

Data Flow in Macromolecular Crystallography (MX) Experiments

Manfred S. Weiss

NP-GMX

Helmholtz-Zentrum Berlin, Germany

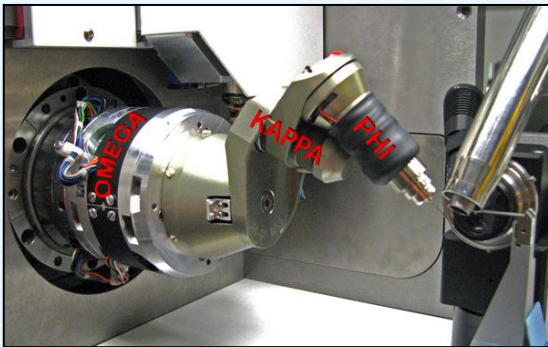
The HZB-MX beam lines

Beamline BL14.1

- Tunable energy: 5 – 16 keV
- Photon flux: 5.0×10^{11} Ph / s



- PILATUS 6M-detector
- CATS robotic sample changer
- High-performance multi-axis goniometer MD2 with MK3

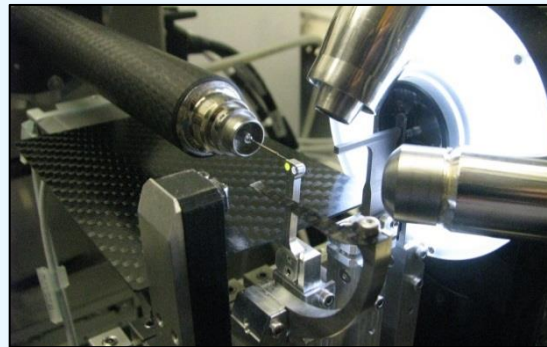


Beamline BL14.2

- Tunable energy: 5 – 16 keV
- Photon flux: 4.5×10^{11} Ph / s

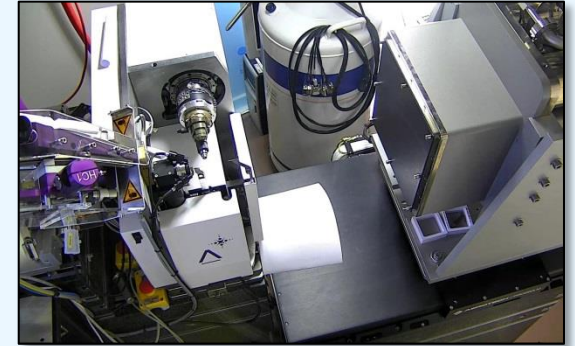


- PILATUS3S 2M-detector
- G-ROB sample changer
- Nanodiffractometer with fast air-bearing axis

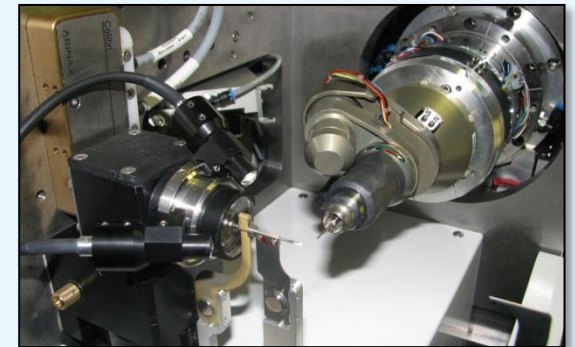


Beamline BL14.3

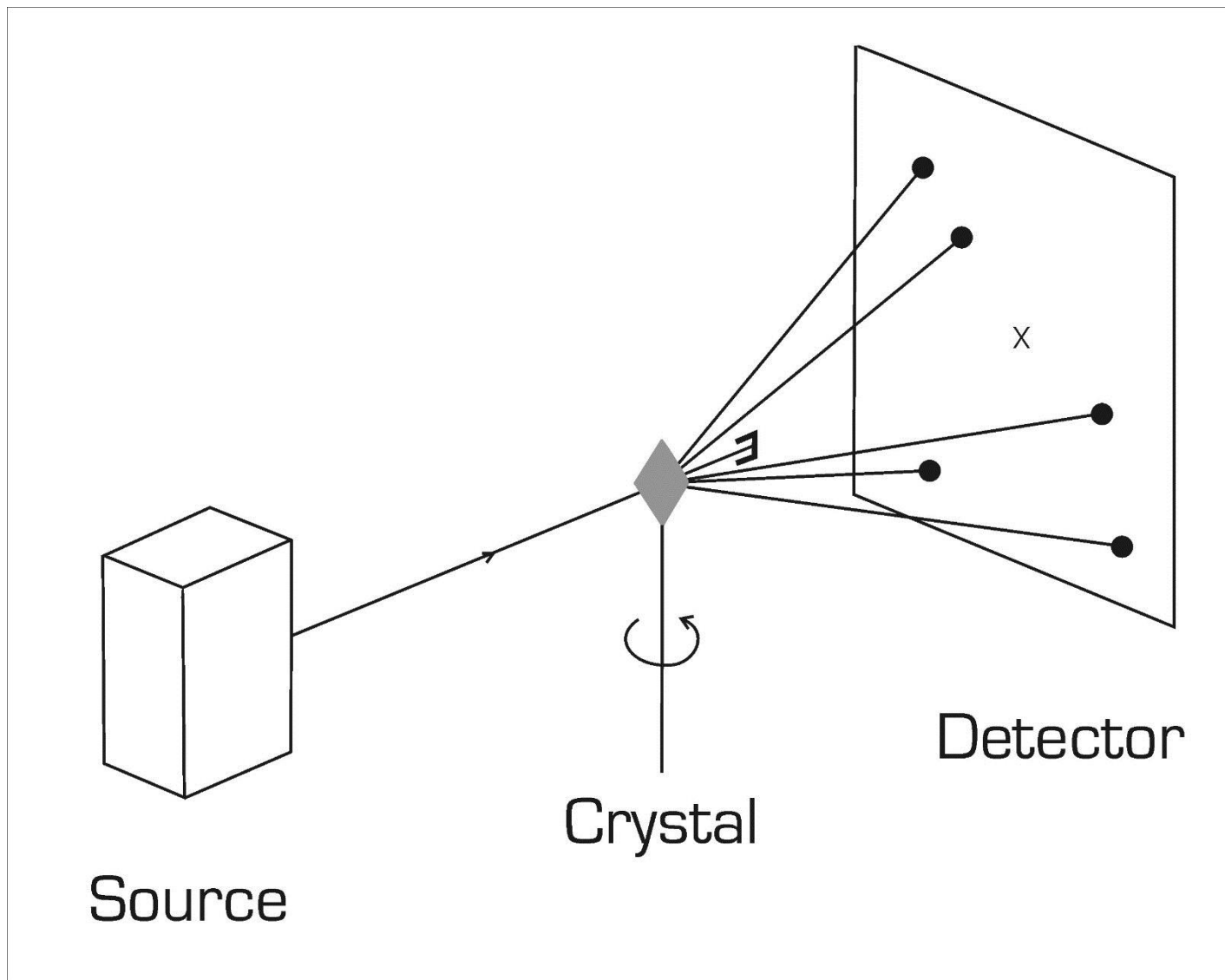
- Fixed photon energy: 13.8 keV
- Photon flux: 1.2×10^{11} Ph / s



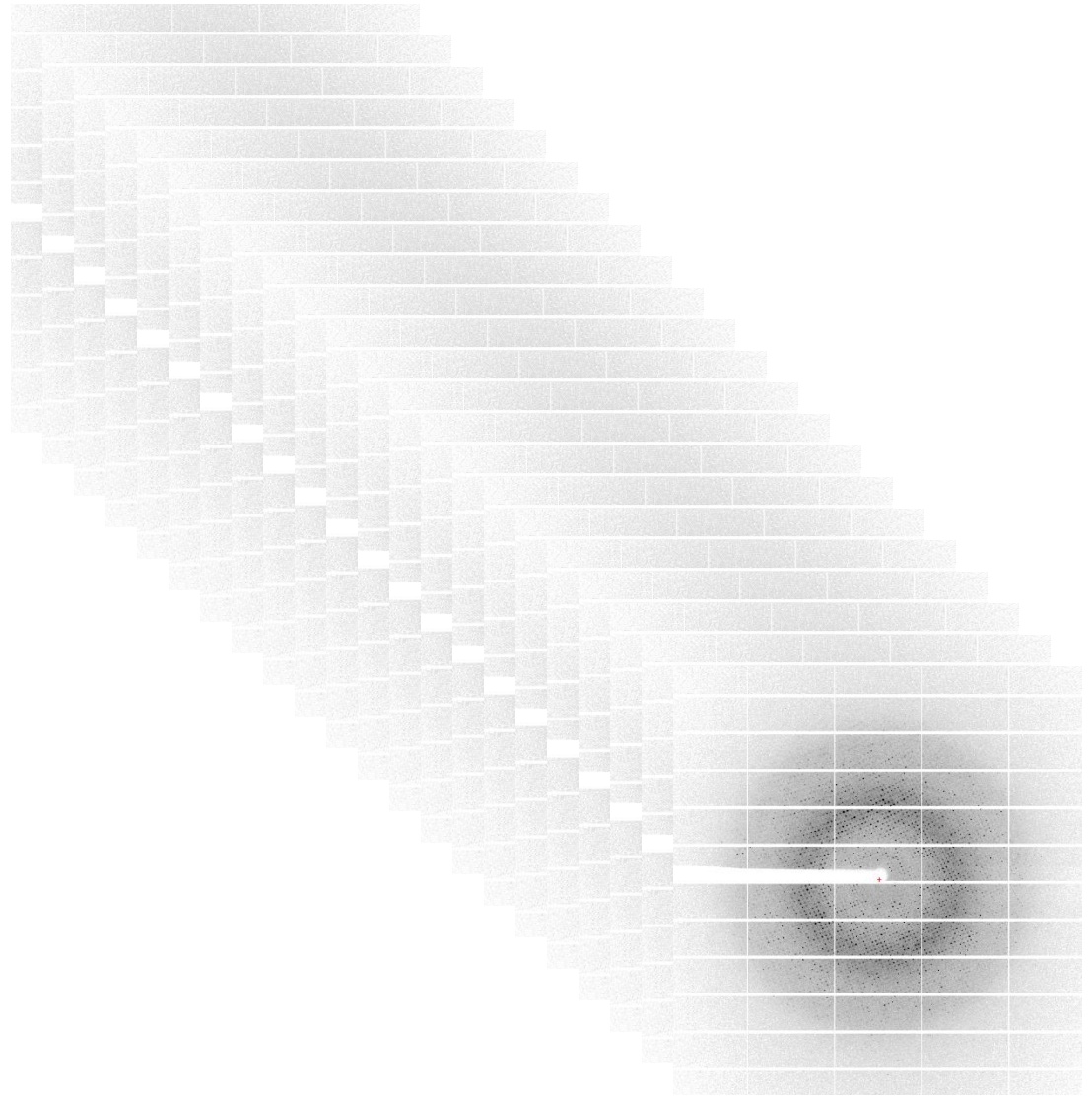
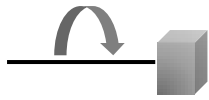
- Rayonix MX225 CCD-detector
- Manual sample mounting
- High-performance multi-axis goniometer MD2-S with MK3



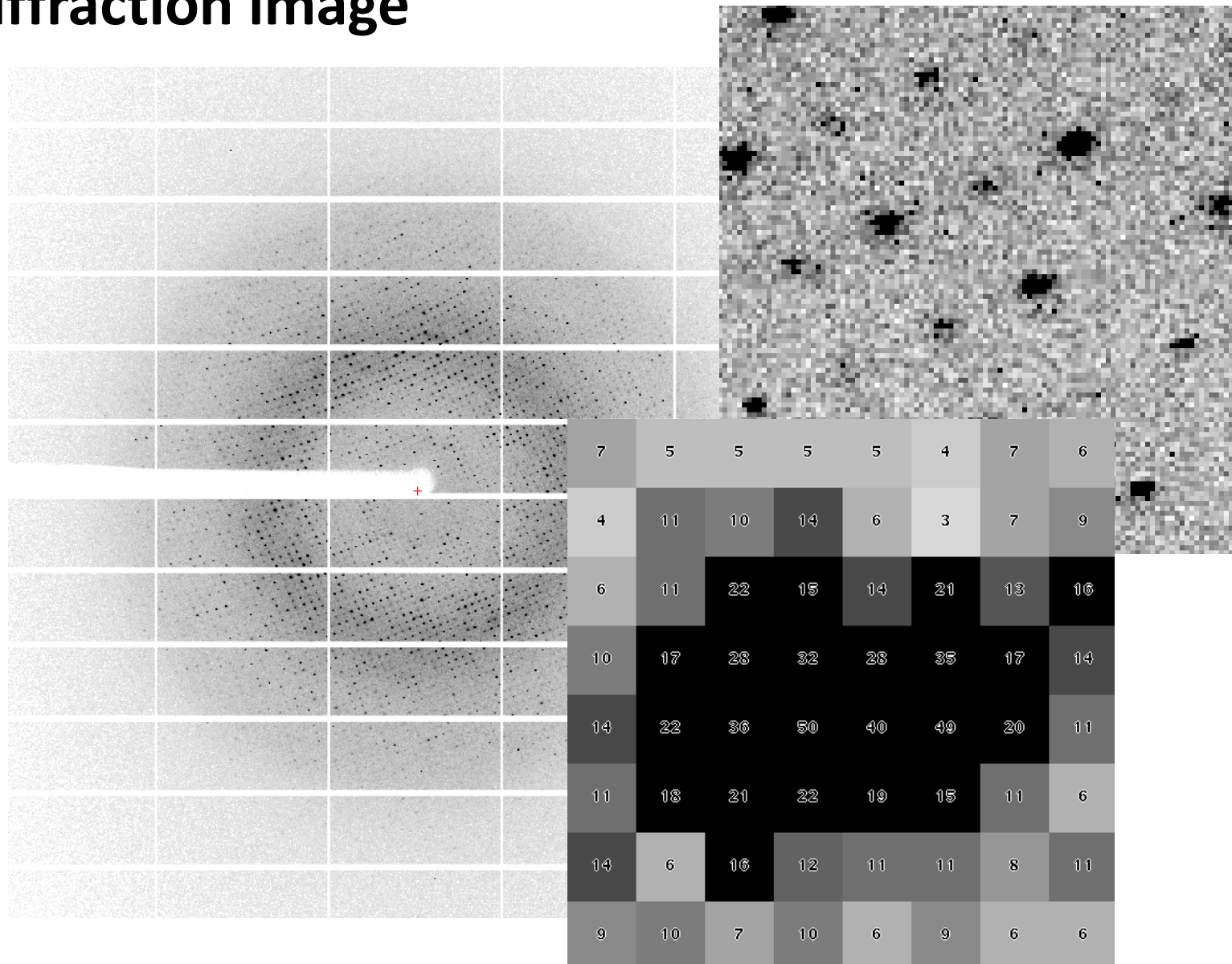
The standard MX experiment



A data set



A diffraction image

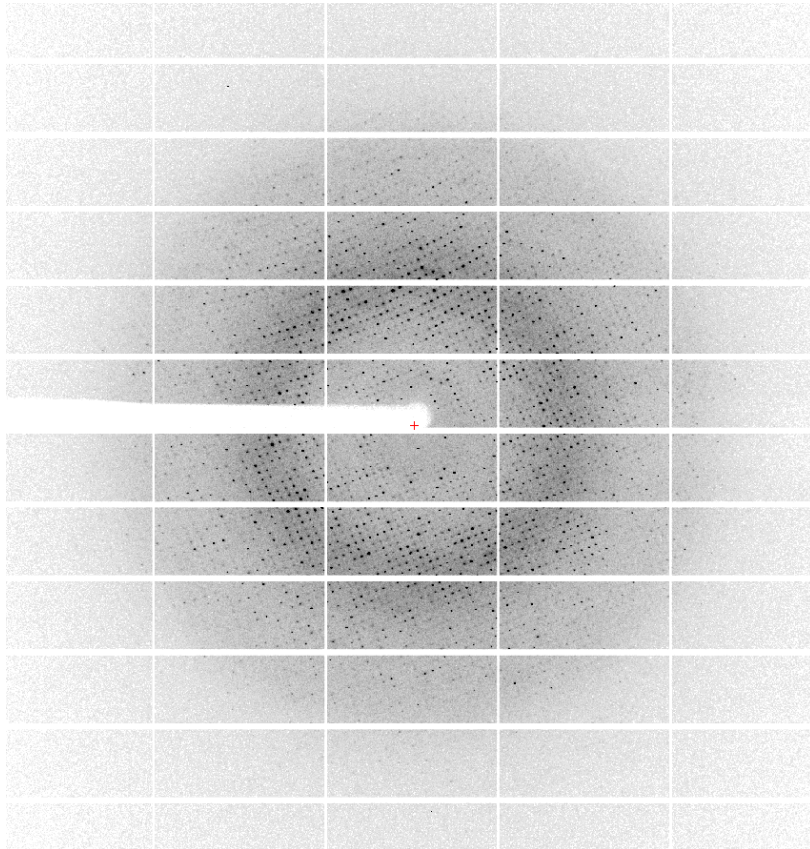


A typical data set then and now

	PILATUS (2013 ff)	CCD (until 2013)
Total rotation range [°]	180	180
No. of images	1800	180
Image size [MB]	6	18
Data set size [GB]	11	3
Total exposure time [s]	360	900
Readout time [s]	4	450
Total time [s]	364	1350
Data sets per 24 hrs	144	48
Data rate [MB/s]	30	2.4
Total data per 24 hrs [TB]	1.5	0.15

Data processing

Data processing is the step in which the collected diffraction images are reduced to a list of individual X-ray reflections (h, k, l) and their intensities $I(h,k,l)$ and estimated errors $\sigma I(hkl)$.

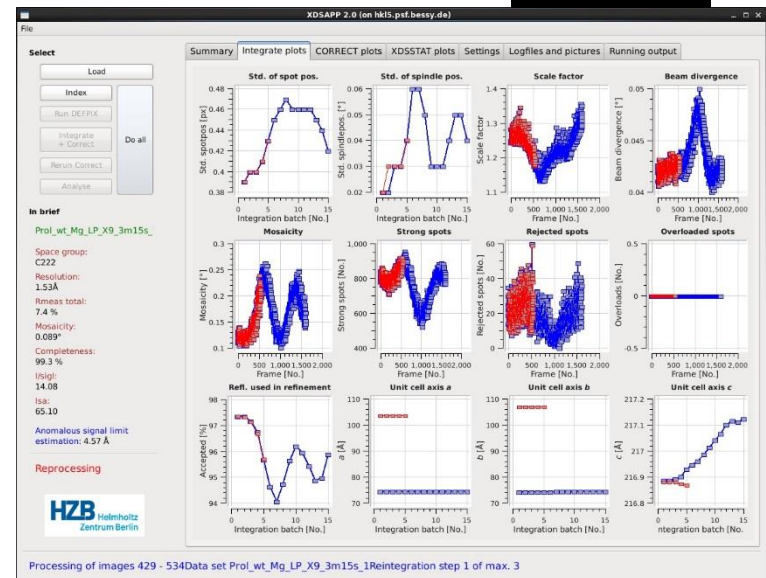


h	k	l	I (hkl)	σI (hkl)
0	0	6	24324	1276
0	0	12	12440	144
0	0	18	5111	123
0	0	24	7350	235
0	1	0	650	89
0	1	1	222	75
0	1	2	899	130
0	1	3	1250	95
0	1	4	2479	148
0	1	5	175	75
.
.
.

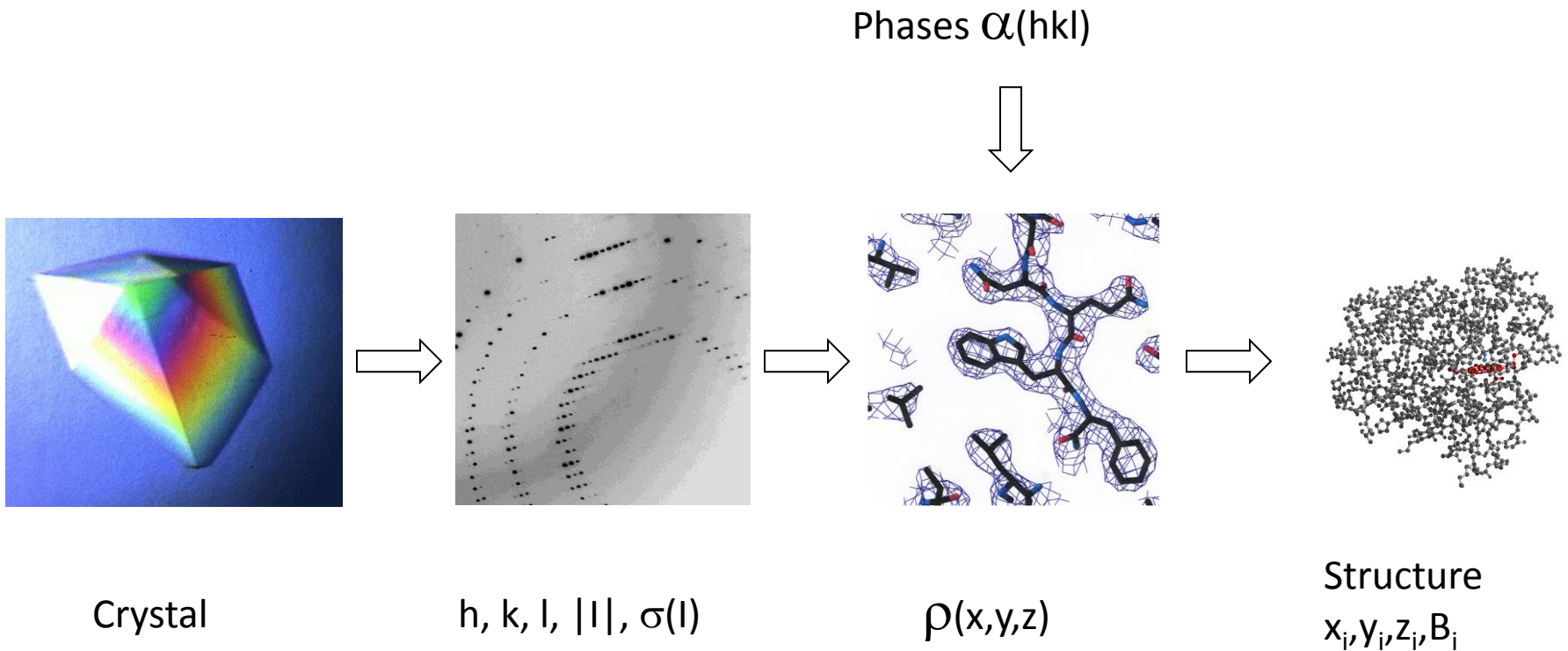
Users wish to have their data processed in real time

- dedicated multi-processor server per beamline
- large data storage areas
- fast (parallel) data processing program
- more or less automated interface
- heavy network traffic

XDSAPP



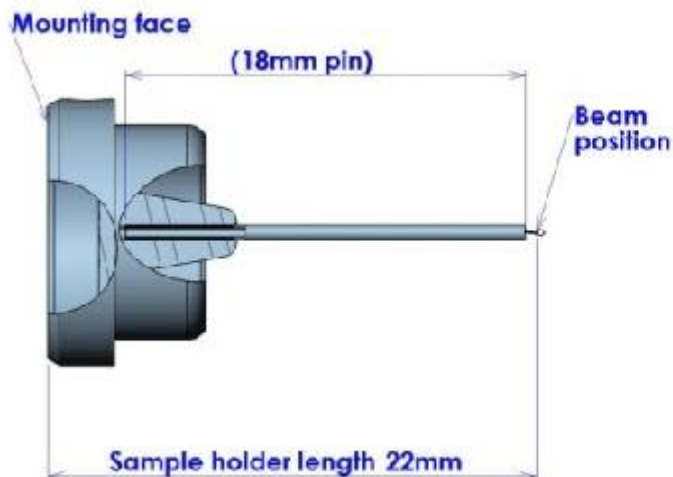
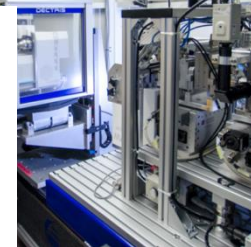
The standard MX project



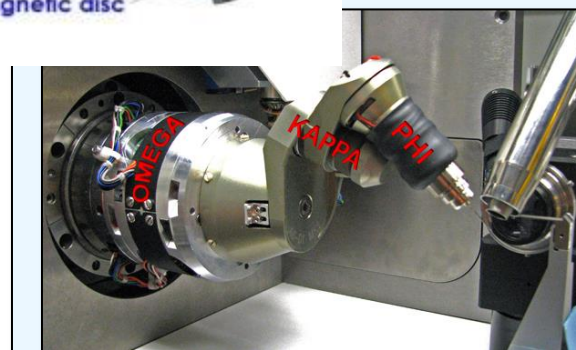
Standardization in MX

Beamline BL14.1

- Tunable energy: 5.0 – 15.5 keV
- Photon flux: 3.6×10^{11} Phot. / s

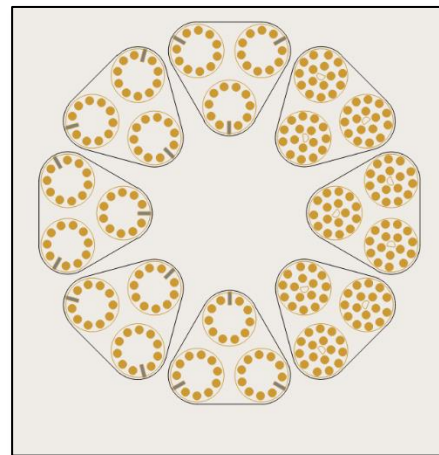


ector
mple changer
nce multi-axis
D2 with MK3



Standardization in MX

- users bring 100-200 samples per 24-hrs of beam time



Metadata in MX

Relevant for the experiment

Wavelength of incident radiation

Orientation of rotation axis

Rotation direction

Rotation increment per image

Crystal-detector distance

Detector type and orientation

Direct beam coordinates

Location and name of data set

Metadata in MX

Relevant for the experiment	Nice to have for the experiment
Wavelength of incident radiation	Date and time
Orientation of rotation axis	Flux on sample
Rotation direction	Beam size and profile
Rotation increment per image	Exposure time per image
Crystal-detector distance	Experimenter
Detector type and orientation	Size and shape of sample
Direct beam coordinates	
Location and name of data set	

Metadata in MX

Relevant for the experiment	Nice to have for the experiment	Relevant for the project
Wavelength of incident radiation	Date and time	Composition of sample
Orientation of rotation axis	Flux on sample	
Rotation direction	Beam size and profile	
Rotation increment per image	Exposure time per image	
Crystal-detector distance	Experimenter	
Detector type and orientation	Size and shape of sample	
Direct beam coordinates		
Location and name of data set		

The problem with Sample Composition

- Macromolecule production
 - Macromolecular purification
 - Macromolecule crystallization
 - Treatment of crystal before shipment to synchrotron
- Data collection
 - Data processing
- Structure determination and refinement
 - Analysis



The problem with Sample Composition

- Macromolecule production
 - *Source organism*
 - *DNA source*
 - *Forward primer*
 - *Reverse primer*
- Macromolecular purification
 - *Extraction from host organism*
 - *Chromatographic techniques*
- Macromolecule crystallization
 - *Buffer compositions*
 - *Concentration of sample*
 - *Purity assessment*
 - *Biophysical conditions*
 - *Method of crystallization*
- Treatment of crystal before shipping
 - *Method*
 - *Plate type*
 - *Temperature*
 - *Concentration of macromolecule*
 - *Buffer composition of sample solution*
 - *Composition of reservoir solution*
 - *Volume and ratio of drop*
 - *Volume of reservoir*
- Data collection
- Data processing
- Structure determination and refinement
- Analysis
 - *Method crystal fishing*
 - *Device used for fishing*
 - *Soaking conditions*
 - *Soaking time*
 - *Cryo-protection*
 - *Method of cryo-cooling*

A Further Complication of the Situation

- Only one in about 100 experiments ends up in the public record
 - *No two samples are the same*
 - *Samples are slightly OR systematically different*
 - *Slight differences are often not documented*

A Further Complication of the Situation

- Only one in about 100 experiments ends up in the public record
- Some data sets are assembled from more than one sample

A Further Complication of the Situation

- Only one in about 100 experiments ends up in the public record
- Some data sets are assembled from more than one sample
- Extreme case: serial crystallography

FAIR data in MX

Findable Accessible Interoperable Re-usable (FAIR) diffraction data are coming to protein crystallography

John R. Helliwell,^a Wladek Minor,^b Manfred S. Weiss,^c Elspeth F. Garman,^d
Randy J. Read,^e Janet Newman,^f Mark J. van Raaij,^g Janos Hajdu^{h,i} and
Edward N. Baker^j

May 2019: Joint editorial in IUCr J. Acta Cryst. D, Acta Cryst. F & J. Appl. Cryst.

Data Flow in MX Experiments

Thank you for your attention