Development of a tool to create a database of first-principles calculations using high-throughput calculations

Kazuyoshi YOSHIMI[1], Tatsumi AOYAMA[1], Yuichi MOTOYAMA[1],
Masahiro FUKUDA[1], Kota IDO[1], Tetsuya FUKUSHIMA[2], Shusuke KASAMATSU[3],
Takashi KORETSUNE[4], Taisuke OZAKI[1]

[1]*Institute for Solid State Physics, University of Tokyo, Kashiwa-no-ha, Kashiwa, 277-8581*,
[2]*National Institute of Advanced Industrial Science and Technology,*
Tsukuba, 305-8568,
[3] *Faculty of Science, Academic Assembly, Yamagata University,*
Yamagata 990-8560
[4] *Department of Physics, Tohoku University, Sendai 980-8578*

In recent years, approaches to predicting physical properties and designing materials using machine learning (materials informatics) have become popular, in which the generation of a large amount of supervised data is essential for highly accurate prediction and design. From this perspective, databases such as MaterialsProject [1], which store crystal structures, experimental measurement results and first-principles calculations, have been developed and are actively used. On the other hand, machine learning requires the preparation of materials and physical quantities as teacher data, but in many cases the desired materials and physical quantities do not exist in existing databases. If tools and environment can be created to quickly prepare these teacher data, it will provide a research stage for the field of materials informatics and is expected to make a significant contribution to its progress.

We developed HTP tools [2] for exhaustive data generation from crystal structures using first-principles calculations under the support of "Project for advancement of software usability in materials science". Specifically, we developed a script cif2x [3] to generate input files from crystal structures for the first-principles calculation software VASP [4], Quantum ESPRESSO [5], OpenMX [6], and AkaiKKR [7]. Samples and tutorials are provided for each software to explain how to use cif2x with each software.

We also developed a tool to generate batch job scripts, moller [8], which can easily realize exhaustive calculations by bulk execution of jobs on the ISSP supercomputer. Since moller is independent of cif2x in nature, it can be used for general exhaustive calculations with various

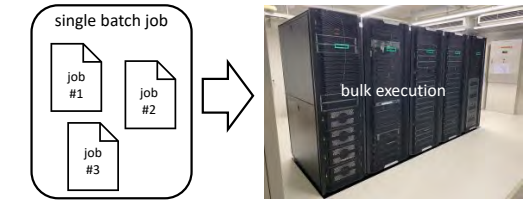

Figure 1: a series of small jobs are executed within a job of large batch classes as a bulk job execution.

solvers. Samples and tutorials are provided that apply HΦ [9] and DSQSS [10] to illustrate the implementation in bulk jobs when multiple calculation scenarios are prepared.

These two software packages, cif2x and moller, are developed as Open Source software and distributed under GPL v.3.0. They have already been pre-installed on the ISSP supercomputer systems. For future plans, we would like to create a computational materials science database using cif2x and moller. It would be interesting to make moller compatible with supercomputing systems other than ISSP supercomputer, to become a standard for exhaustive computations at HPCI. With these tools, we hope that users will make use of them to generate various databases.

References

[1] https://next-gen.materialsproject.org

[2] https://www.pasums.issp.u-tokyo.ac.jp/htp-tools/en/

[3] https://github.com/issp-center-dev/cif2x

[4] https://www.vasp.at

[5] https://www.quantum-espresso.org

[6] https://www.openmx-square.org

[7] http://kkr.issp.u-tokyo.ac.jp

[8] https://github.com/issp-center-dev/moller

[9] https://github.com/issp-center-dev/HPhi

[10] https://github.com/issp-center-dev/dsqss