WIR SCHAFFEN WISSEN – HEUTE FÜR MORGEN

**Jun Zhu :: Paul Scherrer Institute**

TOMCAT data reduction and processing pipelines development

**August 17, 2022**

Outline

- Overview

- **TOMCAT live reconstruction pipeline (tomcat-live)**
  on-the-fly data reduction and processing

- General data reduction and processing pipeline prototype (drappp)

- Joined AWI in January

- Main task: SLS2.0 data processing and reduction
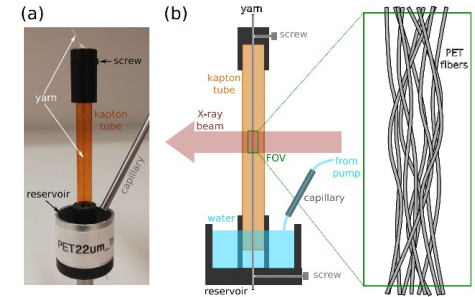
- TOMCAT beamline as a starting point
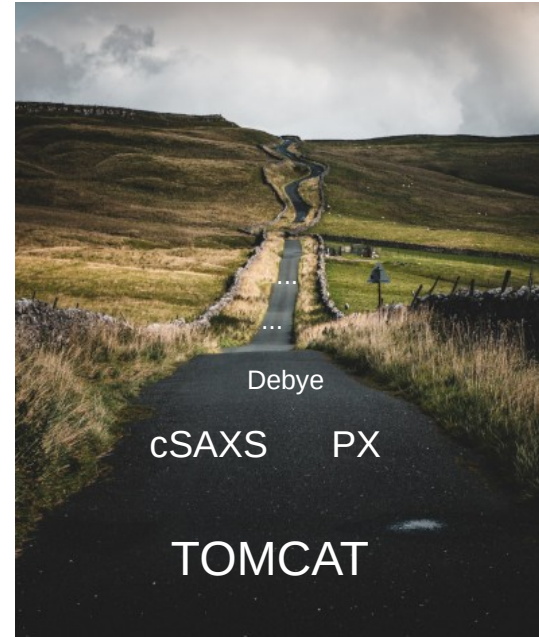
Retrospective data readout based on real-time tomographic reconstruction

Peak data rate > 100 TB/day (at the start of SLS2.0 operation?)
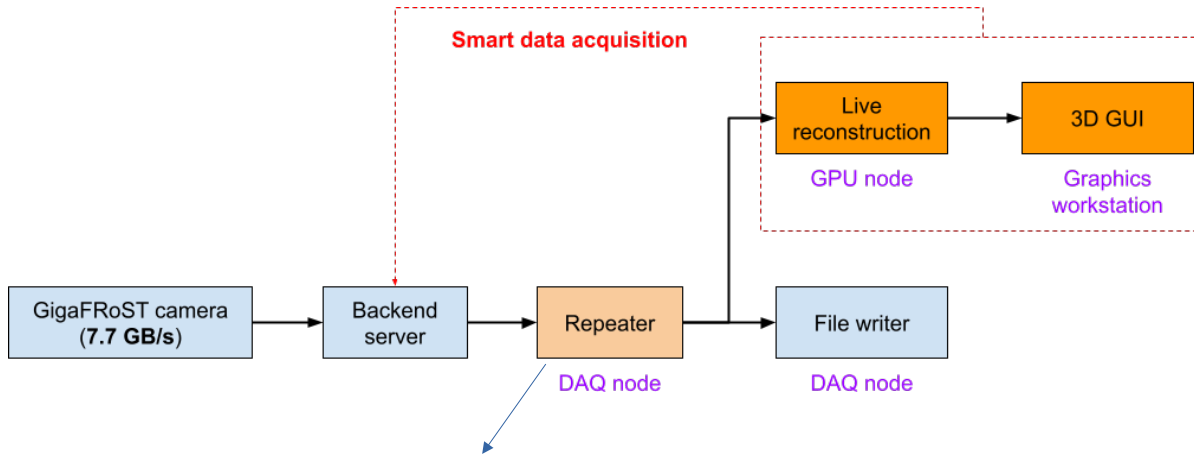
Credit: Leonardo Hax

Buurlage, JW., et al. Sci Rep 9, 18379 (2019)

projections ➡ tomograms

Live reconstruction pipeline – starting point

- RECAST3D

  https://github.com/cicwi/RECAST3D

  > **Not** dedicated for TOMCAT

  > client-server architecture

   client: 3D GUI in OpenGL

   server: preprocessing (CPU) and reconstruction (GPU, ASTRA Toolbox)

- Proof-of-principle test was performed in 2019

  Buurlage, JW., et al. Sci Rep 9, 18379 (2019)

  > Server: temporary solution with code from RECAST3D

  > Not performed in a continuous way

   The existing GigaFRoST backend was not optimized for live data processing.

  > Throughput

   410 MB/s on NVIDIA Quadro K6000 (12 GB)

   Up to 700 MB/s on the GPU node at TOMCAT

- Pure C++17 library (Python plug-in removed)

  performance-critical code

- Replaced "home-made" dependencies with standard libraries

  parallel programming → *oneTBB*, logging → *spdlog*, program options → *boost*, ...

- Extensive code clean-up

- Added unit tests

- Identified and fixed issues with the infrastructure

  Credit: Leonardo Sala and others

- Performance optimization → **throughput up to 2.1 GB/s (3 x)**

  GPU node: NVIDIA V100 32 GB, Intel Xeon Gold 6248R (24 cores / 48 threads)

- Integrated **tomcat-live** (new name) with the existing DAQ

  Ready for test during beamtime

Live reconstruction pipeline – DAQ

- **Existing DAQ**
  - > Frame IDs are wildly unordered: … 6000, 2001, 2010, 2005, 6008 …
  - > A workaround was implemented in tomcat-live to cope with this problem

  A number of projection groups waiting to be filled …

Projection group m	Projection group m + 1	……	Projection group m + n

- **Standard DAQ**
  - > Frame IDs are expected to be ordered.
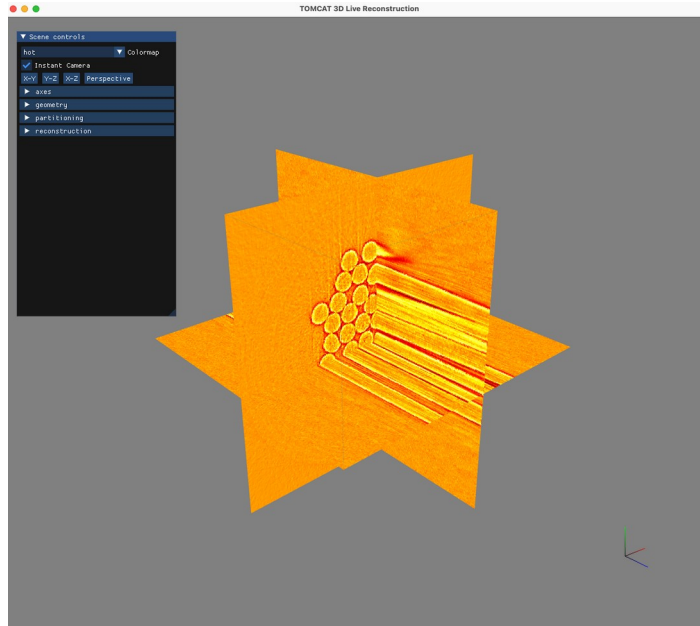
- Three high-resolution slices and a low-resolution volume are reconstructed in real-time.

- Slices can be reconstructed on-demand.
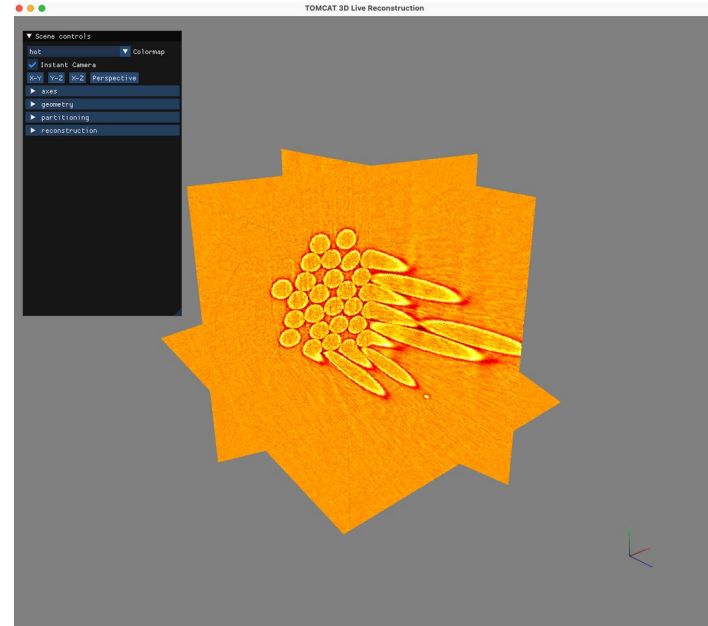
Arbitrary orientations of the slices can be selected.

Live reconstruction pipeline – data reduction

SDSC machine learning proposal: **Smart Data Acquisition for Tomoscopy Experiments**

- "Compression ratio" 4 – 40 *vs* < 4 for safe lossy compression
- Team up with TOMCAT and Controls
- Passed the pre-proposal phase
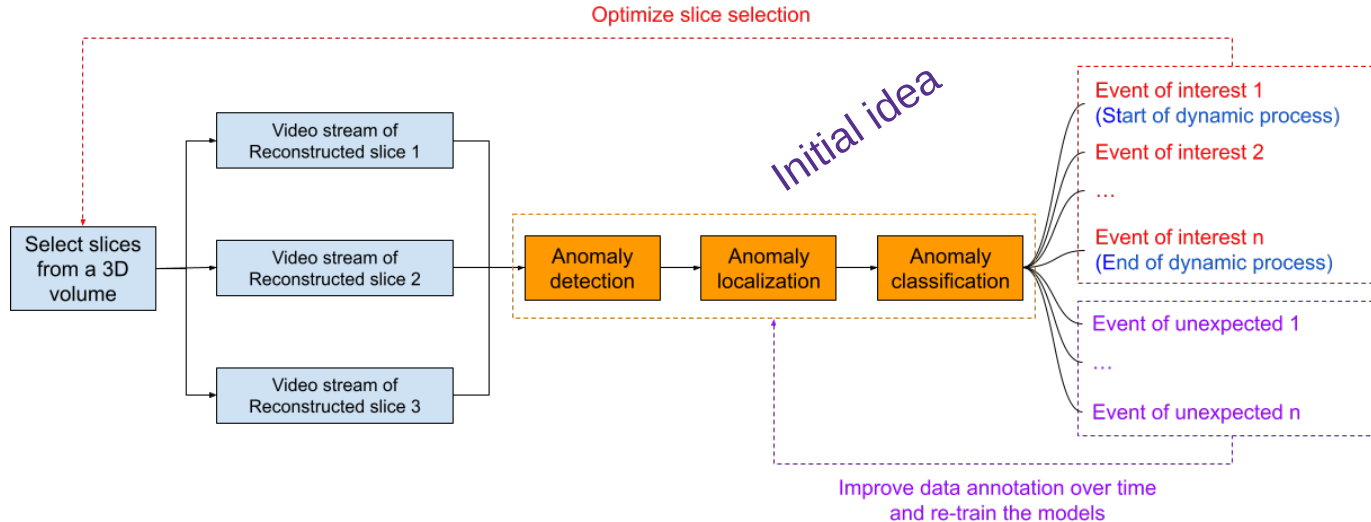
Live reconstruction pipeline – next steps

- Tests during in-house beam times (August and September)

- Deployment

- Hand it over to users

- Improve user experience and add more features

- Finish the full proposal of the SDSC project (by October 24)

- Integrate it with the standard DAQ (after November)

General data reduction and processing pipeline prototype

Requirements (TOMCAT):

- Fast and flexible (plug and play).
- Run on laptops and HPC clusters.
- Save and load the result for each step.

Current status:

- First draft functioning (interfaces to be defined).
- Collaborating with Alain to integrate the existing pipeline.
- Briefly explored other frameworks such as AiiDA and Zocalo.
- Starting to collect requirements from the other beamlines.

My thanks go to

- Markus Janousch

- Leonardo Sala

- Leonardo Hax Damiani

- Christian Schlepütz

- Alain Studer

- Xingxing Yao

- Tadej Humar

- Goran Lovric

- Federica Marone

- ...