Satellite workshop on Data Analysis and Karabo

XFEL User Meeting 2018, 23 January 2018

- Agenda
  - 14:00 Welcome (S. Brockhauser)
  - 14:05 Overview (H. Fangohr)
  - 14:15 Introduction to Karabo (G Flucke)
  - 14:30 Detectors and Calibration (S Hauf)
  - 15:15 Break (Coffee)
  - 15:45 Offline and Online Data analysis at XFEL (T Michelat)
  - 16:30 Discussion
  - 17:00 Close
- Online agenda and URLs at
  - http://bit.ly/2dayxfel

Overview: Data Analysis and Karabo

Hans Fangohr
Control and Analysis Software Group
Senior Data Analysis Scientist

DESY, FLASH seminar room, 23 January 2018

Outline

- European XFEL status
- Karabo
- Overview data analysis infrastructure
- Online data analysis
- Offline data analysis
- Reproducible science
- Outline of the day


European XFEL research campus Schenefeld
North Sea
Denmark
United Kingdom
Edinburgh
Isle of Man
Dublin
Ireland
Manchester
Liverpool
London
Amsterdam
Netherlands
Brussels
Belgium
Cologne
Germany
Frankfurt
Luxembourg
Paris
France
Berlin
Poland
Warsaw
Prague
Czechia
Munich
Vienna
Slovakia
Austria
Budapest
Hungary
Switzerland
Milan
Slovenia
Zagreb
Croatia
Bosnia and Herzegovina
Serbia
Belgrade
Београд
Sarajevo
Monaco
Italy
Rome
Bay of Biscay
Andorra
Barcelona
Porto
Madrid
Portugal
Lisbon
Spain
Valencia
Tyrrhenian Sea
Riga
Latvia
Lithuania
Vilnius
Minsk
Мінск
Belarus
Moscow
Москва
Voronezh
Воронеж
Kiev
Київ
Kharkiv
Харків
Ukraine
Moldova
Chisinau
Odessa
Одеса
Romania
Bucharest
Black Sea
Krasnodar
Краснодар
Sofia
София
Bulgaria
Kosovo
Podgorica
Подгорица
Macedonia (FYROM)
Tirana
Albania
Greece
Athens
Istanbul
Bursa
Ankara
Izmir
Turkey


Hamburg
Schenefeld
Osdorfer Born
3400m
DESY-Bahrenfeld
0
500 m

European XFEL

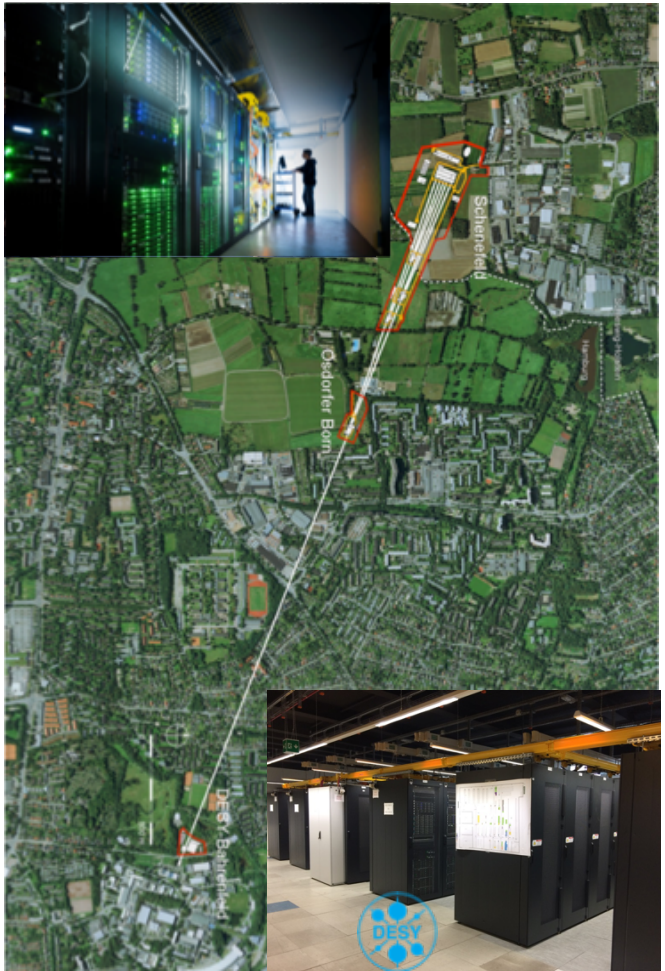
- Official opening 1 September 2017
- 2 of 6 scientific instruments live
- First experiments started 14 Sept 2017
- 12 proposals collected ~700 TB data
- Positive feedback

Prof. Dr. Johanna Wanka, Bundesministerin für Bildung und Forschung, visits SPB hutch

Data analysis infrastructure

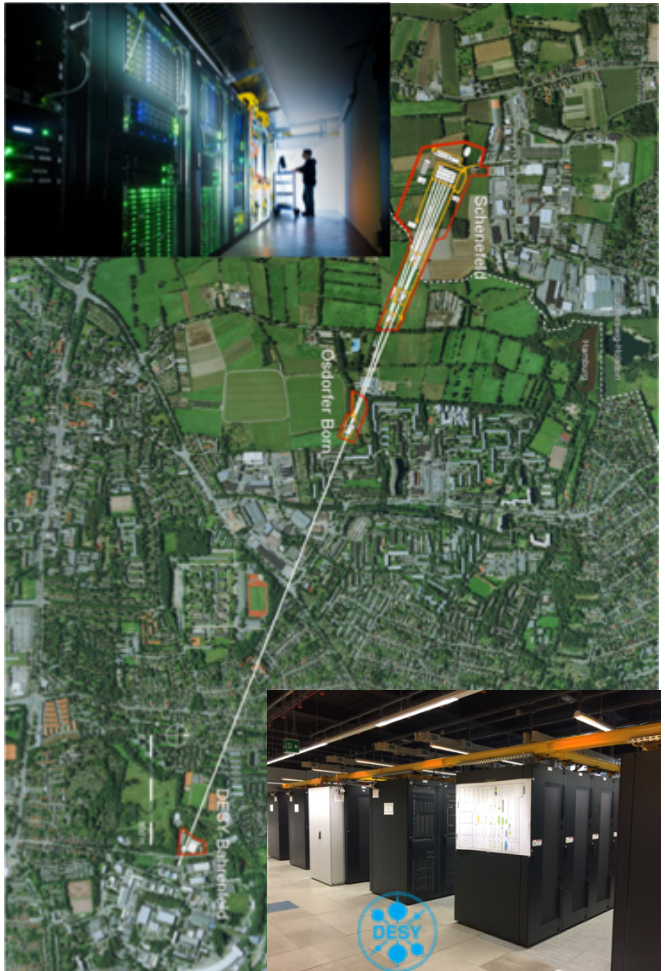
- Hardware: “Online cluster”,
  - 8 nodes x (20 cores, 256GB RAM) dedicated to users
  - Additional nodes for control and XFEL provided calibration and processing

- Hardware: “Offline cluster” = Maxwell cluster (DESY)
  - 80 nodes/3200 cores (Intel Xeon E5-2698v4)
  - ~112 TFlops
  - 512GB RAM each node
  - +20 nodes with other spec
  - 7 GPU nodes available



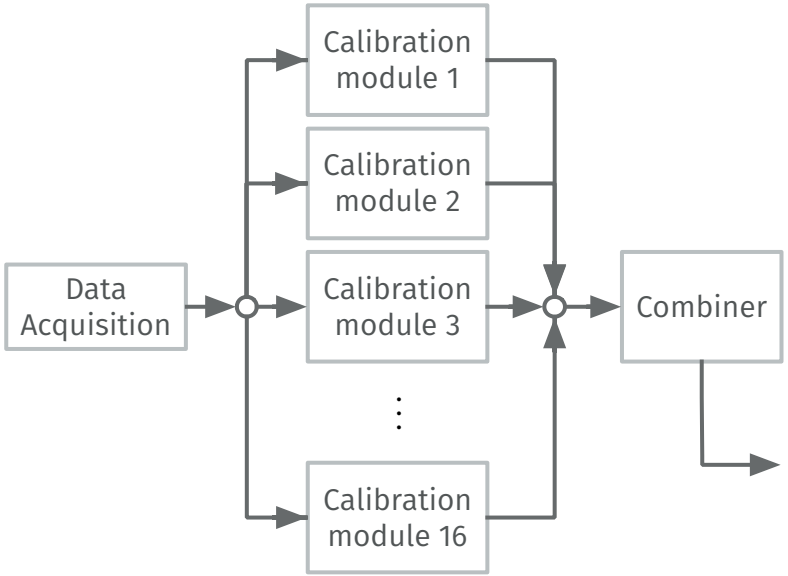
Data availability

- During a measurement (run)
  - Calibrated and raw data available in hutch (GUI, online)
- Data migration after each run
  - After each run, data manager decides on quality of the data: “good”, “unclear”, “bad”
  - “good” and “unclear” data transferred to “Offline cluster”
  - Migration triggers computation of calibrated data at online cluster
- After experiment
  - Raw and calibrated data available
  - Analysis on ”Offline cluster” (Maxwell @ DESY)



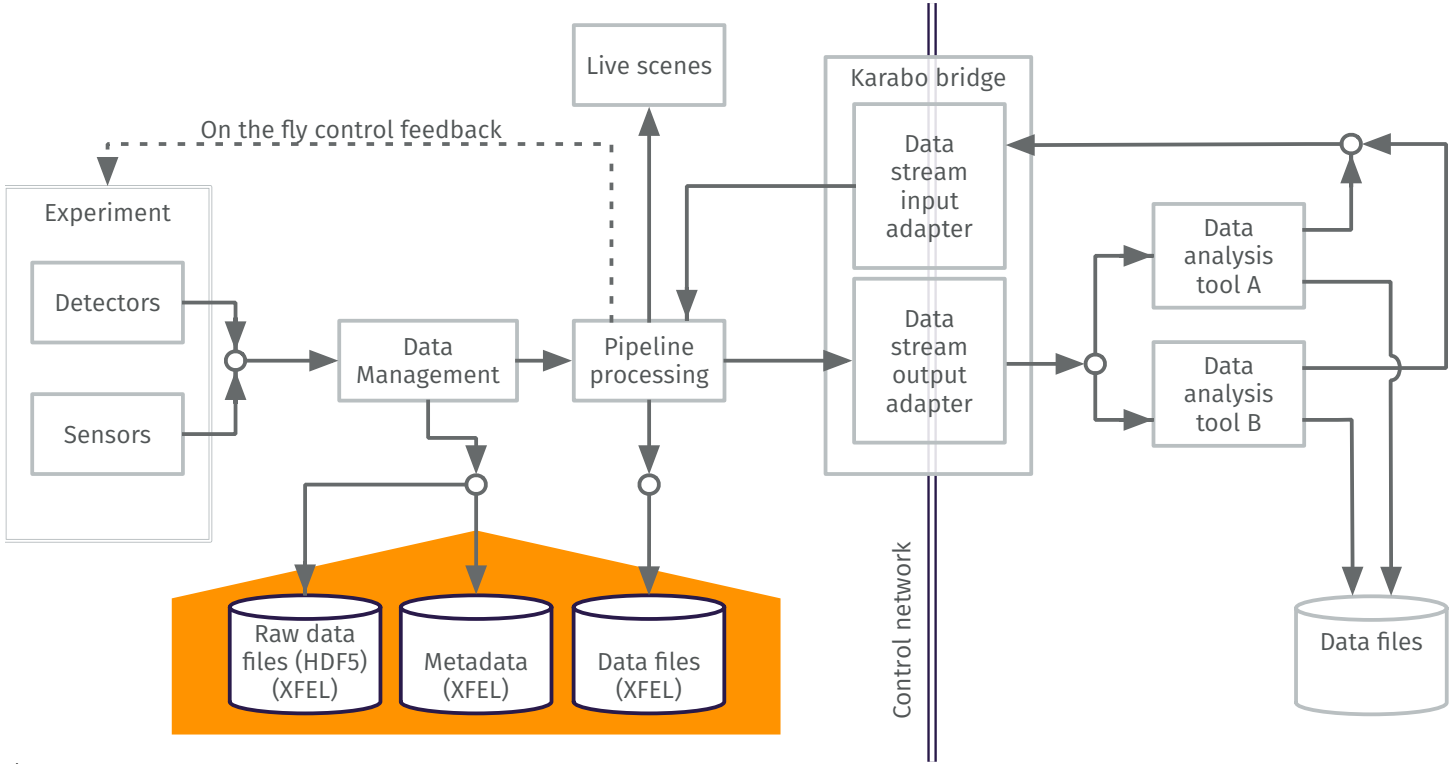
Karabo & Karabo processing pipeline example

- Karabo is framework for control and data
  - Data tokens pass through pipeline
  - Processing units called “devices”
  - Devices can be distributed over hardware
  - Simplified example in figure: calibration for detector modules carried out in parallel



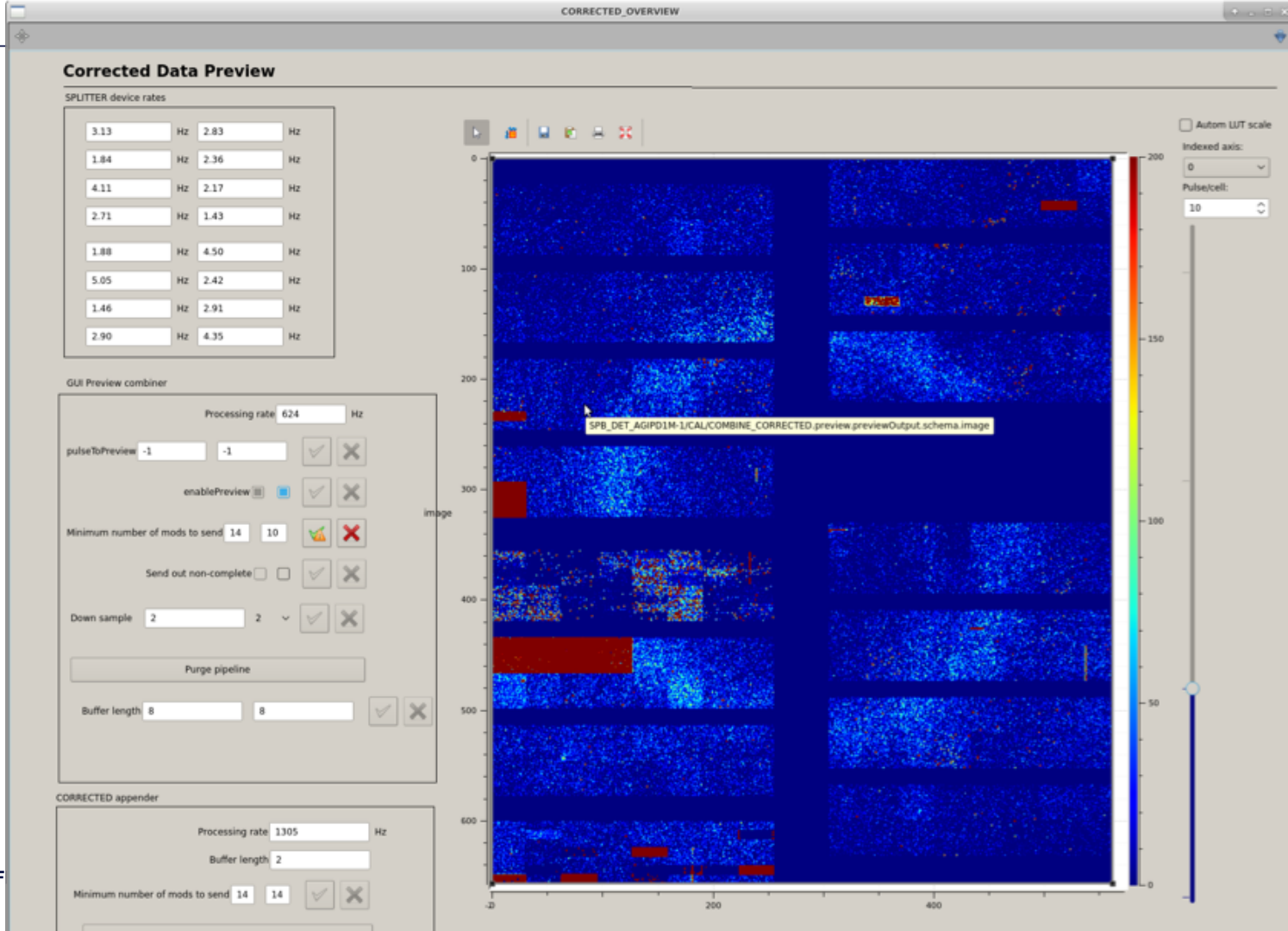
- More details:
  - 14:15 G. Flucke: “Karabo”
  - 14:30 S. Hauf: “Detectors and Calibration”

Online data analysis

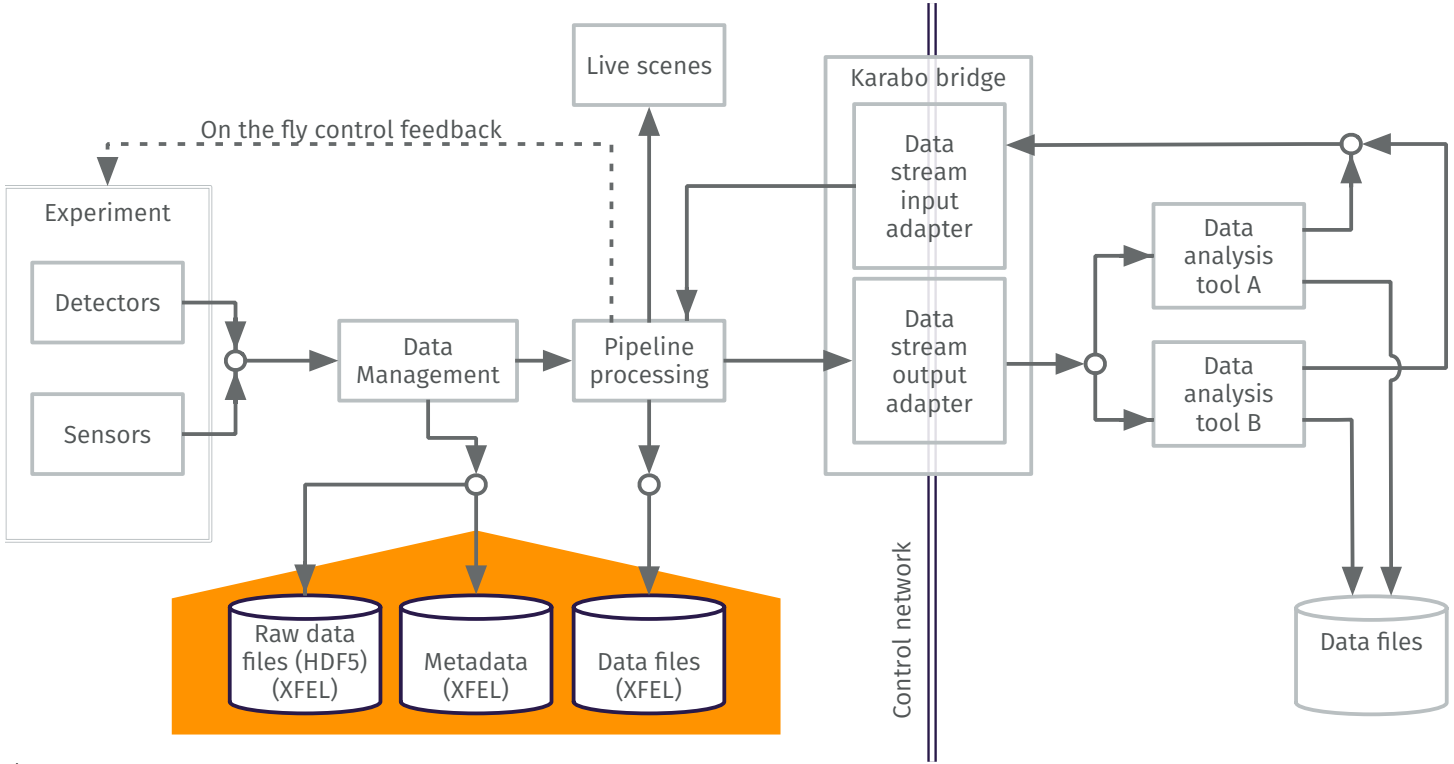


Online data analysis

Online data analysis: Rapid feedback through GUI



Online data analysis



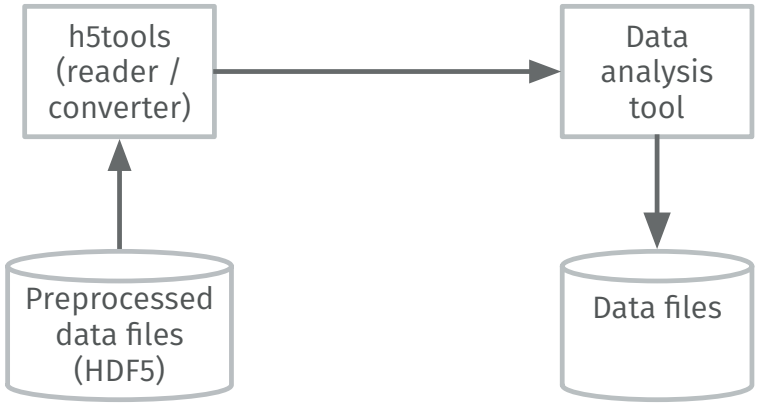
Online data analysis

More details:

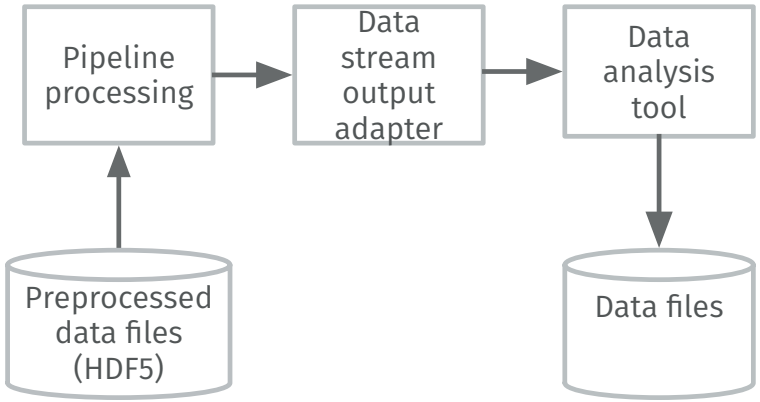
15:45 T. Michelat: “Online Data analysis”

Offline data analysis

- Processing HDF5 files
  - Using European XFEL's h5tools (https://github.com/European-XFEL/h5tools-py)
  - Or read files directly

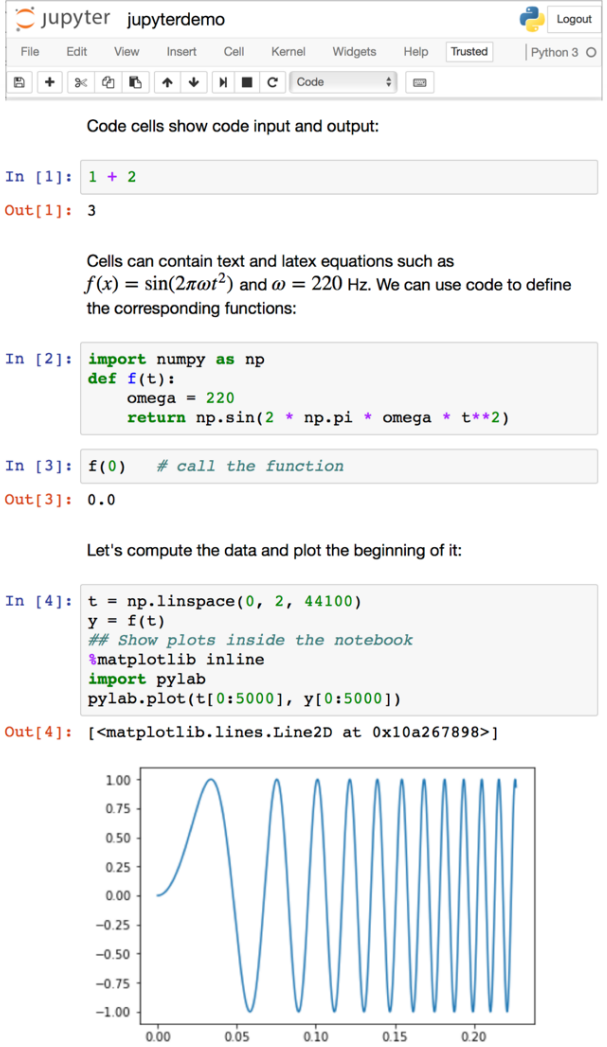


- Sending HDF5 files through the Karabo bridge
  - Imitates online setup
  - Good for re-use of interface
  - Under development – ask if interested



Reproducible Science and Jupyter

- Jupyter Notebook (Morning session)
  - Executable document
  - Code, output, interpretation
- Jupyter Ecosystem
  - Docker, Binder
  - Reproducibility -> better science
- XFEL tools integrate in Notebook
  - grow library of analysis tools and recipes with community
- [1] http://github.com/European-XFEL
  [2] https://in.xfel.eu/readthedocs/docs/pydetlib/en/latest/index.html



Summary

- Karabo, infrastructure, calibration, online and offline analysis
- User support
  - Support before, during and after experiment
  - Growing set of open source tools
  - Collaboration with users and other facilities desired
- Online agenda and URLs at
  - http://bit.ly/2dayxfel
- Agenda
  - 14:15 Introduction to Karabo (G Flucke)
  - 14:30 Detectors and Calibration (S Hauf)
  - 15:15 Break (Coffee)
  - 15:45 Offline and Online Data analysis at XFEL (T Michelat)
  - 16:30 Discussion
  - 17:00 Close
- Contact
  - hans.fangohr@xfel.eu
  - http://fangohr.github.io
  - @ProfCompMod