Usage Report for Fiscal Year 2024

**Project Title:**

# Molecular dynamics simulations and structure modeling of biological macromolecules

**Name:**
**○Osamu Miyashita (1), Wenyang Zhao (1)**

**Laboratory at RIKEN:**
**(1) Center for Computational Science, Computational Structural Biology Research Team**

## 1. Background and purpose of the project, relationship of the project with other projects

Phase retrieval is a critical step to analyze the 3D structure of biological macromolecules when processing data from advanced X-ray diffraction imaging experiments. As illustrated in Fig. 1, the 3D density map of a biological macromolecule can be represented by its complex Fourier components, which consist of both Fourier magnitude and Fourier phase. However, since the Fourier phase cannot be measured, it must be computationally retrieved using the experimentally measured Fourier magnitude and an assumed support that loosely encloses the molecular particle to reconstruct the 3D density map of the biological macromolecule. This process is known as phase retrieval.

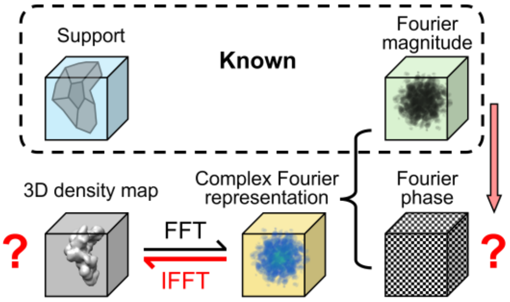

Figure 1. Schematic illustration of the phase retrieval problem, modified from Fig. 1 in Ref. [1].

In recent years, the demand for handling large-scale 3D volumes with fine voxels in phase retrieval has been increasing, driven by the need to process experimental data or, for theoretical studies, synthetic data. However, performing phase retrieval on large-scale 3D volumes is memory-intensive. Consequently, it is essential to employ parallel computing using multiple processes to utilize distributed memory in clusters or supercomputers. In this context, we developed an open-source software package called xphase3d. In this project, we implemented the software on HBW2 and tested its performance.

## 2. Specific usage status of the system and calculation method

Xphase3d is developed in ANSI C. Its core program performs MPI memory-distributed phase retrieval. To achieve this, it segments the 3D arrays during computation and distributes the segments across multiple processes. Broadly, the computations can be categorized into two types: operations on local segments and MPI-based 3D Fast Fourier Transform (FFT). The former is straightforward and does not require any inter-process communication, whereas the latter involves significant all-to-all communication between processes. Xphase3d uses the FFTW library to perform MPI 3D FFTs. Additionally, it utilizes the HDF5 library to read and write 3D arrays on disk. So far, we have successfully compiled xphase3d with its dependent libraries on HBW2.

## 3. Result

To evaluate the compatibility, functionality, and performance of xphase3d on HBW2, we conducted a phase retrieval task on a 3D volume consisting of $1024^3$ voxels. This task requires a minimum of 154

GB of memory, exceeding the 128 GB memory capacity of each node in HBW2's Massively Parallel Computing (BWMPC) system. For this reason, the task had to be distributed across two or more nodes.

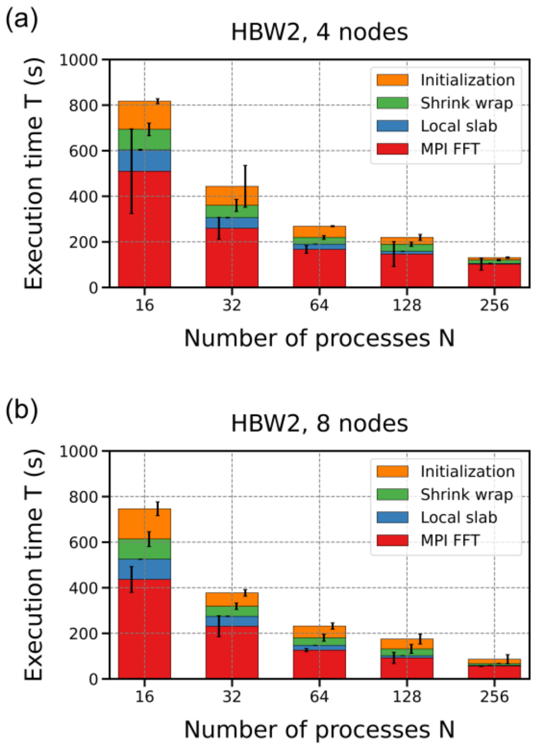

Figure 2. Performance of xphase3d on HBW2. Subfigure (b) is modified from Fig. 3(a) in Ref. [1].

Figure 2 shows the execution time for the task using different numbers of nodes and processes. Since threading parallelism is currently unavailable in xphase3d, the number of processes corresponds to the number of cores used. The bars in different colors represent different computational portions in the phase retrieval task. The error bars indicate the standard deviations among ten repeated measurements. As expected, execution time decreases as the number of processes increases, demonstrating the scalability of xphase3d implemented on HBW2. Additionally, it is observed that, for the MPI FFT portion, using 8 nodes is slightly faster than using 4 nodes, even with the same number of processes. This behavior results from the characteristics of the FFTW library. It also aligns with our performance test of xphase3d on the supercomputer Fugaku, despite their significant differences in system environments.

## 4. Conclusion

We have developed a software package called xphase3d to perform MPI parallel processing and memory-distributed phase retrieval on large-scale 3D volumes. So far, the software has been tested on various platforms, including individual workstations, clusters, and supercomputers. As part of the testing, it has been implemented on HBW2, and its compatibility, functionality, and performance on HBW2 have been confirmed.

## 5. Schedule and prospect for the future

After successfully implementing xphase3d on HBW2, we plan to utilize its computing resources to process a set of synthetic experimental data related to the characterization of biological macromolecular structures via X-ray free-electron laser (XFEL) techniques. In addition, our team is continuously developing a suite of software for structural biology research. These software are specially optimized for supercomputers. We will test them sequentially on HBW2.

## References

[1] W. Zhao, O. Miyashita, M. Nakano, and F. Tama, "Xphase3d: Memory-Distributed Phase Retrieval for Reconstructing Large-Scale 3D Density Maps of Biological Macromolecules," in *2024 IEEE International Conference on Cluster Computing (CLUSTER)*, 2024, pp. 394–402.

# Fiscal Year 2024 List of Publications Resulting from the Use of the supercomputer

## [Conference Proceedings]

[1] W. Zhao, O. Miyashita, M. Nakano, and F. Tama, “Xphase3d: Memory-Distributed Phase Retrieval for Reconstructing Large-Scale 3D Density Maps of Biological Macromolecules,” in *2024 IEEE International Conference on Cluster Computing (CLUSTER)*, 2024, pp. 394–402.

## [Oral presentation]

[1] W. Zhao, O. Miyashita, M. Nakano, and F. Tama, “Xphase3d: Memory-Distributed Phase Retrieval for Reconstructing Large-Scale 3D Density Maps of Biological Macromolecules,” in *2024 IEEE International Conference on Cluster Computing (CLUSTER)*, Sept. 27, 2024, Kobe, Japan.