UPPSALA UNIVERSITET

X-ray free-electron lasers and the CXIDB

ExPaNDS PaNOSC

Community Symposium
2020-11-09

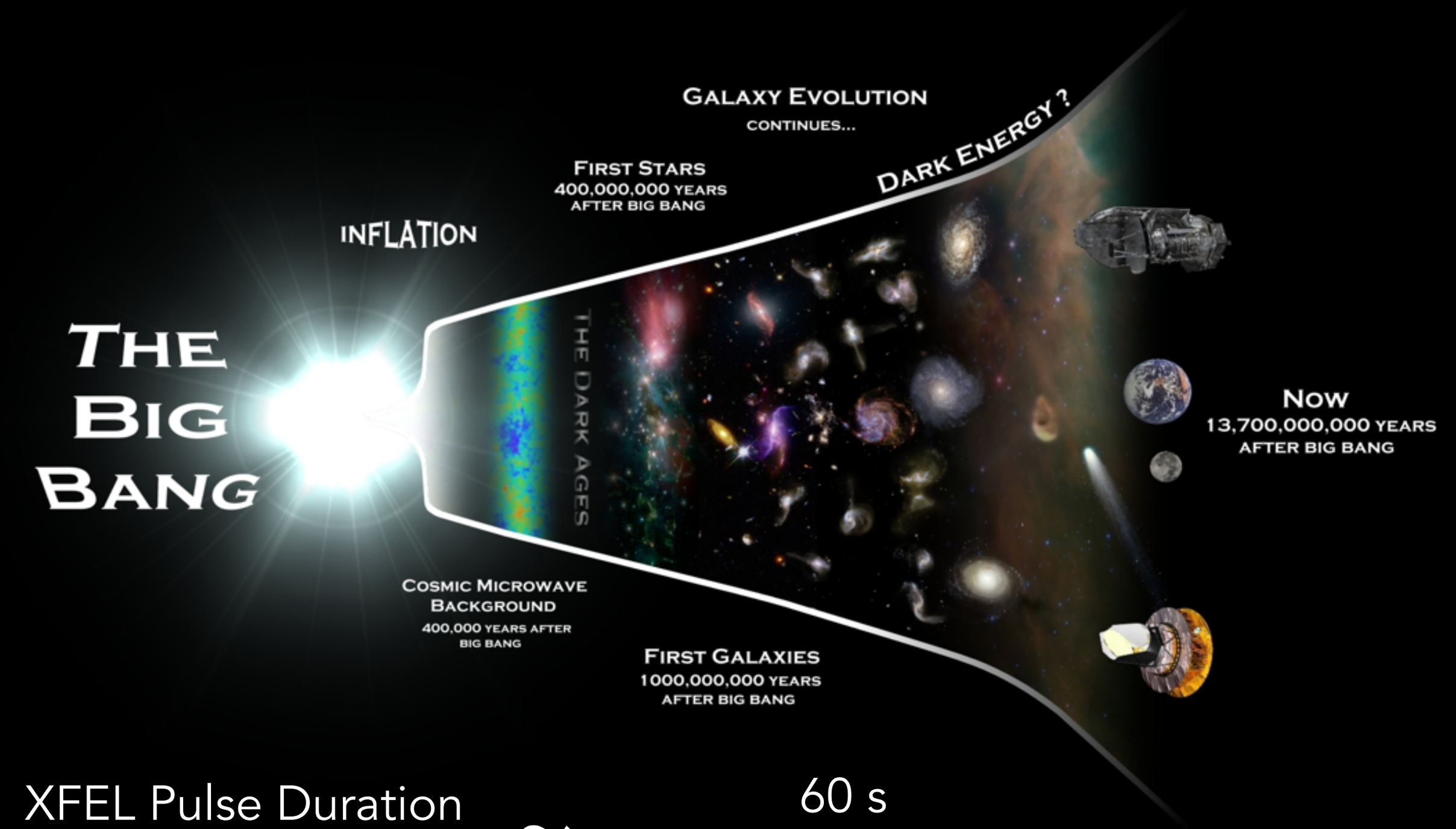
Filipe R N C Maia

Lab. of Mol. Biophysics
Uppsala University

X-ray Free-electron Laser: a light-source of superlatives

- Longest
- Most expensive
- Fastest
- Most powerful

The Speed of XFELs



$$\frac{\text{XFEL Pulse Duration}}{60\ \text{s}} \approx \frac{60\ \text{s}}{\text{Age of the Universe}}$$

The Power of XFELs

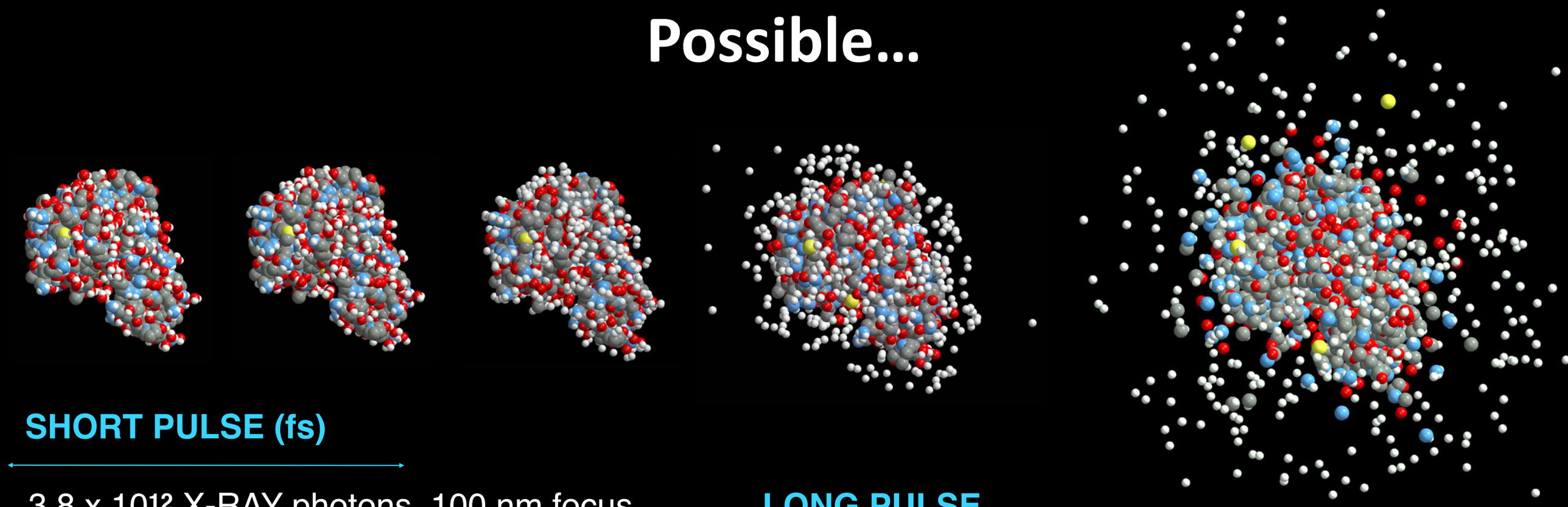
Total irradiated power
= 1.740 × 10^{17} W

10 Billion Fold Increase In Peak Brilliance

1 mm^2

The peak power of the LCLS with a 1 μm^2 focus

This Makes New Experiments Possible…



SHORT PULSE (fs)

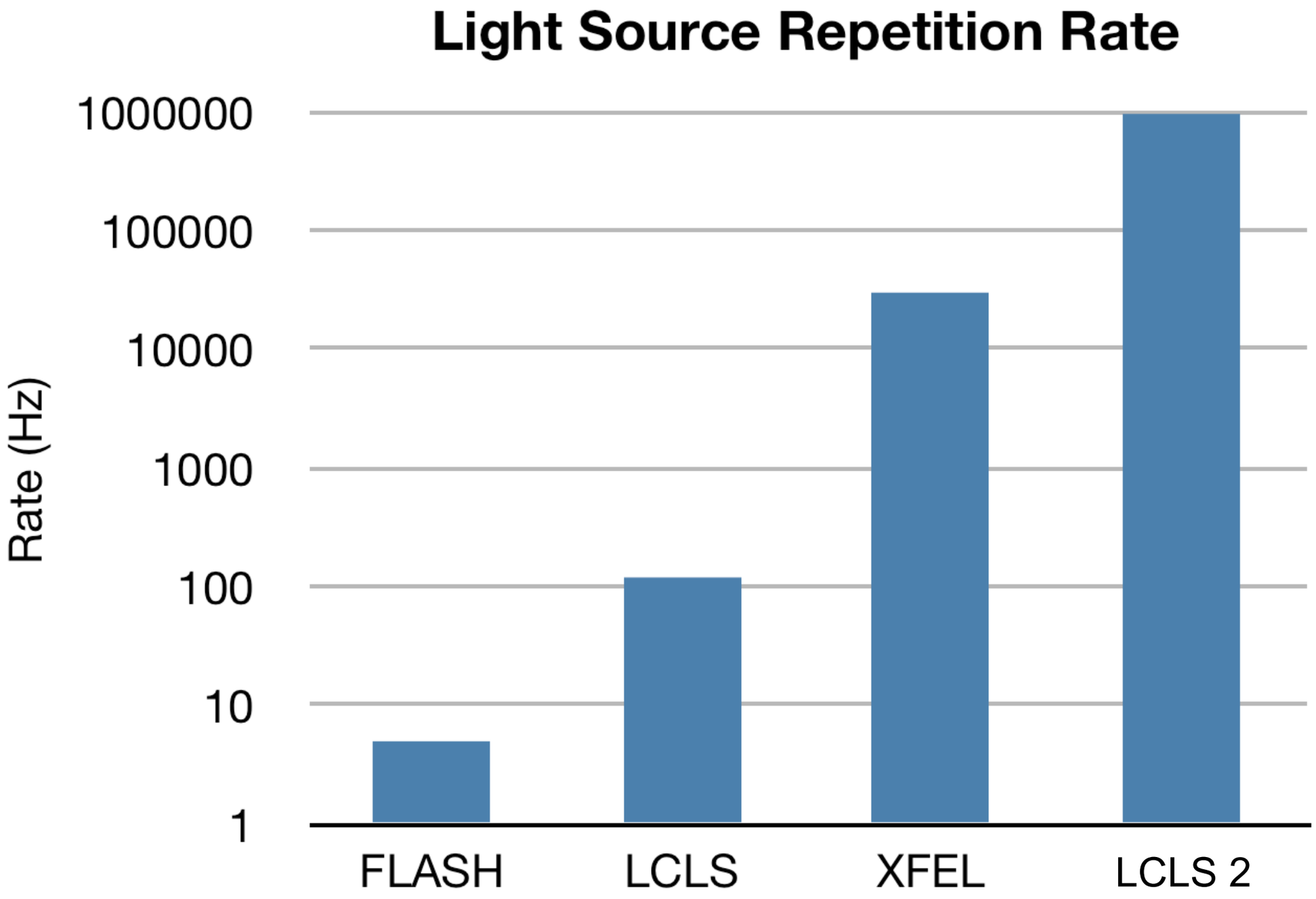
3.8 x 10^{12} X-RAY photons, 100 nm focus

LONG PULSE

SPEED OF LIGHT *vs.*
SPEED OF A SHOCK WAVE

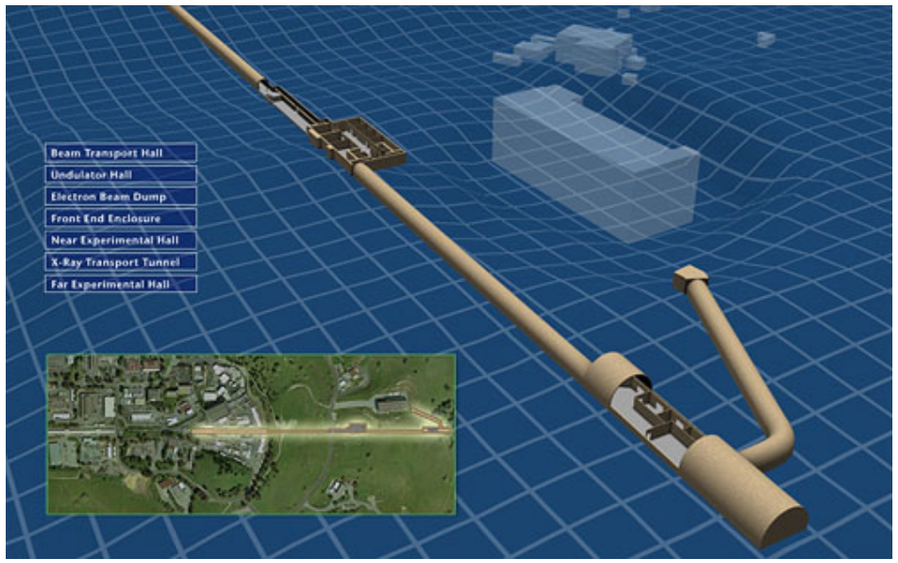
Neutze, R., Wouts, R., van der Spoel, D., Weckert, E., Hajdu, J., *Nature* 406, 752-757, (2000).

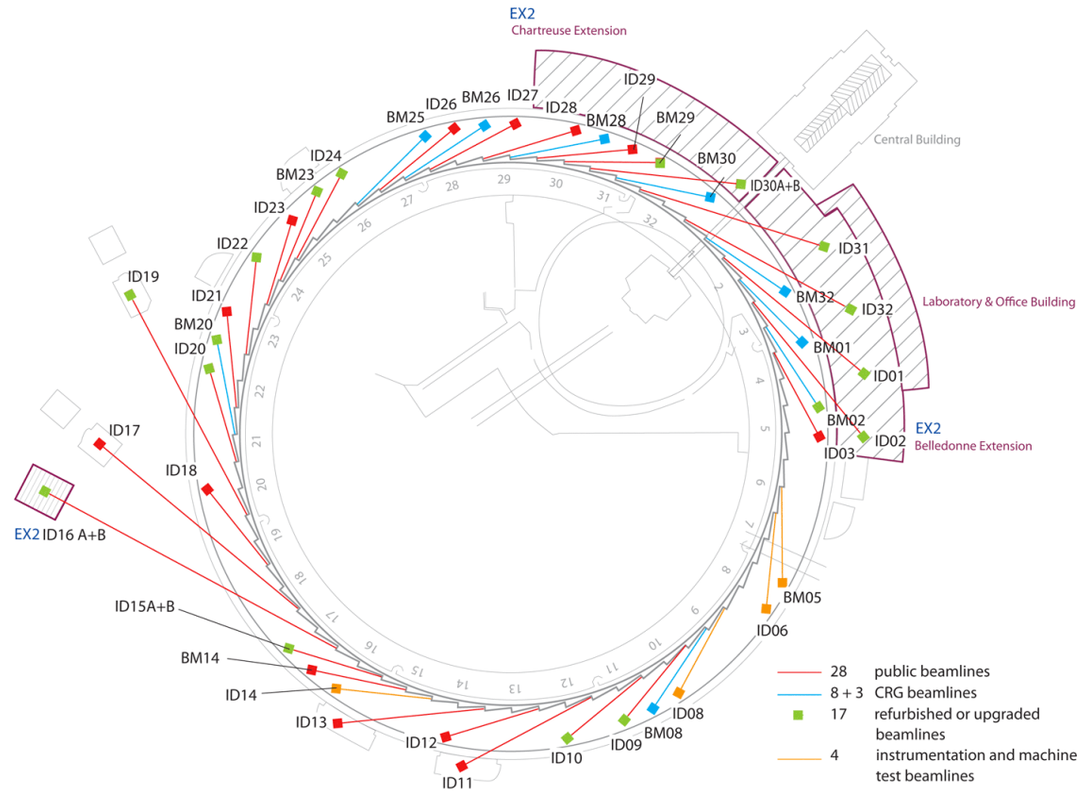
The repetition rate of XFELs is still growing exponentially



Maximize Efficiency of the Facilities

- XFEL produce mountains of data
- They are serial, as opposed to synchrotron's parallel nature.
- The result is a few groups with too much data.
- While others are data starved.
- Sharing is a clear solution.





The birth of the Coherent X-ray Imaging DataBank (cxidb.org)

Goals

- Foster reproducible research
- Enable the test of ideas on real data
- Preserve datasets for future analysis.

View from my office at Lawrence Berkeley Lab while thinking how to create CXIDB, Jan. 2011.

The birth of the Coherent X-ray Imaging DataBank (cxidb.org)

Practical challenges

- Where to get funds?
- Where to get storage?
- How to create the infrastructure?

Just do it, as simple as possible!

View from my office at Lawrence Berkeley Lab while thinking how to create CXIDB, Jan. 2011.

Challenges to Data Sharing

- Sociological challenges
- Lack of adequate rewards.
- Not viewed as a valuable scholarly endeavour
- Afraid to be scooped
- Certain disciplines manage to overcome this: astronomy, oceanography, genomic.

Challenges to Data Sharing

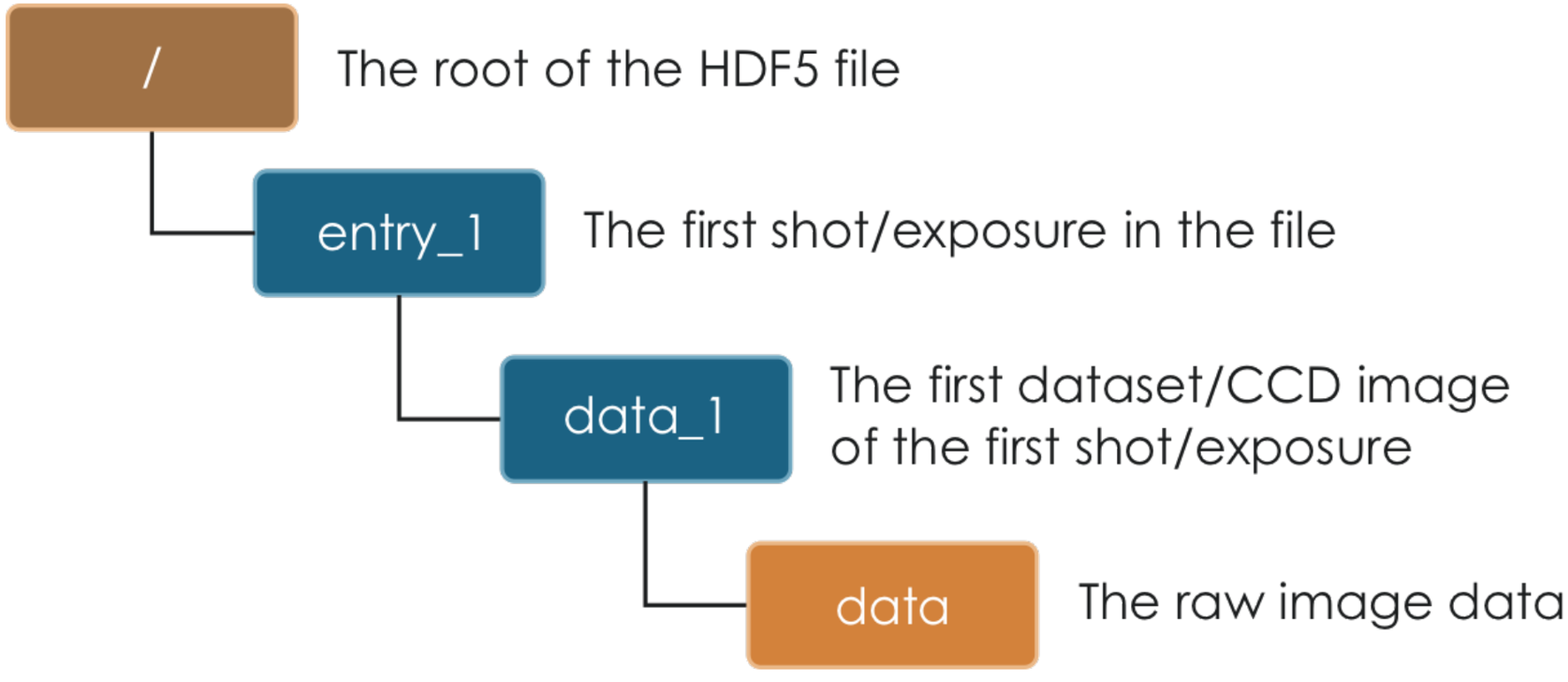
- Data sets should be publishable and citable.
- Data sharing should be part of funding agreements.
- Funding agencies are requiring data management plan in proposals.
- Publications should reward data sharing.

Common File Format

- Data sharing is crucial in this age of large collaborations.
- A common file format is necessary for easy data sharing.
- CXIDB needs a uniform file format
- No existing file format meets our needs
- Creation of the CXI file format.

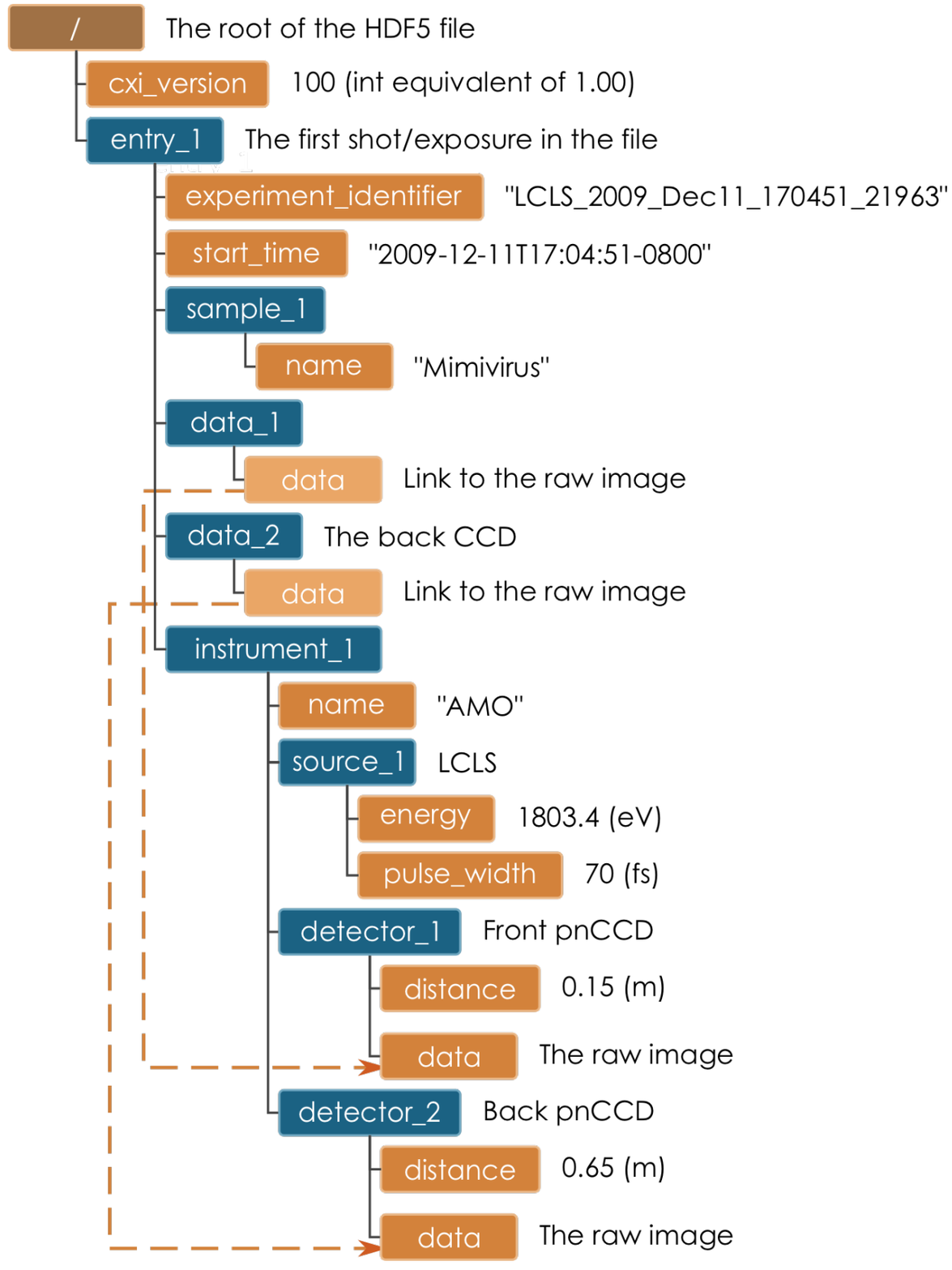
CXI File Format

- HDF5 based.
- Simple.
- High performance



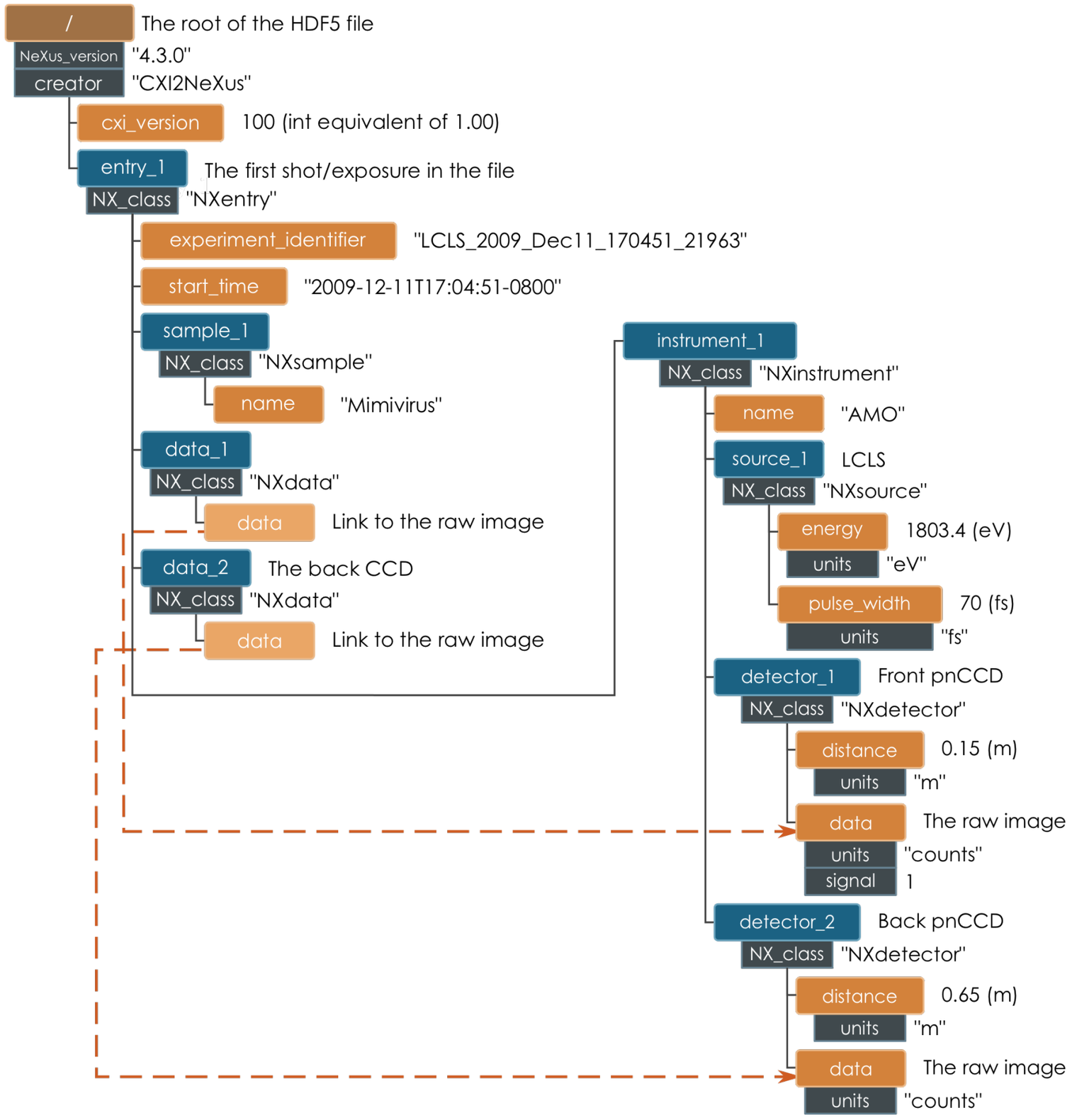
CXI File Format

- Flexible
- Well documented
- Extensible
- Compatible with NeXus



CXI File Format

NeXus Style



Reproducible Data Analysis

CXIDB ID 56

Deposition Summary	
Depositor:	Benedikt Daurer
Contact:	ben...@xray.bmc.uu.se
Deposition date:	2016-12-09
Last modified:	2017-04-10
DOI:	10.11577/1349716

Publication Details	
Title:	Experimental strategies for imaging bioparticles with femtosecond hard X-ray pulses
Authors:	Benedikt J. Daurer, Kenta Okamoto et al.
Journal:	IUCrJ
Year:	2017
DOI:	10.1107/S2052252517003591

Experimental Conditions	
Method:	Single Particle X-ray Diffraction Imaging
Sample:	Omono River Virus
Wavelength:	2.25 Å (5.5 keV)
Lightsource:	LCLS
Beamline:	CXI

Data Files	
Diffraction Patterns:	cxidb_56_hits.tar.gz
Background:	cxidb_56_background.tar.gz

Auxiliary Files	
Metadata:	cxidb_56_metadata.tar.gz
Analysis Description:	github.com/FXIhub/cxic9714-analysis

Description

Facilitating the very short and intense pulses from an X-ray laser for the purpose of imaging small bioparticles carries the potential for structure determination at atomic resolution without the need for crystallization. In this study, we explore experimental strategies for this idea based on data collected at the Linac Coherent Light Source from 40 nm virus particles injected into a hard X-ray beam.

Reproducible Data Analysis

README.md

This repository provides a description of the data analysis tools used for a Flash X-ray Imaging (FXI) experiment which was performed at the Linac Coherent Light Source (LCLS) and is described in

> Daurer B.J., Okamoto K., *et al.* Experimental strategies for imaging bioparticles with femtosecond hard X-ray pulses. *IUCrJ* **4**, 3 (2017). https://doi.org/10.1107/S2052252517003591.

The data has been deposited in the Coherent X-ray Imaging Data Base (CXIDB) with ID **56** and can be downloaded from here: http://cxidb.org/id-56.html

List of available files:

File name	Name	Description
http://cxidb.org/data/56/cxidb_56_hits.tar.gz	**HITS**	Diffraction hits saved as CXI files.
http://cxidb.org/data/56/cxidb_56_background.tar.gz	**BKGR**	Diffraction background saved as CXI files.
http://cxidb.org/data/56/cxidb_56_metadata.tar.gz	**META**	Auxiliary files.

Inspecting CXI files

The easiest way to inspect CXI files is to use the viewing tool Owl (http://github.com/FXIhub/owl), but any inspection tool for HDF5 files can be used.

Requirements

In order to be able to run all the provided scripts and jupyter notebooks, the following has to be installed:

- python 2.7
- numpy

Current Status

CXIDB
Coherent X-ray Imaging Data Bank

- About 170 entries in the data bank.
- 650 TB of data, growing ~50% a year.
- Data from around the world (LCLS,FLASH, European XFEL, SACLA, SwissFEL, FERMI, PAL-XFEL, etc…)
- Storage and connectivity provided by NERSC/LBNL
- Recommended repository of Nature's *Scientific Data,* Gates Open Research, PLOS, etc…

SCIENTIFIC DATA
OFFICIAL RECOMMENDED REPOSITORY

Gates Open Research

PLOS

eLife

Future Challenges

CXIDB
Coherent X-ray Imaging Data Bank

- Storage and network resources
- Universal data standardisation
- Tighter integration with data stored at XFELs
- Easier data access, e.g. with Globus Online
- Online inspection and basic analysis tools