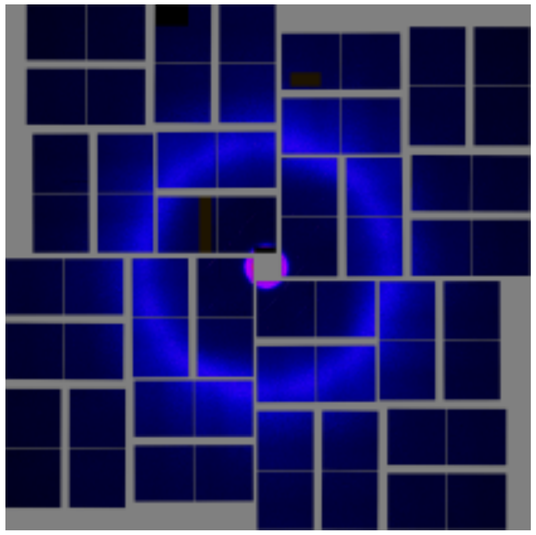
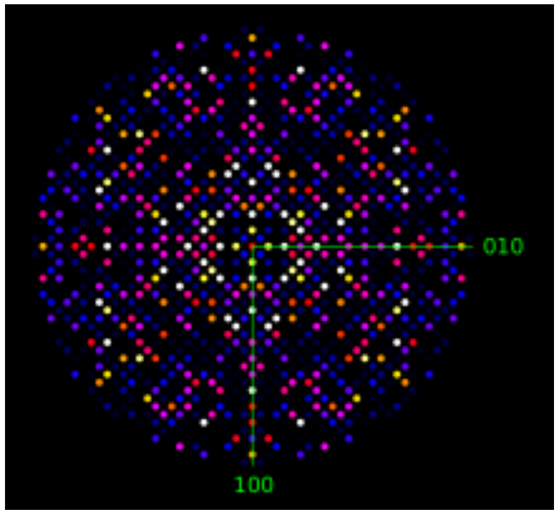
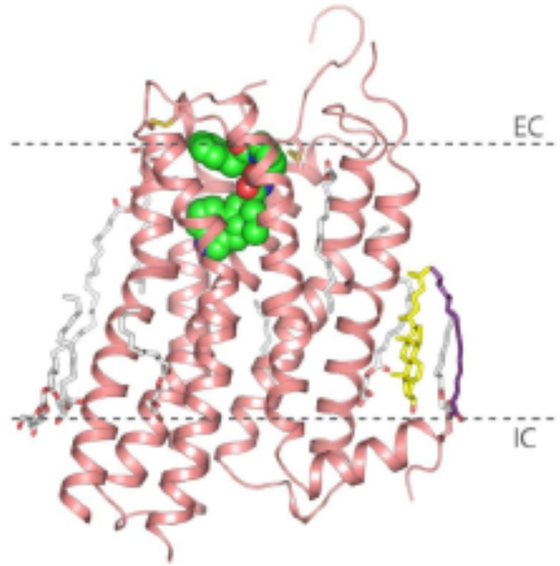
Free-electron laser data.

Problems and challenges







Thomas A. White
IUCr Crystallographic Computing School
Bengaluru, India
19. August 2017

HELMHOLTZ ASSOCIATION

CFEL SCIENCE

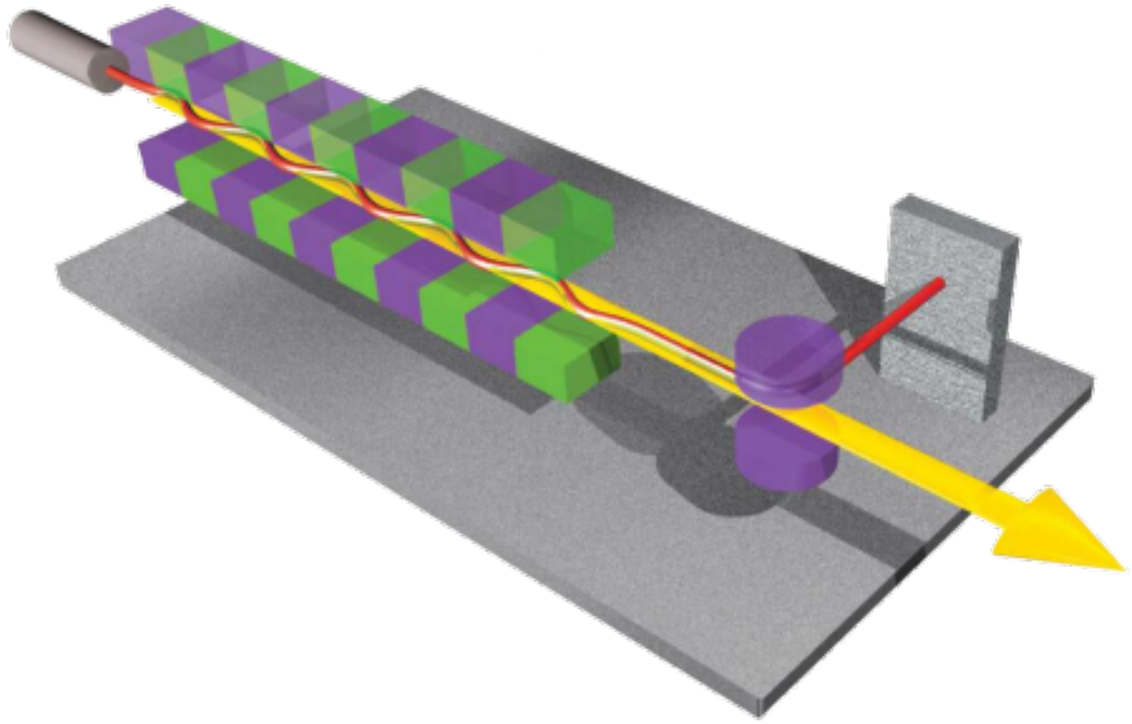
DESY

XFELs

- Free-Electron Laser (FEL) - source of very short, very intense, X-ray pulses.

Currently in operation:

- Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory, USA.
- Spring-8 Angstrom Compact Laser (SACLA), Harima, Japan
- PAL-XFEL, Pohang, Korea
- European XFEL, Hamburg, Germany (only just)
- SwissFEL, Villigen, Switzerland (very soon)



Images: © SLAC National Accelerator Laboratory, European XFEL

Serial femtosecond crystallography

- Study the **structure and function of proteins** ...
- ... using a **hard X-ray free-electron laser** ...
- ... by using **many small crystals** of the protein.

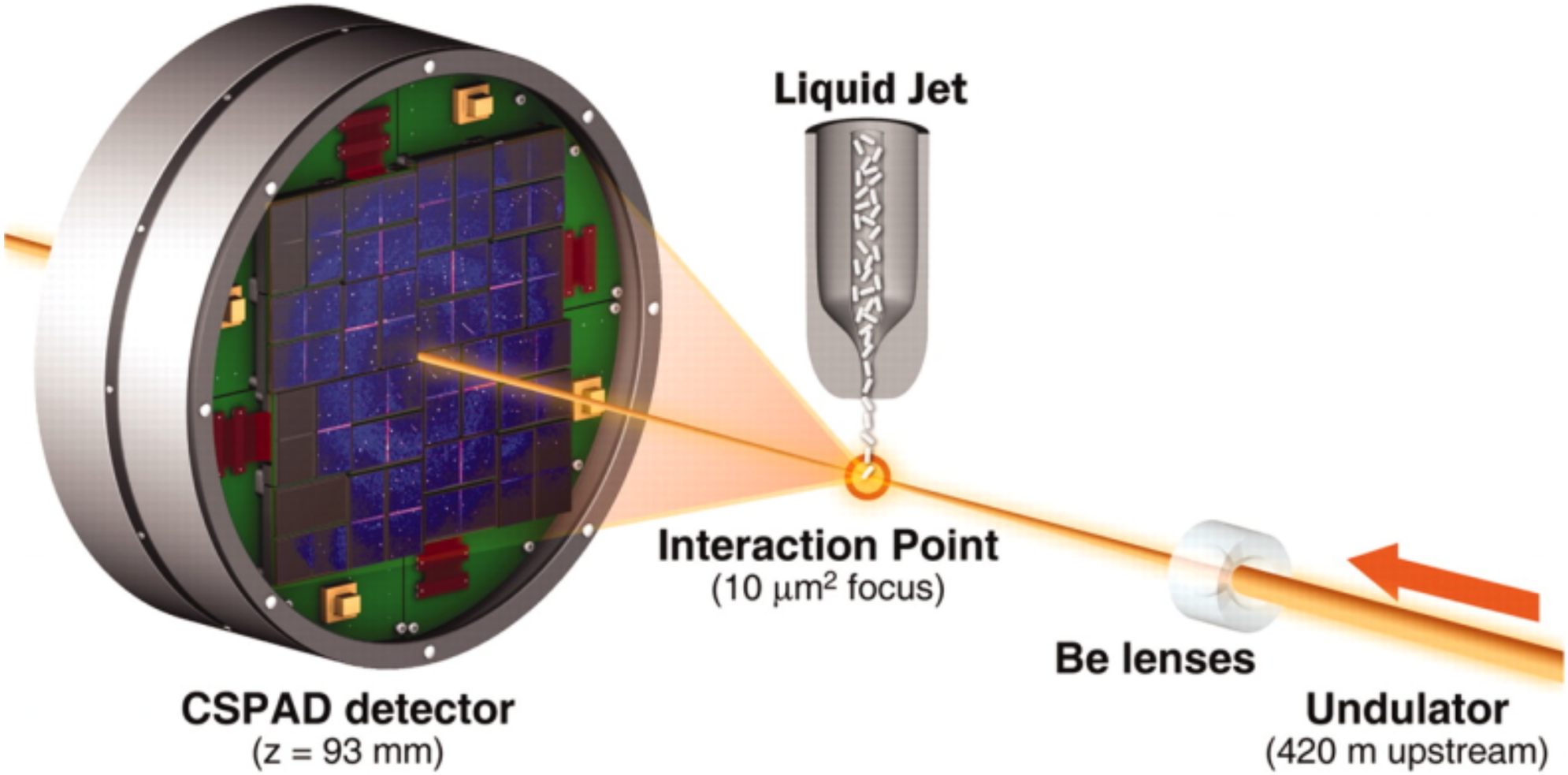
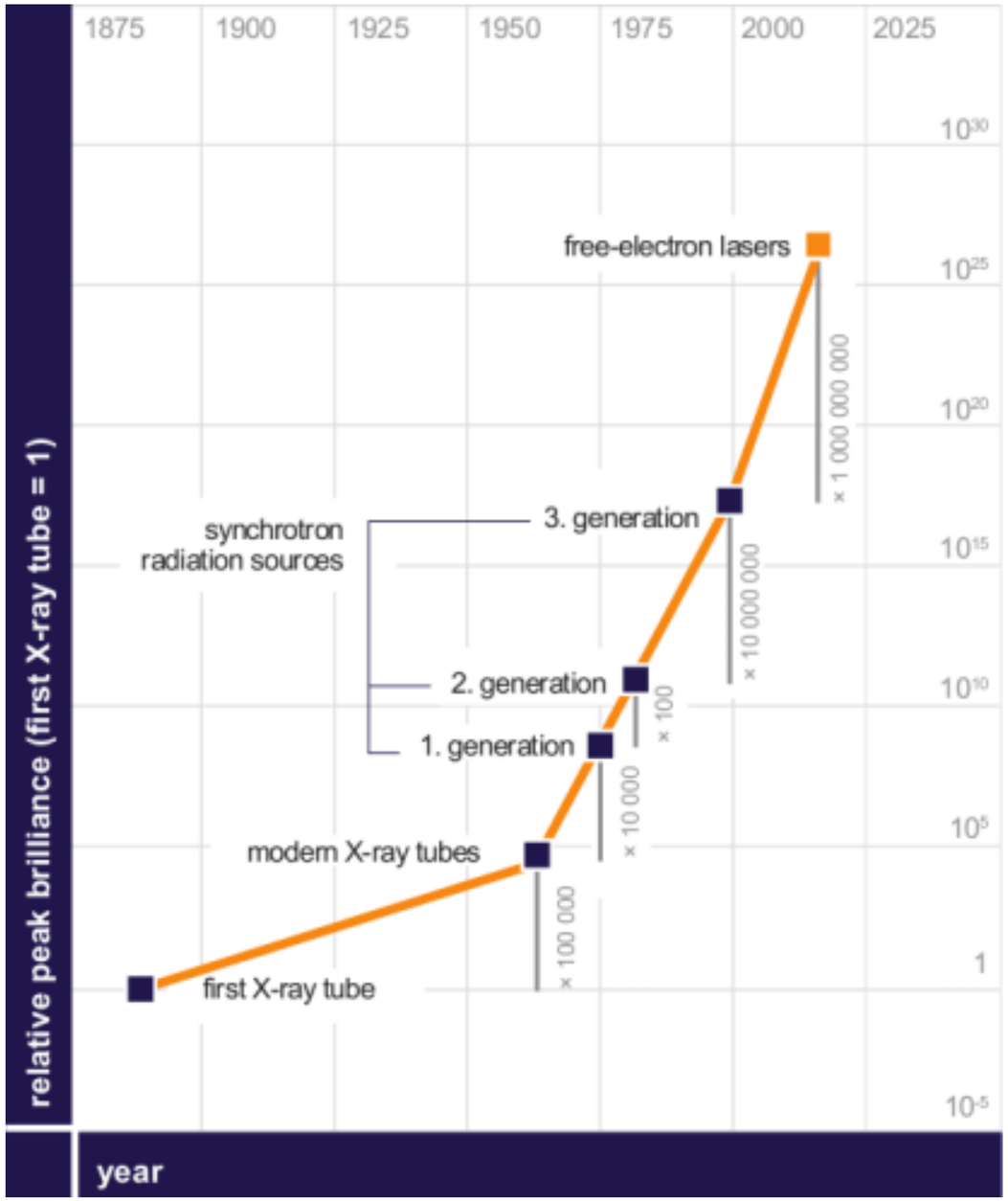
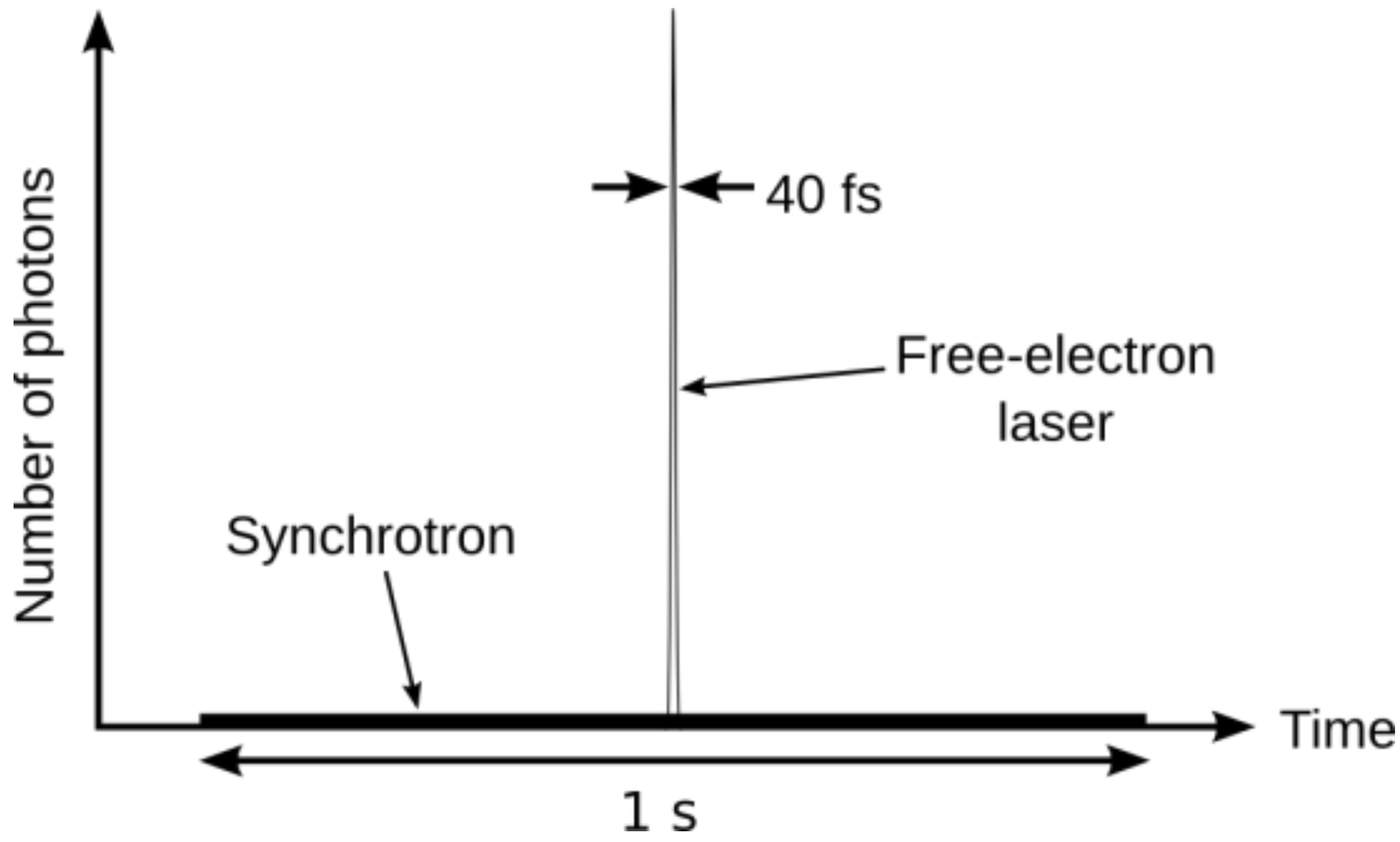


Image: Boutet et al., Science 337 (2012) p362

Why bother?



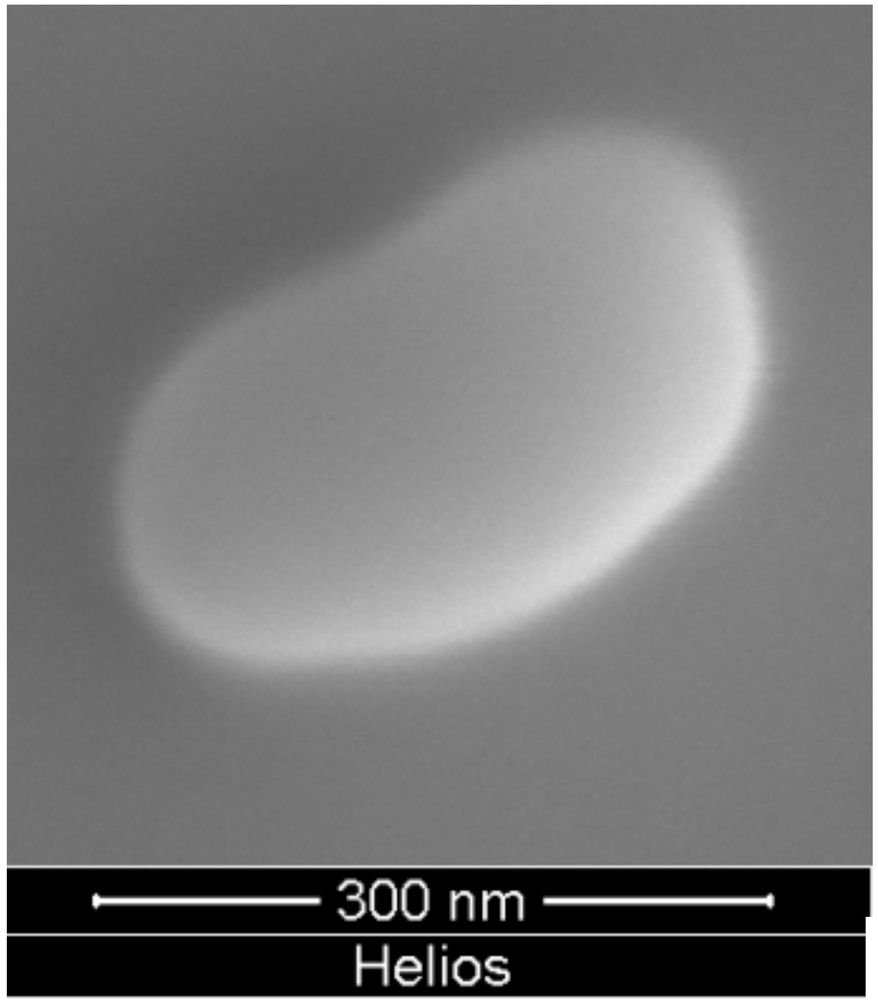
> Don't say: "The high intensity allows diffraction to be recorded from small crystals"



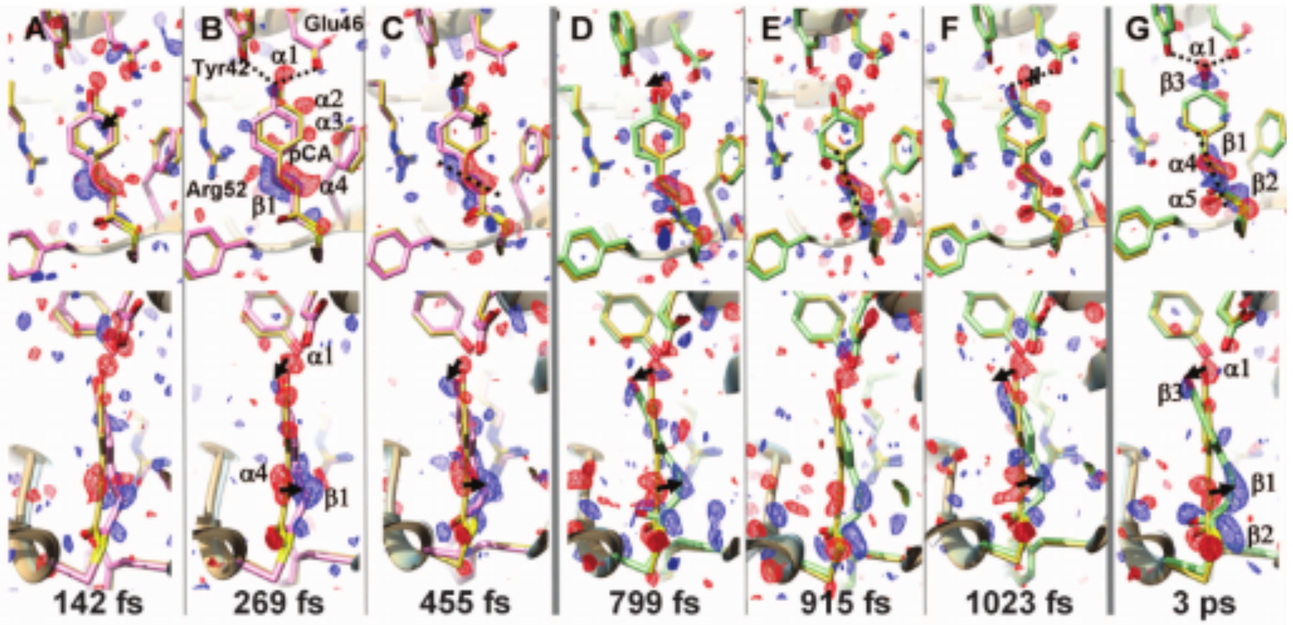
> Do say: "The short pulse length allows radiation damage to be circumvented"

Image left: European XFEL

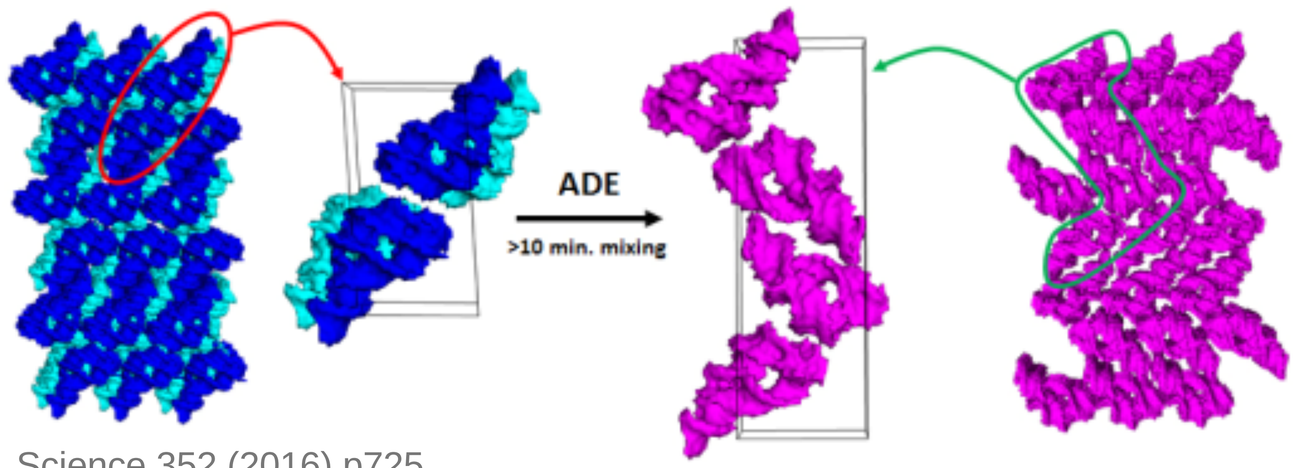
Why bother?



> Small crystals
⟶ Need high intensity
(otherwise no signal)
⟶ Radiation damage



> Time-resolved studies of dynamics, especially on short timescales, or if crystals are destroyed by reaction



Images: Peter Metcalf; Stagno et al., Nature (2016); Pande et al., Science 352 (2016) p725

Challenges of XFEL data

Structural biology challenges

Every crystal is different
Crystals may stick together
Indexing ambiguities

Technical challenges

Data storage - large amount of data
File formats - fast detectors, panels come separately?
Multiprocessing for high throughput

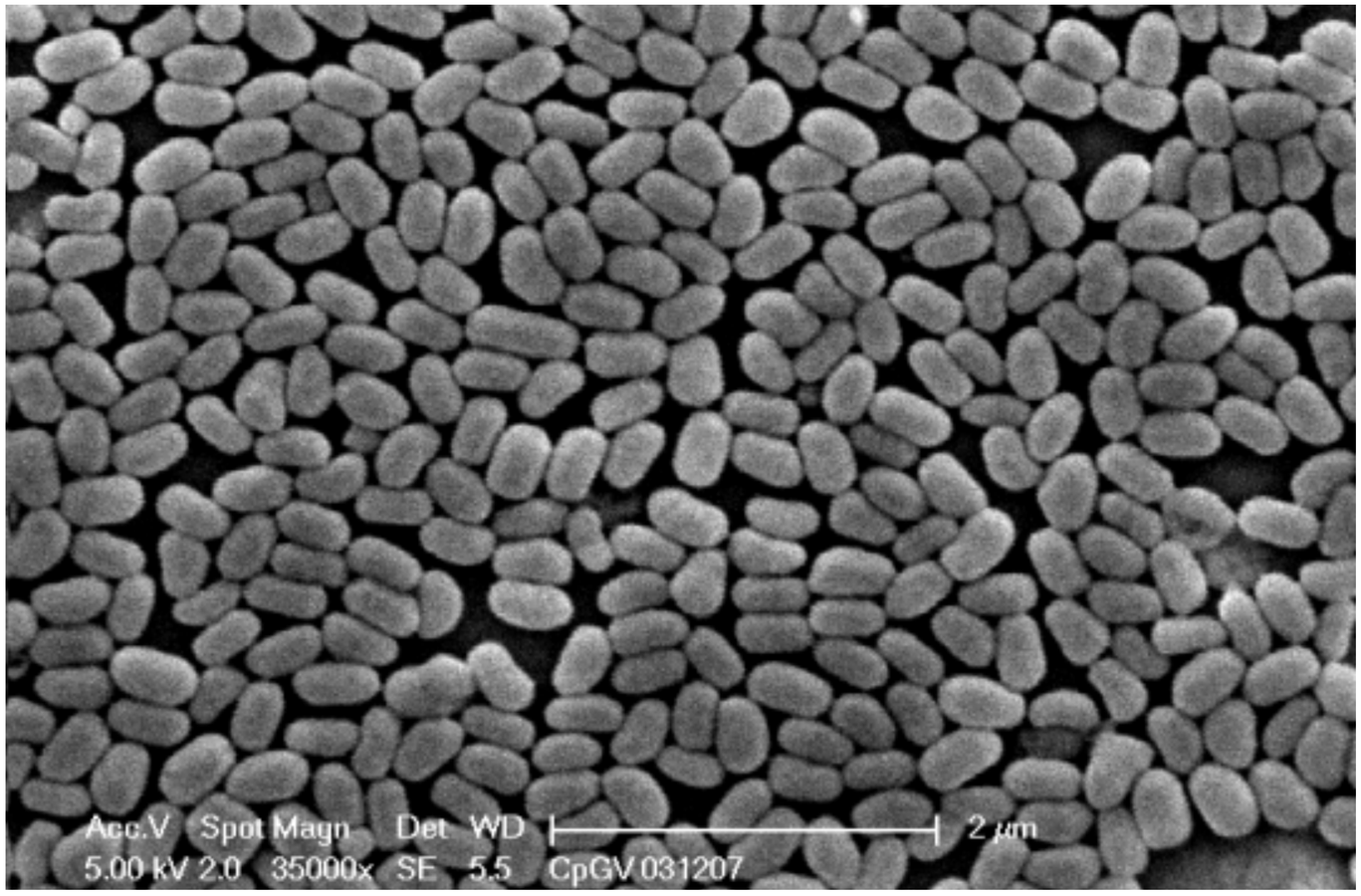
Physics challenges

Segmented detector geometry
Partiality etc

Social challenges

Developing software that everyone can use
"Marketing" a scientific software project

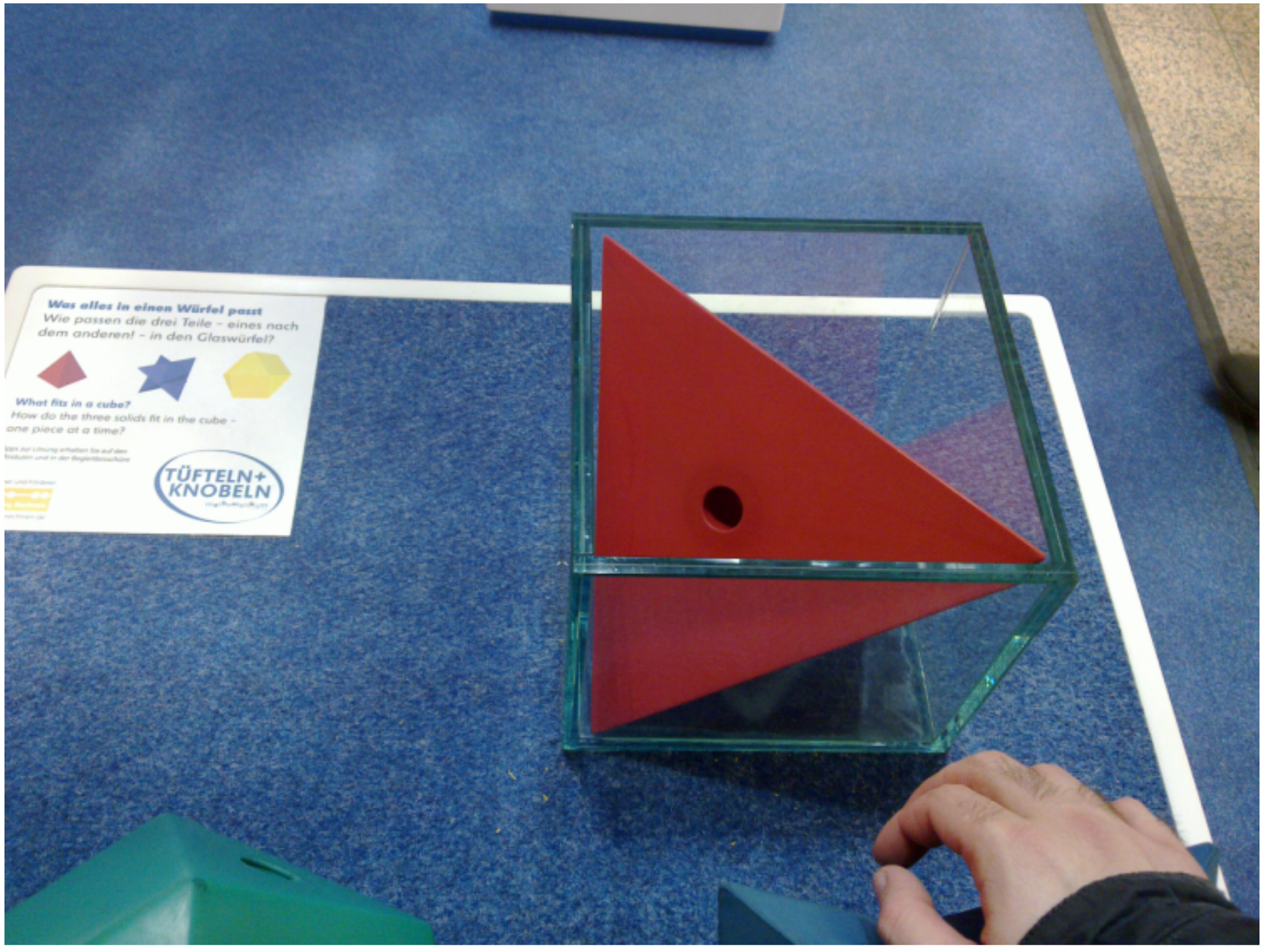
Crystal variability



Granulovirus particles
Virus inside crystalline protein shell
Narrow size distribution
Well-ordered crystal lattice
Available in large quantities

Image: Peter Metcalf

Indexing ambiguities



Technical challenges

Technical challenges

Data storage

File formats

Multiprocessing for high throughput

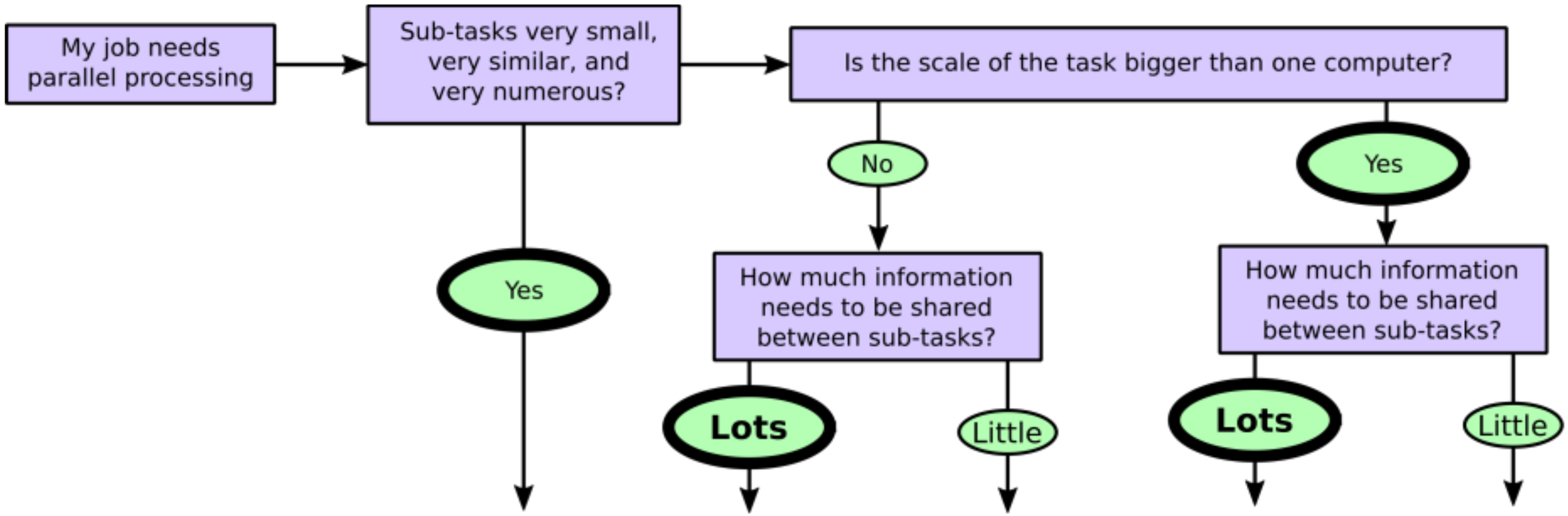
CFEL petabyte data system

Image: Steve Aplin

- 3 PB capacity
- 1200 hard drives
- 64 GBit/sec
- Fibre channel connections

- Cost: you don't want to know (six digits)

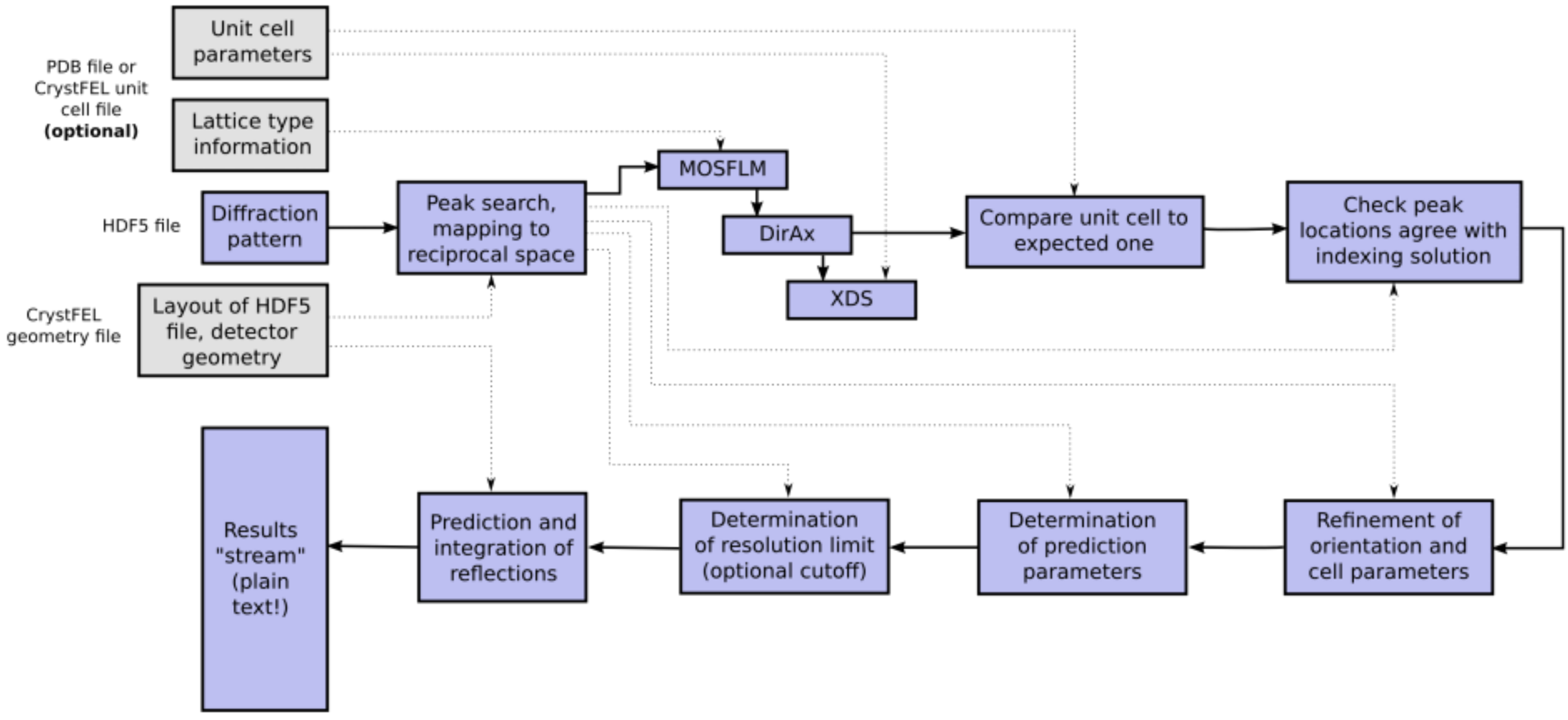
Ways to multiprocess



	GPU	Threads	Processes	MPI	Batch system
	OpenCL CUDA	pthreads OpenMP	fork() Just run many copies GNU parallel	MPICH	SLURM (sbatch, squeue) LSF (bsub, bjobs) Just run many copies
Difficulty:	Use a high-level library Otherwise: ☠	Difficult. Many tricky considerations	It depends.	Not too difficult, good documentation available, even for non-"computer people"	

Don't forget about combining methods.

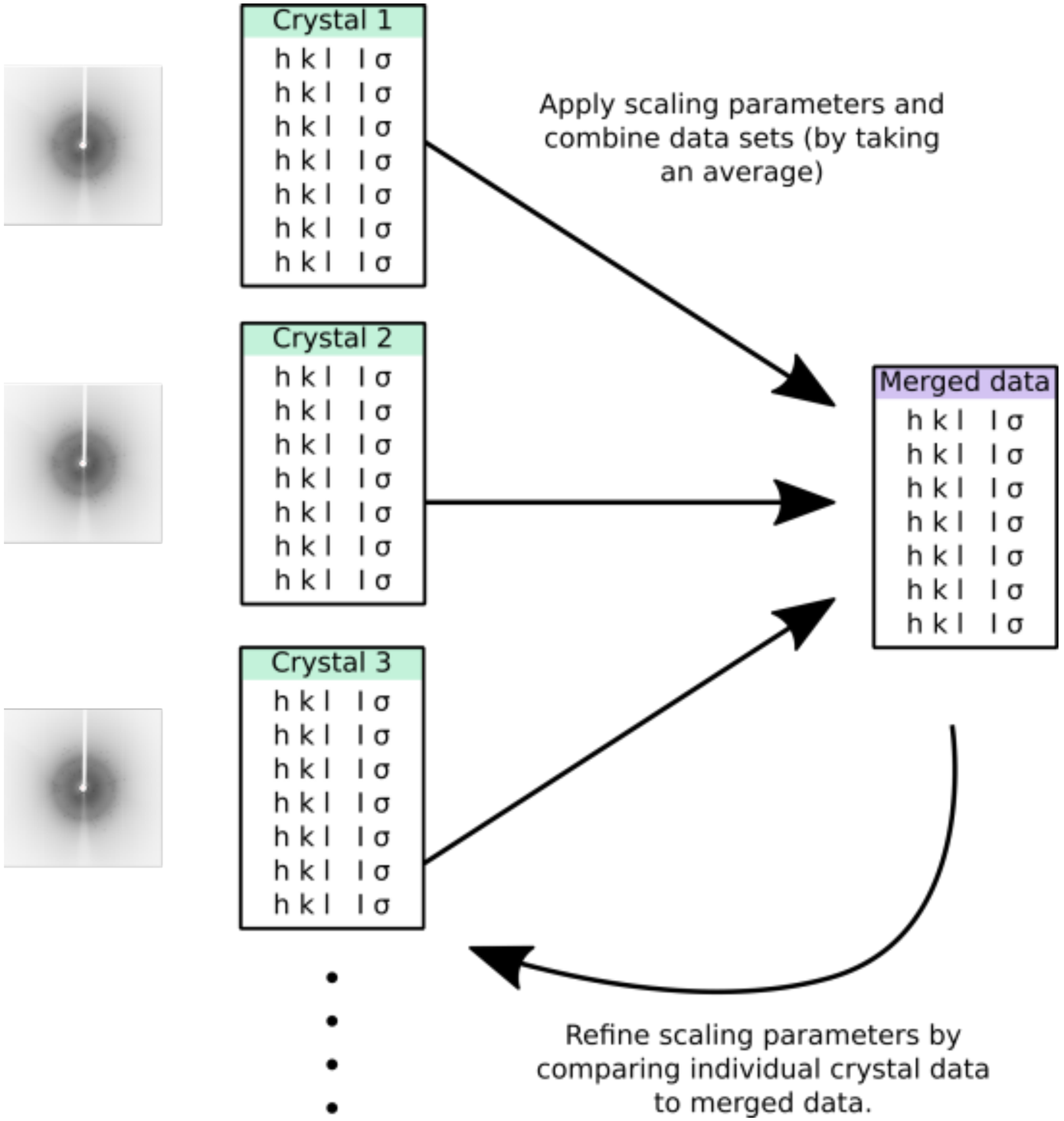
Example: indexing/integrating patterns



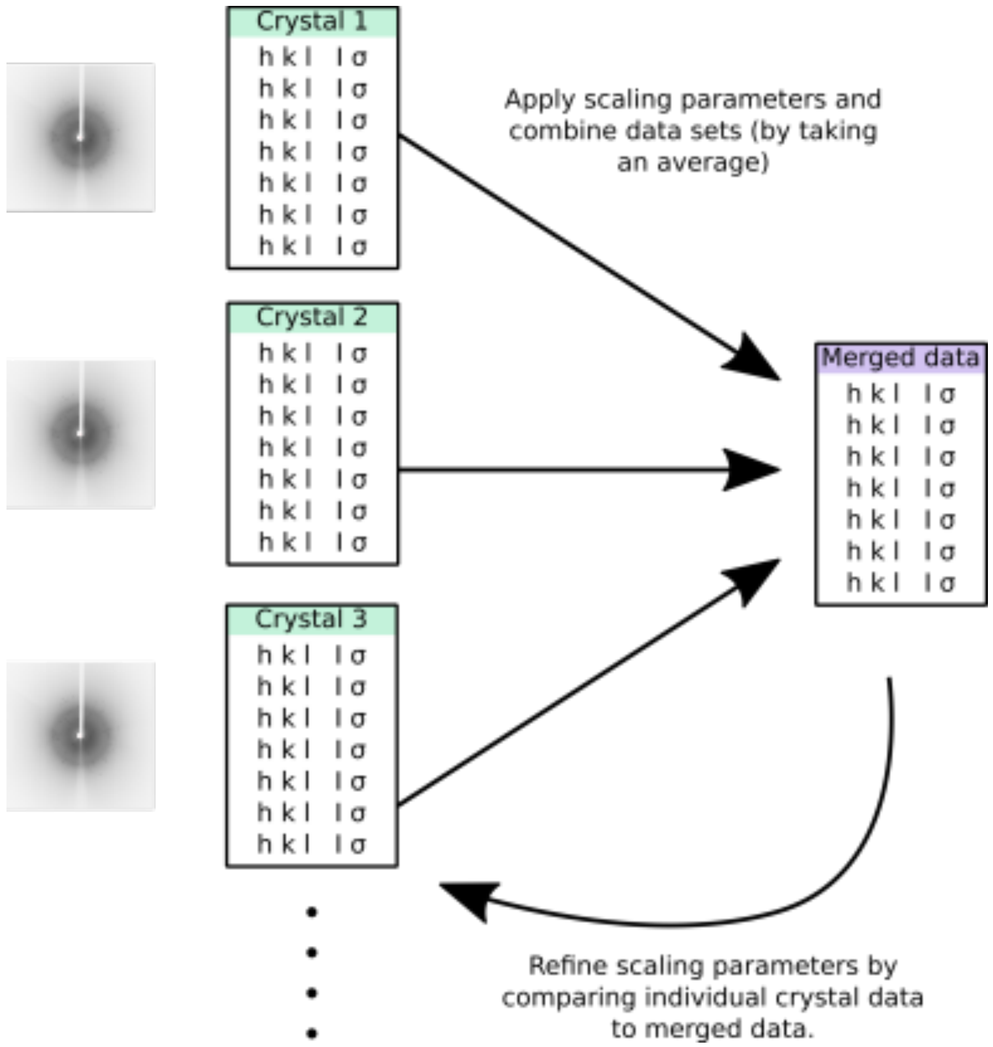
- Scale: Larger than one computer
- Amount of shared data: practically none

⟶ Batch system (multiple processes also available)

Example: merging intensity measurements

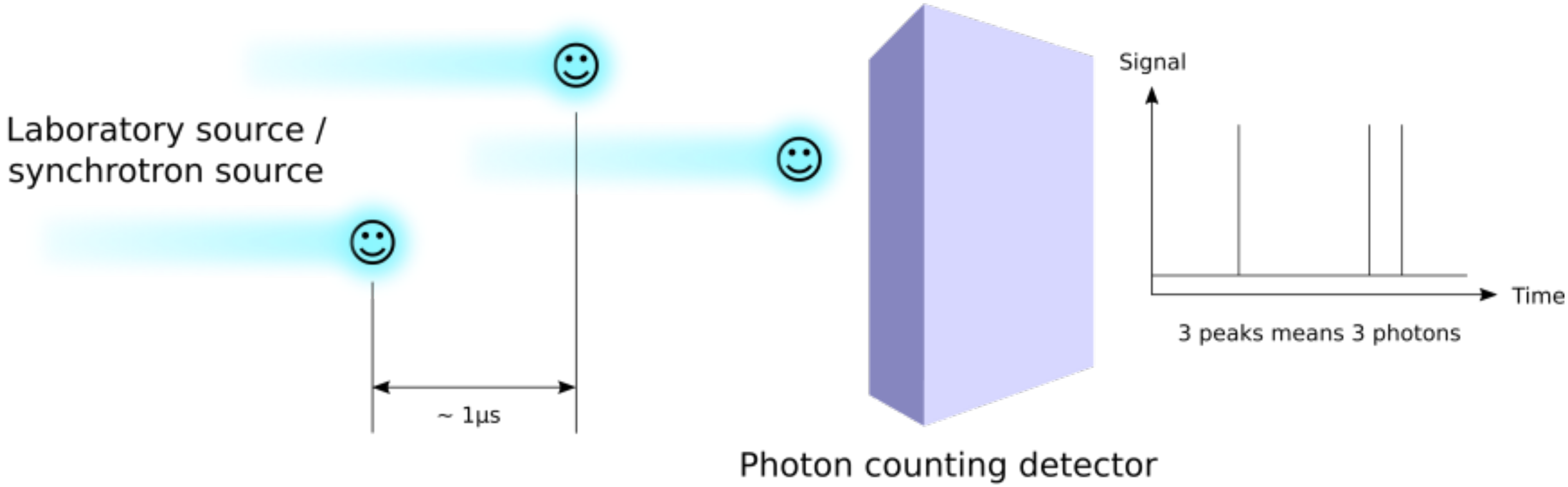


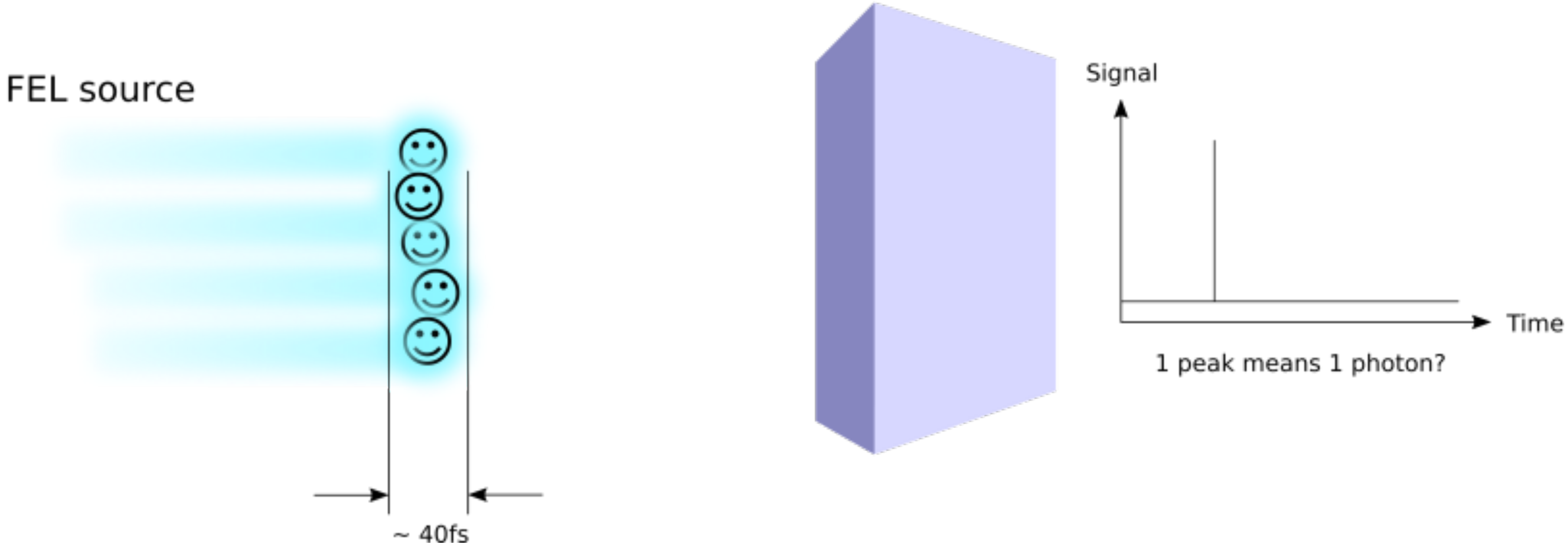
Example: merging intensity measurements



> Scale: about 1 machine
> Shared data: the master list of reflections
⟶ Multithreaded program running on one computer

Types of detector





Cornell-SLAC Pixel Array Detector

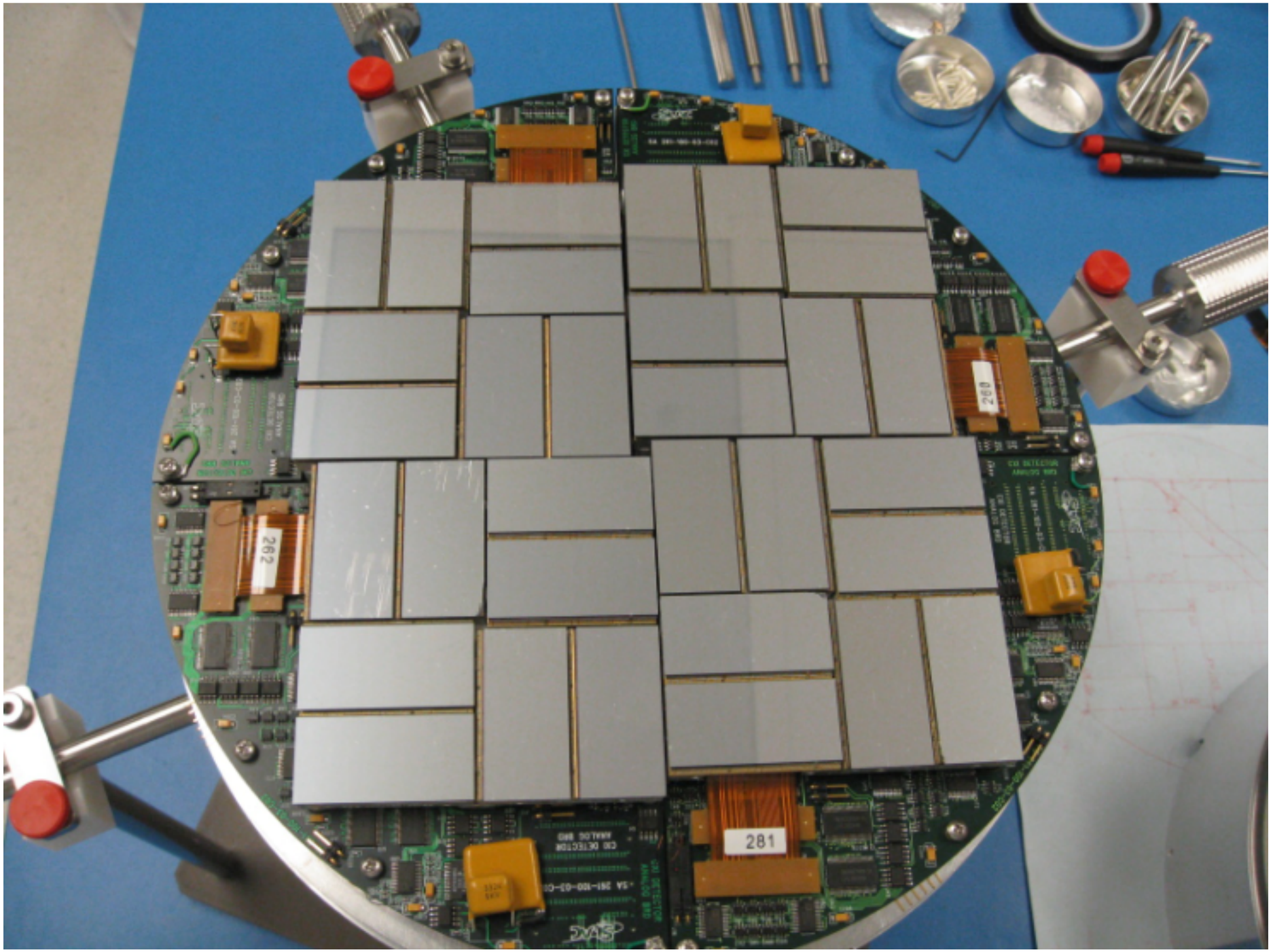


Image: SLAC National Accelerator Laboratory

Adaptive Gain Integrating Pixel Detector

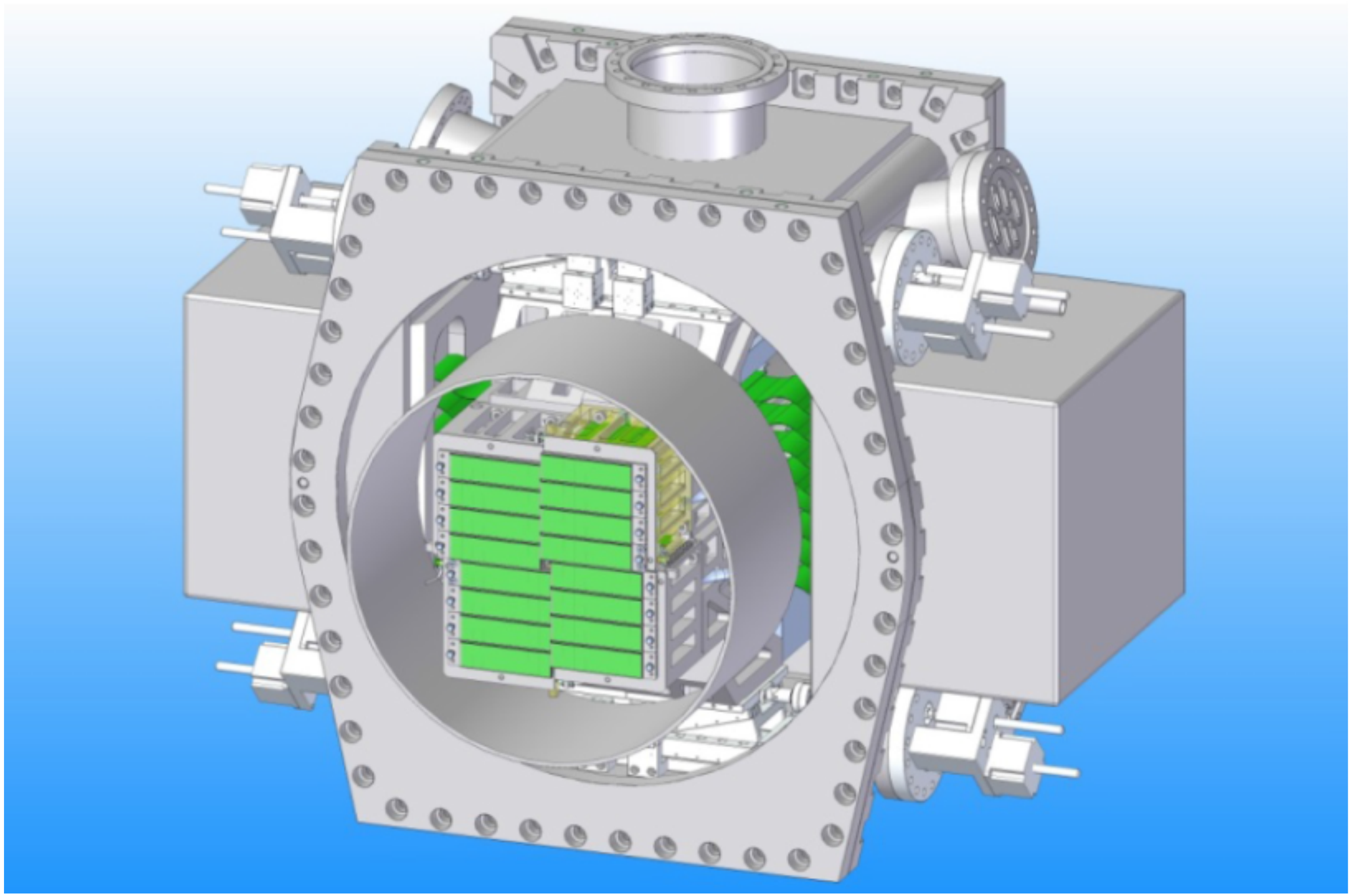
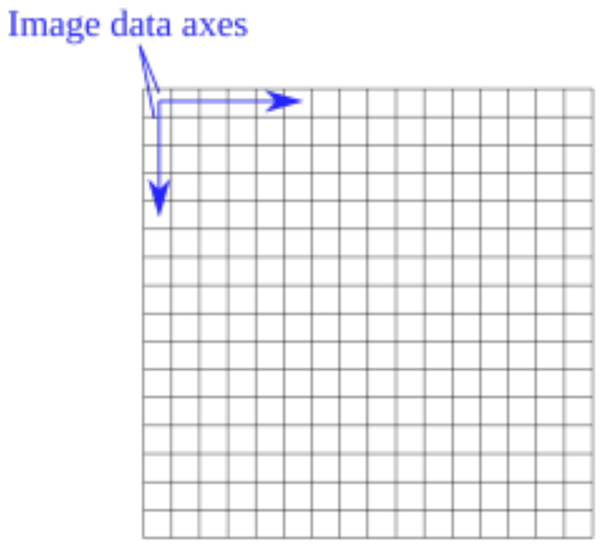
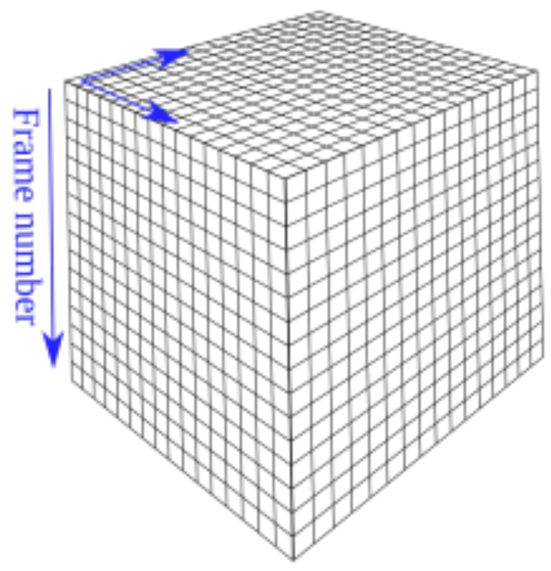
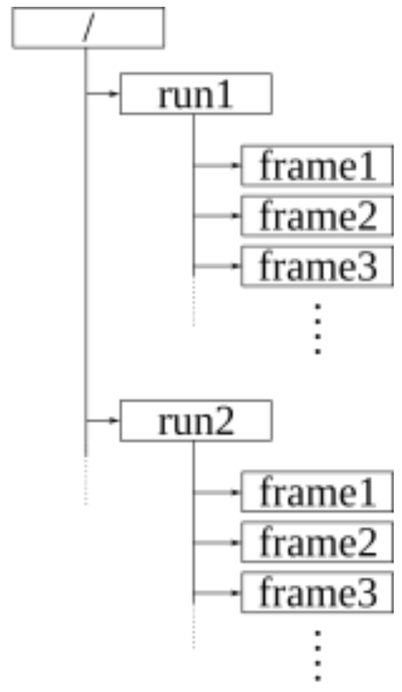


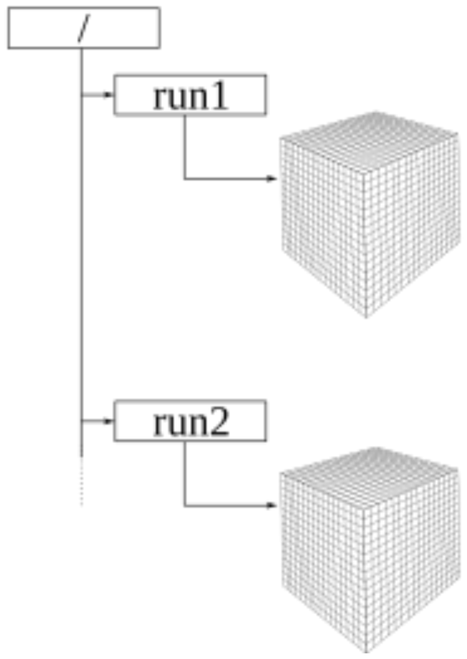
Image: sfx-consortium.org / DESY FS-DS

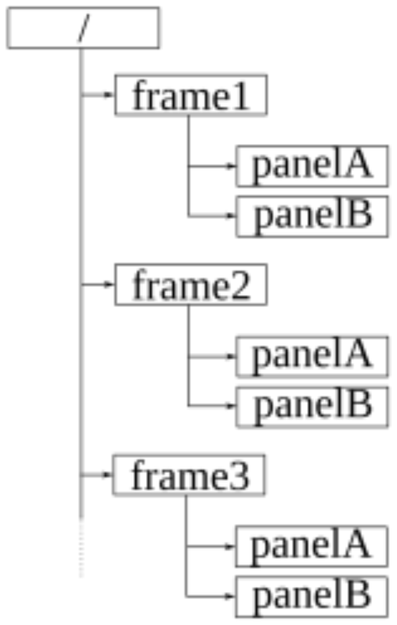
File formats











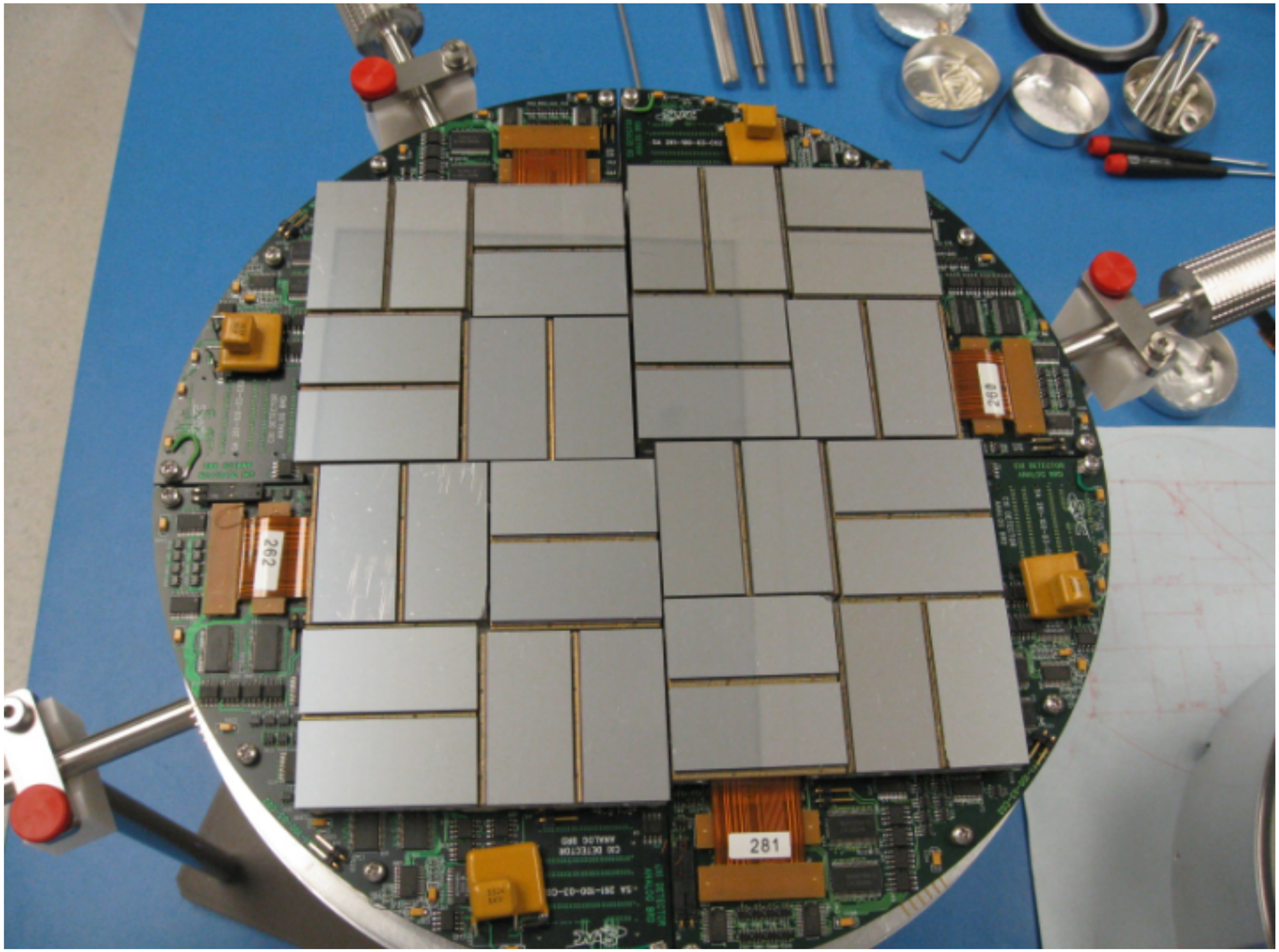
Physics challenges

Physics challenges

Segmented detector geometry

Partiality

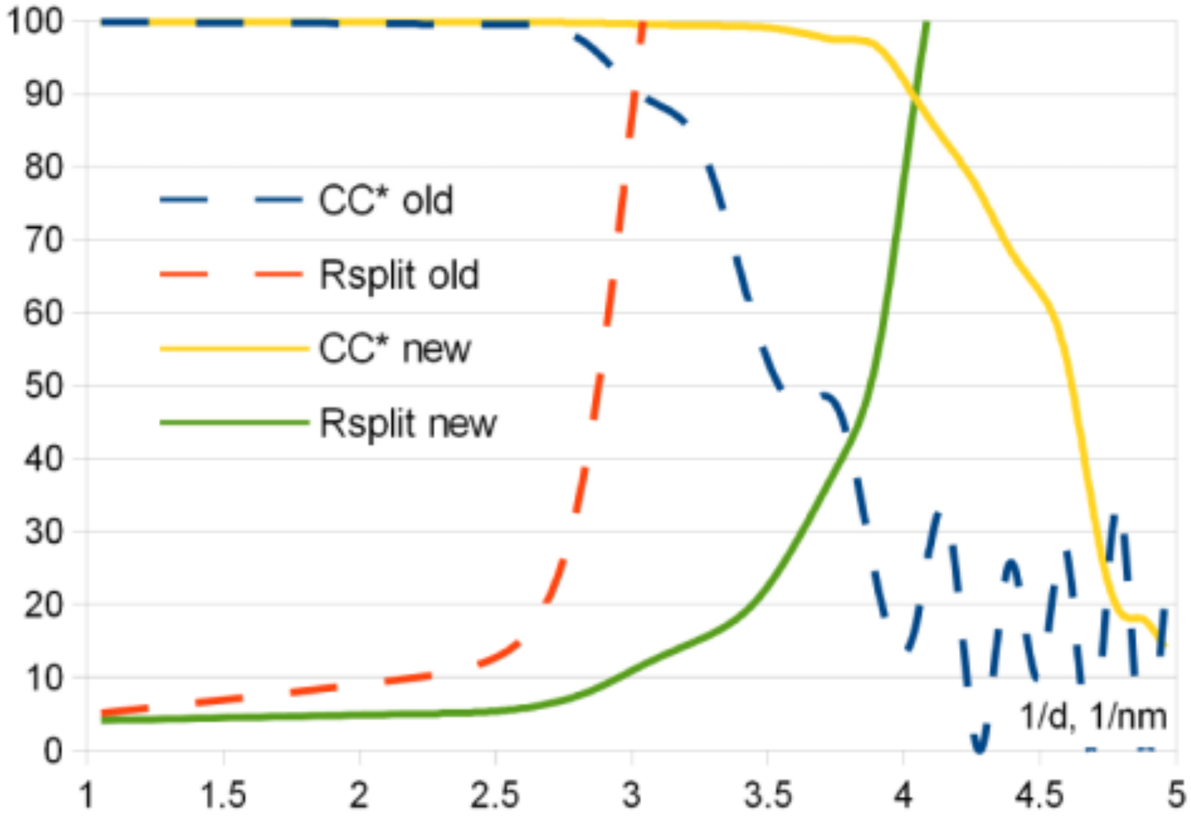
Cornell-SLAC Pixel Array Detector

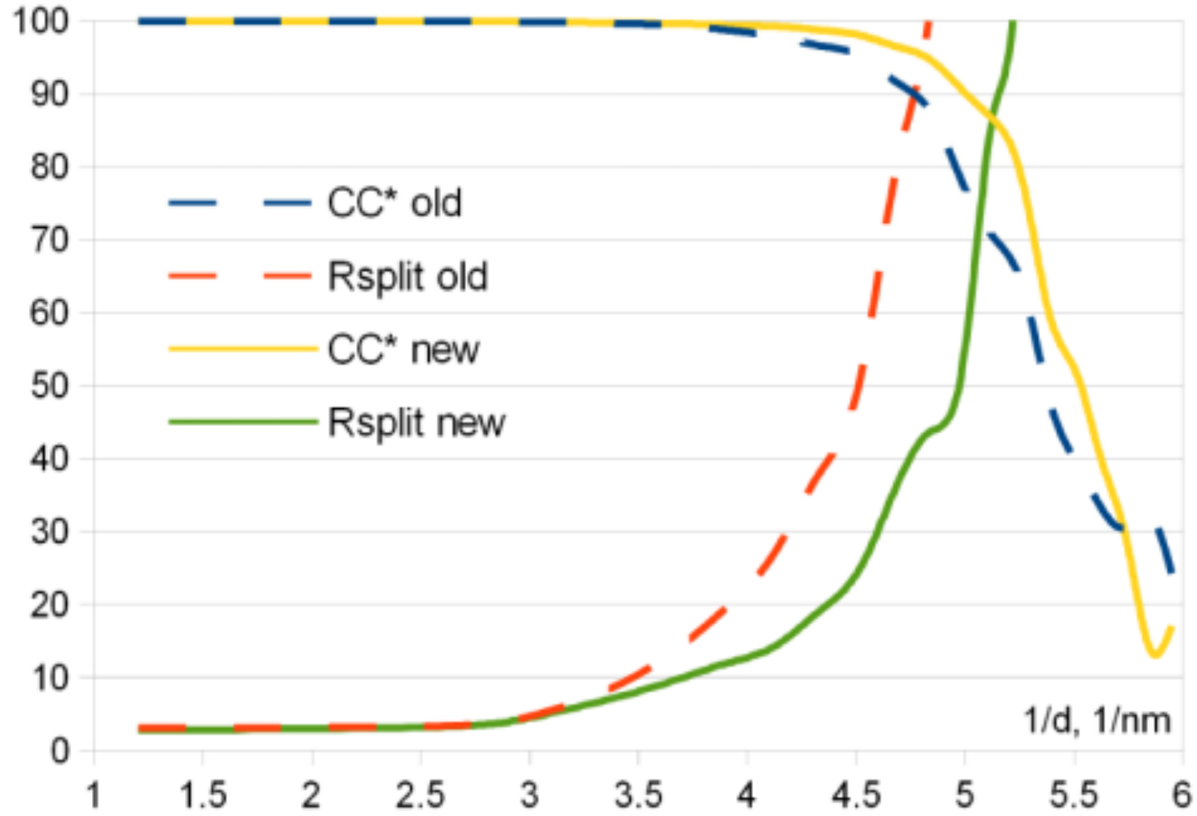


SLAC National Accelerator Laboratory

Importance of detector geometry

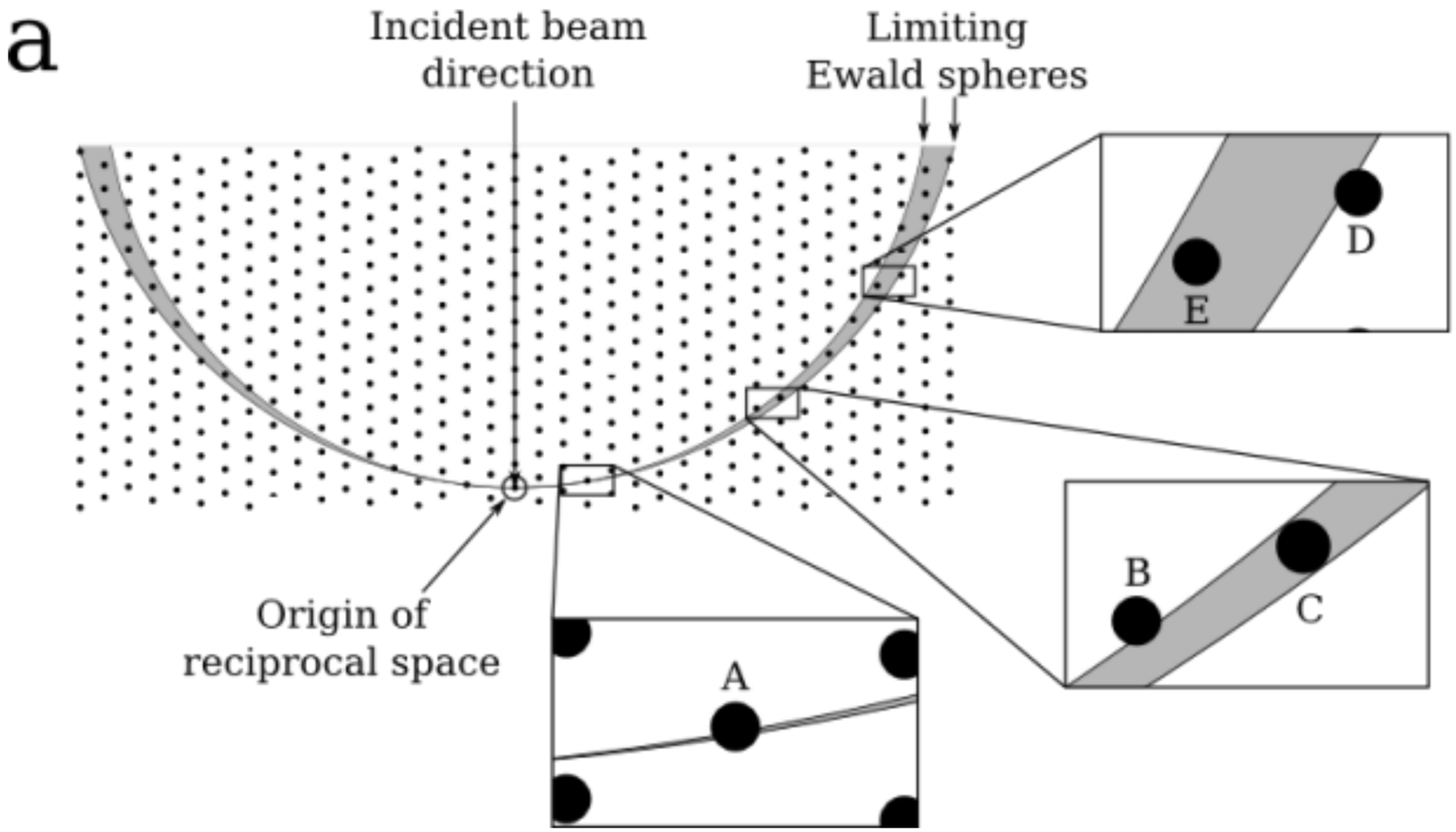
> Even if the geometry looked good, a refinement of a fraction of a pixel can improve the data significantly.

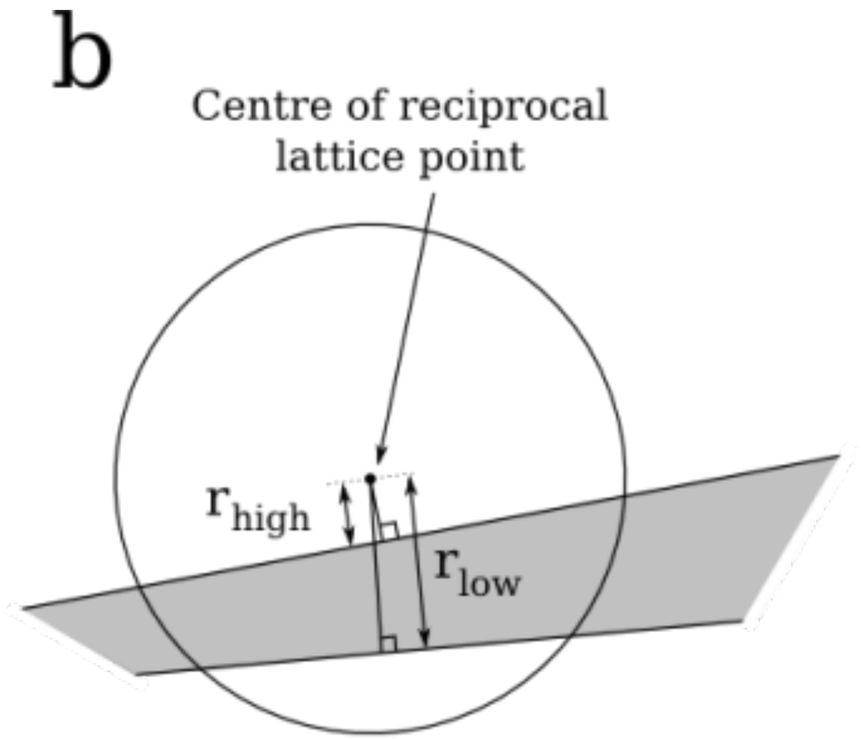




Yefanov et al., Optics Express 23 (2015) p28459

Partially recorded reflections



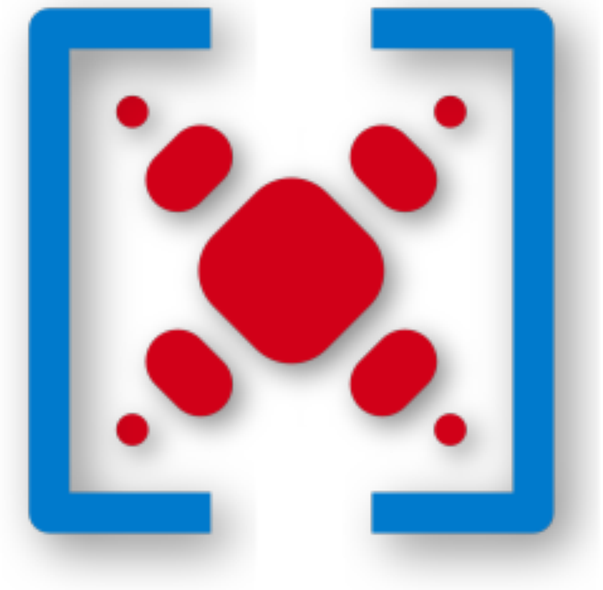


Social challenges

Developing software that everyone can use
"Marketing" a scientific software project

What is CrystFEL?

> **CrystFEL** software suite for FEL crystallography, and also for experiments using other types of light source.



- "Clean diffraction patterns to merged intensities"
- Free and open-source software: GPLv3
- Development since late 2009
- Project led by Thomas White (DESY/CFEL)
- Current version: 0.6.3
- User-oriented development.
- Documentation.

https://www.desy.de/~twhite/crystfel
(or just search for "CrystFEL")

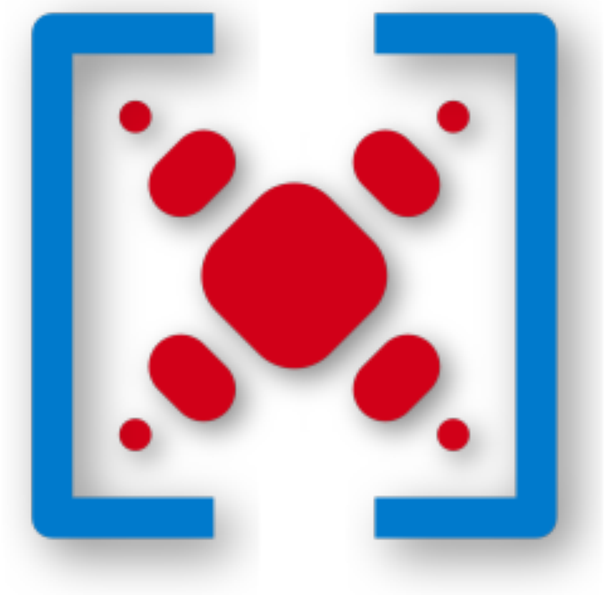
Documentation

More than just a manual:

- "man" pages
- indexamajig --help
- Website:
  - Tutorial
  - Best practice
  - Web versions of "man" pages
  - API documentation
  - Old presentations
  - Change log
  - Installation instructions
  - Citation list
  - FAQ (frequently asked/anticipated questions)
- Test data (thanks to cxidb.org)
- Papers describing the software directly
- Papers describing use of the software (accurately!)

What is CrystFEL?

> **CrystFEL** software suite for FEL crystallography, and also for experiments using other types of light source.



- "Clean diffraction patterns to merged intensities"
- Free and open-source software: GPLv3
- Development since late 2009
- Project led by Thomas White (DESY/CFEL)
- Current version: 0.6.3
- User-oriented development.
- Documentation.

https://www.desy.de/~twhite/crystfel
(or just search for "CrystFEL")

The European XFEL

- https://www.xfel.eu/
- First X-ray beam in experimental hall: 23 June 2017
- First user experiments: September 2017

- 27,000 pulses per second

Images: European XFEL

European XFEL pulse structure

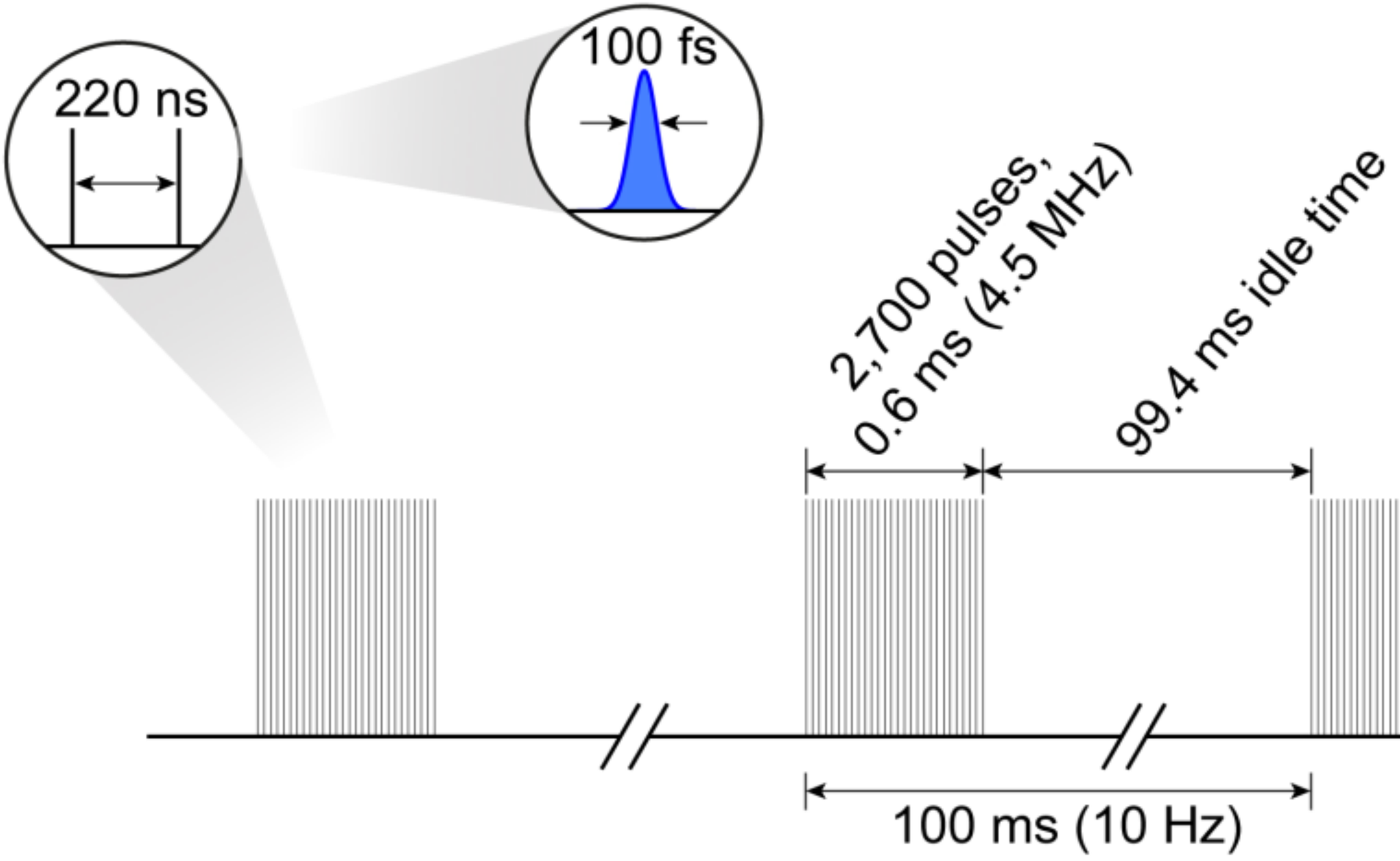


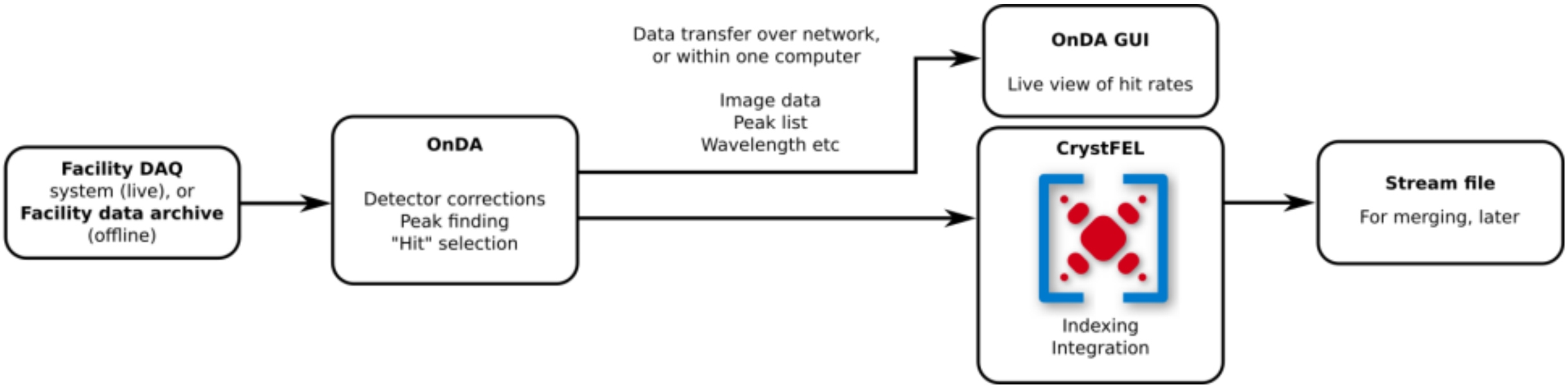
Image: European XFEL

Pipeline without intermediate files 1

> Put our code inside "blocks" the fit into the facility data processing system.
> At European XFEL: "Karabo"

Pipeline without intermediate files 2

> Instead of embedding code in framework, have completely separate programs communicating via a protocol



Acknowledgements: people

Contributors to CrystFEL:

Richard Kirian <rkirian@asu.edu>
Kenneth Beyerlein <kenneth.beyerlein@desy.de>
Andrew Aquila <andrew.aquila@cfel.de>
Andrew Martin <andrew.martin@desy.de>
Lorenzo Galli <lorenzo.galli@desy.de>
Chun Hong Yoon <chun.hong.yoon@desy.de>
Kenneth Beyerlein <kenneth.beyerlein@desy.de>
Karol Nass <karol.nass@desy.de>
Nadia Zatsepin <nadia.zatsepin@asu.edu>
Anton Barty <anton.barty@desy.de>
Thomas Grant <tgrant@hwi.buffalo.edu>
Cornelius Gati <cornelius.gati@desy.de>
Fedor Chervinskii <fedor.chervinskii@gmail.com>
Alexandra Tolstikova <alexandra.tolstikova@desy.de>
Wolfgang Brehm <wolfgang.brehm@gmail.com>
Valerio Mariani <valerio.mariani@desy.de>
Parker de Waal <Parker.deWaal@vai.org>
Takanori Nakane <nakane.t@gmail.com>
Keitaro Yamashita <k.yamashita@spring8.or.jp>
Oleksandr Yefanov <oleksandr.yefanov@cfel.de>
Helen Ginn <helen@strubi.ox.ac.uk>
Steve Aplin <steve.aplin@desy.de>

+ many more for feedback, words of wisdom etc!

https://www.desy.de/~twhite/crystfel
(or just search for "CrystFEL")

Acknowledgements: €€€

- Helmholtz Association via Program-Oriented Funding
- PIER Helmholtz Graduate School
- European Union's 2020 Research and Innovation Programme - Marie Skłodowska-Curie grant agreement 637295 "X-Probe")
- X-ray Free Electron Laser Priority Strategy Programme (MEXT, Japan)
- SACLA HPC and Mini-K
- BMBF German-Russian Cooperation "SyncFELMed", grant 05K14CHA
- BioStruct-X, Seventh Framework Programme of the European Commission