Data Analysis at the European XFEL using Karabo

T. Michelat[a], A. Barty[d], M. Beg[a], M. Bergemann[a], V. Bondar[a], C. Danilevski[a], M. Duarte Trevisani[a], W. Ehsan[a], S. Esenov[a], R. Fabbri[a], G. Flucke[a] , D. Fullà Marsà[a], G. Giovanetti[a], D. Göries[a], S. Hauf[a], D. Hickin[a], E. Kamil[a], D. Khakhulin[a], Y. Kirienko[a], A. Klimovskaia[a], T. Kluyver[a], M. Kuhn[d], A. Lemos[a], D. Mamchyk[a], V. Mariani[d], A. Parenti[a], H. Santos[a], A. Silenzi[a], C. Youngman[a], J. Zhu[a], H. Fangohr[a,c] and S. Brockhauser[a,b]

[a]European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany, [b] Biological Research Center (BRC), Hungarian Academy of Sciences, Temesvri Krt. 62, Szeged, 6726, Hungary, [c]University of Southampton, Southampton, SO17 1BJ, United Kingdom, [d] Center for Free Electron Laser Science, DESY, Hamburg, Germany

The European XFEL has gathered about 1.5PB of raw data in the first year of operation. We describe the data analysis capabilities [1] of the Karabo distributed control system [2] and related software developed at the European XFEL to supporting scientific experiments. The range of requirements for data analysis includes near-real time during experiments and offline analysis.

Data Structure

The European XFEL facility generates coherent and intense **X-ray** pulses by bunches of up to **2700 pulses** repeating every **100 ms**.

The data generated by instruments and detectors for each of these pulses are distributed **at 10 Hz**, which corresponds to the pulses train structure. One **train data** container consist of many images taken for each pulse in a train.

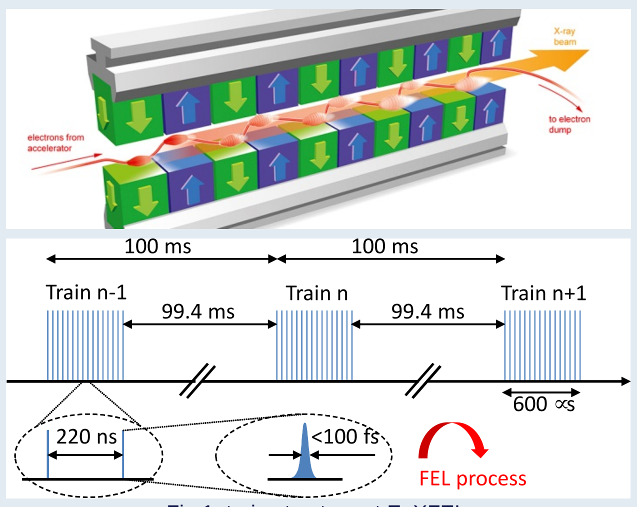

Fig 1. train structure at EuXFEL

Online Processing

During experiments, **near real-time** feedback to users and beam scientists is important in order to optimize the experiment setup and collect the best data possible. We provide solutions to process data and give visualization feedback in the order of seconds.

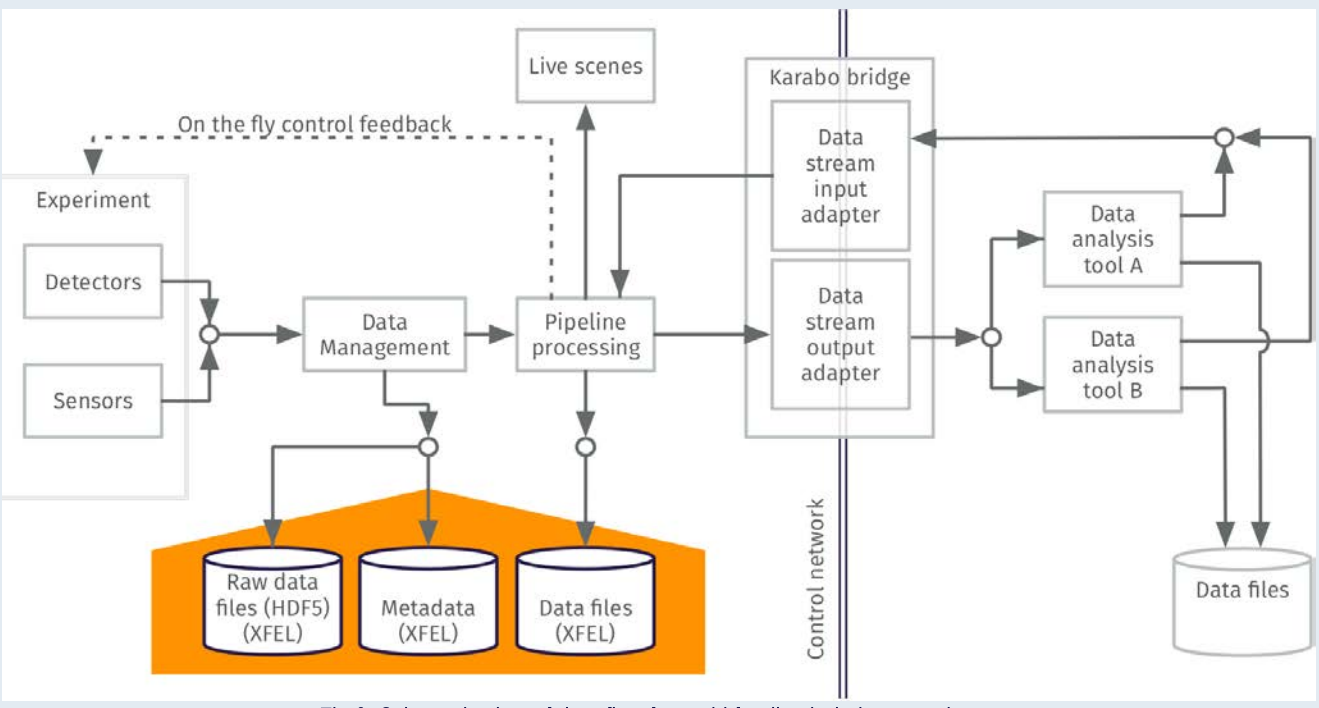

Fig 2. Schematic view of data flow for rapid feedback during experiment

Data Pipelines

The Karabo framework support **peer-to-peer** messaging between devices, allowing the implementation of processing pipelines with **high data rates**. Pipelines implemented in the Karabo environment can easily **parallelize** and **distribute** heavy processing tasks on many physical machines.
Any processing algorithm implemented as a Karabo device can use this feature and benefit from the **Karabo integration** (results permanently stored via **DAQ**, GUI control and **live view**, …)

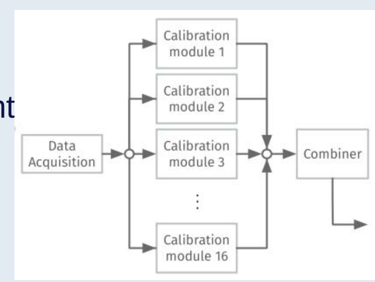

Fig 3. example of pipeline processing

Online Detector Calibration

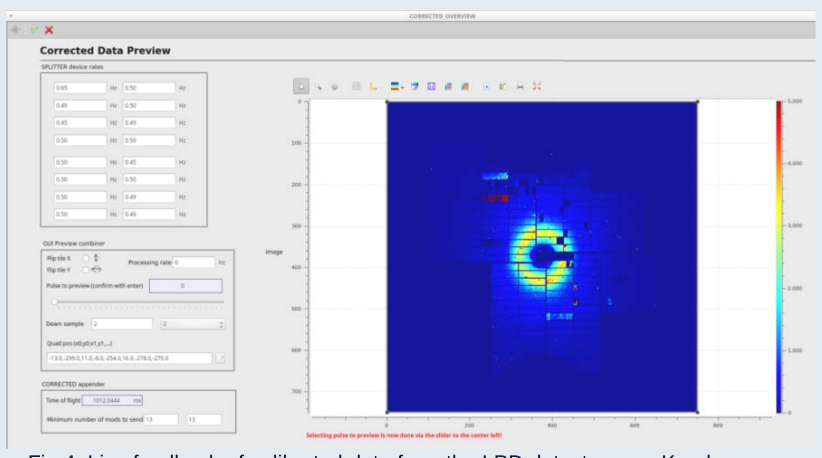
Fig 4. Live feedback of calibrated data from the LPD detector on a Karabo scene

The **large detectors** used at the European XFEL need appropriate **correction** and calibration before further processing can be applied on the data they produce. This is carried out using the data pipelines **distributing** the data over 8 GPU processing nodes on a HPC cluster.

Karabo Bridge

The **Karabo bridge** allows data analysis tools to **receive data** from the Karabo control system without having to be build against the framework. It uses **ZeroMQ** and **msgpack** to send data efficiently. In the future, this will be extended to send data back to Karabo as well. The European XFEL provides **C++** [3] and **Python** [4] clients.

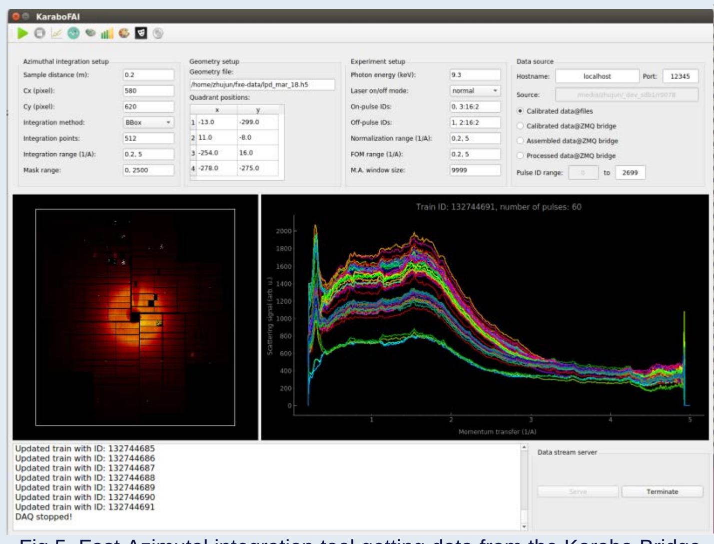
Fig 5. Fast Azimutal integration tool getting data from the Karabo Bridge

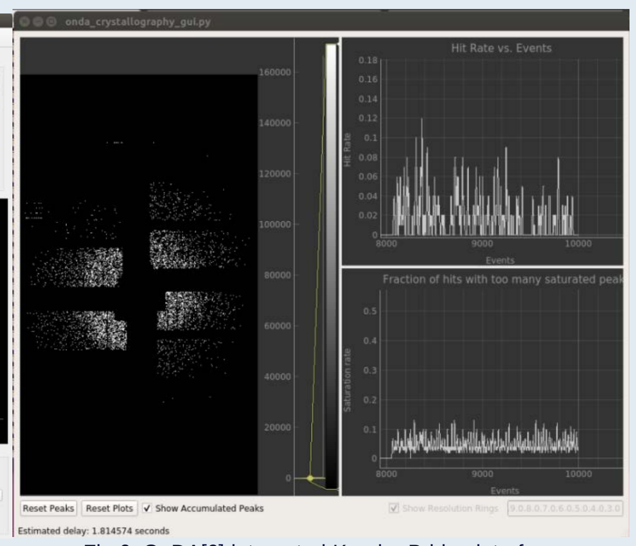
Fig 6. OnDA[6] integrated Karabo Bridge interface

The **Karabo Bridge** has been used to **feed** analysis tools such as **OnDA** [6], **Hummingbird** [7], **CASS** [8], **Dozor** [9] and **PyFAI** [10] based software for near-real time feedback.

Offline Processing

We develop and provide a Python package for data analysis: **Karabo-data** [5]. It is designed for **reading data** stored by the **Karabo data acquisition** service and offer a range of function and routines to **support** users **offline data analysis**.
Karabo-data is an open source project and is in active development. Feature requests and contributions are welcome. Some of the current features:

- Run data reading and exploration
- Data filtering (sources, trains, …)
- Data correlation (Pandas, Xarray)
- Data views to fit external tools
- Stream data as the Karabo Bridge
- Apply detector geometry
- Data conversion (from HDF5)

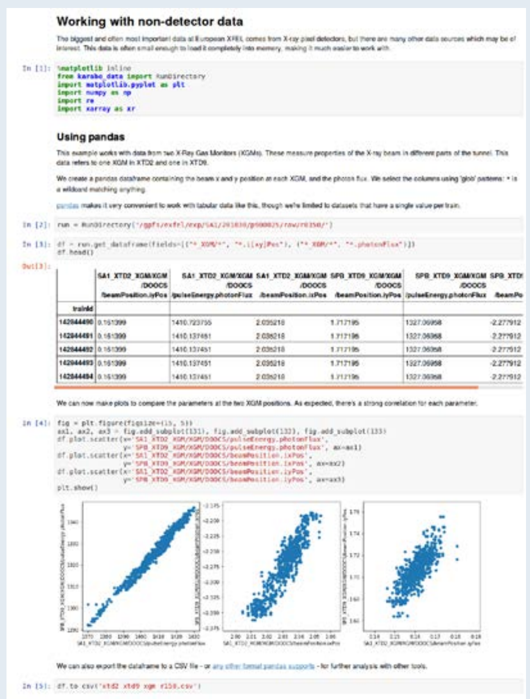
Fig 7. karabo-data example using Pandas dataframe

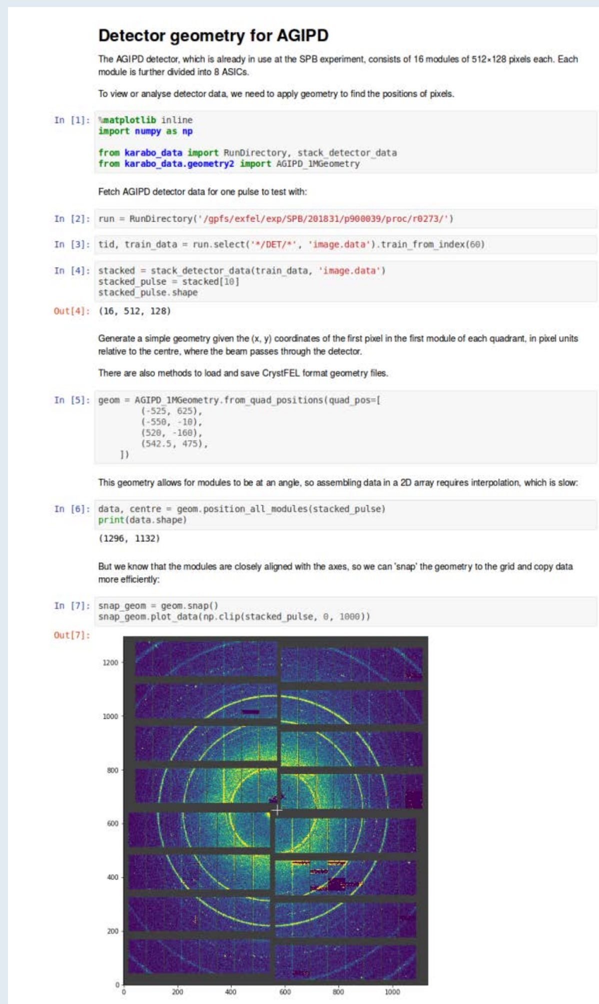
Fig 8. applying geometry correction on detector data

References

[1] H. Fangohr *et al.*, "Data Analysis Support in Karabo at European XFEL", in *Proc. 16th Int. Conf. on Accelerator and Large Experimental Control Systems (ICALEPCS'17)*, Barcelona, Spain, Oct. 2017, paper TUCPA01, pp. 245-252, ISBN: 978-3-95450-193-9, https://doi.org/10.18429/JACoW-ICALEPCS2017-TUCPA01, 2018.
[2] B. Heisen, D. Boukhelef, S. Esenov, S. Hauf, I. Kozlova, L. Maia, A. Parenti, J. Szuba, K. Weger, K. Wrona *et al.*, "Karabo: An integrated software framework combining control, data management, and scientific computing tasks," in *14th International Conference on Accelerator & Large Experimental Physics Control Systems, ICALEPCS2013. San Francisco, CA*, 2013.
[3] https://github.com/European-XFEL/karabo-bridge-cpp
[4] https://github.com/European-XFEL/karabo-bridge-py
[5] https://github.com/European-XFEL/karabo_data
[6] http://journals.iucr.org/j/issues/2016/03/00/zf5001/index.html
[7] http://journals.iucr.org/j/issues/2016/03/00/zd5007/index.html
[8] http://journals.iucr.org/j/issues/2016/04/00/zw5003/index.html
[9] http://journals.iucr.org/d/issues/2015/11/00/tz5083/index.html
[10] http://journals.iucr.org/j/issues/2015/02/00/fv5028/index.html

Acknowledgements

We acknowledge contributions from all staff and collaborators at European XFEL who helped shaping and realizing the presented framework, and financial support from the OpenDreamKit Horizon 2020 European Research Infrastructures project #676541, and the Gordon and Betty Moore Foundation through Grant GBMF #4856, by the Alfred P. Sloan Foundation and by the Helmsley Trust.

European XFEL GmbH, Thomas Michelat, Holzkoppel 4, 22869 Schenefeld, Germany, Phone +49 40 8998-6739, Fax +49 40 8998-1905, thomas.michelat@xfel.eu
www.xfel.eu