Data-driven Planning of Computational Workflows for Materials Modeling

Kolmakov A. A.[1,2,@] **and Shapeev A. V.**[2,3]

[1] Moscow Institute of Physics and Technology, Institutskiy Pereulok 9, Dolgoprudny, 141701, Russia
[2] Digital Materials LLC, Kutuzovskaya 4A, Premises 1, Odintsovo, Odintsovsky District, Moscow Region, 143001, Russia
[3] Skolkovo Institute of Science and Technology, Skolkovo Innovation Center Bldg 3, Moscow, 143026, Russia

[@] kolmakov.aa@phystech.edu

Modern materials modeling increasingly relies on complex computational workflows that integrate quantum-mechanical (QM) calculations, molecular dynamics, and machine-learning interatomic potentials. These approaches enable the prediction of a wide range of material properties at a fraction of the cost of experimental measurements. However, certain properties — such as the liquidus temperature — remain computationally demanding and can even exceed experimental costs. This challenge motivates the continued development of strategies to further reduce the computational expense of such workflows.

There is an often overlooked opportunity to improve the efficiency of such workflows through more effective use of computational resources and improved handling of non-convergent QM calculations. We present a data-driven approach for the automatic planning of computational workflows. Gaussian process regression is used to predict efficient calculation settings, identify unreliable calculations, and select appropriate computational resources. Our approach is validated on melting-point and thermodynamic property calculation workflows based on active learning of the Moment Tensor Potential [1], demonstrating improved robustness and reduced time-to-solution.

[1] Shapeev A V 2016 *Multiscale Modeling amp; Simulation* **14** 1153–1173 ISSN 1540-3467 URL `http://dx.doi.org/10.1137/15M1054183`