



U.S. DEPARTMENT OF
ENERGY



**UNIVERSITY OF
CALIFORNIA**





BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY



U.S. DEPARTMENT OF
ENERGY

The Gold Standard at XFELs

Aaron S. Brewster

Workshop on MX raw image data formats, metadata and validation

Aug 22nd, 2020



research papers

IUCrJ

ISSN 2052-2525

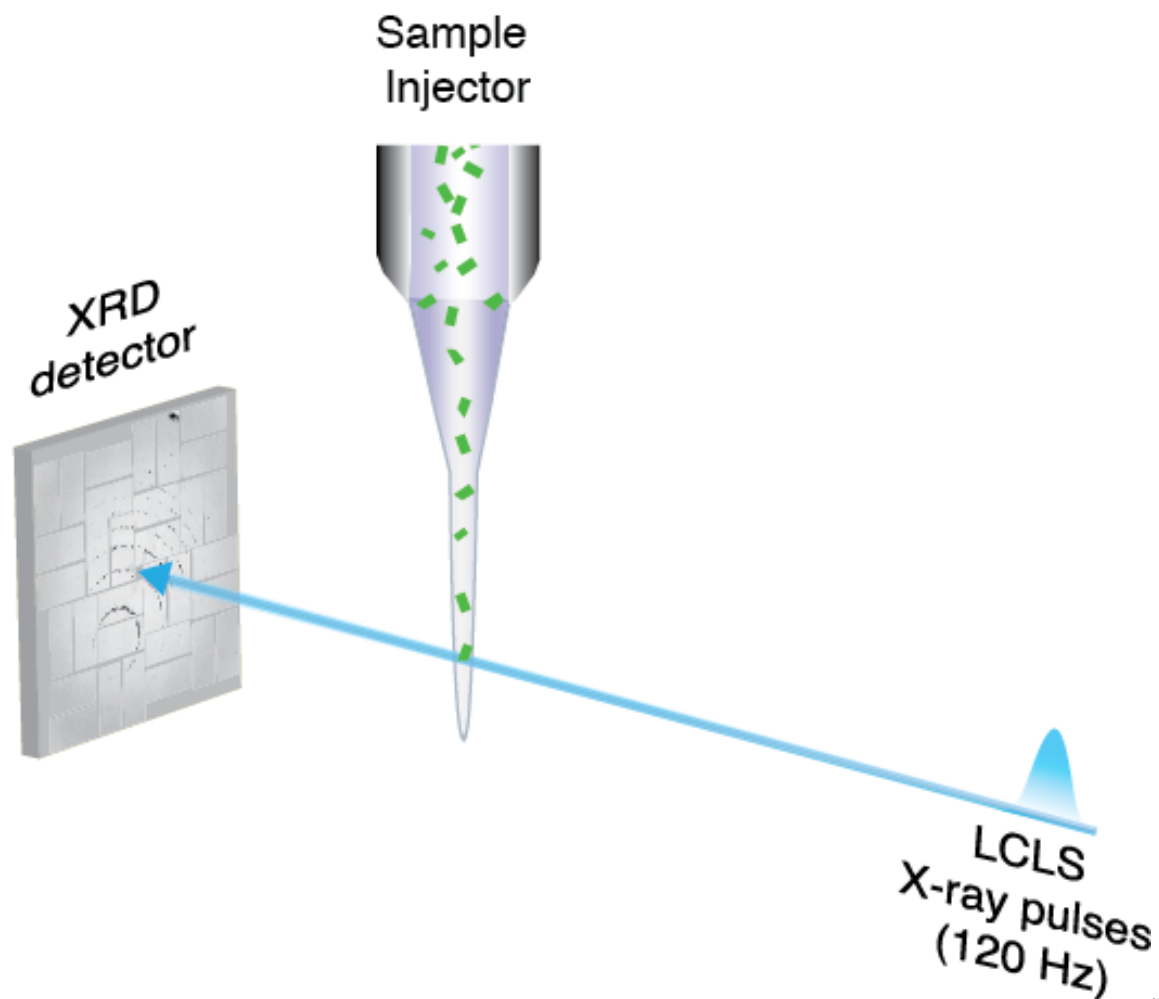
BIOLOGY | MEDICINE

Gold Standard for macromolecular crystallography diffraction data

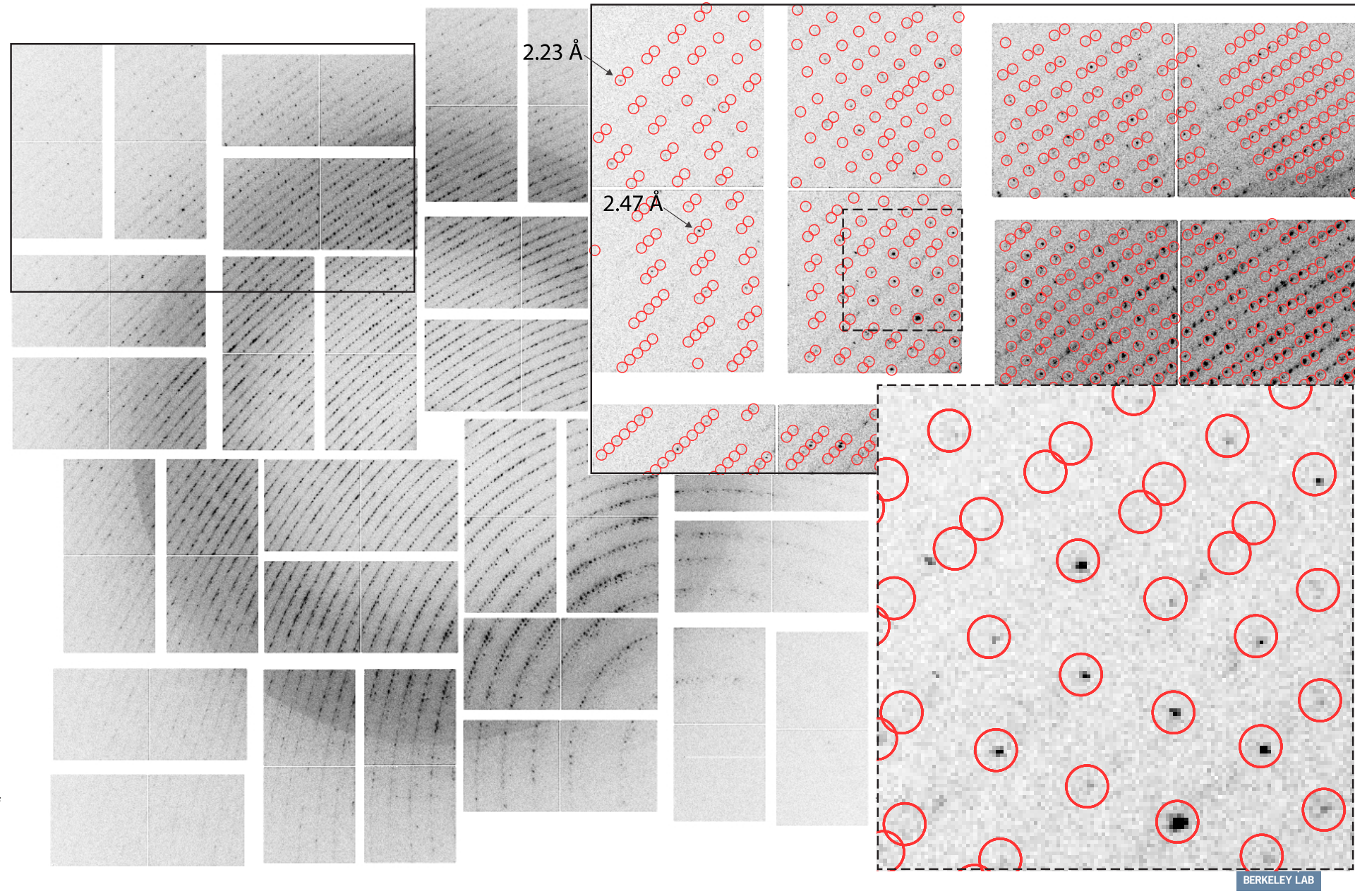
**Herbert J. Bernstein,^{a*} Andreas Förster,^b Asmit Bhowmick,^c Aaron S. Brewster,^c
Sandor Brockhauser,^{d,e,f} Luca Gelisio,^g David R. Hall,^h Filip Leonarski,ⁱ Valerio
Mariani,^g Gianluca Santoni,^j Clemens Vornrhein^k and Graeme Winter^h**

<https://doi.org/10.1107/S2052252520008672>

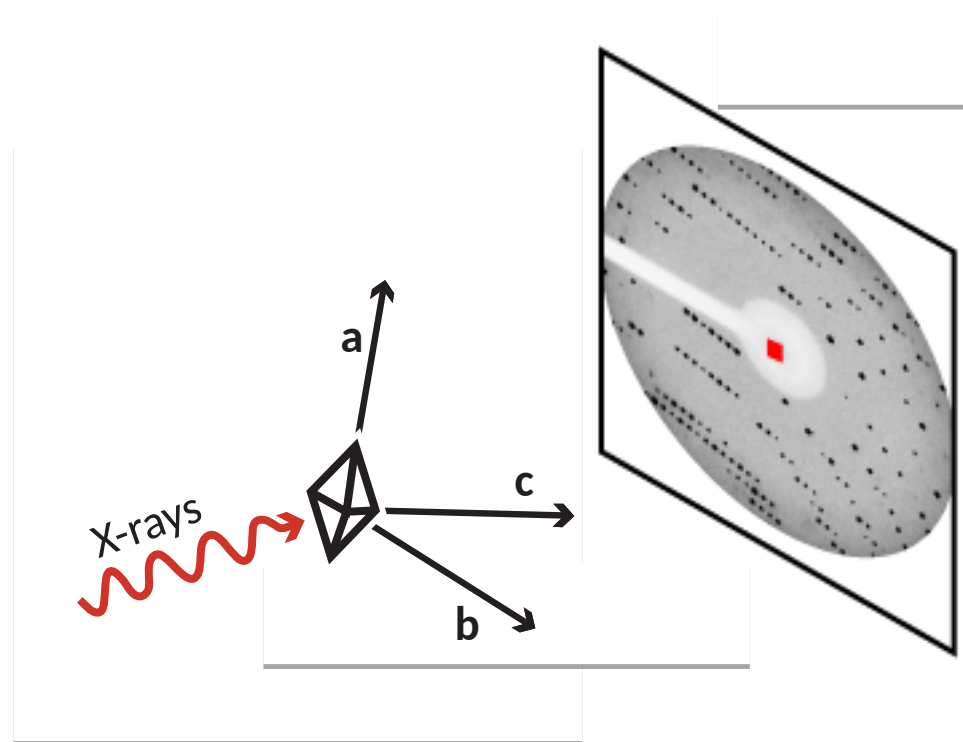
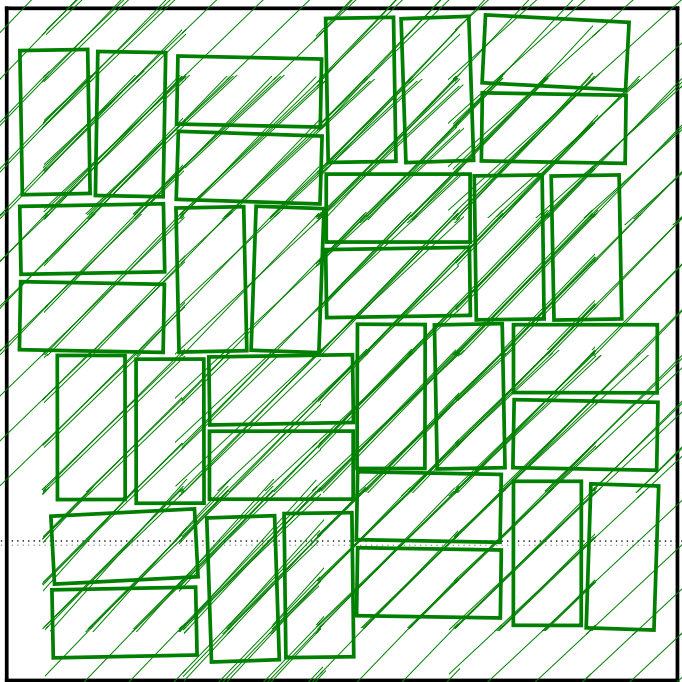
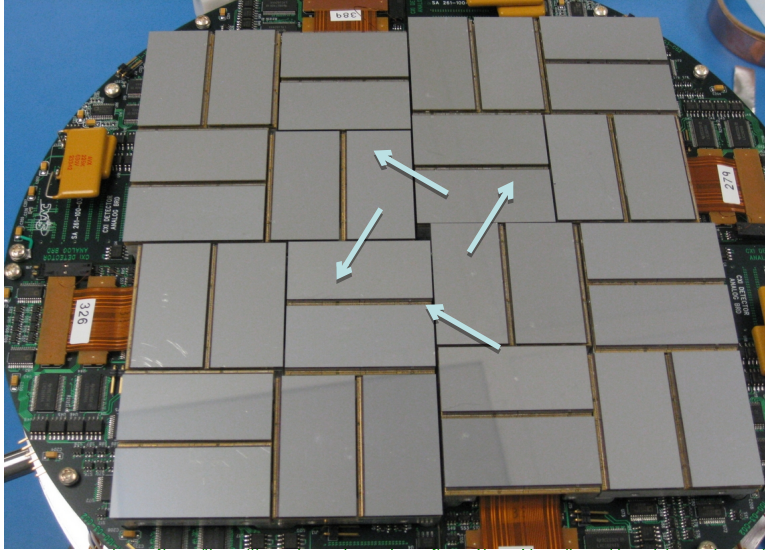
Serial crystallography



Photosystem II diffraction: Young et al (2016) Nature 540: 453

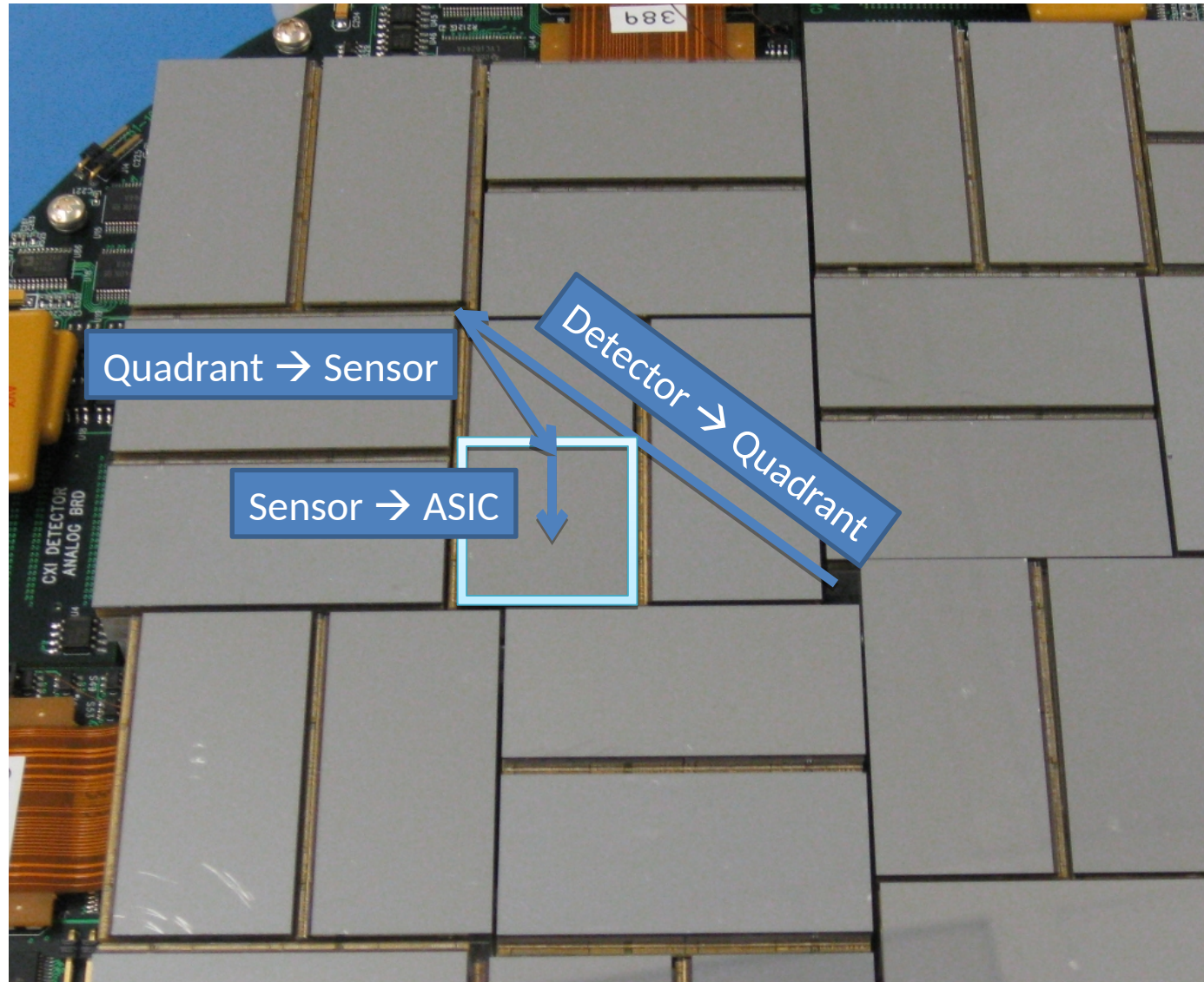


CSPAD: Cornell-SLAC Pixel Array Detector

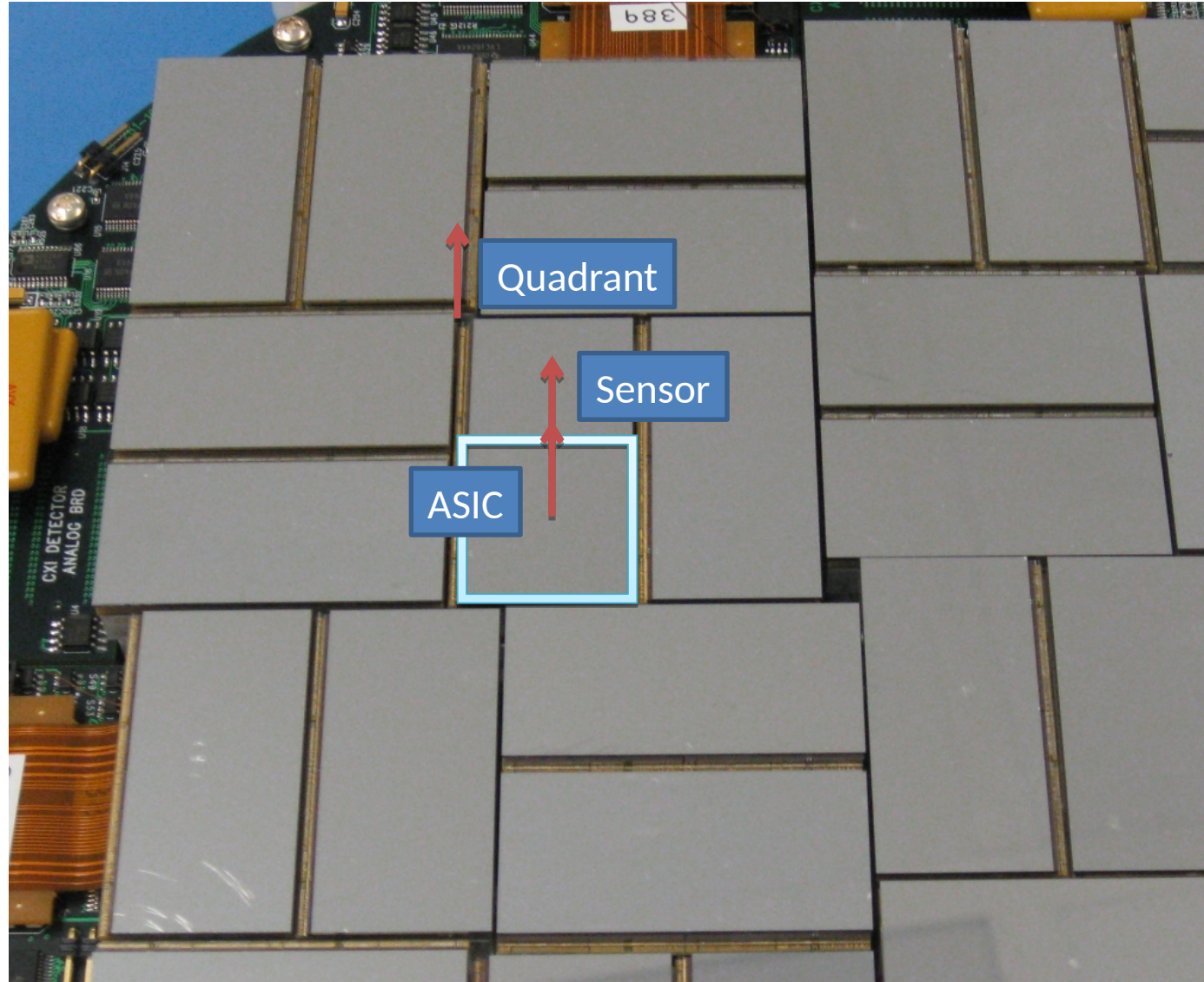


- Tiled construction of the CSPAD detector
- Similar to AGIPD, Octal8 MPCCD and JF16M

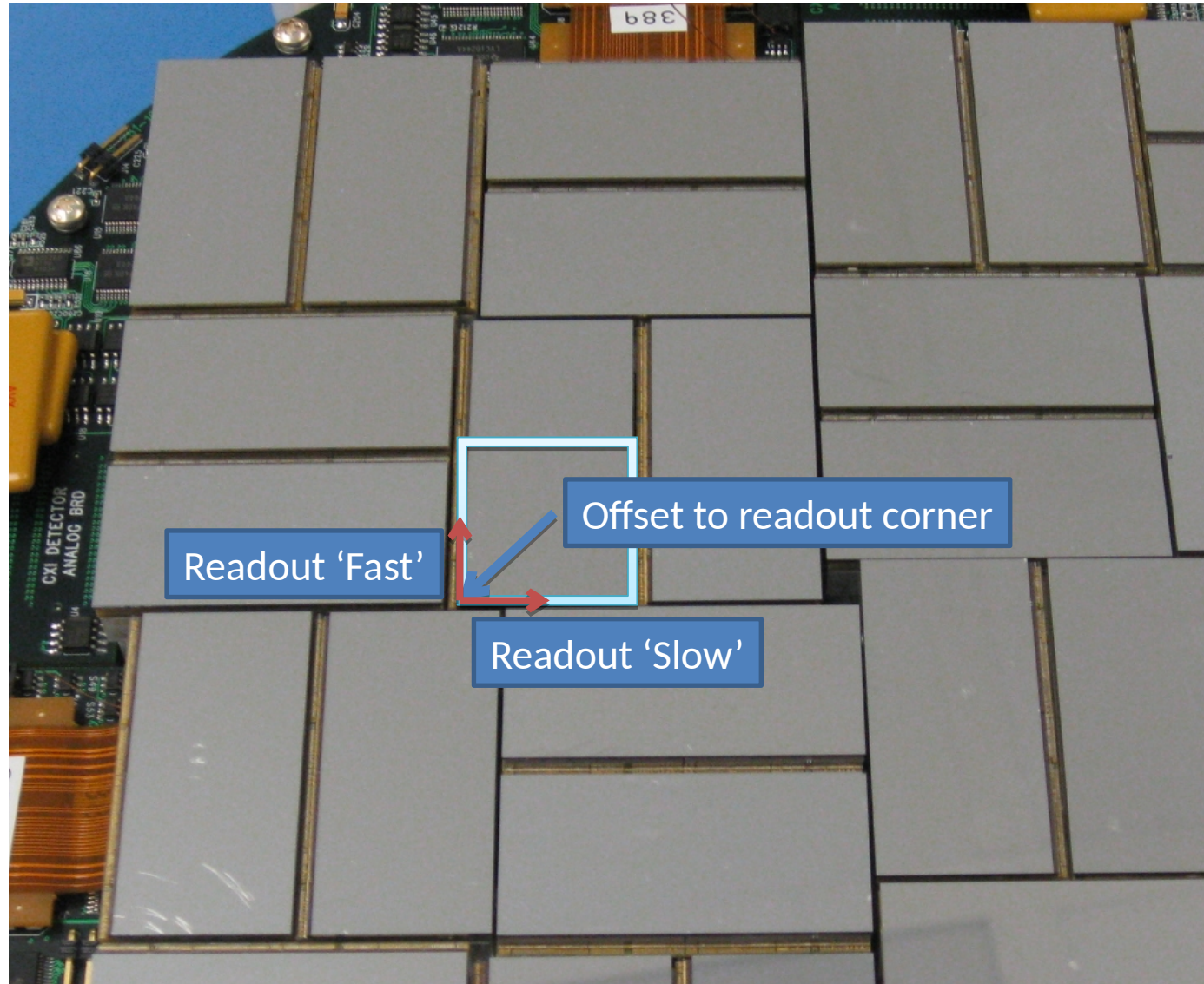
CBF Header: Translation Axes



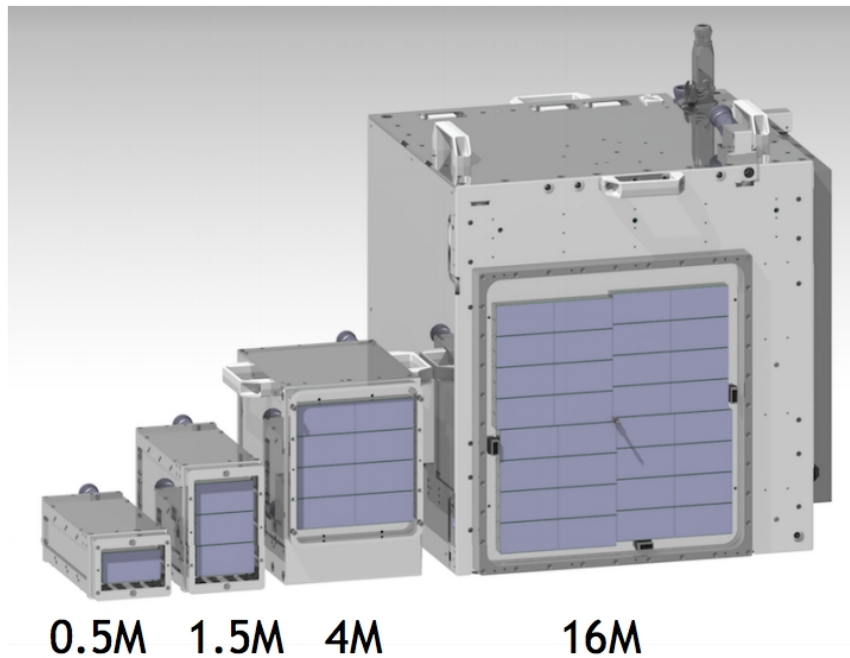
CBF Header: Rotation Axes

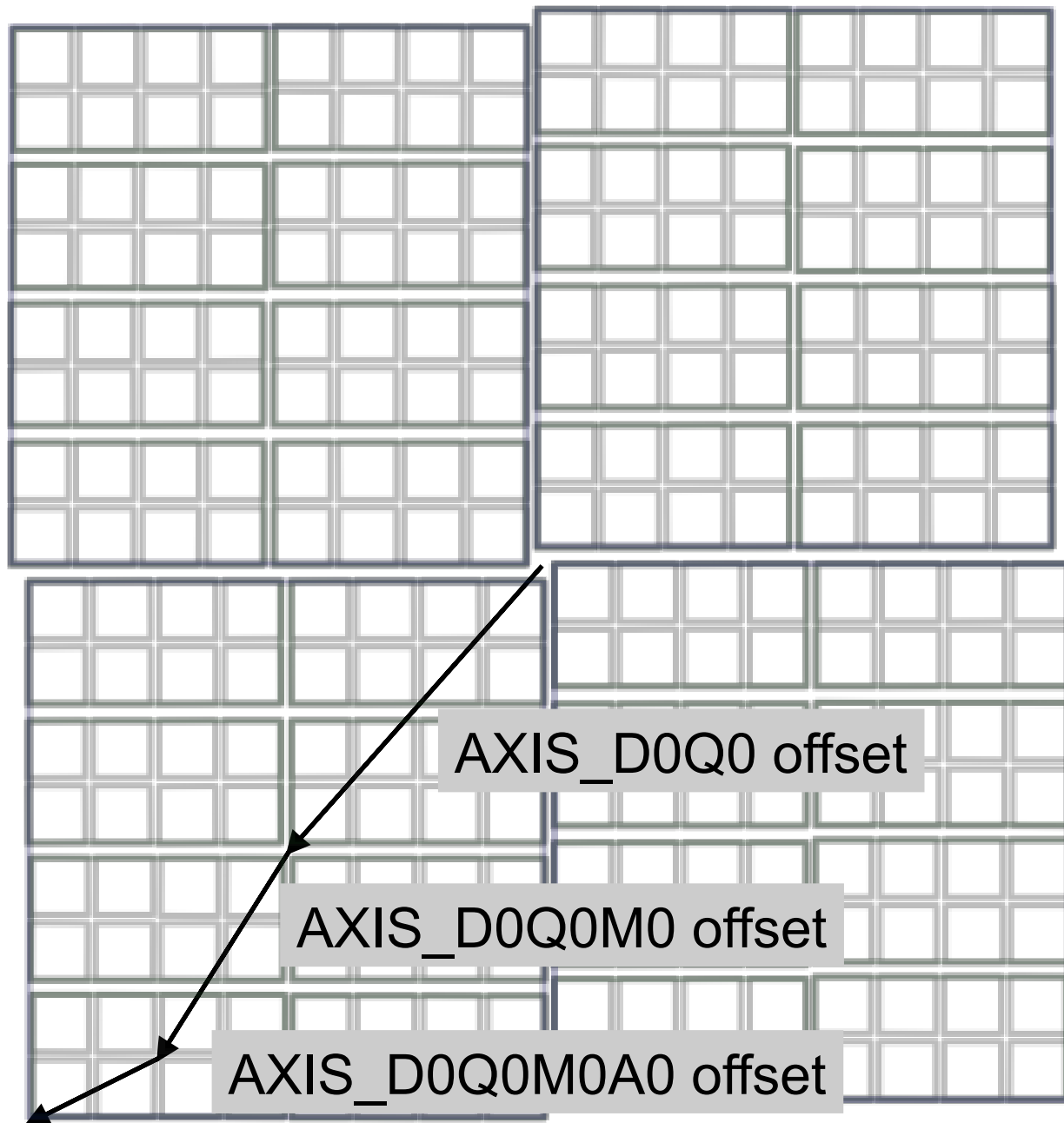


CBF Header: Data Axes



Jungfrau 16M at SwissFEL





D0Q0M0A0

```
fast_pixel_direction
@offset=0,0,0
@vector=-0.9999984140169291,-0.0017810007373656254,0.0
@depends_on = /entry/instrument/ELE_D0/transformations/AXIS_D0Q0M0A0

slow_pixel_direction
@offset=0,0,0
@vector=-0.0017810007373656254,0.9999984140169291,0.0
@depends_on = /entry/instrument/ELE_D0/transformations/AXIS_D0Q0M0A0
```


D0Q0M0A0: depends_on chain

AXIS_D0Q0M0A0

@depends_on = AXIS_D0Q0M0

@offset = 38.58423053508638,-19.131359799099098,0.0

@vector = 0.0,0.0,-1.0

AXIS_D0Q0M0

@offset = 38.310714303741115,-61.1137814011917,0.0

@vector = 0.0,0.0,-1.0

@depends_on = AXIS_D0Q0

AXIS_D0Q0

@offset = 75.1333334212039,-84.65967506456187,0.0

@vector = 0.0,0.0,-1.0

@depends_on = AXIS_D0

AXIS_D0

@offset = -0.4290327180241036,0.12102061135945519,-0.294

@vector = 0.0,0.0,-1.0

@depends_on = AXIS_RAIL

AXIS_RAIL

@vector = 0.0,0.0,1.0

@depends_on = .

Zenodo example data



The image is a screenshot of a Zenodo dataset page. At the top is a blue header bar containing the Zenodo logo on the left, a search bar with the placeholder text 'Search' and a magnifying glass icon, and the words 'Upload' and 'Communities' on the right. Below the header, the date 'July 26, 2019' is displayed on the left, and two buttons labeled 'Dataset' and 'Open Access' are on the right. The main title of the dataset is '68 image lysozyme dataset recorded on the Jungfrau 16M detector at SwissFEL and formatted as a NeXus file'. Below the title, the authors are listed as 'Brewster, Aaron; Wang, Meitian; Bernstein, Herbert J', each preceded by a small green circular icon with a white 'iD'. At the bottom of the snippet, a line of text states: 'Data provided by Meitian Wang at PSI and master file revised May 2020 for full NXmx compliance.'

zenodo Search Upload Communities

July 26, 2019 Dataset Open Access

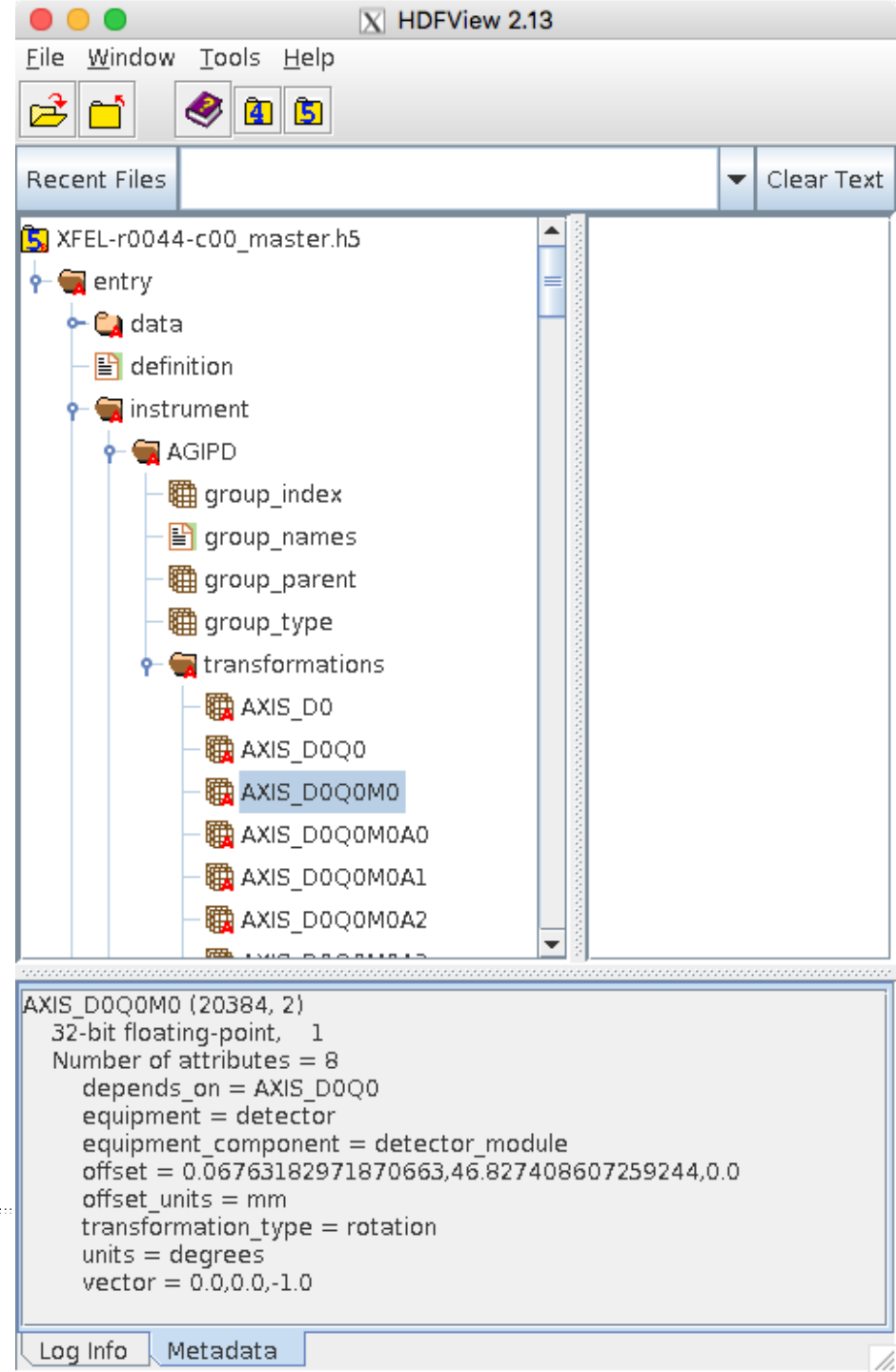
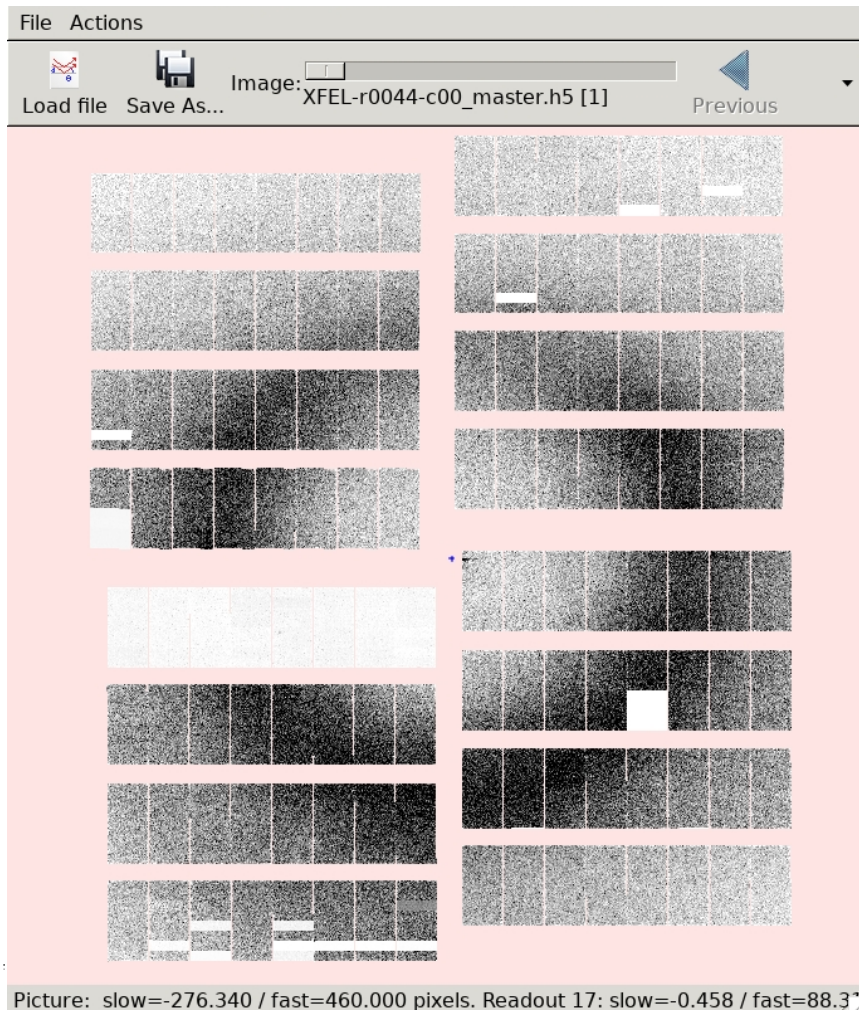
68 image lysozyme dataset recorded on the Jungfrau 16M detector at SwissFEL and formatted as a NeXus file

 Brewster, Aaron;  Wang, Meitian;  Bernstein, Herbert J

Data provided by Meitian Wang at PSI and master file revised May 2020 for full NXmx compliance.

<https://doi.org/10.5281/zenodo.3352357>

NeXus at the EuXFEL



***cctbx.xfel* scripts**

- Two scripts for creating NeXus header files with pointers to calibrated data
- https://github.com/cctbx/cctbx_project/blob/master/xfel/swissfel/jf16m_cxigeom2nexus.py
- https://github.com/cctbx/cctbx_project/blob/master/xfel/euxfel/agipd_cxigeom2nexus.py

What's missing in NeXus and NXmx?

- Event synchronization
 - 16 module detector with 16 data stacks from 16 recorders
 - Not guaranteed to have all frames in all stacks
 - Synchronize with timestamps → expensive lookups
 - Use virtual datasets → performant?

What's missing in NeXus and NXmx?

- Raw data calibration
 - JF16M and AGIPD have 3 gains, each with a pedestal
 - 16 bit raw data: GGXXXXXXXXXXXXXXXX
 - 2 gain bits
 - 00 Low gain/pedestal
 - 01 Med gain
 - 10 Invalid pixel
 - 11 High gain/pedestal
- Current pipeline:
 - Separate jobs apply gain/pedestal to each pixel, write out calibrated data
 - Downside: extra processing resources and data is copied and promoted to float
- Goal: save raw data in NeXus
 - Needs way to express formulas in standard way
 - More discussion: <https://github.com/nexusformat/definitions/issues/711>

Acknowledgements

Berkeley National Lab PSI

Nicholas Sauter
Asmit Bhowmick
Derek Mendez
Dan Paley
Elyse Schriber
Paul Adams
Peter Zwart
Billy Poon
Johannes Blaschke
Vittal Yachandra
Junko Yano
Jan Kern
James Holton

Dmitry Ozerov
Karol Nass
Meitian Wang
Filip Leonarski
John Beale

EuXFEL

Fabio Dall'Antonia
Sandor Brockhauser
Yury Kirienko
Hans Fangohr

Dectris

Andreas Förster

LLNL

Michael Wall

Ronin Institute

Herbert Bernstein

Nebraska

Limei Zhang

Alums

Nat Echols
Tara Michels-Clark
Lee O'Riordan
Muhammed Amin
Oliver Zeldin
Robert Bolotovskiy
Sheraz Gul

LCLS

Uwe Bergmann
Alberto Lutman
Franklin Fuller
Uwe Bergmann
Roberto Alonso-Mori
Sebastien Boutet
Mengning Liang
Amedeo Perazzo
Chuck Yoon
Chris O'Grady
Mark Hunter
Jason Koglin
Ray Sierra
Mona Uervirojnangkoorn

UCSF

James Fraser
Iris Young
Michael Thompson
Alex Wolff
Rahel Woldeyes

CCP4

David Waterman

Diamond Light Source

Gwyndaf Evans
Graeme Winter
David Stuart
Jonathan Grimes
James Parkhurst
Luis Fuentes-Montero
Allen Orville
Pierre Aller
Agata Butryn
Danny Axford

UCLA

David Eisenberg
Duilio Cascio
Michael Sawaya
Jose Rodriguez
Luki Goldschmidt
Johan Hattne

Charité-Universitätsmedizin Berlin

Andrea Schmidt
Michał Szczepek
Patrick Scheerer

IBS

Jacques-Philippe Colletier

Stanford University

Axel Brunger
Bill Weis
Soichi Wakatsuki
Henry van den Bedem

SSRL

Mike Soltis
Ana Gonzalez
Ashley Deacon
Aina Cohen
Artem Lyubimov
Scott McPhillips

NIH/NIGMS grants 1R01GM117126
DOE/Office of Science contract DE-AC02-05CH11231

