Improvements to Hardware-Accelerated 3D Single Particle Imaging Data Reconstruction

Niteya Shah[1,2], Mentor: Christine Sweeney[1], Co-Mentor: Vinay Ramakrishnaiah[1]

[1] Los Alamos National Laboratory, Applied Computer Science Group (CCS-7)
[2] Virginia Tech, Department of Computer Science

Introduction

We present results from using hardware acceleration to reconstruct 3D images of the molecular structure of bio-macromolecules. We process simulated X-ray diffraction data representative of what is gathered at experimental light source facilities, such as the Linac Coherent Light Source. These performance improvements will enable efficient use of these experimental facilities by enabling processing and decision making in real time.

Highlights

- Fast reconstruction of 3D Single Particle Imaging (SPI) data using supercomputers equipped with graphics processing units (GPUs)
- SPI key for innovations in electronics, microbiology, optics, and nanotechnology.
- Optimizations made to improve per-node computational efficiency and accuracy of SPI by using better algorithms, improving data movement and access, reusing data structures and reducing memory fragmentation.
- **50% improvement in reconstruction accuracy.**
- **4X speedup in reconstruction time and 485X speedup in resolution calculation.**

Single Particle Imaging

- Technique for determining the structure of bio-macromolecules using diffraction patterns gathered from directing an X-ray beam at single particles before they are destroyed.
- SPI via free electron lasers (XFELs) is key for understanding the time evolution of enzymatic reactions.

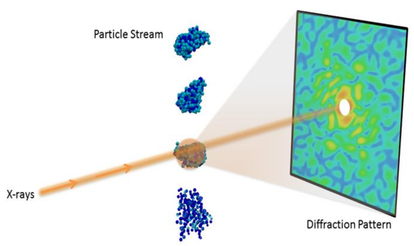
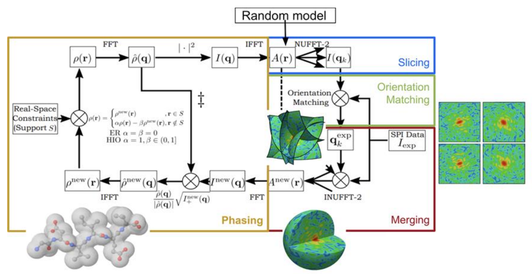


Figure 1. Experimental setup of SPI

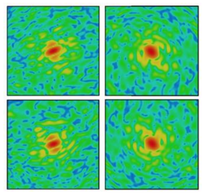
Computational Imaging Requirements and Implementation:

- Multi-Tiered Iterative Phasing (MTIP) algorithm determines the states, orientations, intensities, phases, and underlying structure [1]
- **SpiniFEL** [2] code implementation of MTIP for project ExaFEL: Data Analytics at Exascale for Free Electron Lasers



Quantities		Grids	
$\rho(r)$	Electron Density of Sample	r	N^3 Cartesian Real-Space Grid (N = 100 – 500)
$A(r)$	Autocorrelation of Electron Density	q	N^3 Cartesian Fourier-Space Grid (N = 100 – 500)
$I(q)$	3D Intensity Function / Diffraction Volume	q_k	Nonuniform Model Data (10^5 points)
$I(q_k)$	Model Diffraction Images / Model Orientations		
I_{exp}	Experimental Diffraction Images		

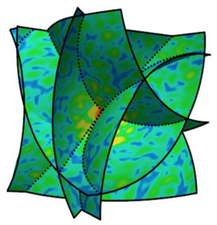
Methods



Slicing

Create M 2D model images from the current electron density estimation.

- Precomputed the einsum and stored it in pinned memory to reduce wasted calculations and improve memory transfer.

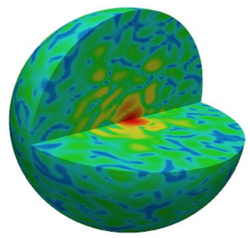


Orientation Matching

Match the model images to the reference images by finding the pair with the smallest Euclidean distance.

$$||(X_{A,n} - Y_{B,n})||^2 = \sum_{i=1}^{n} X_{A,i}^2 \oplus \sum_{i=1}^{n} Y_{B,i}^2 + X_{A,n} \cdot Y_{B,n}^T$$

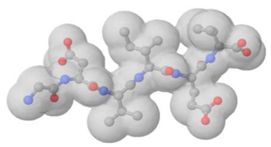
- Reformulated Euclidean distance to a GEMM operation to take advantage of Tensor Cores.
- Identified closest pairs using a GPU optimized tree-style segmented reduction.



Merging

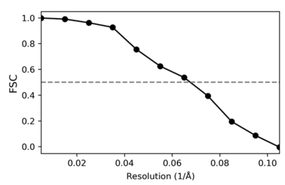
Combine the reference orientations into a 3D diffraction volume. Solve as a linear system using the conjugate gradient (CG) algorithm

- Replaced CPU implementation with GPU version.
- Created persistent data structures that can be reused to reduce memory management costs and fragmentation.
- Used double precision for CG to overcome its instability, leading to improved convergence.



Phasing

Convert 3D diffraction volume into a molecular structure by applying real-space and reciprocal space constraints on the electron density estimate of the sample.



Resolution Computation

Use Fourier shell correlation (FSC) to calculate the achieved resolution.

- Offloaded computations to the GPU.
- Plotted results in another thread to not stall the main runtime.

[1] Jeffrey J Donatelli, James A Sethian, and Peter H Zwart. Reconstruction from limited single-particle diffraction data via simultaneous determination of state, orientation, intensity, and phase. Proceedings of the National Academy of Sciences, 114(28):7222–7227, 2017.
[2] https://gitlab.osti.gov/mtip/spinifel
[3] We acknowledge the help and guidance of Chun Hong Yoon (SLAC), Ariana Peck (SLAC), Jeffrey Donatelli (LBNL), Wu-chun Feng (Virginia Tech) as well as Perlmutter and Cori, the computing resources provided by NERSC.

Results

Tests were run on NERSC Perlmutter supercomputer GPU nodes, each of which have a 3rd Gen AMD EPYC CPU, 4 A100 Nvidia GPUs and use Slingshot10 interconnect for MPI communication.

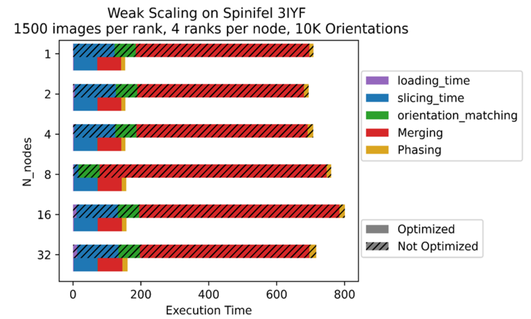


Figure 3. Weak scaling studies for SpiniFEL shows constant execution time per node.

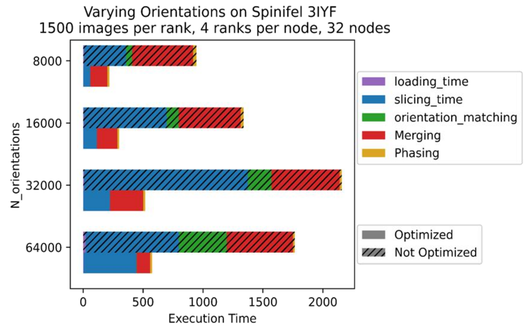


Figure 4. SpiniFEL shows speedup for varying number of orientations with better scalability than previous implementation.

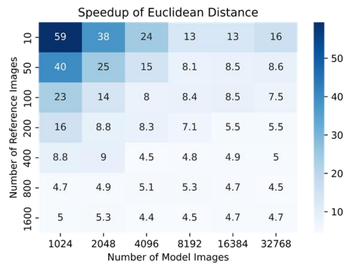


Figure 5. Speedup* for Euclidean distance for 128x128 sized images

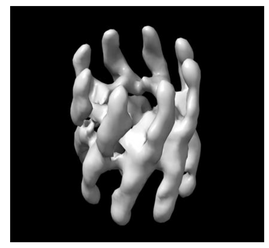


Figure 6. 3D electron density map shows structure of 3IYF protein. Image generated by SpiniFEL reconstruction.

Summary and Future Work

We see significant gains in performance that scale up as the problem size grows. Additionally, the changes also improve time taken to measure convergence and improve the final resolution achieved. This work helps meet Exascale Computing Project goals for maximizing the use of GPUs on emerging exascale machines.

We plan to use shared memory regions in MPI within each node for storing common data structures used in various parts of the algorithm. This should reduce the overall memory utilization and allow the algorithm to scale up to larger problem sizes.

Los Alamos NATIONAL LABORATORY
NNSA National Nuclear Security Administration
Managed by Triad National Security, LLC, for the U.S. Department of Energy's NNSA.
LA-UR-22-28196
ExaFEL project funded by the DOE Office of Science and NNSA Exascale Computing Project.