Large-scale Parallel Design for Cryo-EM Structure Determination on Heterogeneous Many-core Architectures

Liang Qiao*, Hongkun Yu*, Kunpeng Wang*, Ruixin Sun‡, Wenlai Zhao*¶, Haohuan Fu†, Guangwen Yang*
*Department of Computer Science and Technology, Tsinghua University, Beijing, China
†Department of Earth System Science, Tsinghua University, Beijing, China
‡School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu, China
¶Corresponding Author, email: cryinlaugh@gmail.com

***Abstract*—Cryo-EM structure determination is the most important research area in structural biology. With the development of cryo-electron microscopy, the resolution has been enhanced significantly, which leads to the huge computation to reconstruct the biomolecule in recent years. In this paper, we present a large-scale parallel design for Cryo-EM structure determination on heterogeneous many-core architectures. A novel task parallel strategy is proposed to reduce the redundant computation and improve the scalability on large-scale systems. Further, We distribute the reconstruction model to each node and rearrange the data layout to achieve high parallel efficiency and reduce the memory footprint. The proposed comprehensive parallel design shows highly parallel efficiency and scalability on large-scale heterogeneous architectures, which could significantly accelerate the whole period of Cryo-EM structure determination process.**

***Index Terms*—heterogeneous many-core, Cryo-EM, structure determination**

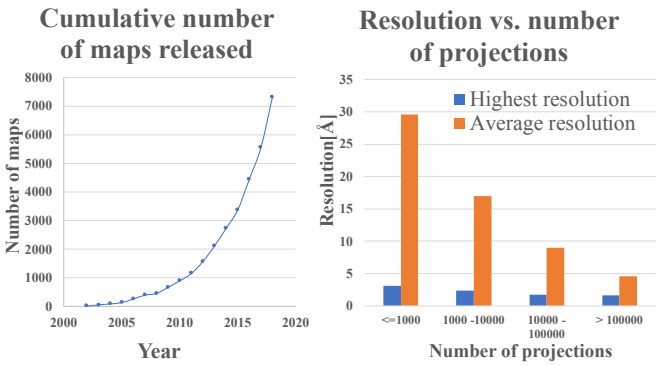

Fig. 1. Left: The cumulative number of maps released. Right: The resolution vs. number of images.

I. Introduction

Structure determines function. Understanding the biological functions of biomacromolecules in living cells is the most fundamental topic in the field of structural biology as well as pharmacy industry. Hence, the method of high-resolution structure determination of macromolecules is always a key problem for researchers to address. Honored with the 2017 Nobel Price, the cryo-electron microscopy (cryo-EM) technique is one of the most powerful tool to determine near atomic-resolution structure of biomacromolecules, which uses electron beam as light to take images for biological samples at high resolution.

Since 2013, atomic-resolution structures of small-size and middle-size biological macromolecules [1]–[5] are successfully to be determined by cryo-EM techniques as shown in the left of Figure 1, which is called the resolution revolution [6]. Besides the technical breakthroughs in image recording and electron optics [7], [8], computing technologies for three-dimensional (3D) reconstruction also play a key role in structure determination of cryo-EM. The newest method of cryo-EM 3D reconstruction is based on the Expectation-Maximization algorithm with empirical Bayesian approach [9], which builds the 3D electron-density model of a target biomolecule from a large set of integral projections. Unlike the computed tomography(CT), the orientation of each projection is unknown, leading to extra workloads to do orientation determination. Hundreds of thousands or even millions of 2D projections with non-deterministic orientations are essential to cover all orientations in space for the further reconstruction and to improve the signal-to-noise ratio (SNR) by averaging the noise of particles to build atomic-resolution models, as shown in the right of Figure 1.

Therefore, to satisfy the huge and increasing computational demand in cryo-EM 3D reconstruction tasks, HPC platforms, even supercomputers, become attractive choices in cryo-EM field. Modern HPC platforms is tend to be heterogeneous many-core systems such as CPU+GPU or SW26010 [10], [11]. Even though much optimizations are proposed for improve the parallel efficiency [12] [13] [14] on heterogeneous systems, they just focus on the data parallel and can not be effective for Maximization phase. Further, in majority reconstruction algorithm, the memory needs to hold several entire models for the computation, which can hardly stored in a single node on modern heterogeneous many-core architectures due to the increasing model size.

In this paper, we propose a highly efficient and scalable parallel design for Cryo-EM structure determination based

978-1-7281-1867-3/19/$31.00 ©2019 IEEE

on particle filter algorithm [15], which is capable of accelerating the reconstruction process on heterogeneous many-core systems. To improve the performance, a comprehensive task parallel design is presented, which effectively reduce the computation and improve the scalability. Then, we parallelize the Maximization phase in structure determination process with the distributed model strategy. Further, we implement our parallel design and optimization on the Sunway TaihuLight supercomputer and common Intel cluster, which achieve high parallel efficiency and scalability.

In summary, we present a large-scale parallel design for the Cryo-EM structure determination on heterogeneous many-core architectures. The main contributions of this work are as follows:

- A task parallel design for the Expectation phase is proposed with the optimized MPI communication and data reusing, which can significantly reduce the computation and improve the scalability.
- A data parallel design for the Maximization phase is proposed with the distributed model, which can parallel the Maximization phase on large-scale systems efficiently.

II. Background and Related Works

A. Expectation-Maximization 3D Reconstruction Algorithm

The key mathematical principle in 3D reconstruction of cryo-EM images is **central slice theorem**. The theorem introduces that the results of the the Fourier transformation of a 2D orthographic integral projection of 3D model in real space by the given orientation, and the though-the-origin slice of the 3D model in Fourier space parallel to the given orientations are equal.

The method of iterative refinement is commonly adopted in the cryo-EM 3D reconstruction task as the basic approaches [16], [17]. Based on an initial estimate probability distribution of 3D model, we first determine the most likely orientations for each particles, and then interpolate the particle images to generate a new refined model. A major advance in recent years was the implementation of maximum-likelihood (ML) approaches. Scheres et al. firstly implements the bayesian approaches for 3D estimation and classification, which treat the projection orientations of the particle images as latent variables to solve numerically by a full batch Expectation-Maximization(EM) algorithm in the RELION [9]. This Bayesian-based method exhibits good ability to identify and characterize multiple structural states for high-resolution structure determination.

Particle-filter cryo-EM 3D reconstruction algorithm, which is proposed by Hu et al. [18], is one of the most state-of-art algorithm, to do robust cryo-EM 3D reconstruction works with self-adaptive parameter adjustment, tolerance to bad particles and per-particle defocus refinement. Basicly, the particle-filter reconstruction algorithm is also based on the full-batch EM algorithm with bayesian approaches to accomplish cryo-EM 3D reconstruction tasks, which involves several stages as shown in Figure 2.

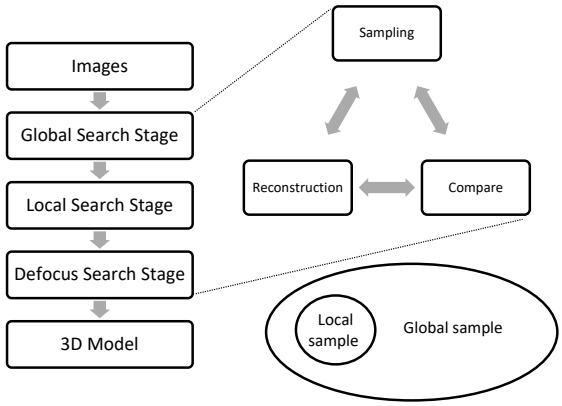

Fig. 2. Particle-filter based 3D Reconstruction Workflow

Generally, the whole workload of 3D reconstruction consist of many rounds of refinement of the 3D model, which can be divided into three stages. The first several rounds, called *Global Search Stage*, which need to take much more global samplings from whole space of orientations in every rounds, to examines all orientations to avoid the locally optimal. The second stage, *Local Search Stage*, will only sample from the relatively small regions of space of orientations with significant higher probabilities in the previous round, which increases the density of sampling to search for more precise orientations for the original images. *Defocus Search Stage*, the third stage, usually executes in the last round of iterations, to refine more parameters such as the *defocus*, by add the value of *defocus* as a new sampling dimension.

In each round, the computation workloads can be divided into *expectation phase(E-phase)* and *maximization phase(M-phase)*. The *E-phase* determines the most likely orientations for each input images, by calculating the probability distributions of massive sampling orientations from the space of orientations via steps called *Sampling* and *Compare*. Massive 2D central slices from different orientations by different sampling methods (i.e. SIS [19], curvilinear quadrilaterals by HEALPix [20]) are calculated from the current estimate of the 3D model in *Sampling* step. Then the differences between these sampling slices and original images are computed to obtain the probability distributions for those orientations in the *Compare* step. In the *M-phase*, all input images are inserted back to an empty 3D volume in Fourier Space according to the probability distributions of orientations, which is called *reconstruction* step. Iterative FFT and iFFT should be processed here to correct the model in Real Space [21], [22].

B. Large-scale Heterogeneous Many-core systems

Modern HPC platforms is tend to be heterogeneous many-core systems by the integration of multi-core CPUs and accelerators(i.e. GPU, Field-Programmable Gate Arrays, Intel Xeon Phis) [10], [23]or consisting of new high performance heterogeneous many-core chips such as SW26010 [11]. The Summit supercomputer [10] is now the No.1 supercomputer in Top 500 List, which has 4,356 nodes, each one housing

two Power9 CPUs with 22 cores each and six NVIDIA Tesla V100 GPUs each with 80 streaming multiprocessors (SM), with totally 2414592 cores. The Sunway TaihuLight supercomputer [11] is previous No.1 supercomputer with 40,960 SW26010 processors and 10 million cores. Each SW26010 processor has 4 symmetric core groups (CG) which can be configured as multiple running modes flexibly and all CGs are interconnected through the network on chip (NoC). Every CG are composed of one management processing element(MPE), 64 computing processing elements (CPE). Each CPE has a 64KB Scratch Pad Memory (SPM) which can be configured as either a user-controlled local directive memory (LDM) or a software-emulated cache. Direct Memory Access (DMA) is supported by the MC for data transportation between the SPM and the DDR memory. These systems pose significant challenges to the large-scale parallel Design for Cryo-EM structure determination algorithms, since the workloads should be divided precisely to map on the whole platforms.

Even lots of efforts have been put to parallelize the 3D reconstruction works [16], [24], there are still obstacles for large scale parallel on these systems. The first is that current works only focus on the task-level parallelization(data parallel on input images). Such a coarse-grained parallel strategy limits exploiting the more fined parallelism of the algorithm (i.e. the Maximization phase). Second, in most reconstruction algorithm, the memory needs to hold several entire models for the computation, which can hardly stored in a single node on modern heterogeneous many-core architectures due to the increasing model size. Besides, the scalability of these programs is not considered as they are not initially developed for ultra-large scale computing systems.

III. Large-scale Parallel Design for E-M Algorithm

To parallelize the Expectation-Maximization algorithm on the large-scale heterogeneous many-core architectures, we analyze the entire workflow and propose the hybrid parallel design not only on data level but also on task level. For task level, we focus on the Expectation phase and explore the pipeline to overlap the redundant computation and communication overhead. Reorganization of the sampling and compare operations can improve the scalability and parallel efficiency on large-scale systems. For data level, we design a parallel scheme based on distributed model, which can solve the memory constraints for accelerators on many-core systems and parallelize the Maximization phase to many computing nodes. According to the bandwidth gap between main memory and local memory on heterogeneous architectures, a SoA to AoS module is proposed to improve the IO efficiency and parallel granularity. In this section, we will introduce the entire parallel design in detail in order of the Expectation-Maximization algorithm workflow.

A. Task Parallel in Expectation Phase

The entire Cryo-EM structure determination workflow is implemented by a pipelined approach generally. In the Expectation phase, each image will be compared with lots of 2D slices extracted from the 3D reference to find the approximate orientation, as shown in Algorithm 1. For most implements of E-M algorithm, the input images are assigned to each node to achieve simple data parallel. With this straightforward workflow, the process can obtain good scalability, but there are lots of redundant computation, which will lower the efficiency on large-scale system. To achieve highly parallel efficiency on a large-scale heterogeneous many-core system, we explore the task parallel workflow among the computing nodes to reduce redundant computation and improve the scalability of the alignment process.

Algorithm 1 Main Processed in Expectation Phase

```
for phase = 0 to Max_Phase do
  if phase == 0 then
    for i = 0 to nImg do
      Sampling(N_S)
      Compare(1,N_S)
    end for
  else
    for i = 0 to nImg do
      Sampling(nS)
      Compare(1,nS)
    end for
  end if
end for
```

In the Particle Filter method, the extracted sampling points for each image will be able to cover the majority of the sample space in a global sampling process. Reusing the random sampling points for each image can reduce the redundant computation, which means that each image can share the same set of slices to obtain the approximate orientation. When we assign the sampling processes to each computing nodes (totally n nodes), the time consumed for sampling step in one iteration will be reduced by n times. Meanwhile, each node to get the whole set of slices need an additional communication operation MPI_ALLReduce. The entire time consumed for the expectation process in one iteration can be expressed as below (N_S represents the number of slices, $t_{sampling}$ represents the time for one sampling operation):

$$T_{expectation} = N_S/n \times t_{sampling} + T_{All\_Reduce} + T_{compare}$$

To obtain effective performance improvements by this scheme, the additional cost T_{All_Reduce} needs to be smaller than the reduced sampling time $(n-1)N_S/n \times t_{sampling}$. Generally, on the large-scale computing system, we can divide nodes into small groups to reduce the cost of MPI_ALLReduce, but it will lower the parallel efficiency. To overcome the trade-off between computation and communication, we design a MPI communication scheme to share the common data of slices between computing nodes. With the above methods, we can achieve the lightweight communication cost and overlap it by the transformed calculation process, as shown in Figure 3.

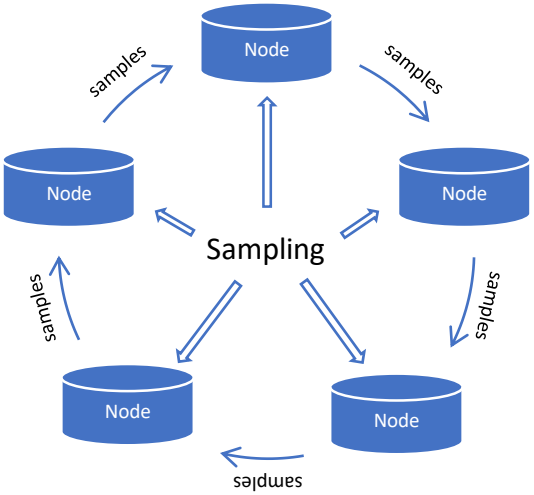

Fig. 3. Task Parallel with MPI Communication

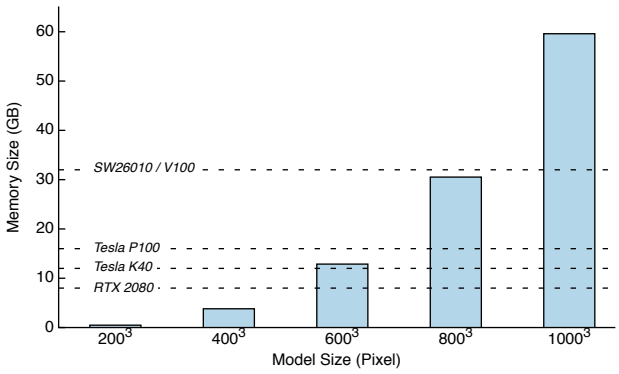

Fig. 4. Memory Usage for Common Model Size

Analyzing the complexity of compare operations, we can find out proper amount of compare operations to execute in one time and overlap the communication overhead. Therefore, the communication overhead can be covered by the following partial compare operation, and the entire time consumed for expectation phase is shown as below:

$$T_{expectation} = N_S/n \times t_{sampling} + T_{compare}$$

The entire Expectation phase is transformed into a comprehensive parallel workflow: the data of slices like the items on the conveyor belt is delivered to next nodes in task parallel and each node executes a data parallel scheme. With the task parallel scheme, we proposed a efficiency workflow and redesign the expectation phase to adapt the large scale parallelization and achieve high scalability.

B. Data Parallel in Maximization Phase

Algorithm 2 Main Processes in Maximization Phase

```
for i = 0 to nImg do
    T, F ← InsertP()
end for
for i = 0 to MAX_Iter do
    if checkC() == 0 then
        C_Fourier = W * T
        C_Real ← iFFT(C_Fourier)
        Correction(C_Real)
        C_Fourier ← FFT(C_Real)
        Update(W)
    else
        break
    end if
end for
```

Maximization is the key phase to assemble the 2D images to a 3D model in Cryo-EM structure determination process, as shown in Algorithm 2. Most of the Cryo-EM structure determination programs are serial in Maximization phase even though parallel in Expectation phase. When we try to implement the program on large-scale heterogeneous platforms to solve a larger problem, this kind workflow will obtain low parallel efficiency and scalability. Moreover, with the larger size and higher resolution of 3D structures, memory constraint will become the bottleneck to limit the problem-solving scale not only for common clusters, but also for large-scale heterogeneous platforms. During the maximization phase, the memory usage will increase to 4 times of the 3D model, which will exceed the memory capacity of popular compute nodes, as shown in Figure 4. To overcome the memory constraints and parallel the Maximization phase, we manage to distribute the large-scale 3D model into many nodes, which can reduce the memory footprint and improve the scalability of Maximization phase significantly.

In the Maximization phase, each node will insert the assigned images into two initialized models F and T by $InsertP()$ function. We then allocate the space of W weight model and C parameter model, which will be involved in the repeated Fourier transform process. According to the entire Maximization phase, the operations on models F, T and W are element-wise, so we can distribute them into n computing nodes. For the parameter model C, it is generated by T and W, and will be changed in repeated Fourier transform process. Therefore, after distribute the models F, T, W, an additional MPI_ALLGather operation needs to be executed in each iteration while the time consumed for computing is reduced to $T_{computing}/n$ and memory overhead is reduced from 4 to $(n+3)/n$ times model size.

To reduce the communication overhead, we explore the data layout of the models. For the data parallel in Maximization phase, 2D images will be inserted into 3D model on each node. When we use large-scale computing nodes, each node only needs to insert fewer images to the model. However, each node still stores an entire model which will take up lots of memory space. We can change the layout of models which means store the inserted point by a key-value pair, which includes the index and value information. Due to the fewer images to be inserted, there are a small number of key-value

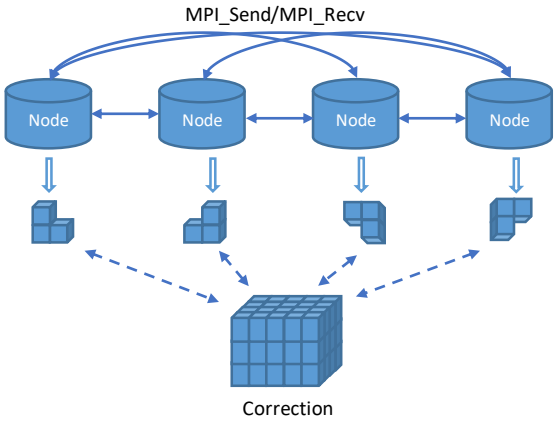

Fig. 5. Distributed Model with MPI Communication

pairs need to be stored. Then, the memory usage is reduced further which can also reduce the communication overhead for the smaller data transfer.

During the Maximization phase, the images and models will be transferred to Fourier space, and the pixel will change to a complex, which will be difficult to use $simd$ method to accelerate the computation. We further design a module to transfer the data structure from complex to two separate doubles and store them in two model space. By this SoA (structure of arrays) to AoS (array of structures) design, we can easily use the $simd$ operations on heterogeneous many-core systems.

IV. Experiments

In this section, we implement our hybrid multi-level parallel design on several Intel compute nodes (each node consist of 2 E5-2670v3 with 12 cores) and the Sunway TaihuLight supercomputer to evaluate the parallel efficiency and scalability. For the experiments data, we use the typical real data sets: CNG (cyclic-nucleotide-gated) and generated cases for various parameters. The data set of CNG channel has 211826 particles and the size of particles is 160×160 (padding size is 320×320). For the large-scale efficiency and heterogeneous accelerators, the experiments are performed on Sunway TaihuLight (SW26010 processor), and for the effectiveness of distributed model scheme, it is evaluated on the Intel cluster.

A. Performance of Task Parallel

To evaluate the performance of the task parallel in Expectation phase, we implement the optimization on Sunway TaihuLight supercomputer. As shown in Figure 6, with the increasing number of computing nodes and fixed number of pixels 256, the optimized communication among computing nodes obtains a good scalability without additional overhead. And for the different number of pixels in global search stage by using 512 nodes, the execution time of Expectation phase is significantly reduced relative to the original serial workflow, which is shown in Figure 7

The scalability of the task parallel scheme is shown in Figure 8. We compare the scalability from 256 nodes to 2048

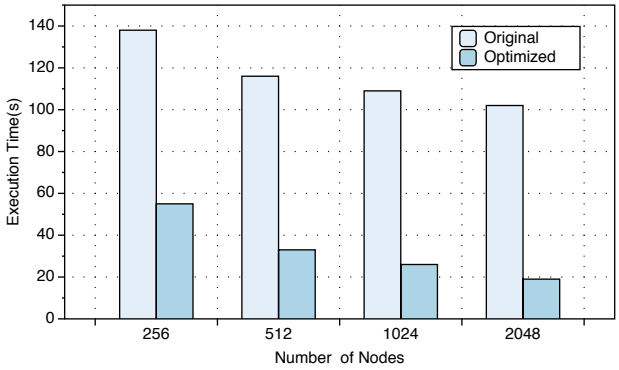

Fig. 6. Execution Time for Different Node Number

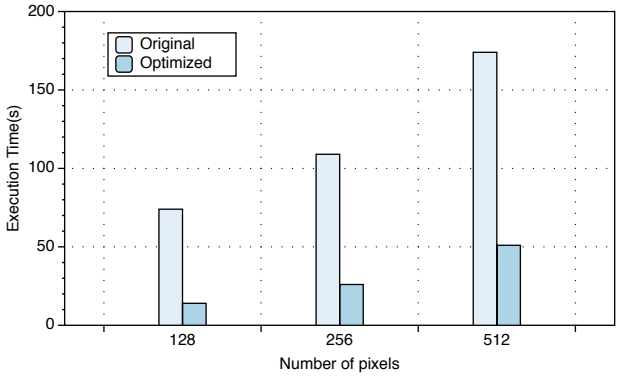

Fig. 7. Execution Time for Different Pixel Number

nodes on the Sunway TaihuLight, and the relative speedup for the task parallel scheme is higher than the original serial workflow in the Expectation phase.

B. Performance of Data Parallel

For the Maximization phase, we manage to distribute the models to each node, which can reduce the memory footprint from 4 times to $1+3/n$ times (n is the number of nodes). We implement the distributed model strategy on 32 Intel nodes with 4 groups. While the memory constraints are overcame,

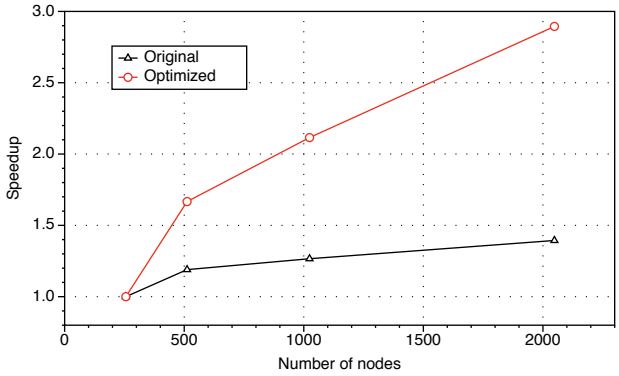

Fig. 8. Scalability for the 3D Pipeline Strategy

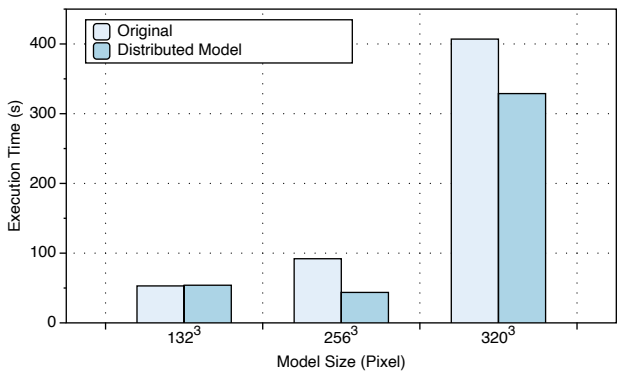

Fig. 9. Performance for Distributed Model

the computation can also be reduced, as shown in Figure 9. For the small model, the execution time is not reduced because the additional MPI_ALLGather communication overhead can not be covered by the computational reduction.

V. Conclusion and Future Work

In this paper, we propose a comprehensive parallel design for Cryo-EM structure determination based on the particle filter algorithm, which is capable of accelerating the reconstruction process on heterogeneous many-core systems. With the pipeline reorganization and distributed model, the Expectation-Maximization algorithm can achieve high scalability and parallel efficiency on large-scale systems. We experiment our parallel design and optimizations on the common Intel cluster architecture and Sunway TaihuLight supercomputer, which successfully obtain better performance and correct result. Our work can significantly shrink the whole Cryo-EM structure determination process, which could bring huge benefits to structural biology researches.

Our future work are aimed at exploring the collaboration between E-M algorithm and hardware architectures. According to the features of accelerators on many-core system, we can assign the computation in computing cores level. Further, based on the parallel design and more memory oriented optimizations, we can develop a large-scale parallel framework on the supercomputers, which can greatly promote the research of structural biology.

References

[1] E. Cao, M. Liao, Y. Cheng, and D. Julius, "Trpv1 structures in distinct conformations reveal activation mechanisms," *Nature*, vol. 504, no. 7478, p. 113, 2013.

[2] P. Lu, X.-c. Bai, D. Ma, T. Xie, C. Yan, L. Sun, G. Yang, Y. Zhao, R. Zhou, S. H. Scheres *et al.*, "Three-dimensional structure of human γ-secretase," *Nature*, vol. 512, no. 7513, p. 166, 2014.

[3] C. Yan, J. Hang, R. Wan, M. Huang, C. C. Wong, and Y. Shi, "Structure of a yeast spliceosome at 3.6-angstrom resolution," *Science*, vol. 349, no. 6253, pp. 1182–1191, 2015.

[4] M. Wu, J. Gu, R. Guo, Y. Huang, and M. Yang, "Structure of mammalian respiratory supercomplex i1iii2iv1," *Cell*, vol. 167, no. 6, pp. 1598–1609, 2016.

[5] M. Li, X. Zhou, S. Wang, I. Michailidis, Y. Gong, D. Su, H. Li, X. Li, and J. Yang, "Structure of a eukaryotic cyclic-nucleotide-gated channel," *Nature*, vol. 542, no. 7639, p. 60, 2017.

[6] W. Kühlbrandt, "The resolution revolution," *Science*, vol. 343, no. 6178, pp. 1443–1444, 2014.

[7] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, and Y. Cheng, "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em," *Nature methods*, vol. 10, no. 6, p. 584, 2013.

[8] Y. Cheng, "Single-particle cryo-em at crystallographic resolution," *Cell*, vol. 161, no. 3, pp. 450–457, 2015.

[9] S. H. Scheres, "Relion: implementation of a bayesian approach to cryo-em structure determination," *Journal of structural biology*, vol. 180, no. 3, pp. 519–530, 2012.

[10] J. Hines, "Stepping up to summit," *Computing in Science & Engineering*, vol. 20, no. 2, pp. 78–82, 2018.

[11] H. Fu, J. Liao, J. Yang, L. Wang, Z. Song, X. Huang, C. Yang, W. Xue, F. Liu, F. Qiao *et al.*, "The Sunway TaihuLight supercomputer: system and applications," *Science China Information Sciences*, vol. 59, no. 7, p. 072001, 2016.

[12] H. Zhou, F. Ni, L. Zhao, and F. Zheng, "Parallel optimization of relion: Performance comparison based on cluster for cpu/gpu and knl," in *Proceedings of the 2018 2nd International Conference on Algorithms, Computing and Systems*. ACM, 2018, pp. 48–52.

[13] J. Zhang, Z. Wang, Y. Chen, Z. Liu, and F. Zhang, "Memory-efficient and stabilizing management system and parallel methods for relion using cuda and mpi," in *International Symposium on Bioinformatics Research and Applications*. Springer, 2018, pp. 205–216.

[14] P. Cossio, D. Rohr, F. Baruffa, M. Rampp, V. Lindenstruth, and G. Hummer, "Bioem: Gpu-accelerated computing of bayesian inference of electron microscopy images," *Computer Physics Communications*, vol. 210, pp. 163–171, 2017.

[15] M. Hu, H. Yu, K. Gu, Z. Wang, H. Ruan, K. Wang, S. Ren, B. Li, L. Gan, S. Xu *et al.*, "A particle-filter framework for robust cryo-em 3d reconstruction," *Nature methods*, vol. 15, no. 12, p. 1083, 2018.

[16] J. De la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldivar, J. Vargas, J. Carazo, and C. Sorzano, "Xmipp 3.0: an improved software suite for image processing in electron microscopy," *Journal of structural biology*, vol. 184, no. 2, pp. 321–328, 2013.

[17] N. Grigorieff, "Frealign: high-resolution refinement of single particle structures," *Journal of structural biology*, vol. 157, no. 1, pp. 117–125, 2007.

[18] M. Hu, H. Yu, K. Gu, Z. Wang, H. Ruan, K. Wang, S. Ren, B. Li, L. Gan, S. Xu *et al.*, "A particle-filter framework for robust cryo-em 3d reconstruction," *Nature methods*, vol. 15, no. 12, p. 1083, 2018.

[19] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

[20] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, "Healpix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere," *The Astrophysical Journal*, vol. 622, no. 2, p. 759, 2005.

[21] V. Abrishami, J. Bilbao-Castro, J. Vargas, R. Marabini, J. Carazo, and C. Sorzano, "A fast iterative convolution weighting approach for gridding-based direct fourier three-dimensional reconstruction with correction for the contrast transfer function," *Ultramicroscopy*, vol. 157, pp. 79–87, 2015.

[22] S. H. Scheres, "Relion: implementation of a bayesian approach to cryo-em structure determination," *Journal of structural biology*, vol. 180, no. 3, pp. 519–530, 2012.

[23] B. Tsay, "The tianhe-2 supercomputer: Less than meets the eye?" 2013.

[24] J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, and S. H. Scheres, "New tools for automated high-resolution cryo-em structure determination in relion-3," *Elife*, vol. 7, p. e42166, 2018.