EMBL

SC26

Megabytes In Murmuration: From 8-MB X-Ray Projections To Interpretable TB-Scale Structure Through Automated HPC

Marina Nikolova, Jonas Albers, Angelika Svetlove, Fabio De Marco, Ksenia Denisova, Stefan Fiedler, Elizabeth Duke, Gleb Bourenkov

European Molecular Biology Laboratory, Hamburg, Germany

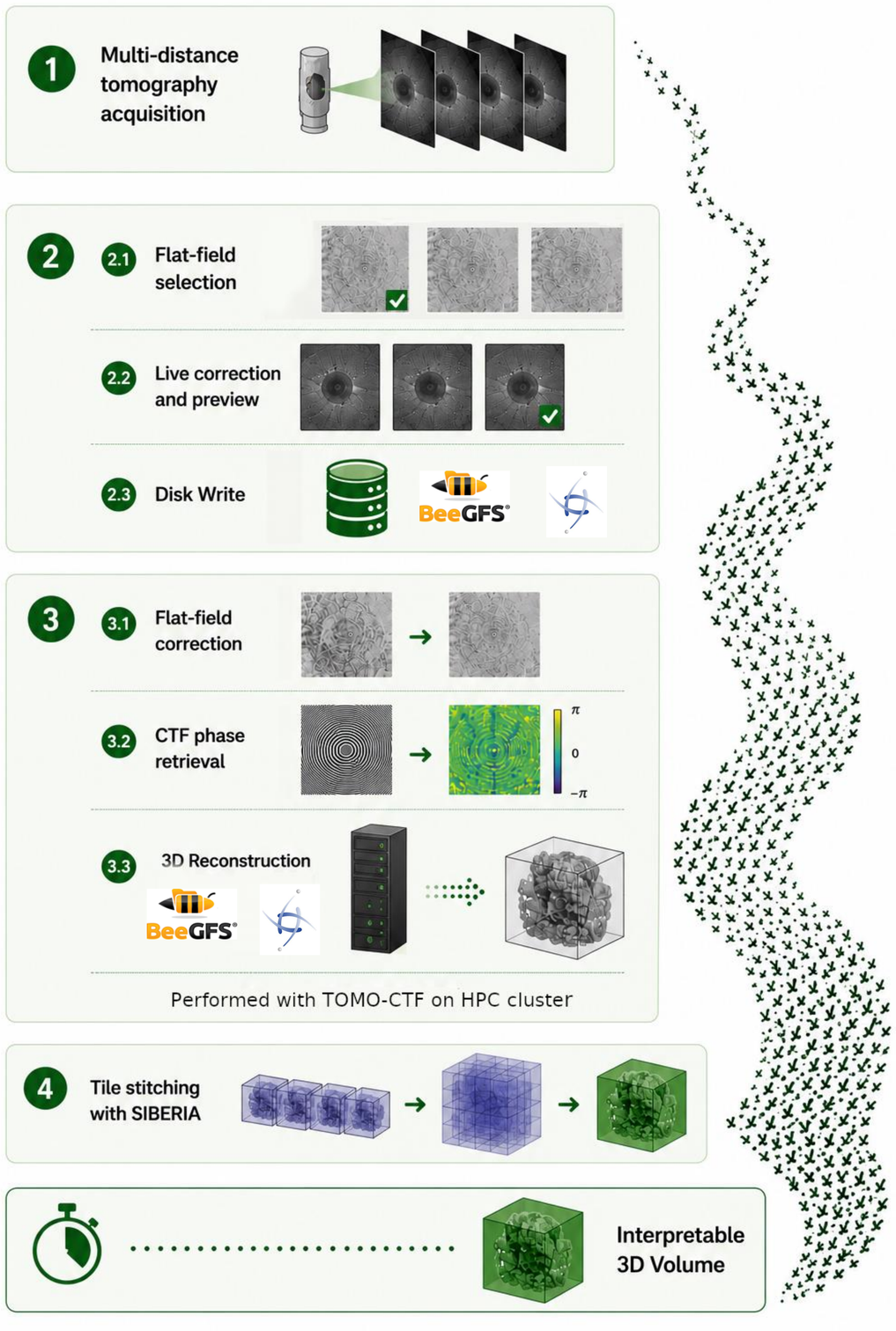
Background

- Synchrotron tomography produces thousands of 2D projections
- Individual projections provide limited biological insight
- Meaningful sample features emerge in reconstructed 3D volumes
- Extended field-of-view and tiled acquisitions increase processing complexity
- Rapid visualization is essential for experimental decision-making

Aims

- Automate the full acquisition-to-volume workflow at the P14 EMBL beamline
- Keep pace with high-throughput beamline data
- Deliver interpretable 3D volumes shortly after exposure
- Support rapid experimental decisions and efficient beamtime use

Automated Acquisition-To-Volume Pipeline



Results

End-to-end Automated Processing

- Fully automated acquisition → preprocessing → reconstruction → stitching
- Acquisition overlapped with reconstruction in tiled data collections
- Immediate feedback during the experiment

Real-time Acquisition Processing

Task	Frame Size	Rate
Camera Acquisition, (100 + 1810) x 4 frames	8 MB	100 Hz
Corrected Frame Live Preview	~ 2 MB	23 Hz
Raw Frame Live Preview	8 MB	25 Hz

Reconstruction Performance

Dataset	Compute Resource	Time
Standard 180° scan	4 × 88-core cluster nodes	30 s
Extended Field of View 360° scan	2 × 88-core cluster nodes	90 s

Raw data in TIFF format (16 bits per pixel) read from a 1.6 TB BeeGFS system
Reconstructed 3D TIFF stack (32 bits per pixel) written onto 44 TB BeeGFS NVMe storage over 56 Gb/s InfiniBand

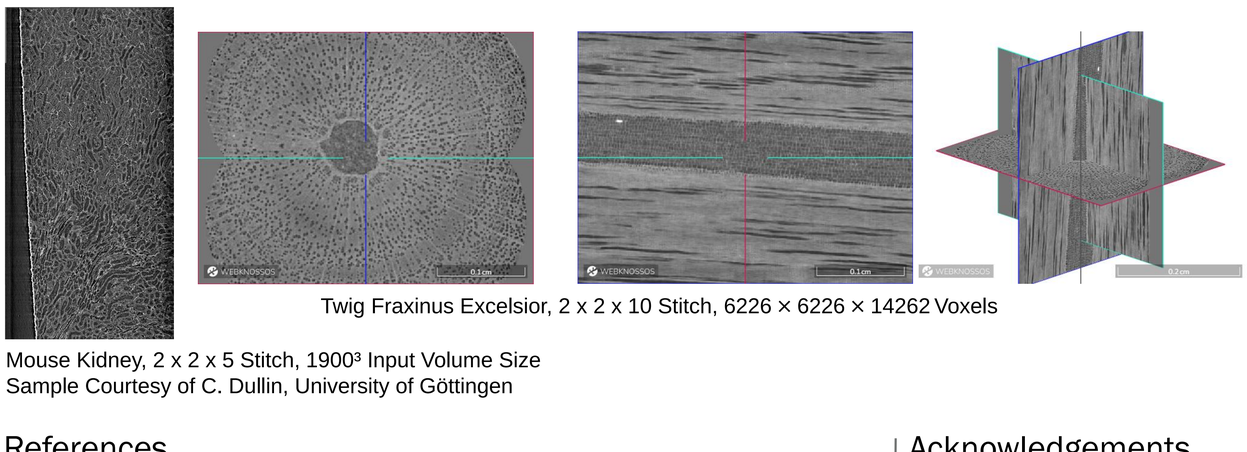
Large-scale 3D Stitching

Volume Dimensions	Volume Count X x Y x Z	Stitched Volume Size	Time
4 MP	1 x 1 x 2	27 GB	20 s
4 MP	2 x 3 x 1	51 GB	45 s
4 MP	1 x 1 x 10	114 GB	2 min
16 MP	1 x 1 x 2	107 GB	1 min 15 s
16 MP	1 x 1 x 4	194 GB	2 min 30 s
16 MP	3 x 3 x 1	207 GB	4 min 20 s
16 MP	2 x 2 x 5	552 GB	31 min
16 MP	2 x 2 x 10	1.1 TB	45 min
16 MP	4 x 4 x 7 (- 7)	2.1 TB	~ 7h

Reconstructed 3D TIFF stacks of ~2000 slices (32 bits per pixel) read from 44 TB BeeGFS NVMe storage
Stitched 3D TIFF stack (16 bits per pixel) written onto 1.6 TB BeeGFS system

Key Outcomes

- 100 Hz image stream with on-the-fly pre-processing
- Multi-stage pipeline with parallel task execution and distributed MPI cluster processing
- Large tiled datasets automatically transform into an interpretable aligned 3D volume
- Pipeline scales from a single dataset to more than 100 automatically reconstructed and stitched 3D volumes



Twig Fraxinus Excelsior, 2 x 2 x 10 Stitch, 6226 × 6226 × 14262 Voxels

Mouse Kidney, 2 x 2 x 5 Stitch, 1900³ Input Volume Size
Sample Courtesy of C. Dullin, University of Göttingen

References

Albers J. et al. High Throughput Tomography (HiTT) on EMBL beamline P14 on PETRA III. Journal of Synchrotron Radiation (2024). DOI: 10.1107/S160057752300944X

Nikolova, M., Albers, J. & Bourenkov, G. (2026). TOMO-CTF. Version 9.0. Tomographic Reconstruction Software for HiTT at P14. https://git.embl.org/grp-ximg/tomo-ctf

Nikolova, M. & Bourenkov, G. (2026). SIBERIA. Stitching Software for HiTT at P14. Version 1.0. https://git.embl.org/grp-ximg/siberia

Acknowledgements

EMBL Hamburg IT

ChatGPT

nikolova@embl-hamburg.de

EMBL Heidelberg
Meyerhofstraße 1 · 69117 Heidelberg · Germany
T+49 6221 123456-78 · www.embl.org