

WP7 Task 3

Felicita Gernhardt, Peter Steinbach

Helmholtz-Zentrum Dresden-Rossendorf, Core Facility for Digital Infrastructure and Computing

3rd May 2022

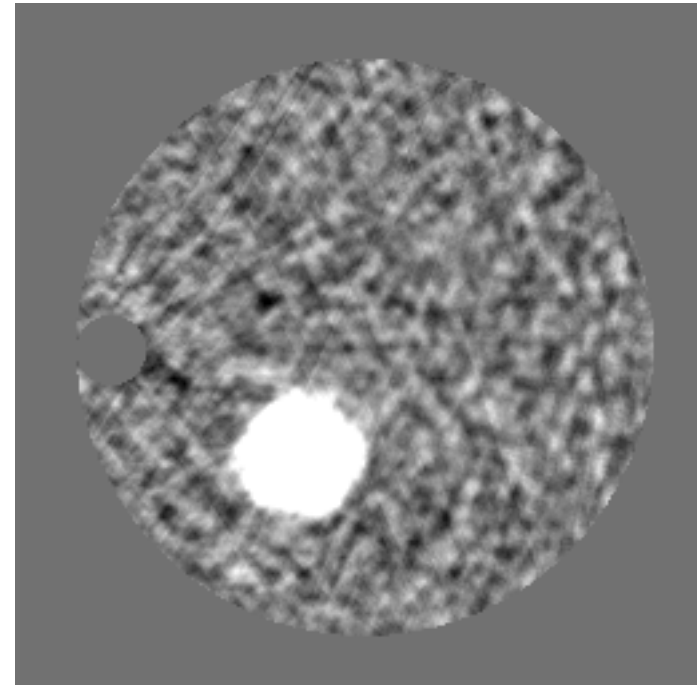
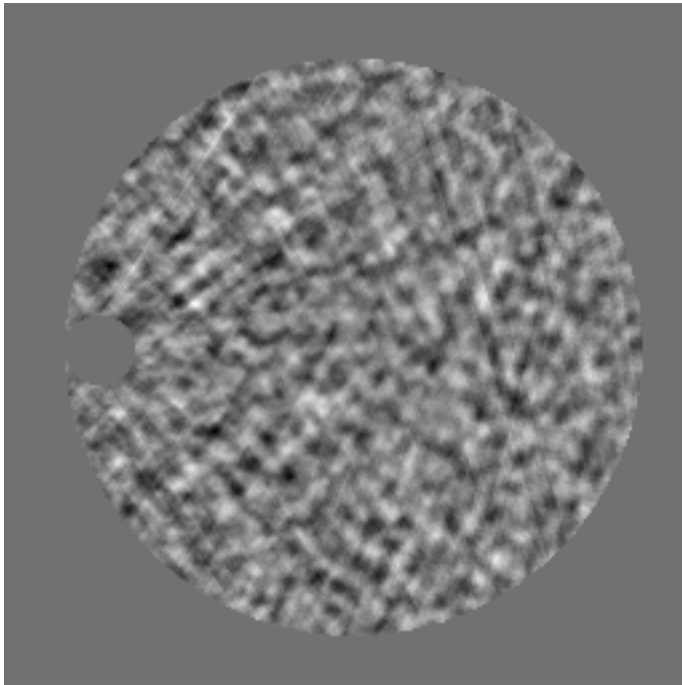


Task

- Adopting new algorithms and technologies for lossless and lossy data compression
- Datasets from domain scientists
- Try out different compression techniques, write custom compression pipelines
- Feedback important

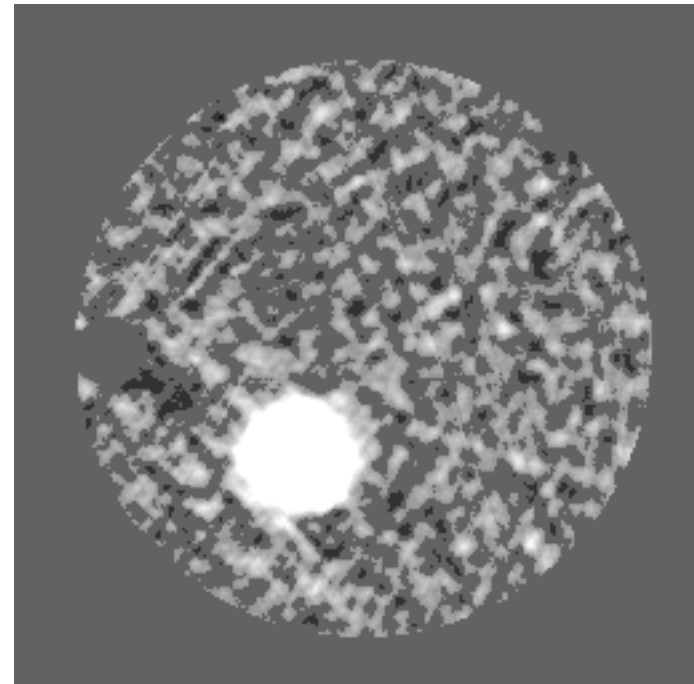
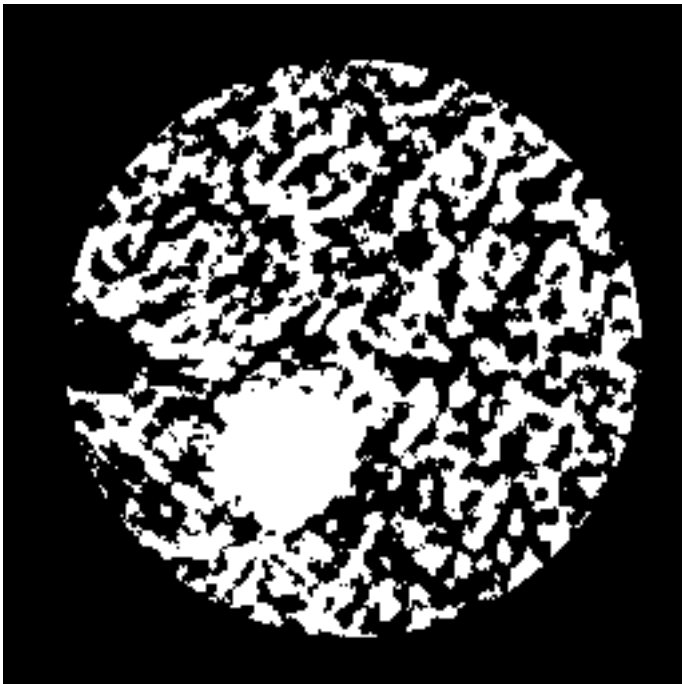
ROFEX, single bubble study

- [Ultrafast electron beam X-ray computed tomography](#)
- Dual-tier-dataset of a single bubble study



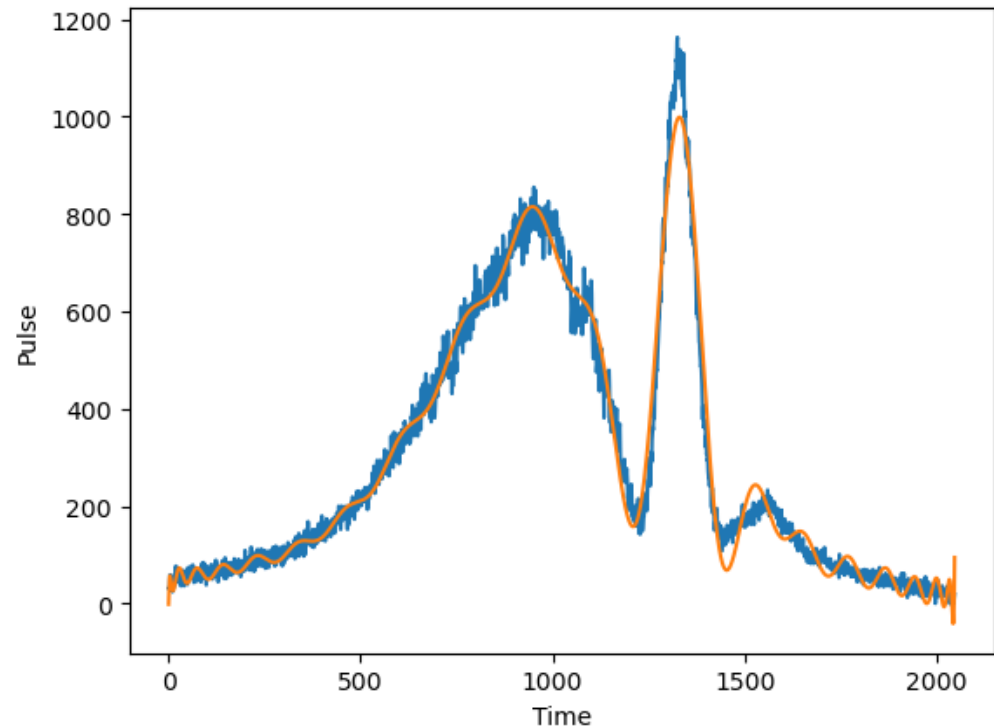
ROFEX, single bubble study

- Otsu thresholding to separate signal and noise
- Conversion from float32 to uint8 (256 buckets)
- Noise quantized into 4 buckets, signal quantized into 252 buckets
- With additional lossless compression (zip), **compression ratio of 10**
 - But feedback about downstream analysis still needed



T-ELBE

- High-field THz-driven Phenomena group
- Signal of pulse represented with 2048 values
- Compression: polynomial fitting
- Polynom degree 50
 - 50 (float64) values instead of 2048 (uint16) values
 - Compression ratio: 10
- Drawback: tends to overshoot
 - Feedback still needed

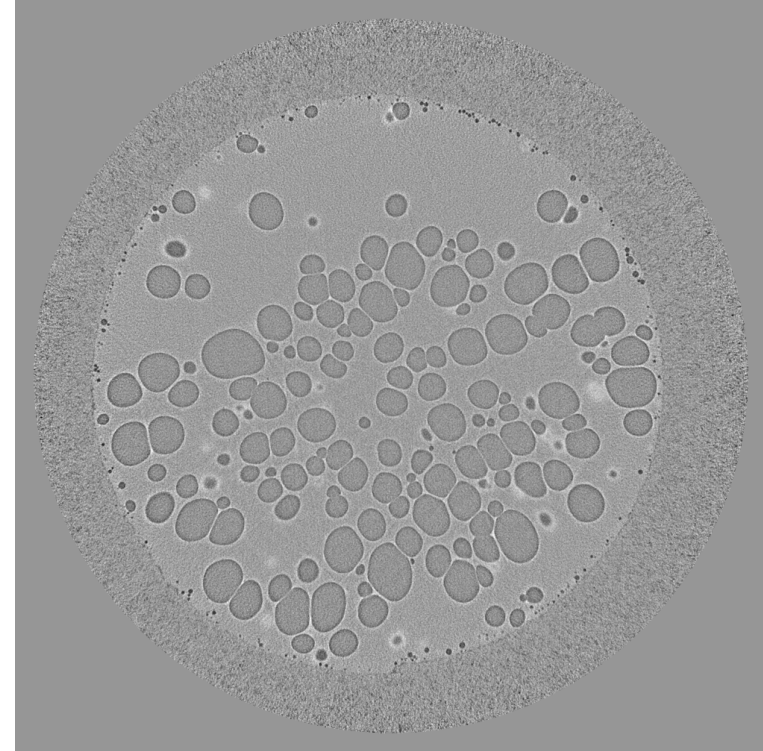


ALBA

- Received datasets from Alessandra Patera
- X-ray, acquired at TOMCAT beamline
- Datasets previously analyzed
 - Important for measuring quality of lossy compression
- Next step: take a look at data, what compression technique is most suitable? Denoising?

PSI

- Federica Marone paper: “[Impact of lossy compression of X-ray projections onto reconstructed tomographic slices](#)”
- Datasets used in paper publicly available
- Shared code for analysis
- Alain Studer shared script for reconstruction
- Next step: denoising, what effect has it on compression results?



Summary

- Received and worked with datasets from within LEAPS and HZDR internal
- Mostly tried out lossy compression
- Challenge: Measuring quality after applying lossy compression
 - Communication with domain scientists important
 - How does downstream analysis look like?
 - What parts of the data are most important, what parts are noise?