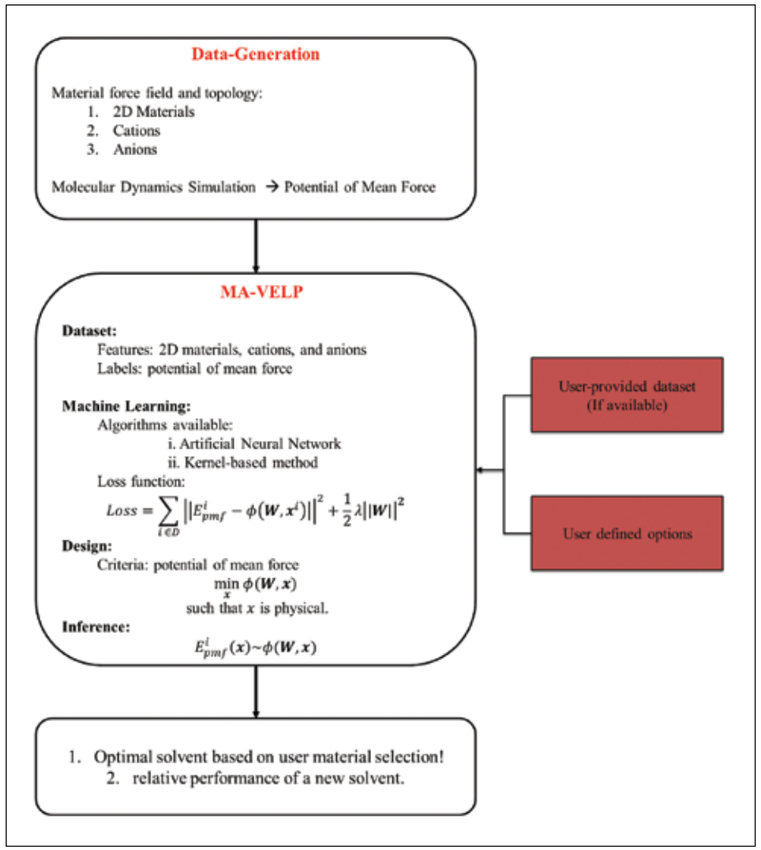

Figure 2: Workflow of current study for the tool development.

RESULTS & IMPACT

The current study uses MD simulations and machine learning methods to help users with computational or experimental background with the selection of an optimal solvent for 2D materials production with the LPE process. The most commonly used solvents for the LPE process are RTILs, also known as "designer solvents," so named because more than a million RTILs could theoretically be synthesized. Considering the number of possible solvents and the costs and risks associated with carrying out experiments, it is necessary to guide experimentalists, especially in the initial steps. Even though MD simulations provide physically consistent and realistic data on the LPE process with a resolution impossible to obtain in experiments, MD data are used to understand the LPE process at the molecular level; the design of solvents requires further development. To make good use of the MD data, the team applied a machine learning algorithm with a high predictive capability to relate solvent composition with MD data, with the objective of facilitating solvent design. The machine learning models can then be used later to optimize solvent compositions, narrowing the number of possible solvents and guiding experimentalists toward the most promising ones.

WHY BLUE WATERS

Increasing the accuracy of machine learning models requires a huge amount of data. Each specific solvent simulation must go through energy minimization, equilibrium in the NVT ensemble and, in the NPT ensemble, peeling and umbrella sampling, followed by data analysis. Such MD simulations are computationally expensive; Blue Waters makes such simulations possible.

PUBLICATIONS & DATA SETS

K. Toussaint, "Machine learning-assisted high-throughput computational design of solvent for liquid-exfoliation," in preparation, 2019.