Towards Autonomous Experiments by Connecting High Performance Microscopy with High Performance Computing

Peter Ercius[1*], Alexander J Pattison[1], Wolfgang Theis[2], Chris Harris[3], Samuel S. Welborn[3], Bjorn Enders[3]

[1] The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America.
[2] School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.
[3] National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America.
* Corresponding author: percius@lbl.gov

The digitization of controls, data, and analysis in microscopy is bringing the idea of autonomous microscopes closer to reality than ever before. Automated transmission electron microscopy (TEM) is already fairly routine for some experiments the only require simple repetitive tasks such as imaging biological macromolecules for single particle cryoEM [1], tilt series for electron tomography [2], and movies for crystallography [3]. The vast majority of TEM experiments are conducted completely by human operators who choose the regions of interest, optimize experimental parameters, and make decisions about data quality visually during an experiment. The field is still a long way from having completely autonomous TEMs that can adapt to sample difficulties and tune experimental parameters based on data quality and desired experimental outcomes. Part of the issue is the lack of capability for feeding information learned from on-line, live data analysis back into the on-going experiment.[4] This presentation will discuss current capabilities for large scale data reduction and analysis using high performance computing (i.e. supercomputing) and progress towards developing a true feed-back loop that places data analysis and theory in the experimental loop.

The Molecular Foundry hosts a very fast direct electron detector called the 4D Camera with an 87,000 Hz readout speed producing a data stream of 480 Gbit/s. A typical raw dataset consists of over 700 GBytes acquired in 15 seconds.[5] The large amount of data produced spurred the connection of the detector data acquisition system with the National Energy Research Scientific Computing Center (NERSC) for data reduction and on-line analysis. Data reduction is accomplished in less than 5 minutes providing rapid feedback to the operator. We have integrated the 4D Camera with a pipeline-based automatic control system capable of long running high resolution experiments.[6] We have also implemented a web-based online control, data transfer, and metadata tracking system for data processing.[7] This provides an ideal platform for designing and testing algorithms for autonomous operation with direct access to large scale data processing. [8]

References:

1. Y.Z. Tan, A. Cheng, C.S. Potter, B. Carragher, *Microscopy*, **65** (2016), p. 43–56. https://doi.org/10.1093/jmicro/dfv369
2. J. Feng, A.P. Somlyo, A.V. Somlyo, and Z. Shao, *Journal of Microscopy*, **228** (2007), p. 406–412. https://doi.org/10.1111/j.1365-2818.2007.01859.x
3. M. Roslova, S. Smeets, B. Wang, T. Thersleff, H. Xu, X. Zou, *Journal of Applied Crystallography*, **53** (2020), p. 1217–1224. https://doi.org/10.1107/S1600576720009590

4. S. R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S. V. Kalinin, M. J. Olszta, R. E. Dunin-Borkowski, N. Salmon, K. Hattar, W.-C. D. Yang, R. Sharma, Y. Du, A. Chiaramonti, H. Zheng, E. C. Buck, L. Kovarik, R. L. Penn, D. Li, X. Zhang, M. Murayama, M. L. Taheri, *Nature Materials*, **20** (2021), p. 274–279. https://doi.org/10.1038/s41563-020-00833-z
5. P. Ercius, I.J. Johnson, P. Pelz, B.H. Savitzky, L. Hughes, H.G. Brown, S.E. Zeltmann, S.-L. Hsu, C. C.S. Pedroso, B.E. Cohen, R. Ramesh, D. Paul, J.M. Joseph, T. Stezelberger, C. Czarnik, M. Lent, E. Fong, J. Ciston, M.C. Scott, C. Ophus, A.M. Minor, A.P. Denes, *arXiv* [physics.ins-det] (2023). https://arxiv.org/abs/2305.11961
6. A.J. Pattison, C.C.S. Pedroso, B.E. Cohen, J.C. Ondry, A.P. Alivisatos, W. Theis, P. Ercius, *Nanotechnology*, **35** (2023). https://doi.org/10.1088/1361-6528/acf938
7. Distiller, https://github.com/OpenChemistry/distiller (accessed February 20, 2024)
[8]. This work was primarily funded by the US Department of Energy in the program "4D Camera Distillery: From Massive Electron Microscopy Scattering Data to Useful Information with AI/ML." Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.