

PAUL SCHERRER INSTITUT



Jun Zhu :: Paul Scherrer Institute

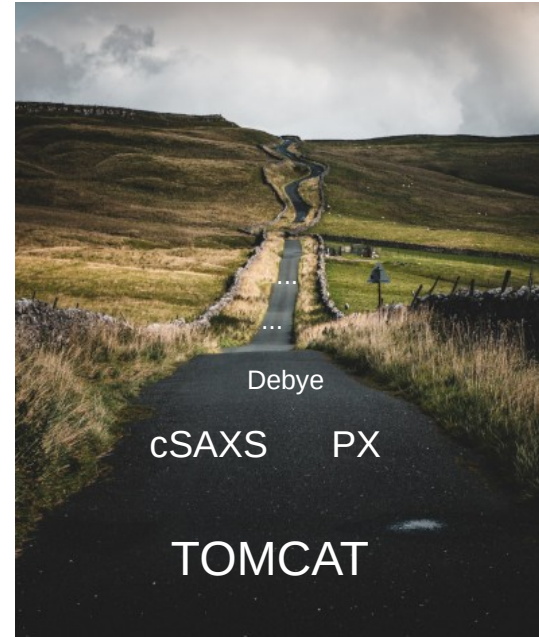
TOMCAT data reduction and processing pipelines development

August 17, 2022



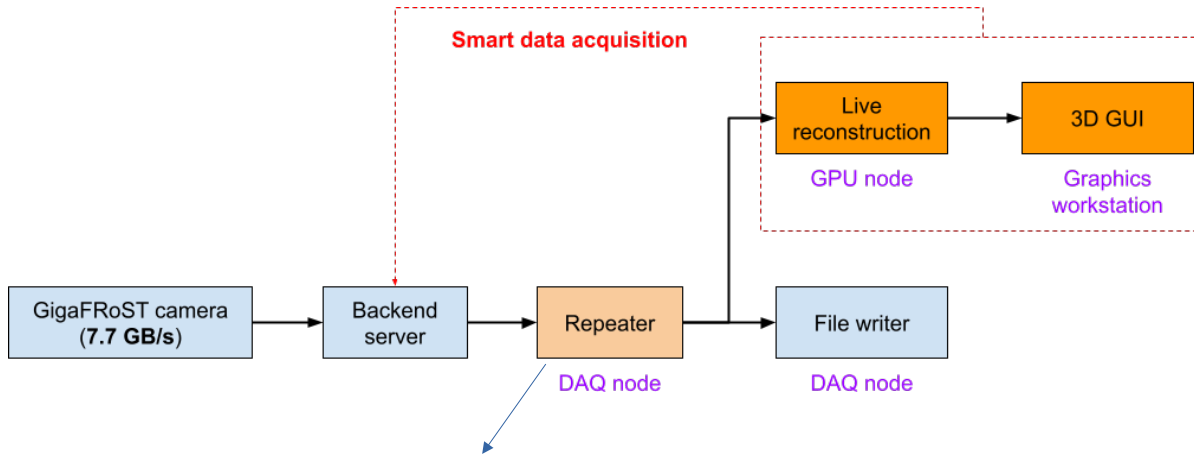
- Overview
- **TOMCAT live reconstruction pipeline (tomcat-live)**  
*on-the-fly data reduction and processing*
- General data reduction and processing pipeline prototype (drappp)

- Joined AWI in January
- Main task: SLS2.0 data processing and reduction
- TOMCAT beamline as a starting point

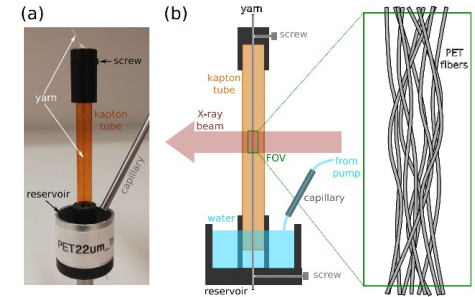


## Retrospective data readout based on real-time tomographic reconstruction

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



Credit: Leonardo Hax



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

projections → tomograms

## - 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

[Burlage, 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*

# Live reconstruction pipeline – current status

- 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

## - Existing DAQ

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

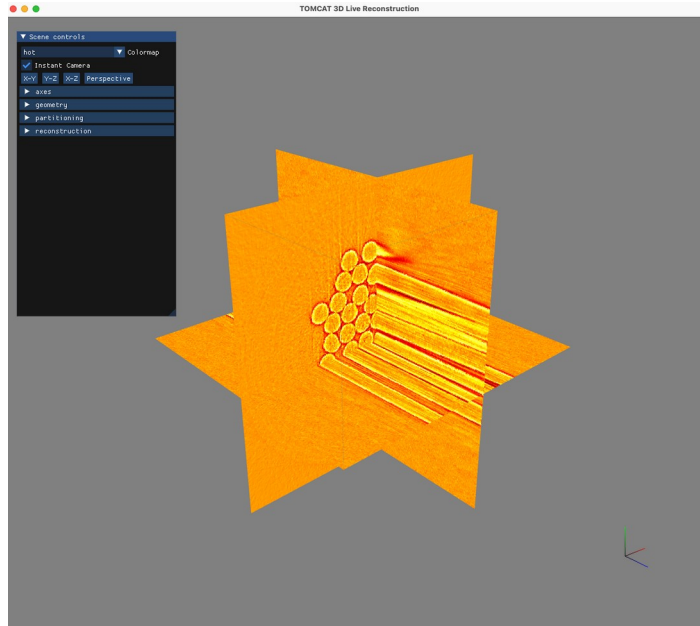
A number of projection groups waiting to be filled ...



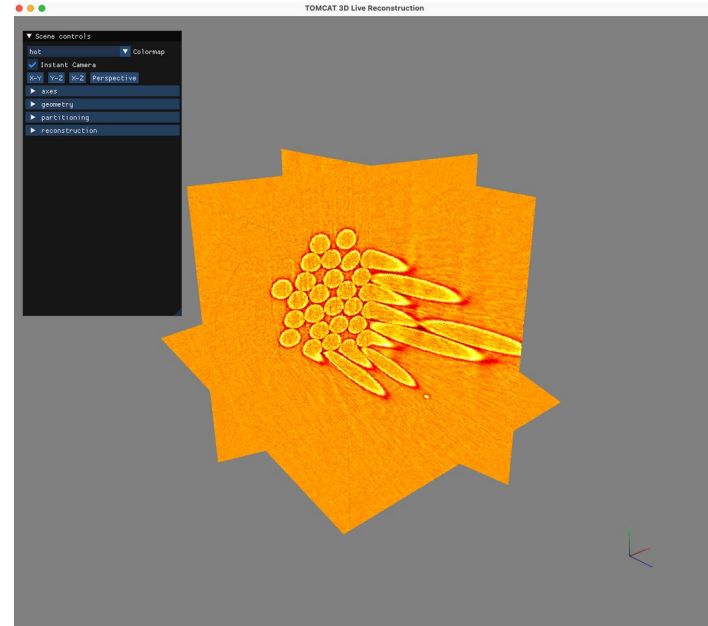
## - Standard DAQ

- > Frame IDs are expected to be ordered.

# Live reconstruction pipeline – GUI



- Three high-resolution slices and a low-resolution volume are reconstructed in real-time.
- Slices can be reconstructed on-demand.

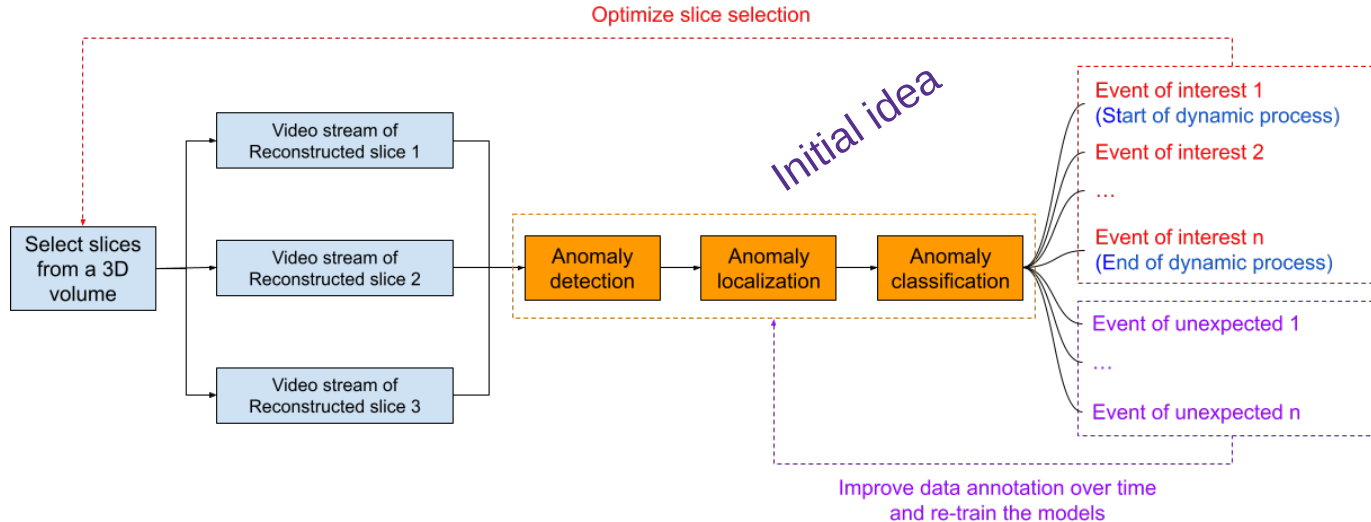


Arbitrary orientations of the slices can be selected.



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



- Bubble formation
- Fluid uptake
- ...

## 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](#))

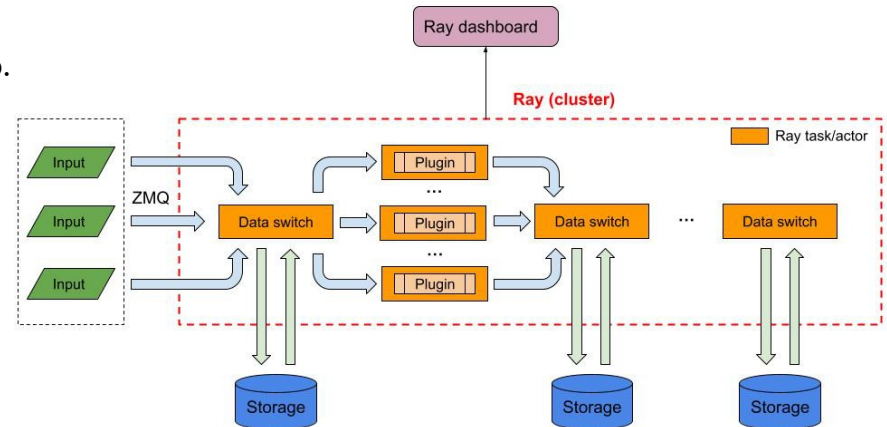
## Requirements (TOMCAT):

- 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.

Host	PID	Uptime (s)	CPU	RAM	Pages	Disk	Send	Received	Logs	Errors
ray-018.gal.ch (128.128.85.76)	32 workers / 32 cores	204 18h 13m 28s	84.3%	28.1 GB / 251.6 GB (11%)	0.0 MB / 72111.3 MB	4.3 GB / 35.0 GB (12%)	0.1 MB/s	0.0 MB/s	<a href="#">View all logs (183 lines)</a>	No errors
ray-019.gal.ch (128.128.85.76)	32 workers / 32 cores	204 18h 13m 42s	81.6%	27.0 GB / 251.6 GB (11%)	0.0 MB / 72110.6 MB	4.3 GB / 35.0 GB (12%)	0.1 MB/s	0.0 MB/s	<a href="#">View all logs (176 lines)</a>	No errors
ray-020.gal.ch (128.128.85.81)	32 workers / 32 cores	204 18h 13m 42s	86.4%	28.0 GB / 251.6 GB (11%)	0.0 MB / 72115.6 MB	4.3 GB / 35.0 GB (12%)	0.1 MB/s	0.0 MB/s	<a href="#">View all logs (183 lines)</a>	No errors
ray-017.gal.ch (128.128.85.71)	32 workers / 32 cores	146 01h 45m 52s	76.0%	27.0 GB / 251.6 GB (11%)	0.0 MB / 72073.5 MB	4.3 GB / 35.0 GB (12%)	0.6 MB/s	0.2 MB/s	<a href="#">View all logs (241 lines)</a>	No errors
ray (PID: 22322)	start_raylog	00h 00m 00s	3.1%	71.3 MB	N/A	N/A	N/A	N/A	<a href="#">View log (1 line)</a>	No errors
ray (PID: 22324)	start_raylog	00h 00m 00s	3.1%	341.3 MB	N/A	N/A	N/A	N/A	<a href="#">View log (1 line)</a>	No errors
ray (PID: 22416)	process_actor	00h 00m 56s	102.6%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22416)	process_actor	00h 00m 56s	102.6%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22469)	process_actor	00h 00m 55s	102.2%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22470)	process_actor	00h 00m 55s	103.0%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22470)	process_actor	00h 00m 55s	102.2%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22470)	process_actor	00h 00m 55s	102.6%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22470)	process_actor	00h 00m 55s	102.6%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22493)	process_actor	00h 00m 55s	102.2%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22484)	process_actor	00h 00m 55s	102.2%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22485)	process_actor	00h 00m 55s	102.2%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22486)	process_actor	00h 00m 55s	102.6%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22487)	process_actor	00h 00m 55s	102.6%	454.8 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22488)	process_actor	00h 00m 55s	103.0%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22489)	process_actor	00h 00m 55s	101.8%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors
ray (PID: 22490)	process_actor	00h 00m 55s	102.6%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (9 lines)</a>	No errors
ray (PID: 22491)	process_actor	00h 00m 55s	102.2%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (9 lines)</a>	No errors
ray (PID: 22492)	process_actor	00h 00m 55s	102.6%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (9 lines)</a>	No errors
ray (PID: 22493)	process_actor	00h 00m 55s	102.2%	452.0 MB	N/A	N/A	N/A	N/A	<a href="#">View log (7 lines)</a>	No errors



## My thanks go to

- Markus Janousch
- Leonardo Sala
- Leonardo Hax Damiani
- Christian Schlepütz
- Alain Studer
- Xingxing Yao
- Tadej Humar
- Goran Lovric
- Federica Marone
- ...

