TomoPy: A framework for the analysis of synchrotron tomographic data

Doğa Gürsoy,* **Francesco De Carlo and ? ?**

Advanced Photon Source, Argonne National Laboratory, 9700 S. Cass Ave., Argonne IL 60439-4837 USA.
Correspondence e-mail: dgursoy@aps.anl.gov

Abstract goes here.

1. Introduction

Analysis of large tomographic datasets at synchrotron light sources is becoming progressively more challenging due to the increasing data acquisition rates that new technologies in x-ray sources and detectors enable. A next generation of synchrotron facilities currently under design and construction around the world, will provide diffraction limited X-ray sources and is expected to boost the current data rates by several order of magnitude stimulating even more the need for efficient analysis tool development and integration. Several synchrotron users, to take advantage of specific instrument characteristics and specifications, already collect data at different facilities, but they are often left alone integrating this data with available or home grown software tools.

Explain important capabilities and vision: Uniting efforts of different facilities and beamlines doing similar tasks. Emphasize key terms: Open-source, collaborative, well-documented and maintained. Python/C++, OS independent, parallelizable, modular and able to support a functional programming style that many people in Academia prefer.

Within the same class of instruments there are often strong similarities that are not systematically utilized to build a coordinated software development platform or framework where each facility and user group can take advantage from or contribute in, ultimately saving on-site resources by sharing the computing tasks with the user community.

Cross-development and integration between labs is not well because of... We aim to overcome this issues by ....

Here we describe in details an attempt to provide such a collaborative framework for the analysis of synchrotron tomographic data that has the potential to unify the effort of different facilities and beamlines performing similar tasks. The basic principles of this Python/C++ based framework, called *tomopy*, are: OS and data format independent, parallelizable, modular and supporting a functional programming style that many researchers in Academia prefer.

2. Background

Which other facilities and beamlines it is (is expected to be) used? Touch the idea of Data Exchange (once this issue is overcome ...). Examples of collaborations with Academia?

When it comes to the digital storage of tomographic experimental data and the development of analysis tools at synchrotron light sources around the world, the situation is very heterogeneous. As different research teams and instruments have grown at various facilities, they have often developed local data and analysis models based on instrument hardware specificity and often drawing upon the particular preferences of a scientist writing software.

A little bit of background history of APS practices. Touch ongoing LDRD project. What is going on at other beamlines and other Labs? Other projects addressing this issues?

Many tomographic analysis tools utilize licensed, and often expensive, software packages like Matlab and IDL, other rely on specific, and often complex to maintain, computing infrastructure like MPI-based CPU or GPU clusters.

The availability of inexpensive multi-core CPU workstation and the reduced cost of computer memory allowed to simplify the computing infrastructure required to process tomographic data (Rivers, 2012), and started the development of single workstation tools able to perform most computing tasks.

This transition is affecting all major synchrotron facilities where new effort is now dedicated into developing new tools that can be deployed at the facility for real time processing as well as distributed to users for off site data processing.

3. TomoPy framework

3.1. Programming paradigm

The differences between academia and industry. Mention common coding practices in Academia. People in Academia are leaned towards functional programming.

Modularity is favored. The overall effect of removing a module on the package is tried to be minimized.

3.2. Structure

Explain code structure. Functions and interfaces with the code-base. Touch Data Exchange and data converters (give a complete list of converters and how they can be implemented). Explain data domain module (preprocessing), data-to-image domain module (recon) and image domain module (post process). Mention about visualization (discuss about Mayavi, Avizo, Amira, etc.) and what is planned.

Put a diagram figure and explain throughly.

3.3. Computing platforms

There are a number of viable options to to digest such high-throughput data and facilitate the computational analysis as a means of utilizing technological advancements. GPUs, CPUs and scalability.

GPU computing: The emergent computational potential of graphical processing units (GPUs) in the recent past has led to a great interest of research and development in ways to leverage GPUs for computationally intensive tasks. However, the memory of a single GPU is still not sufficiently large enough to digest large tomographic datasets without avoiding excessive input/output (I/O) overhead.

CPU computing on dedicated workstations: Current trend in processor design is towards chips with multiple low-power consuming cores which leads to a new layer of sophistication in code development. More hands-on control over data and is preferable in terms of support and development projects.

Scalability with clusters of GPUs or CPUs. APS has access to Mira, BNL uses NERSC. Others? Touch data transfer issues. Users have less control over data. Large computer-clusters generally require an efficient queuing optimization, computation times are highly correlated with the demand.

Discuss input output formats. Mention ongoing attempts :Data Exchange.

We aim an open open source platform that everybody can contribute and install easily.

4. Development

Explain how anyone can contribute. How it is distributed. BSD License. Documentation.

5. Methods

Present here the concept of pipeline for data processing and distribution

5.1. Pre-processing methods

5.1.1. Ring removal: The reconstruction quality can be dramatically degraded when local inhomogeneities consistently appear in every projection data. This is usually caused by the impurities of the scintillation detector and can not be totally removed with the flat-field correction. These measurement artifacts usually stand out as vertical stripes in sinograms and are transformed to reconstructions as ring shaped artifacts. The process of clearing these artifacts are called as the ring-removal.

There are numerous ways to correct these effects. The easiest and computationally most effective version is the median filtering along the sinogram. Median filtering works nicely without compromising reconstruction quality if the stripes are a few pixel width. One of the most powerful methods is the Wavelet-Fourier method which exploits the appearance of the

5.1.2. Phase retrieval Single-step methods.
Touch maybe briefly multi-energy approaches with energy-selective measurements.

5.1.3. Registration Data fusion: Local tomography.
Phase-correlation: Correction of wabbling in images. Alignment for multi-step phase retrieval.

A 180 degrees rotation of the sample is generally sufficient to efficiently solve the reconstruction problem. However for some samples, especially when the sample size is larger than the field of view of the detector, a rotation of 360 degrees can be used as an option for data acquisition.

5.2. Reconstruction

Determination of center of rotation.
Gridrec. ART.

5.3. Post reconstruction

??

5.4. Storing Data

Formats supported. Many are happy with TIFF.
Explain storing provenance and its importance.

6. Discussions

Wrap the important points, vision and aims and conclude.

Appendix A
Appendix title

Appendix text.

Acknowledgements

References

Donath, T., Rissi, M. & Billich, H. (2013). Synchrotron Radiation News, **26**(5), 34–35.
  URL: *http://www.tandfonline.com/doi/abs/10.1080/08940886.2013.832589*
Rivers, M. (2012). In Developments in X-Ray Tomography VIII, vol. 8506. SPIE.