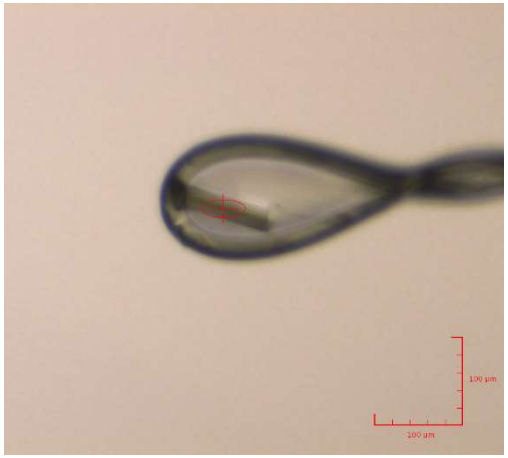
X-ray Diffraction experiment

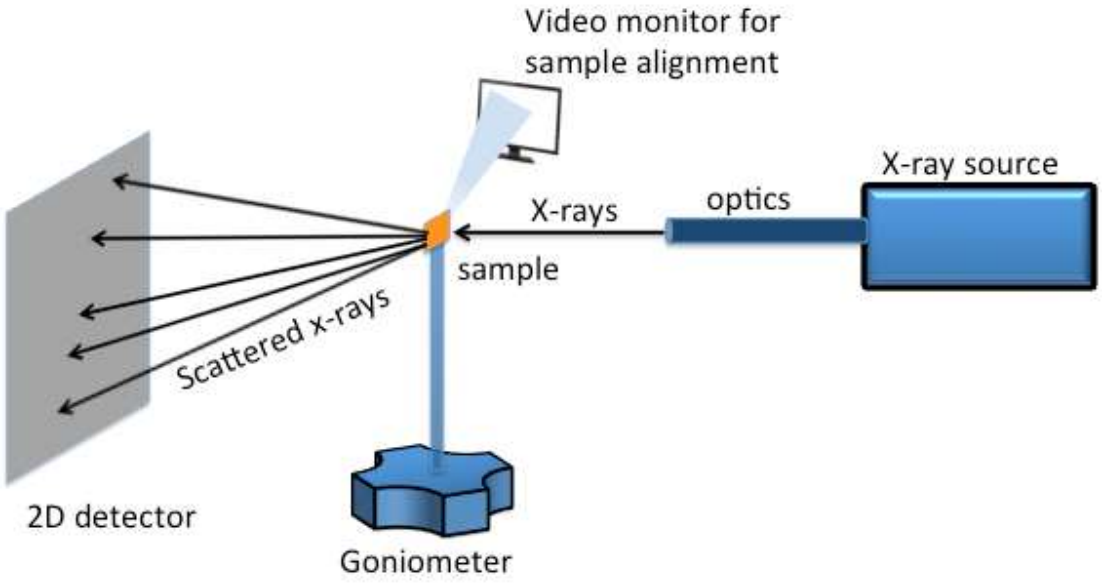


Corso di Biocristallografia e Microscopia Elettronica rdezorzi@units.it

X-ray diffractometer

Essential components:

- X-ray source in the 5-25 keV range
- Monochromator/filters and focusing system
- Goniometer
- Detector



Additional components:

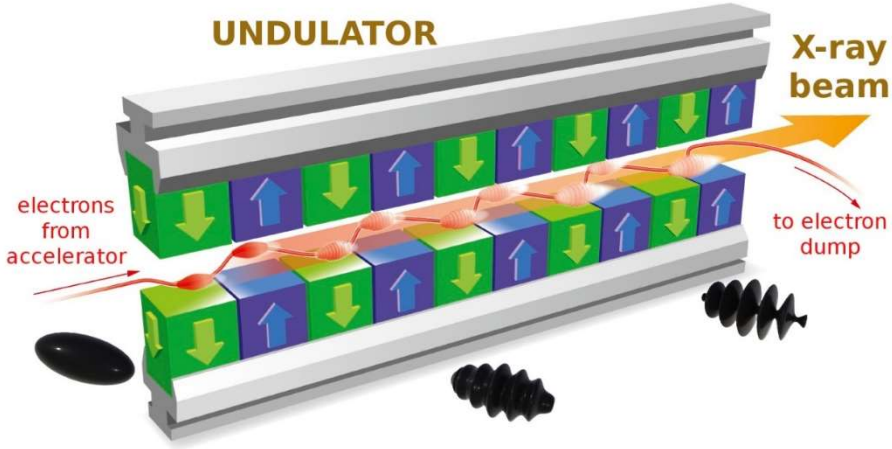
- Cryocooler
- Optical microscope for centering
- Fluorescence detector
- Automated sample changer
- ...

X-ray sources

X-ray sealed tube

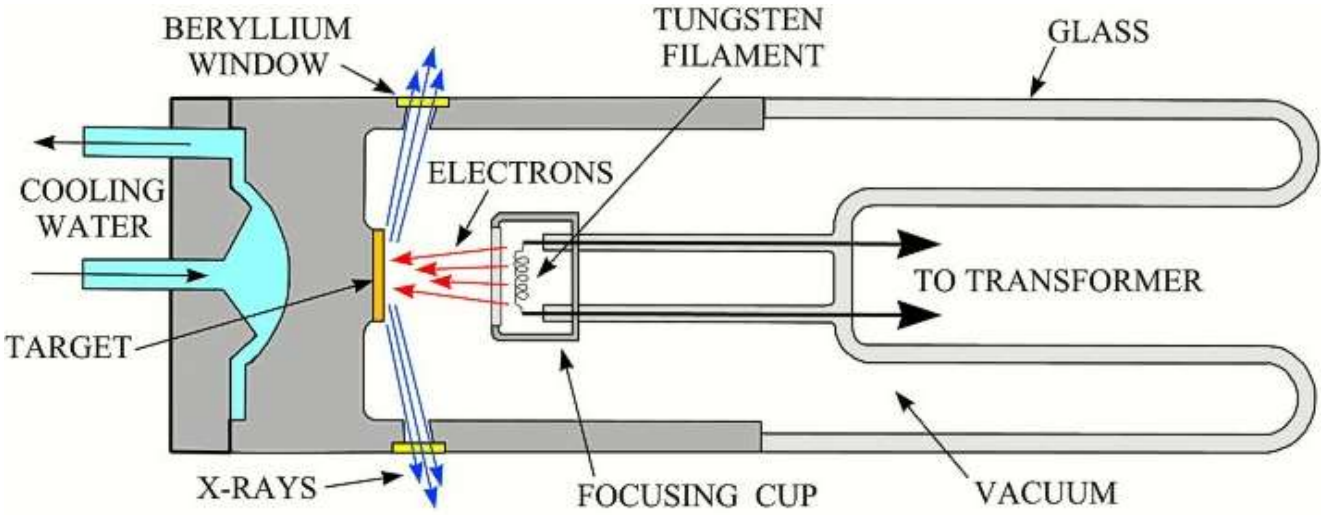
Synchrotron

Rotating anode tube



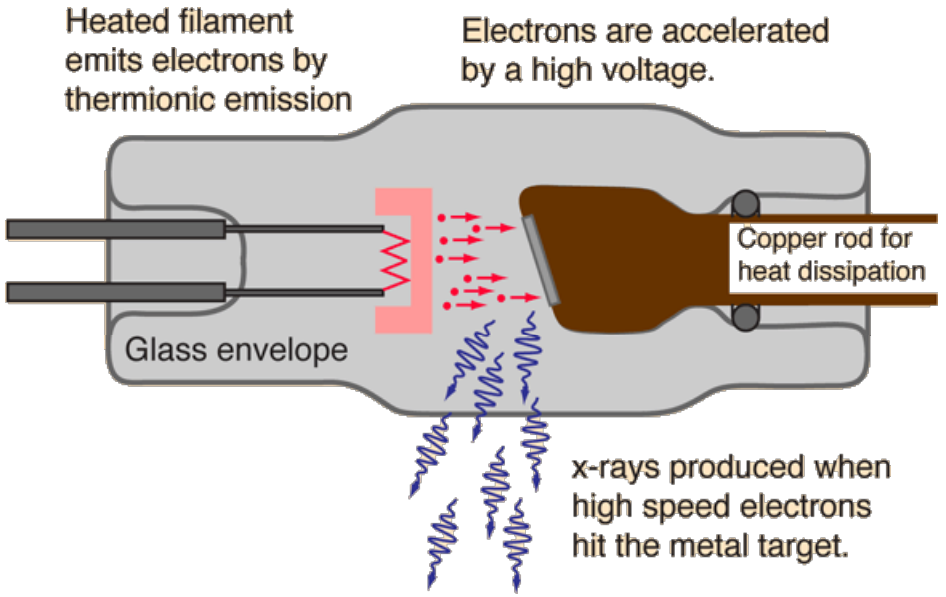
X-FEL
(X-ray Free Electron Laser)

X-ray tubes



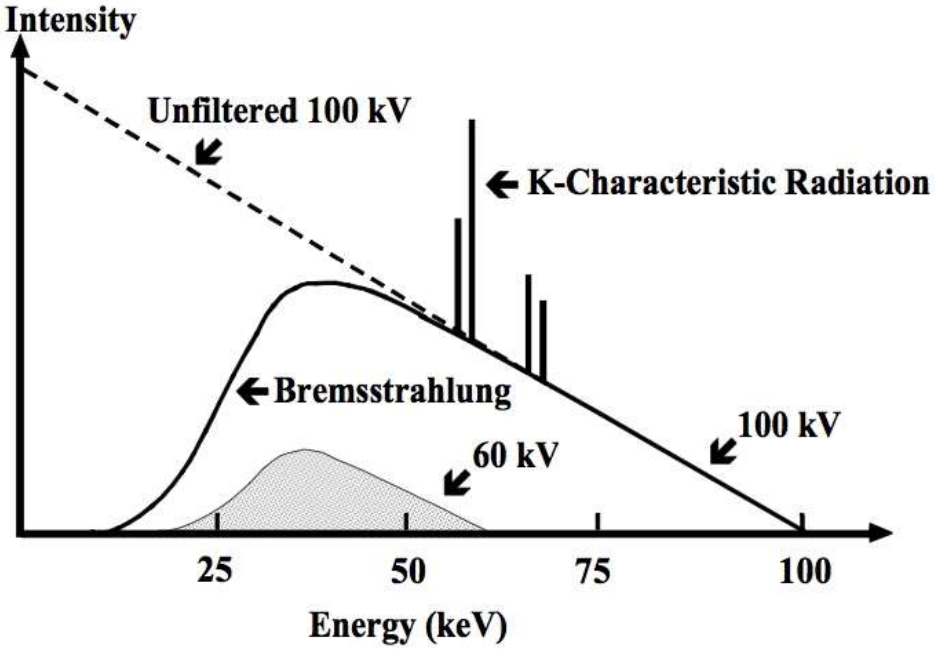
Material of the anode (target) characterizes the wavelength of the emitted radiation.

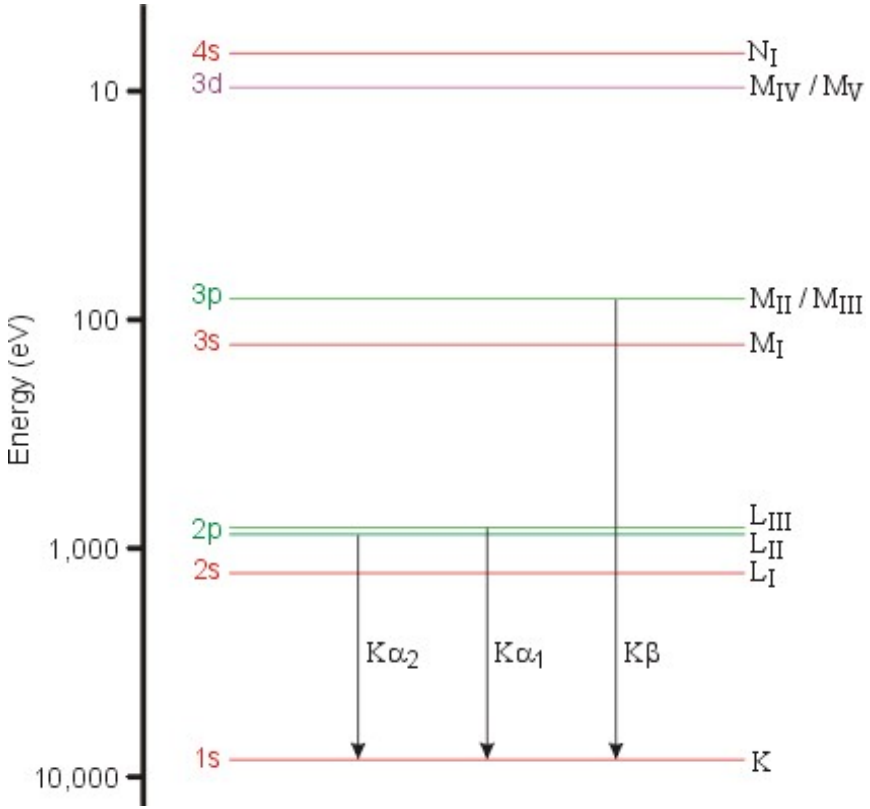
Intensity of the radiation depends on ΔV applied. Maximum intensity depends on efficiency of cooling system.

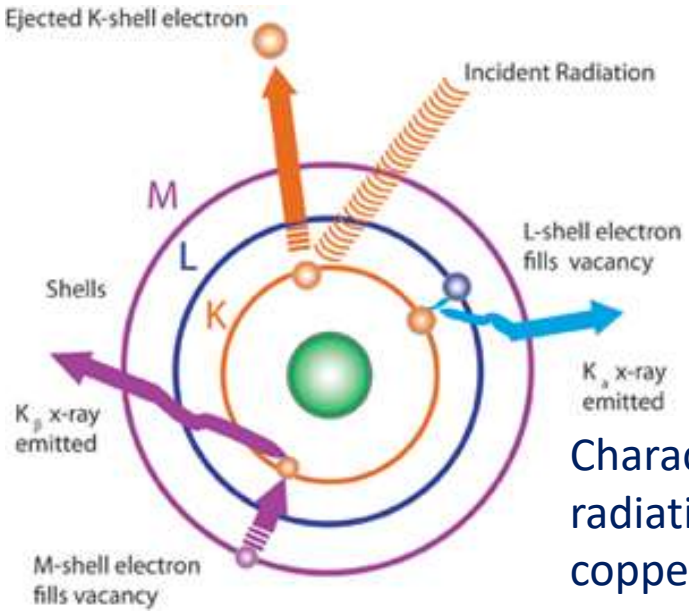


Bremsstrahlung radiation:
due to deceleration of incident electrons in the target metal (e.g. copper).

Characteristic radiation:
corresponding to electronic transitions, with wavelength characteristic of target metal.



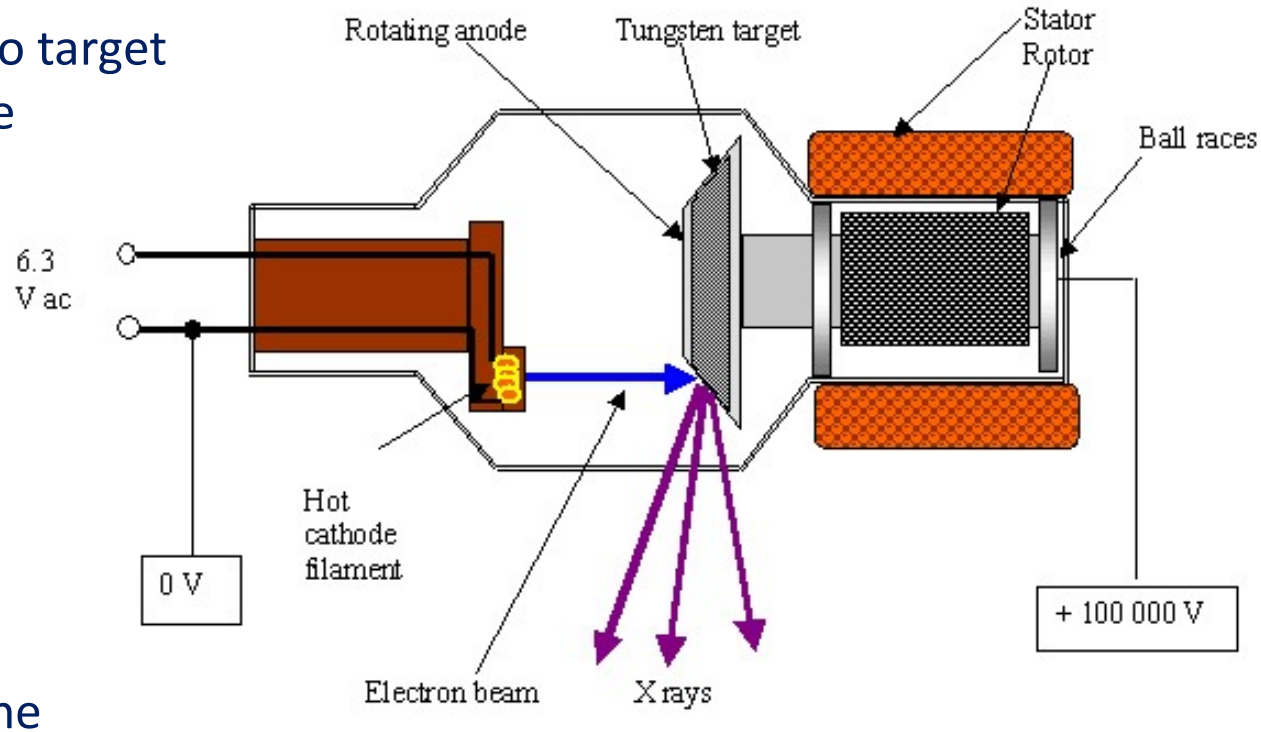




Characteristic radiation for a copper anode:
$K\alpha$ = 1.54184 Å
$K\beta$ = 1.39222 Å

Rotating anode

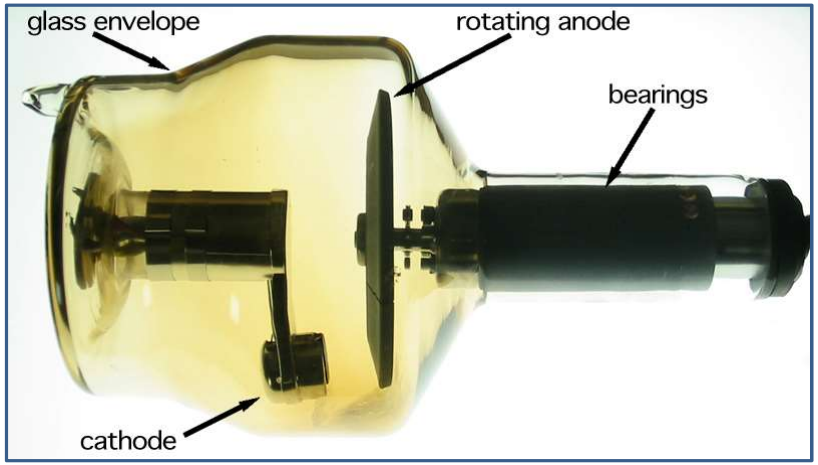
Increased intensity due to target rotation that allows more efficient cooling.



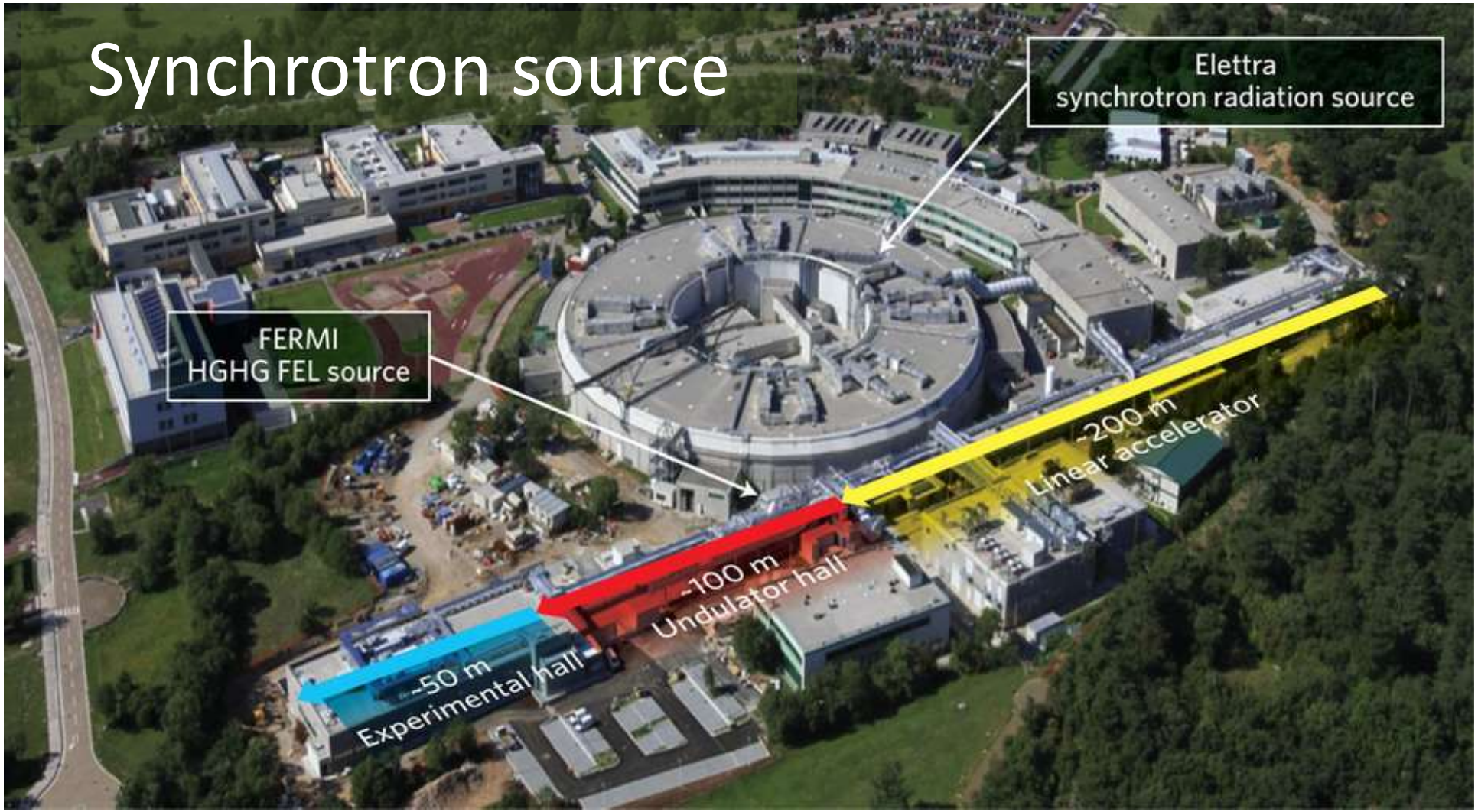
Higher intensity of the emitted radiation,

but…

1) more expensive
2) requires continuous pumping of vacuum

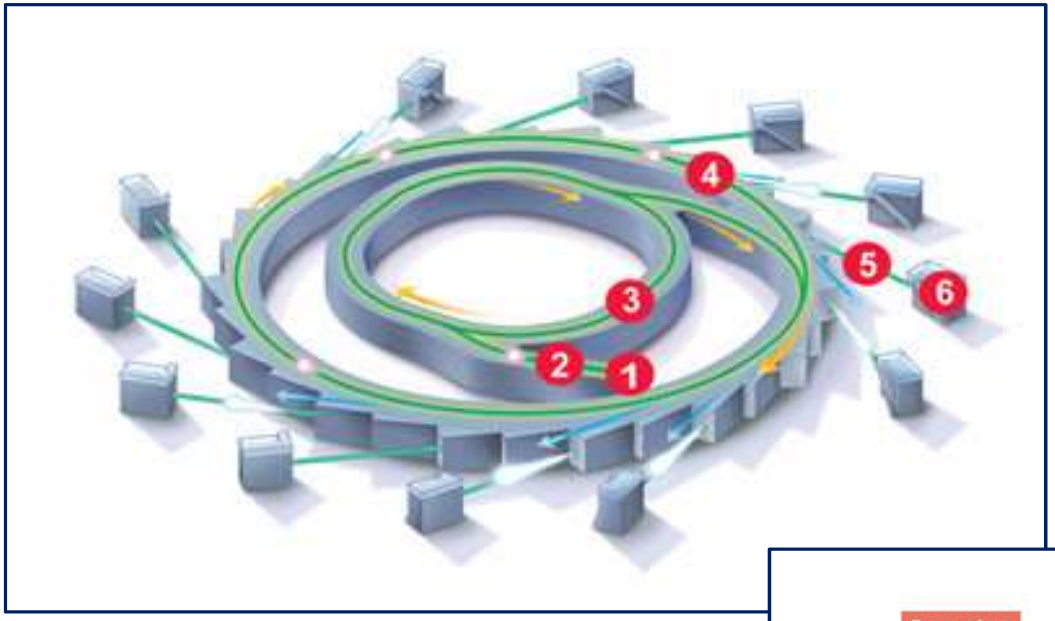


Synchrotron source



Advantages:

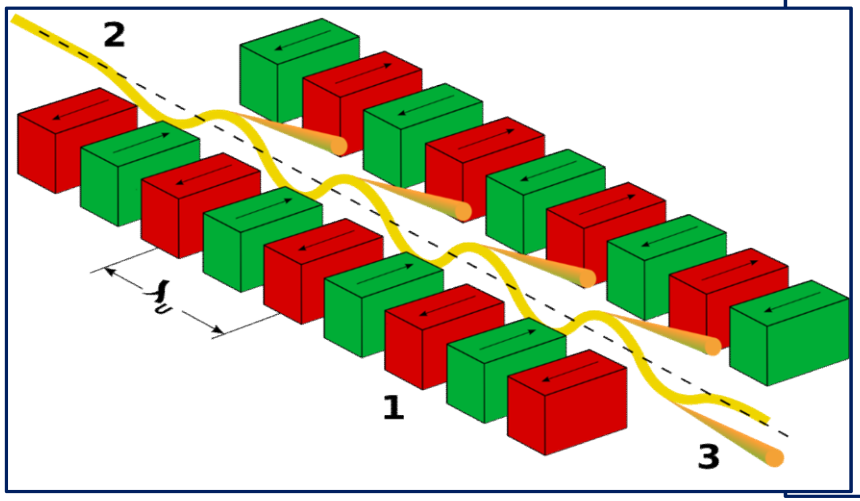
- High brilliance
- Tunable source
- Small beam divergence
- White beam
- Polarization

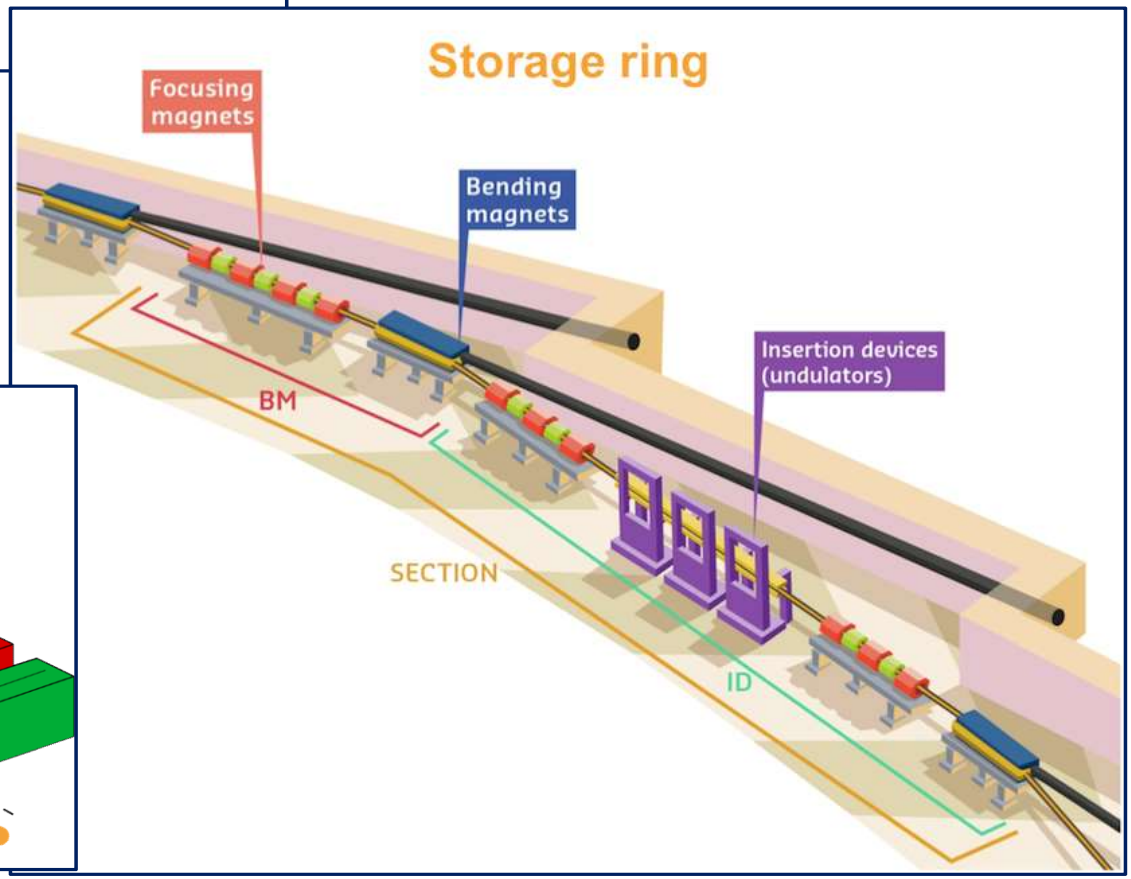


1. Electron gun
2. Linear accelerator (Linac)
3. Booster ring
4. Storage ring
5. Beamline
6. Experimental station

Radiation obtained from:

- Bending magnets
- Insertion devices: undulators and wigglers

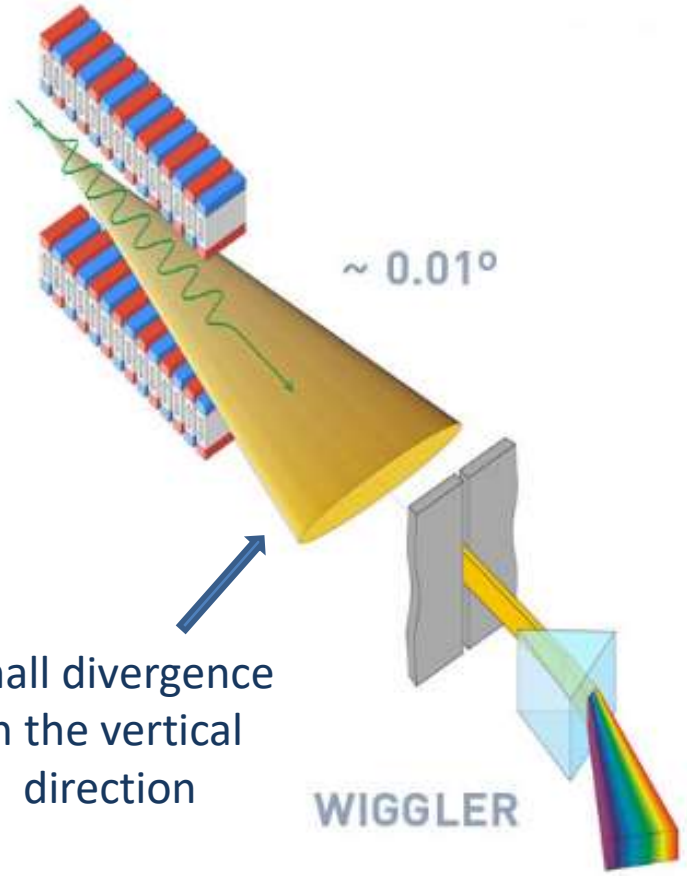




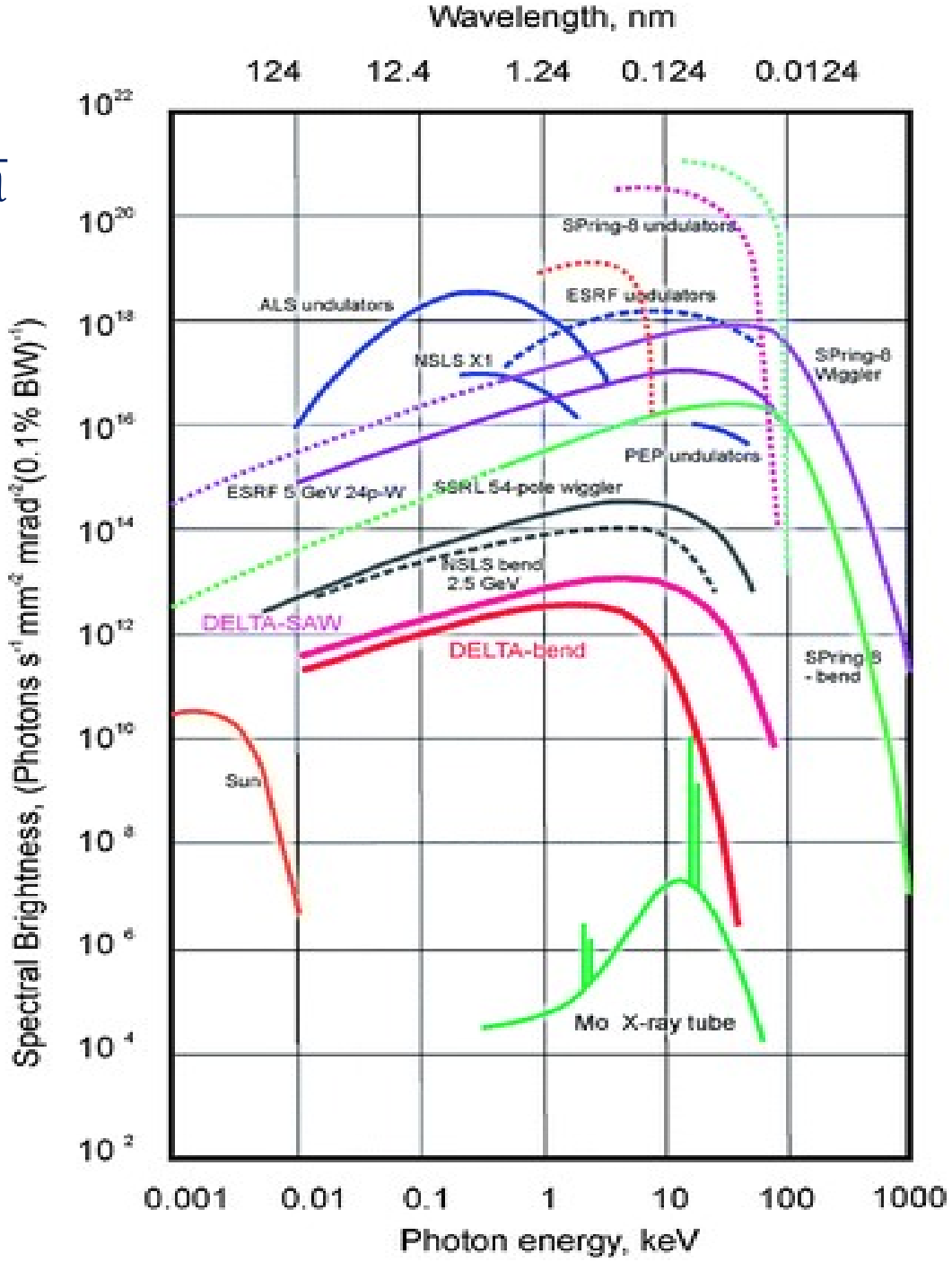
Brilliance

$$Brilliance = \frac{photons/s}{mrad^2 \cdot mm^2 \cdot 0.1 \cdot \delta\lambda/\lambda}$$

Brilliance: number of photons per second, per surface area, per dihedral angle, per bandwidth



Small divergence in the vertical direction



Optics

Monochromator / filters

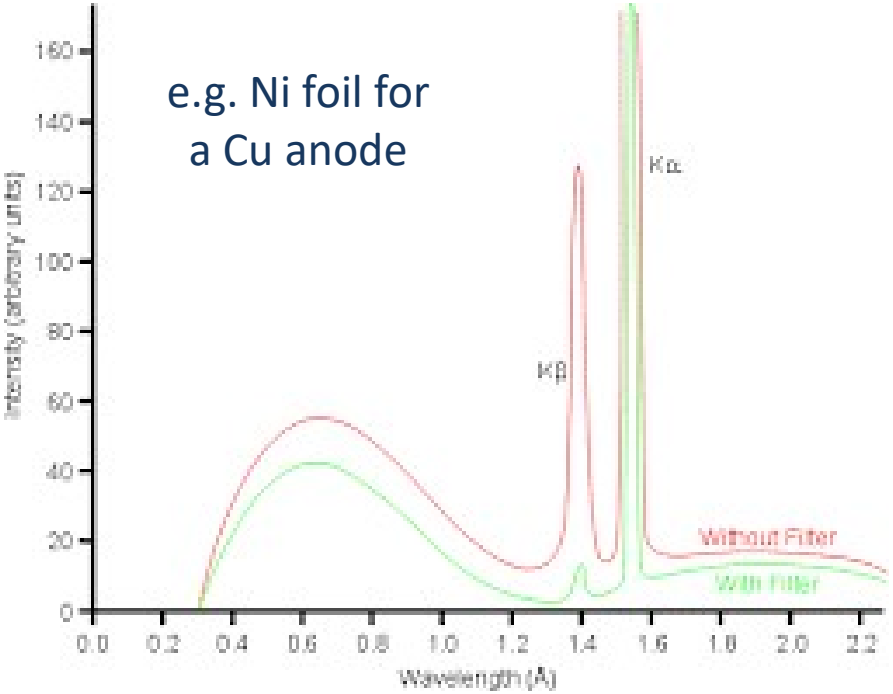
- for traditional sources: metal foil of element preceeding the anode metal
- for synchrotron sources: double crystal, e.g. Si(111)

Focusing mirrors / collimator

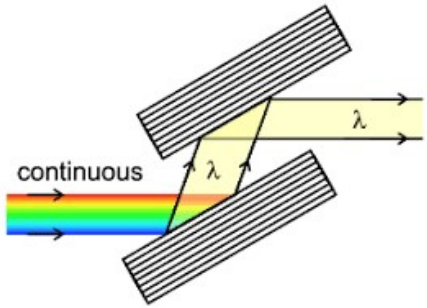
Be windows

Shutters

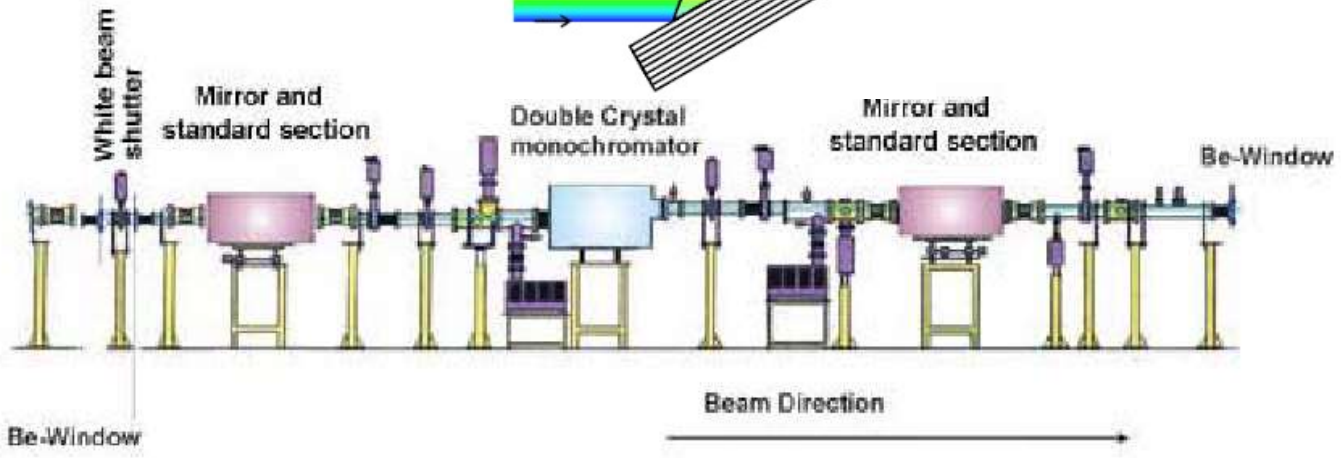
Slits



Resolution:
$10^{-4} - 10^{-5}$ Å



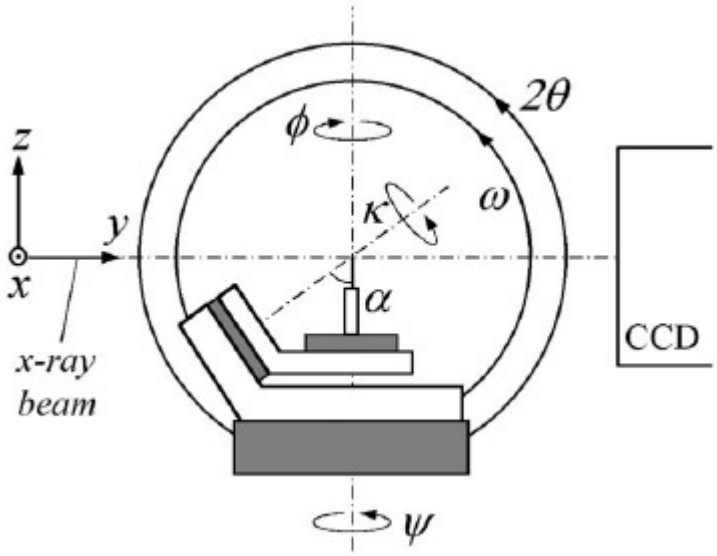
$\lambda = 2d \sin\theta$

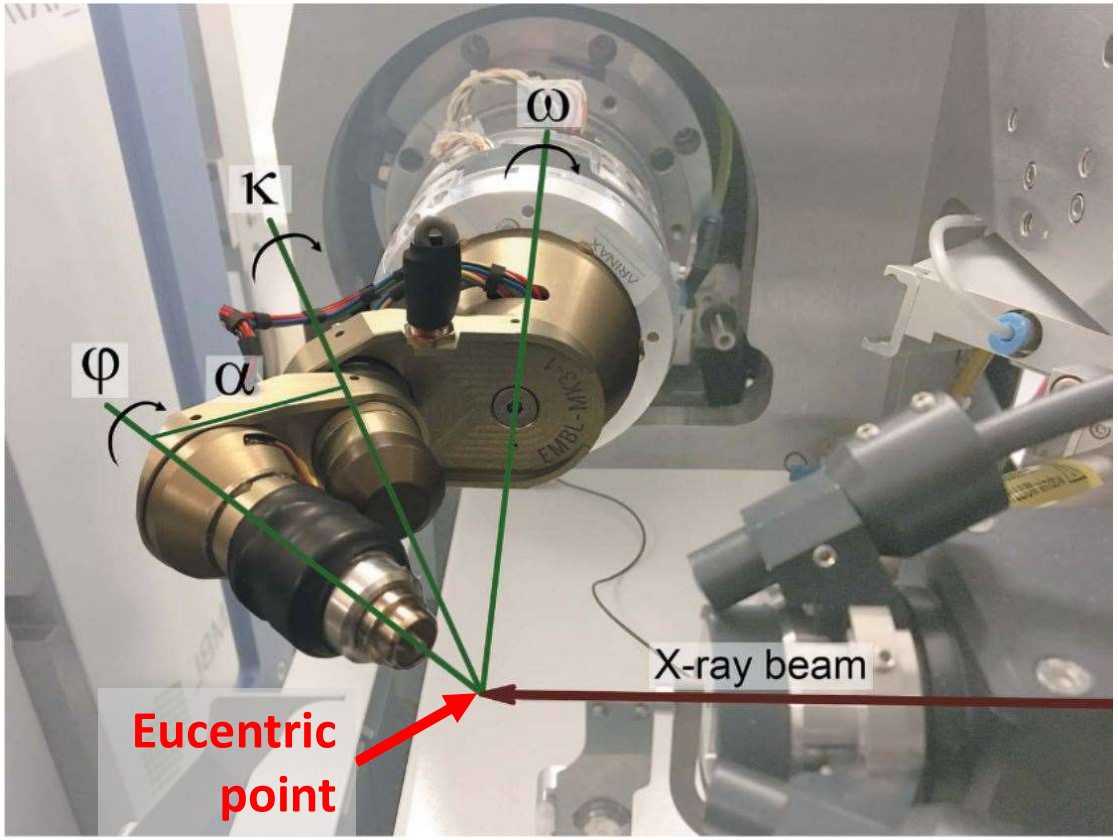


Goniostat

Single axis (φ): rotatation around a single axis, perpendicular to the incident beam

4-circles goniometer: 3 axes of rotation of the crystal and an additional rotational axis of the detector



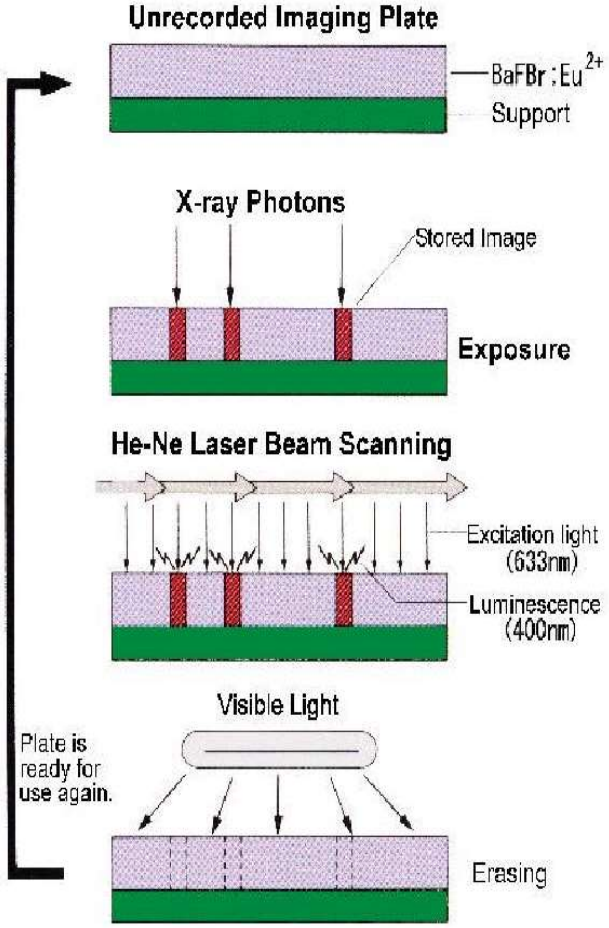


Detectors

Image plate:

(1) X-ray captured by phosphorescent material

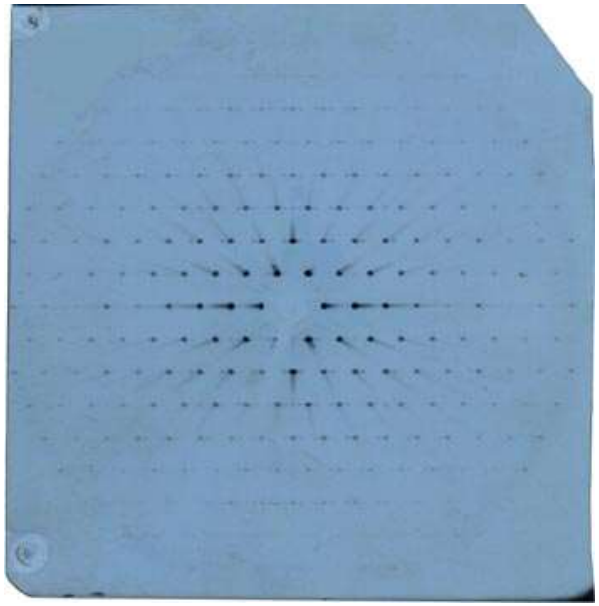
(2) Stimulated emission of visible light



Characteristics:

- Quantum efficiency
- Dynamic range
- Spatial resolution

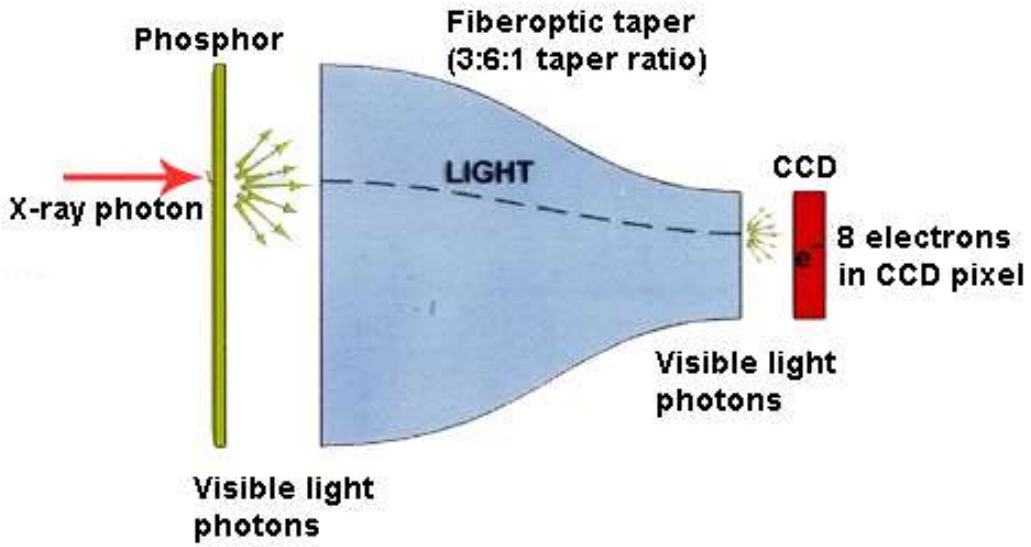
Photographic films



Charged Coupled Device (CCD):

(1) X-ray converted in visible light

(2) Excitation of CCD chip

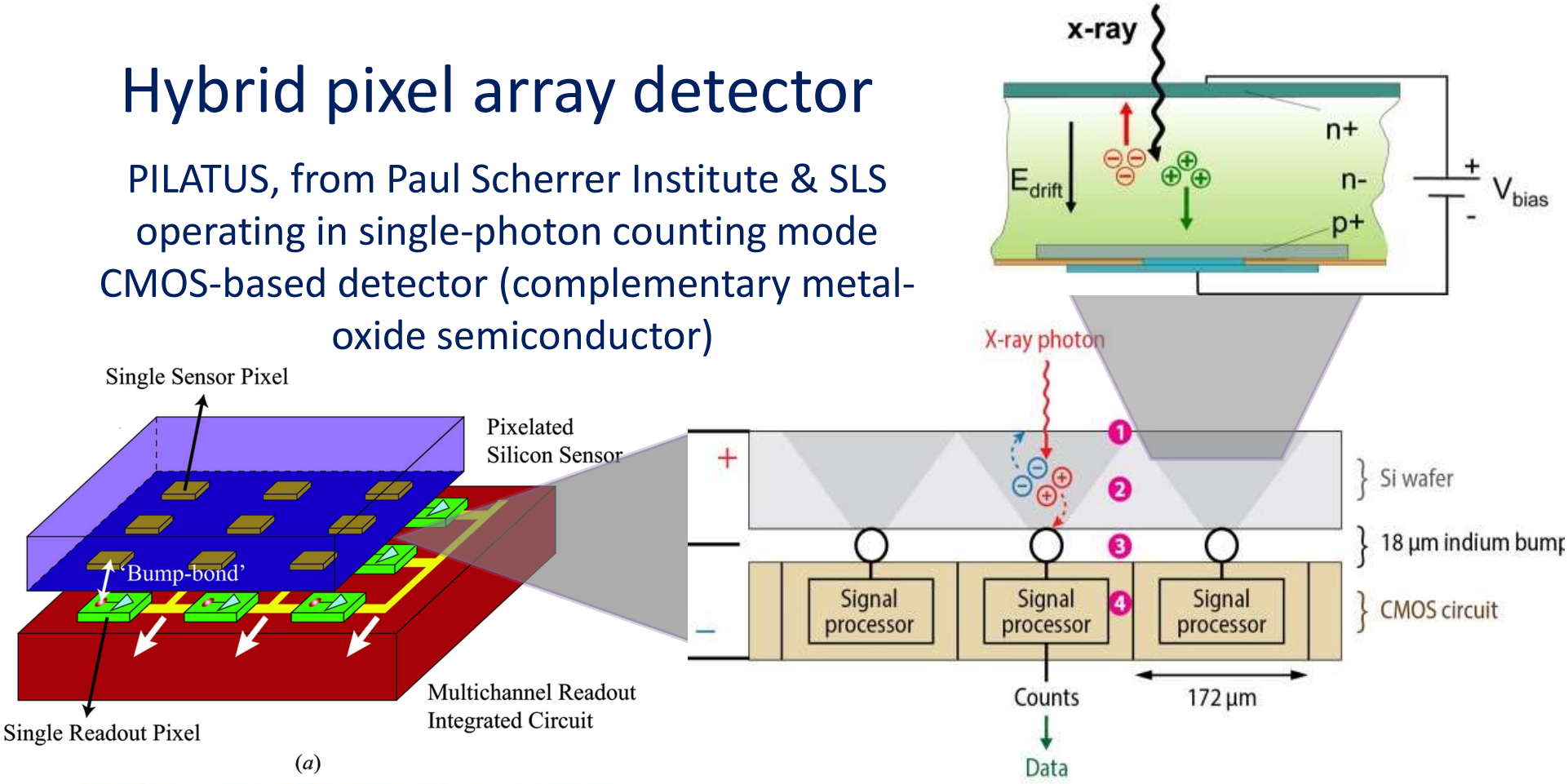


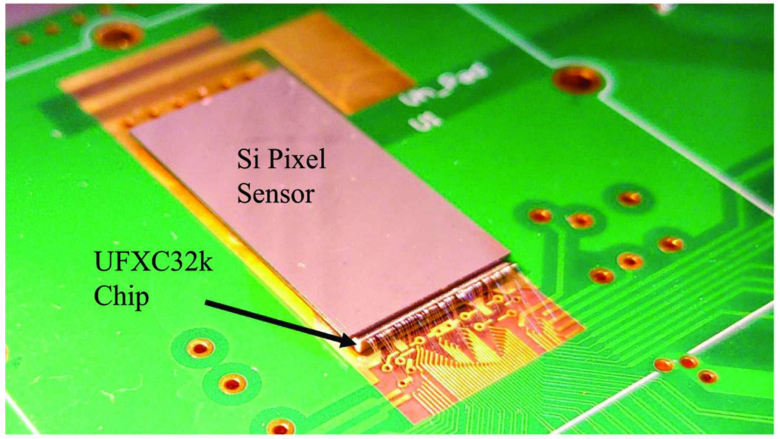
Hybrid pixel array detector

PILATUS, from Paul Scherrer Institute & SLS
operating in single-photon counting mode
CMOS-based detector (complementary metal-oxide semiconductor)

Hybrid pixel array detector

PILATUS, from Paul Scherrer Institute & SLS operating in single-photon counting mode CMOS-based detector (complementary metal-oxide semiconductor)





- Maximum dynamic range
- Excellent Point Spread Function (PSF)
- High Detector Quantum Efficency (DQE)
- No dark-current and readout noise
- Readout time <6.7ms
- Allows to collect fine-φ-sliced data

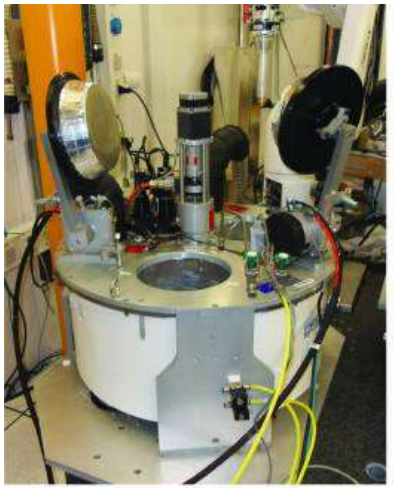
Cryocrystallography

Crystal freezing:

- **reduces radiation damage**
- **reduces thermal vibration**
- **reduces conformational disorder**
- **enhances signal-to-noise ratio**

1. Crystal mounted on the goniometer with a loop on a magnetic base and flash frozen under a stream of nitrogen gas at ≈100K.

2. Crystal pre-frozen in liquid nitrogen and kept in dewar. On the beamline, mounting with the automated system


(a)

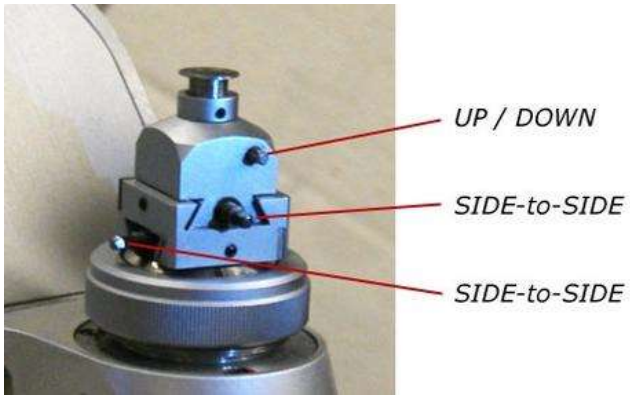

(b)


(c)

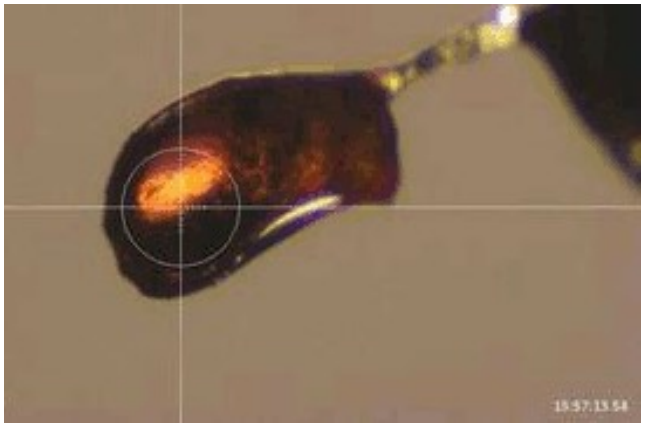
Crystal centering

Crystal should be in the beam path during all data collection (rotation). And ideally beam size should not excede crystal size to reduce background.

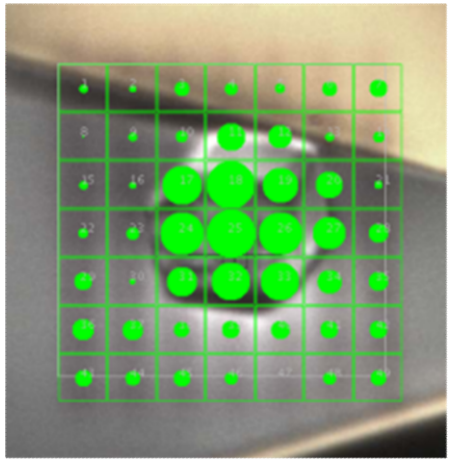
1. Manual centering: center crystal in 2 directions, 90° apart

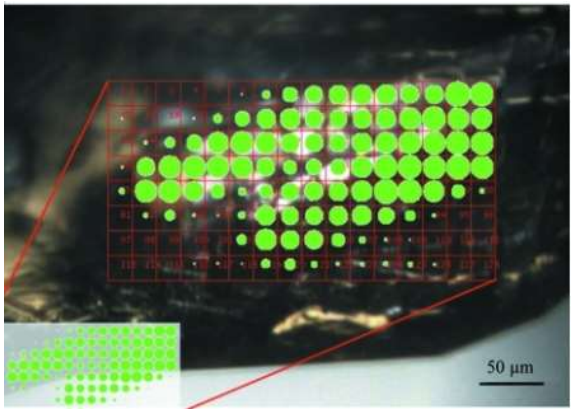


2. Automatic centering (@beamlines): 2- or 3- click procedures



3. Centering with gridding (@microfocus beamlines)





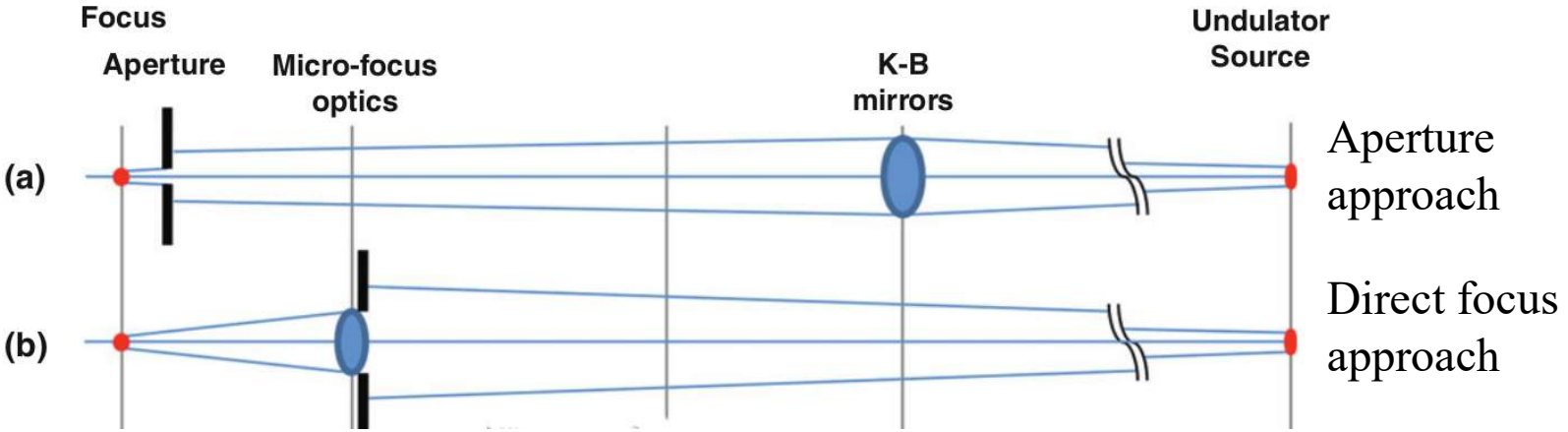
Ideal for centering of small, hard to distinguish crystals, but requires very small beam compared to crystal size.

Microfocus beamlines

Beamlines with beamsize of 1-20μm:

- Ideal for small crystals
- Low background scattering from solvent/loop
- High Signal-to-Noise Ratio (SNR)
- Possible selection of well diffracting region in larger crystals & helical data collection strategy
- "Grid-scan" for crystals difficult to visualize
- *In-situ* diffraction

Beamlines:
ID13 & ID23-2 @ESRF
ID24 @Diamond
23ID-B & D @APS
24ID-E @APS
BL32XU @SPring8
X06SA @SLS
PX2 @Soleil



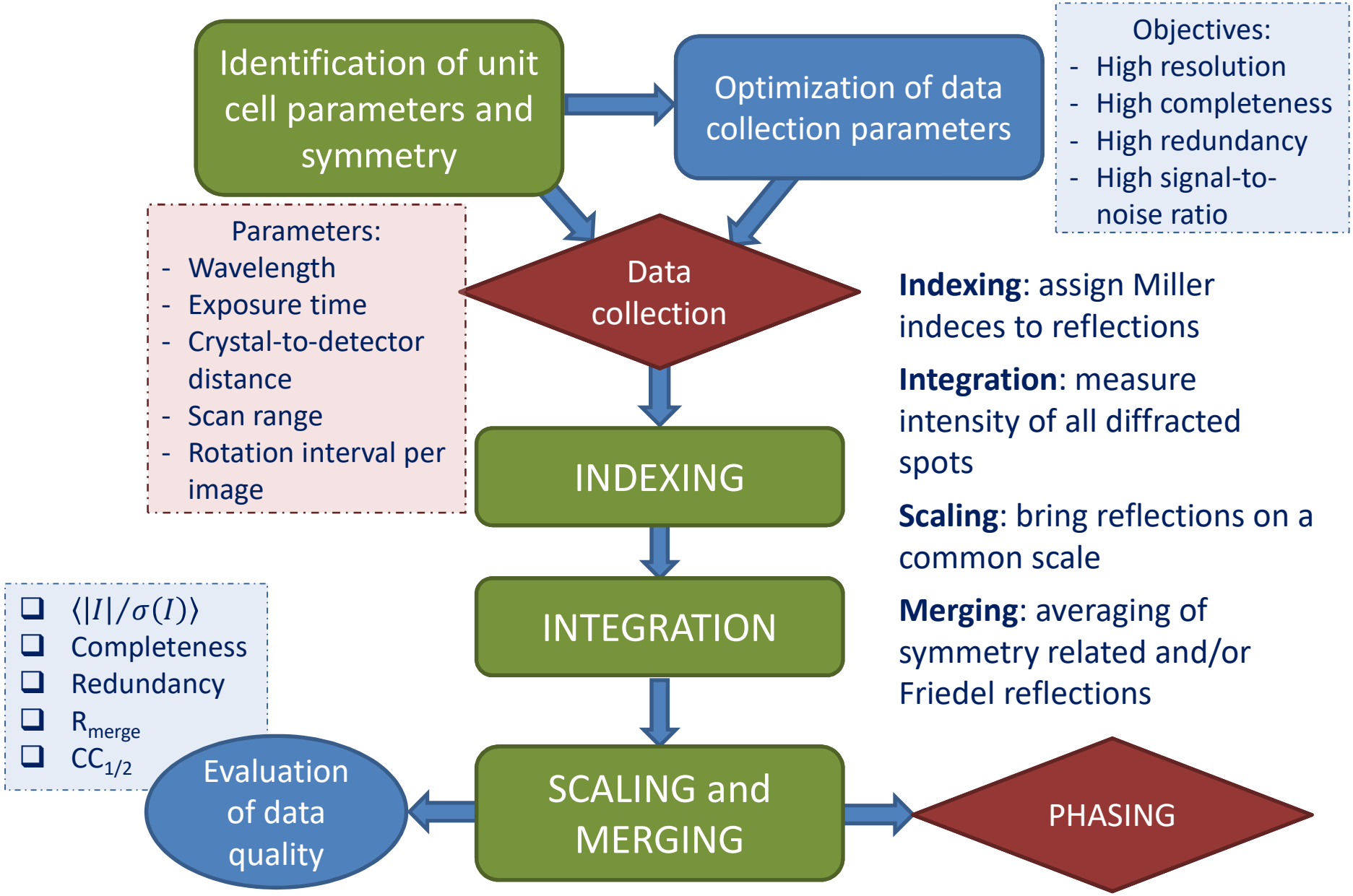
Requirements:

- High brilliance photon source (ondulator)
- Highly stable beam (Top-up mode)
- Thermally stable optics & endstation
- No vibrations on optics
- High precision goniometer

Problems:

- Higher radiation damage on crystals (overcome by helical data collection strategy)
- Usually requires screening for small crystals

X-ray diffraction experiment workflow



X-ray diffraction experiment

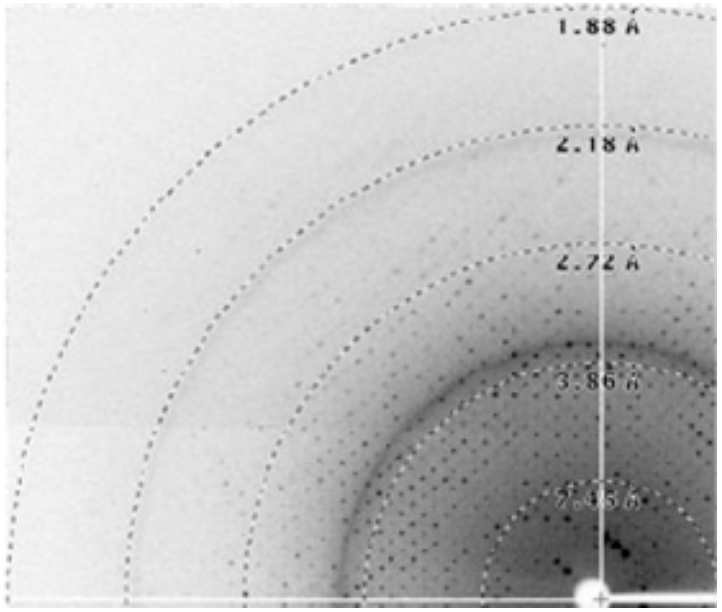
1. Select wavelength (i.e. dimension of Ewald sphere...).
2. Mount crystal on diffractometer. Center.
3. Check crystal quality by taking diffraction images at φ = 0° and at φ = 90° (check anisotropy)

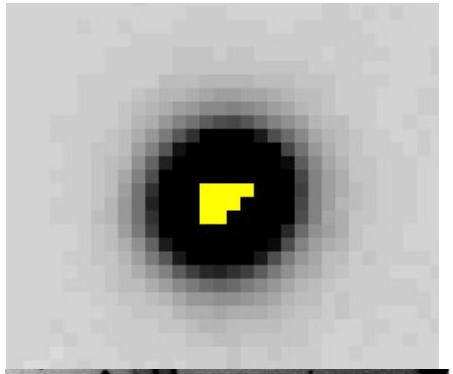
   -> is it salt? Is it a single crystal?

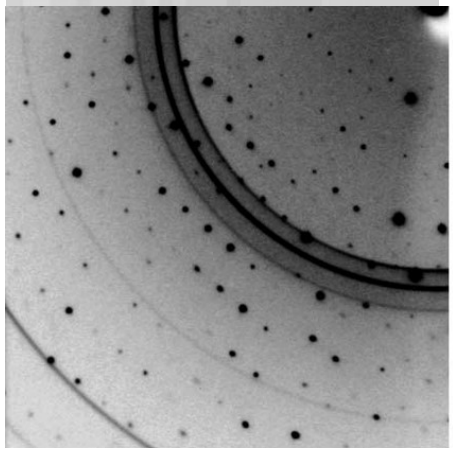
   -> from initial images evaluate **resolution limit** and crystal-to-detector distance: *smaller distance for high resolution, but check sovrapposition of spots in the diffraction pattern. Larger unit cells have small reciprocal space cells, and very close diffraction spots.*

   -> evaluate exposure time: *longer exposure allows to collect stronger high resolution reflections, but check overloads in the low resolution spots; longer exposure causes more radiation damage.*

   -> check presence of ice rings



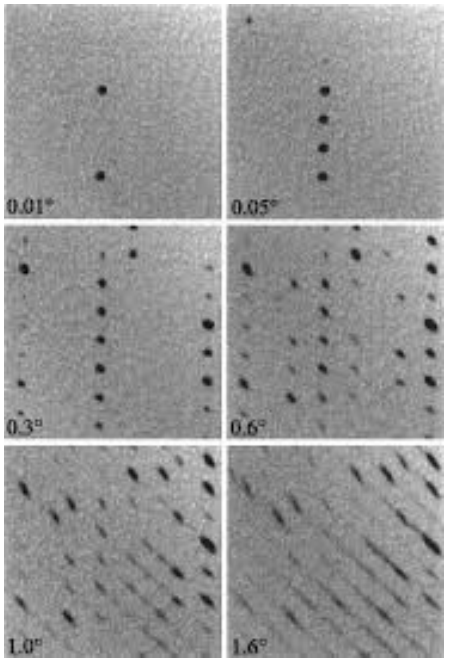




Low resolution data are important for phasing!

3. Collect further images to determine **unit cell parameters, symmetry and crystal orientation** (≈10 images, $\Delta\varphi$ = 1°/image)

   -> evaluate **mosaicity** of the crystal (software): larger reciprocal lattice nodes in diffraction conditions for more than one image

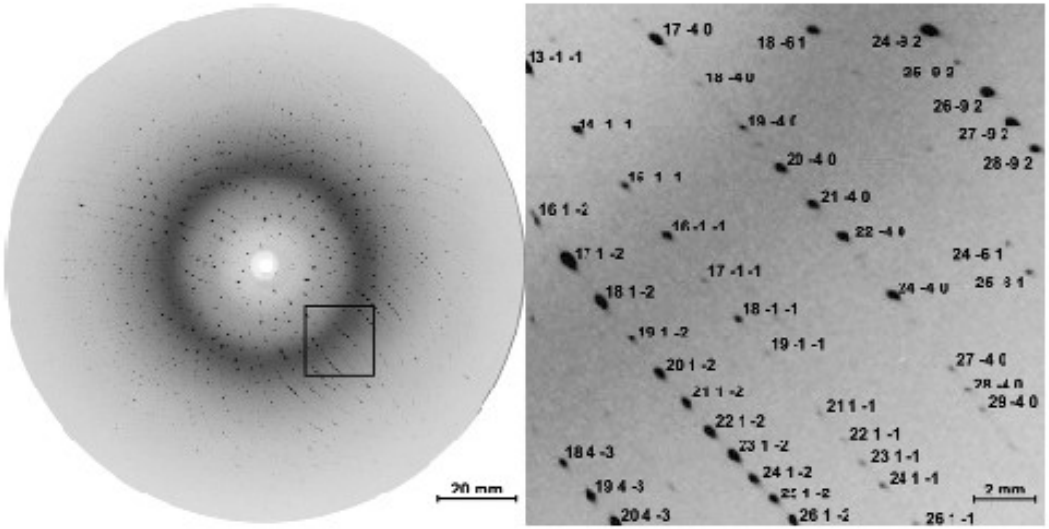


4. **Indexing**: analyze images to identify

(a) crystal orientation,
(b) unit cell parameters,
(c) symmetry (Laue group),
(d) and assign to each reflection the Miller indexed of the family of planes

using knowledge of

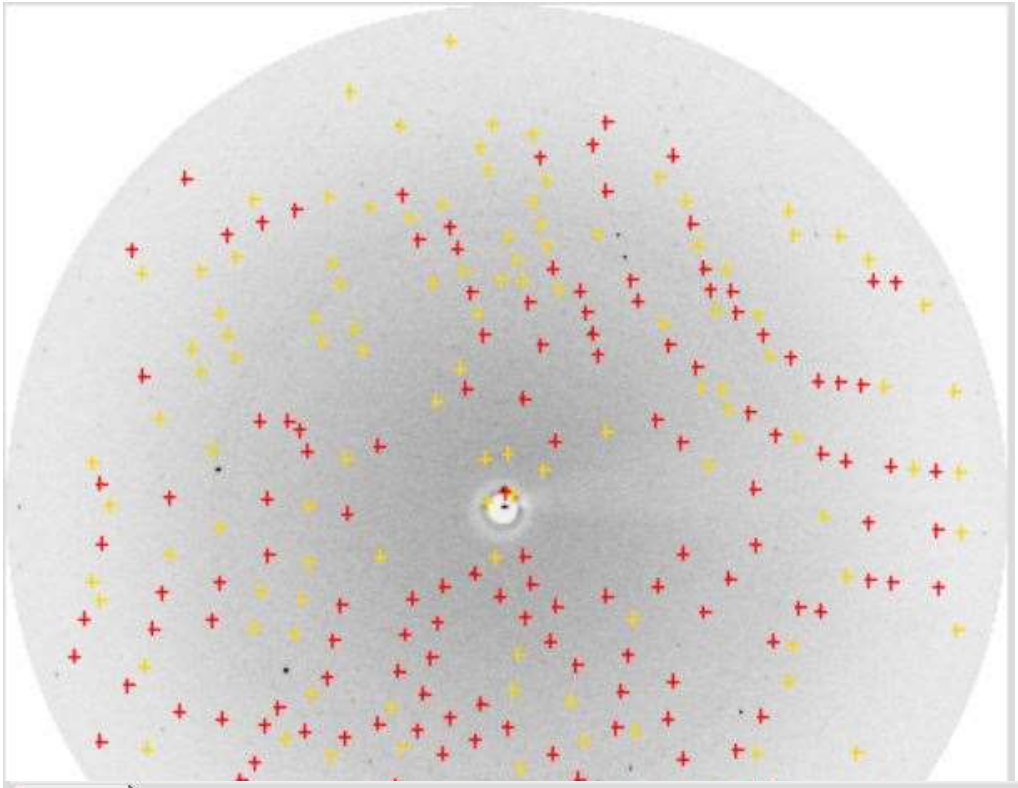
(1) wavelength,
(2) crystal-to-detector distance
(3) beam position



Choice of the unit cell:

(i) Highest symmetry

(ii) Smallest cell



Indexing

(Indexing with Mosflm)

Mosflm finds the diffracted spots.

From the diffracted spots, a list of possible unit cells is obtained. Mosaicity is estimated.

The highest symmetry with low penalty is selected.

Lattice 1

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	Nref	δ beam
1 (ref)	aP	0	58.6	58.6	62.1	90.0	118.1	120.0	0.10	250	0.49 (0.1)
2 (ref)	aP	0	58.7	58.6	62.1	61.8	61.8	59.9	0.10	246	0.49 (0.1)
3 (ref)	mC	1	101.6	58.6	62.1	90.0	123.0	90.0	0.11	250	0.47 (0.2)
4 (ref)	mC	1	101.6	58.6	62.1	90.0	123.0	90.0	0.11	250	0.47 (0.2)
5 (ref)	mC	2	101.4	58.6	62.1	90.0	123.0	90.0	0.09	245	0.48 (0.2)
6 (ref)	mC	2	101.4	58.6	62.1	90.0	123.0	90.0	0.09	245	0.48 (0.2)
7 (ref)	hR	4	58.6	58.6	156.4	90.0	90.0	120.0	0.12	255	0.48 (0.3)
8 (reg)	mC	59	101.4	58.6	62.0	90.0	122.9	90.0	-	-	-
9 (reg)	mC	59	101.4	58.6	62.0	90.0	122.9	90.0	-	-	-
10 (reg)	hR					90.0	90.0	120.0	-	-	-
11 (reg)	mC					90.0	133.3	90.0	-	-	-

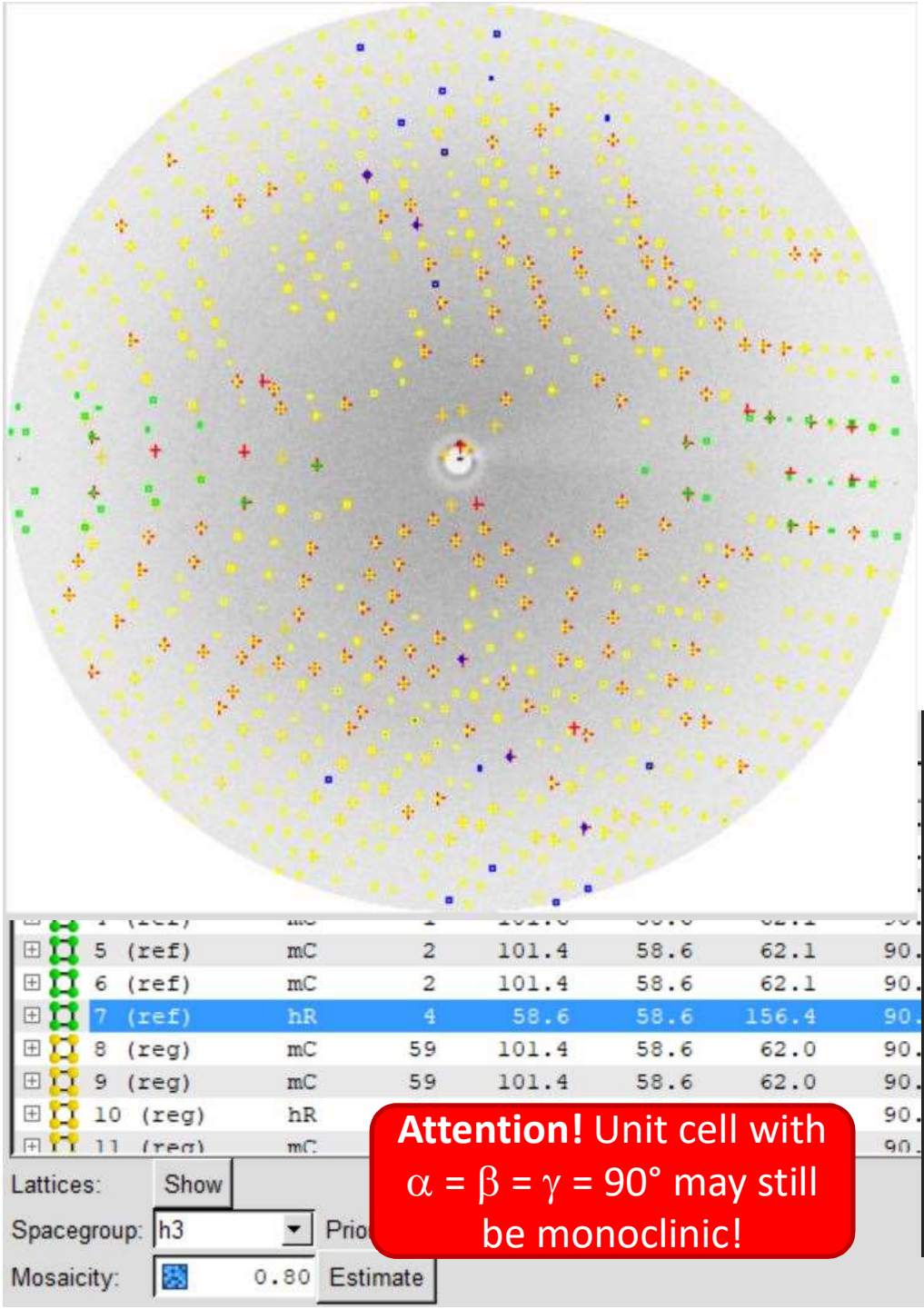
Attention! Unit cell with $\alpha = \beta = \gamma = 90°$ may still be monoclinic!

Lattices: Show

Spacegroup: h3 Prio

Mosaicity: 0.80 Estimate

Start beam search Show



Indexing

(Indexing with Mosflm)

Mosflm finds the diffracted spots.

From the diffracted spots, a list of possible unit cells is obtained. Mosaicity is estimated.

The highest symmetry with low penalty is selected.

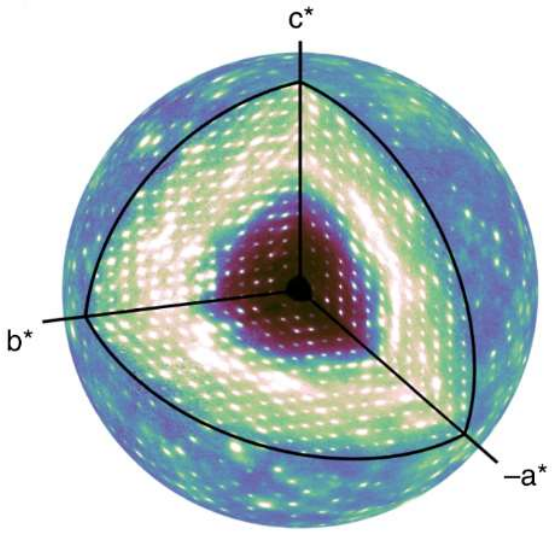
Expected diffraction spots are compared to experimental spots.

□ Full reflections (from nodes of the reciprocal lattice that have been completely crossed the Ewald sphere in this image)

□ Partial reflections (nodes that partially crossed the Ewald sphere)

□ Reflections in the Lorentz exclusion region (nodes in reflection conditions for very long time due to their proximity to the rotation axis)

5. Data collection **strategy**: data collection has to cover all asymmetric unit in the reciprocal space (φ_{start}-φ_{end} depend on symmetry...). If anomalous signal needs to be collected, reciprocal asymmetric unit is NOT centrosymmetric.



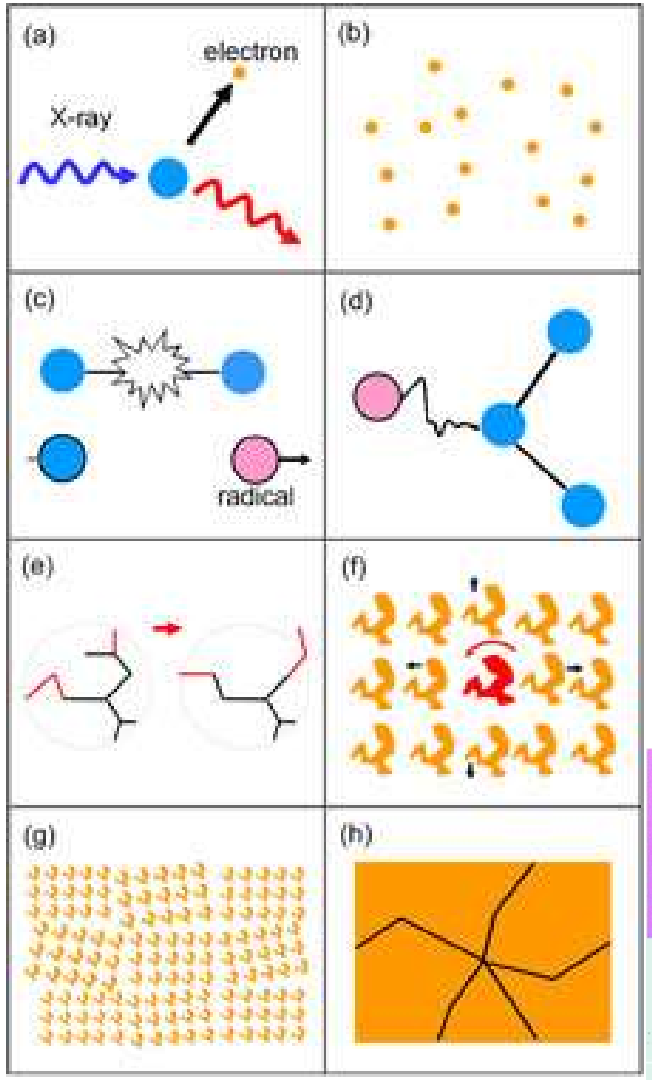
Use maximum precaution: if in doubt, consider lowest symmetry and collect more images.

*Software that optimize strategy for data collection identify the best way to collect the highest amount of unique data (not related by symmetry or Friedel) in the first images and suggest an optimal φ interval to increase **completeness**.* → Ideally, completeness higher than 90%

Identify the best $\Delta\varphi$ to avoid overlap.

Small $\Delta\varphi$ (e.g. 0.1°) may be useful to integrate the spots in 3D: evaluate the shape of nodes in the reciprocal space as they cross the Ewald sphere, by analyzing together subsequent images. In this case, all reflection appear as partials. → Fine slicing

→ **Strategy: wavelength, exposure time, crystal-to-detector distance, $\Delta\varphi$ rotation angle per image, overall range of φ** → Collect data!!



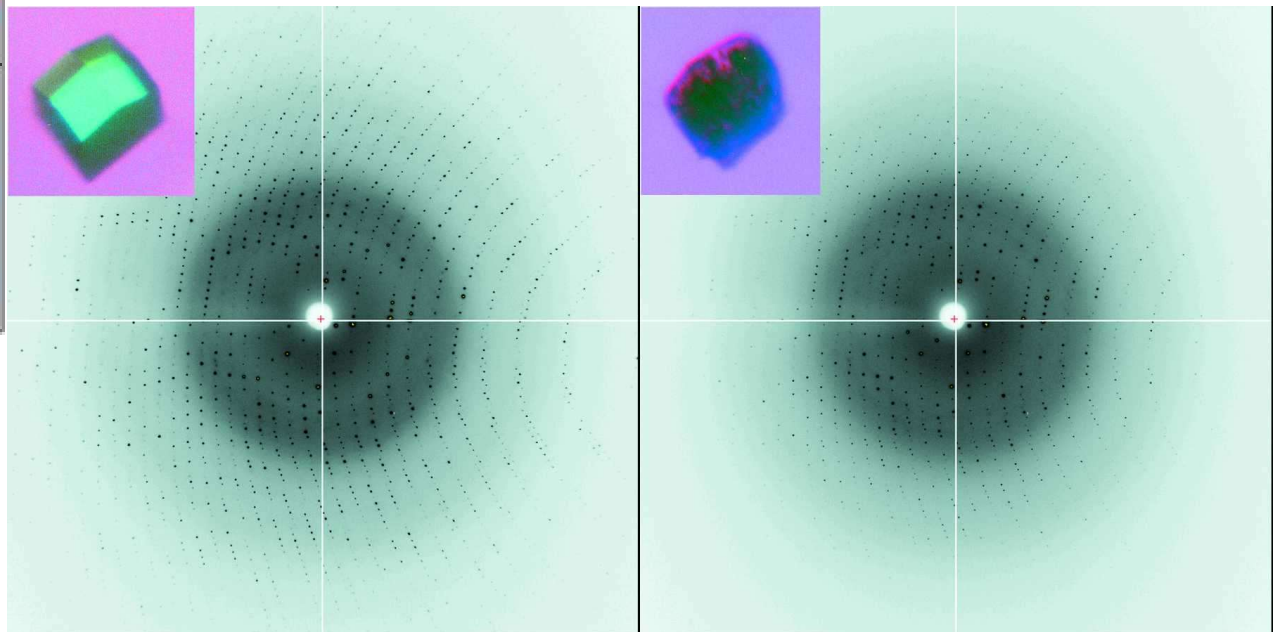
Radiation damage

Caused by interaction of X-ray with protein present in the crystal: secondary electrons generated by X-ray radiation create radicals, break bonds, oxidize species, up to major rearrangements of proteins in the crystal lattice.

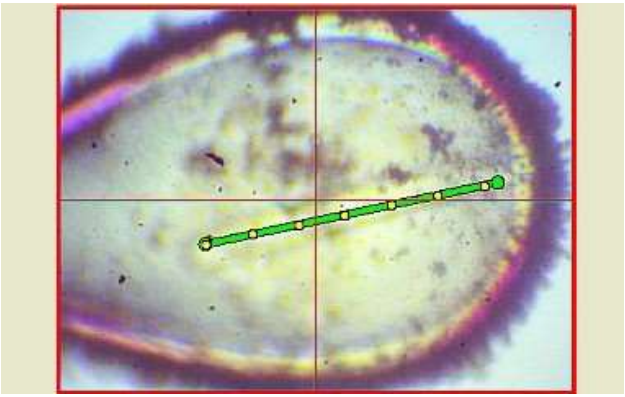
Results: loss of resolution, increase in B-factors, increase in mosaicity, increase in unit cell volume.

Initial image

Final image

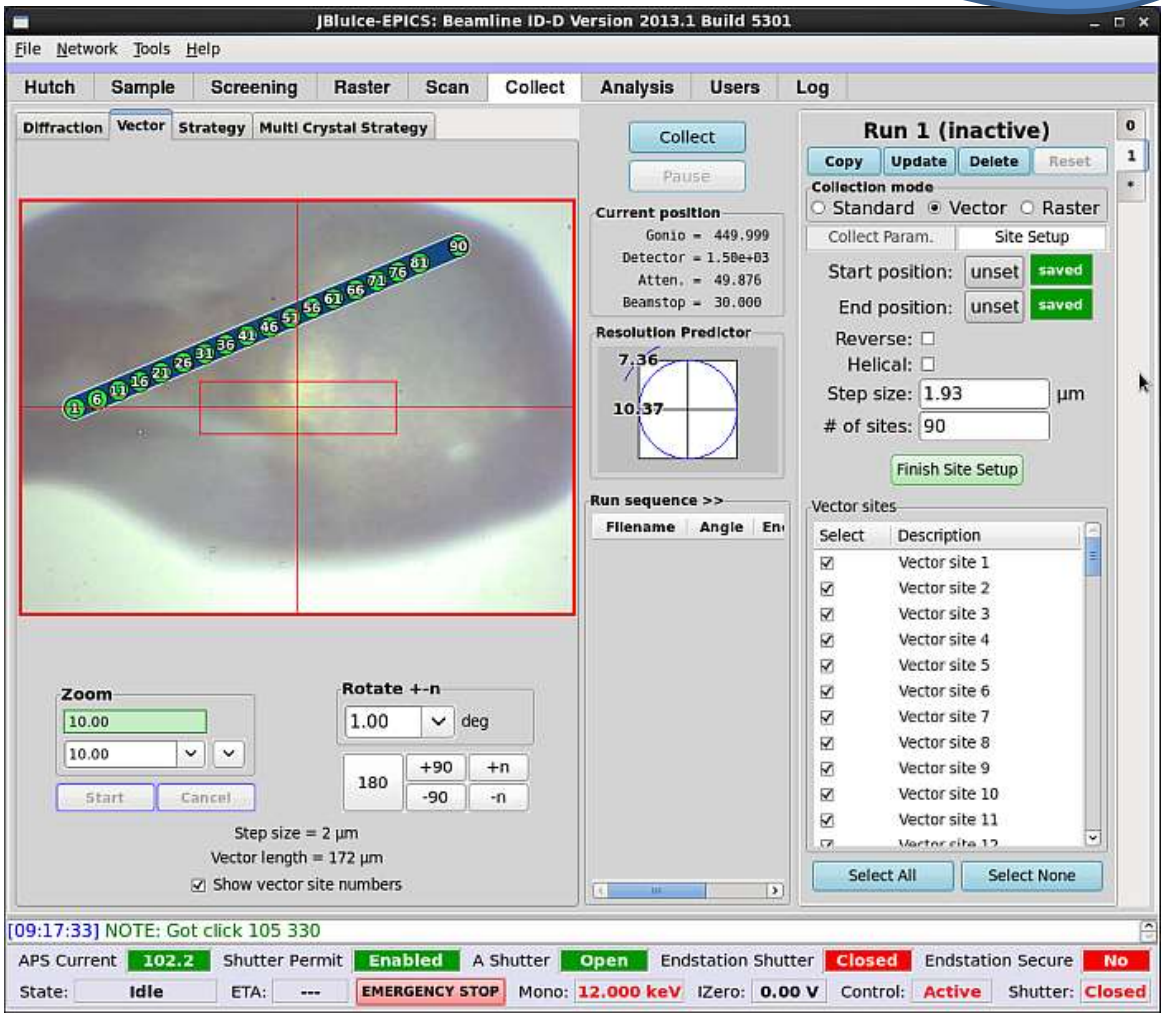


Higher reactivity:
metal ions > disulfide bonds > -COOH groups > -OH groups > S-CH_3 groups



Vector (or helical) data collection

Only for microfocus beamlines



Data collection strategy for radiation sensitive crystals.

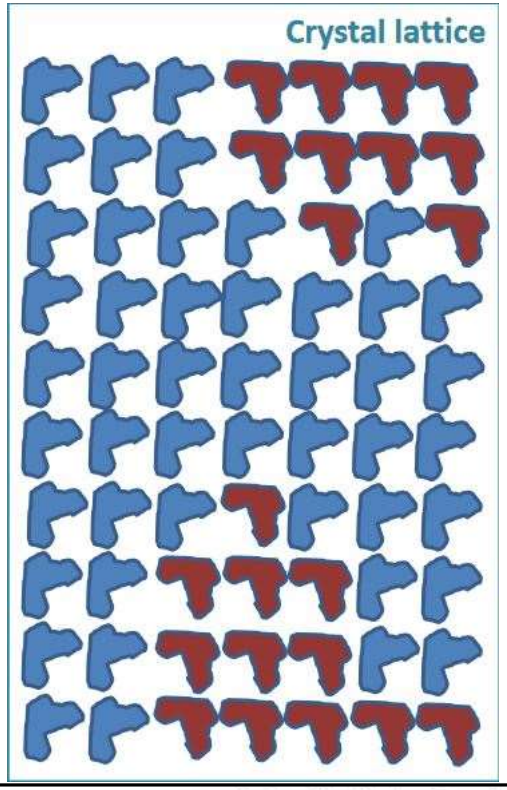
The user defines two points (start & end) and the data collection is performed while crystal is automatically translated between them.

At each frame, fresh sample moves into the beam, allowing longer exposure times per image and significantly reducing radiation damage.

Ideal for rod shaped crystals.

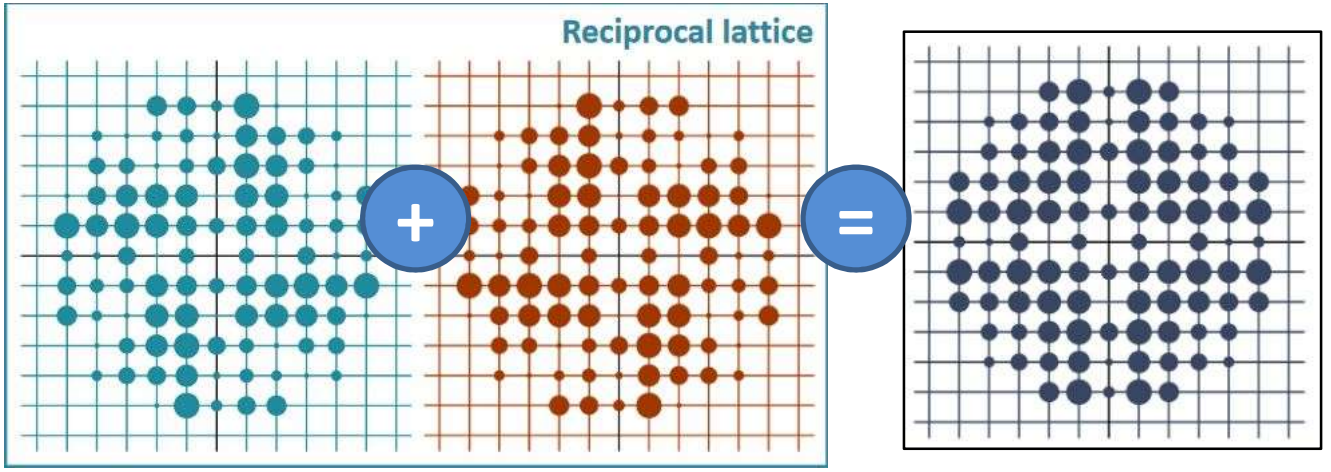
Better B-factors profile and better statistics.

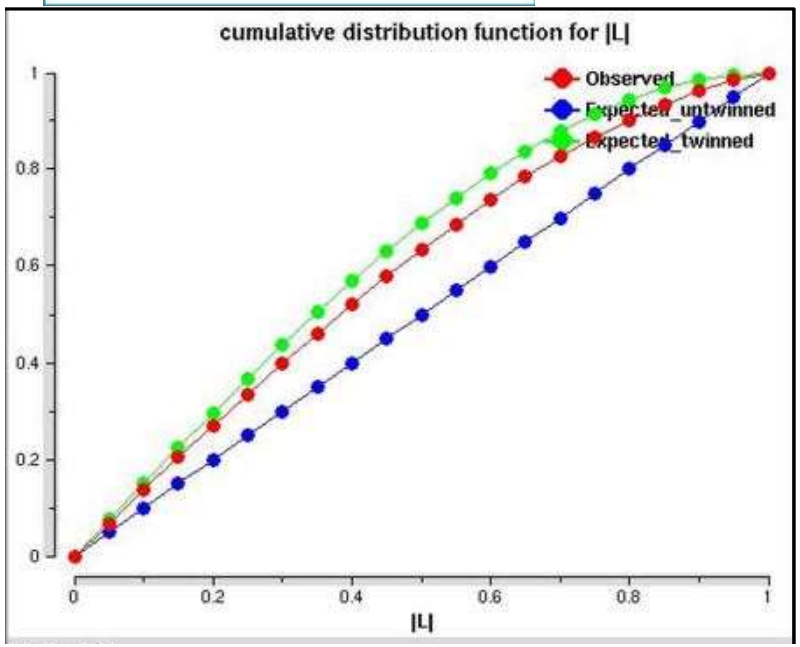
GM/CA CAT @APS



Twinning

Growth defect on crystals: presence of multiple crystal lattices in different domains within the same crystal.



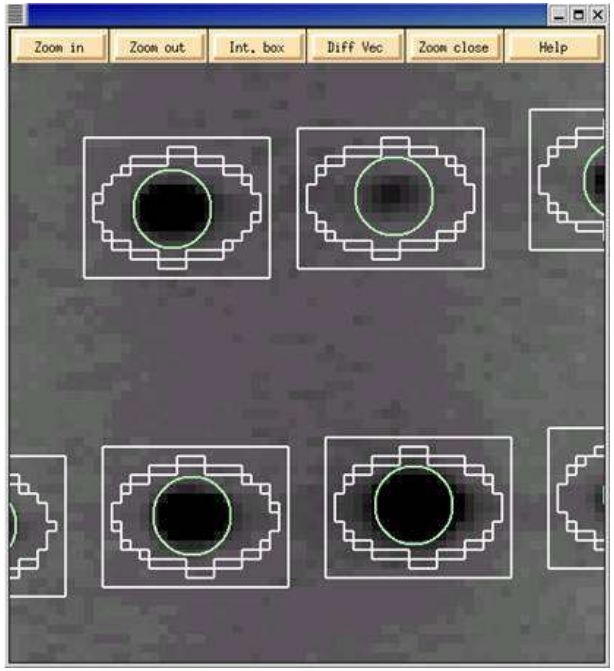


In merohedral and pseudomerohedral twinning, the twin law is a symmetry operation of the lattice and reflections by different domains overlap perfectly, often simulating higher symmetry.

Structures from twinned crystals can be solved if twinning law is known.

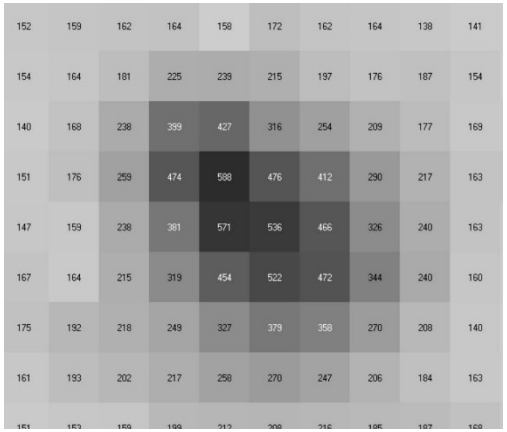
Analysis of cumulative distribution of intensities suggests presence of twinning.

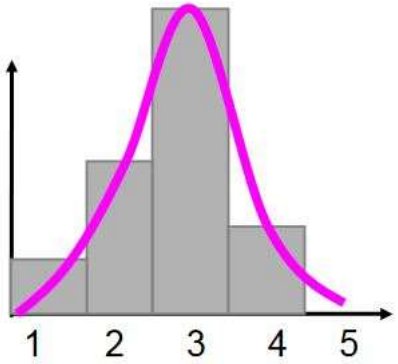
Integration

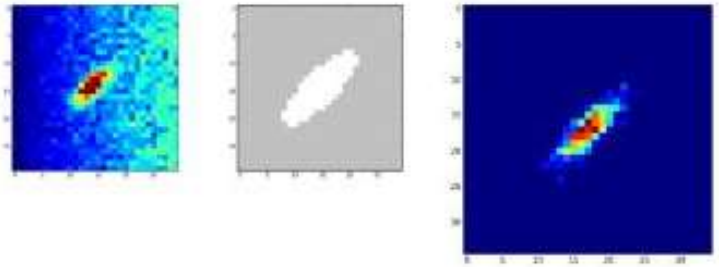


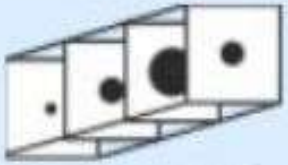
To measure intensities of each spot:

- Analysis of the spot profile, adjusting for spot shape and anisotropy
- From single pixel intensity, fitting of a 2D curve
- Background subtraction









3D integration *eg* XDS, d*trek

With newer software, 3D integration of fine slicing data collection: integration over subsequent images.

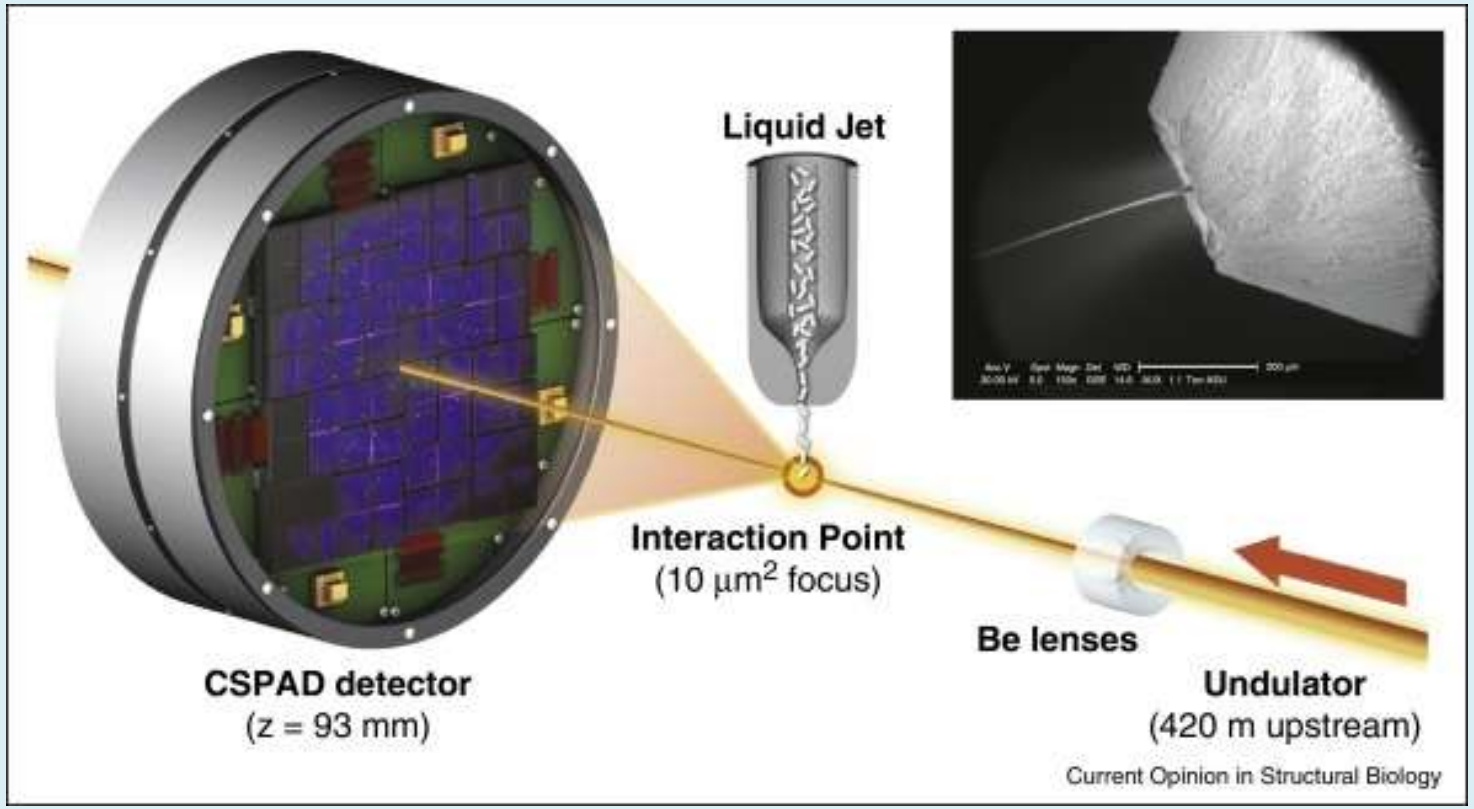
Final goal of integration: obtain a list of *hkl* reflections with corresponding intensities, *I*, and errors, $\sigma(I)$.

1	9	-27	6.50	31.60
-1	-9	-27	13.90	24.10
1	-9	-27	32.00	25.70
-1	9	-27	-10.30	22.60
1	10	0	106.49	21.00
-1	-10	0	107.99	44.50
1	-10	0	132.99	38.90
-1	10	0	139.89	21.80

Serial Femtosecond Crystallography (SFX)

Fast acquisition of diffraction patterns

Fast injection of micro- and nano-crystals



X-ray Free Electron Laser (XFEL):
High intensity of X-ray, more than synchrotron radiation, in pulses

LCLS@Stanford
SACLA@SPring8
XFEL@Hamburg

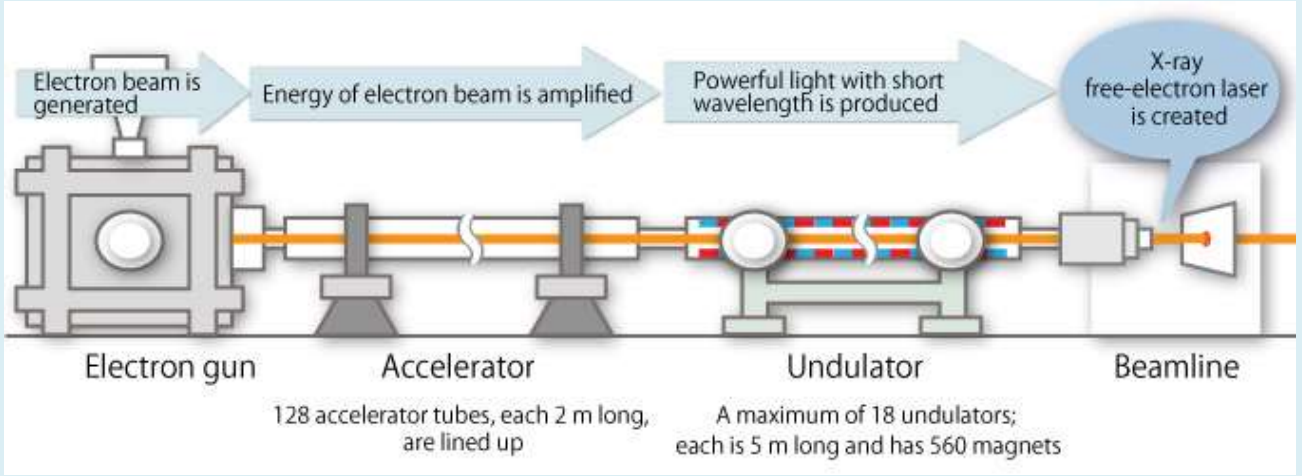
High intensity of radiation causes significant radiation damage to (micro)crystals.

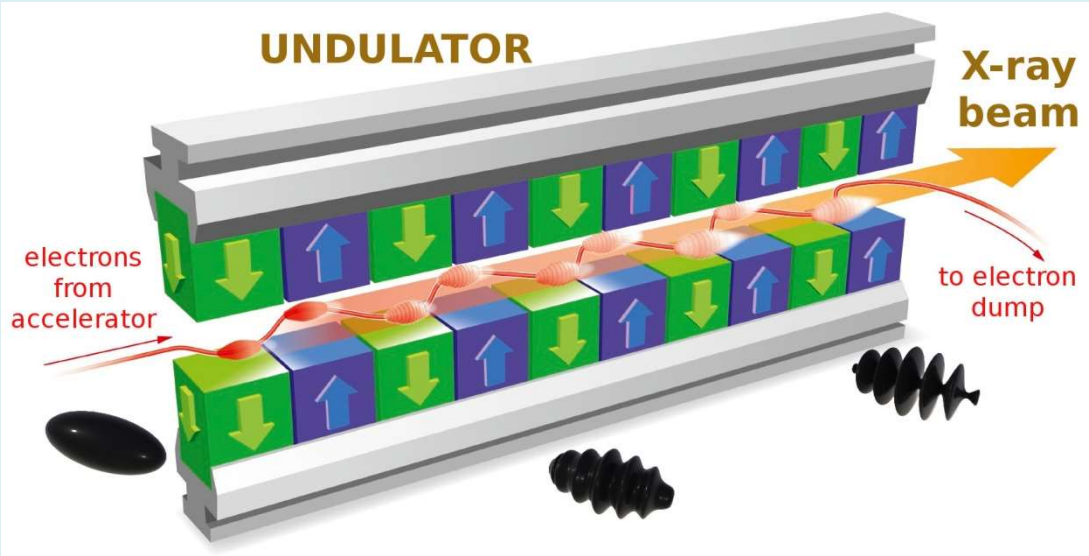
Short pulses (< 50 fs) are too fast to allow radiation damage to be visible.

Diffraction before destruction:
Collect one diffraction pattern before crystal is destroyed by X-ray beam.

X-ray Free Electron Laser (XFEL)

New X-ray source: **extremely brilliant, pulsed** X-ray beams



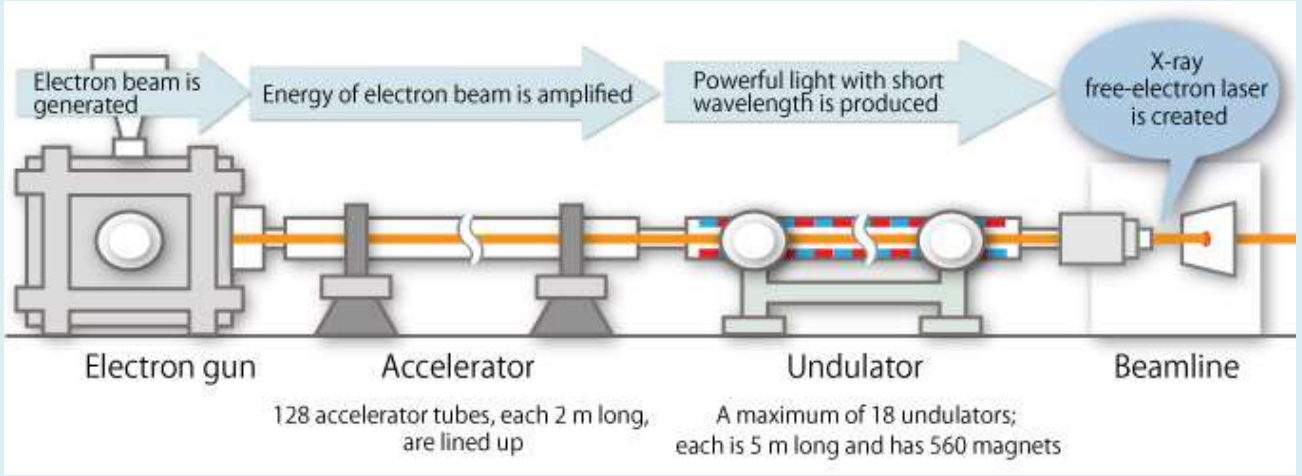


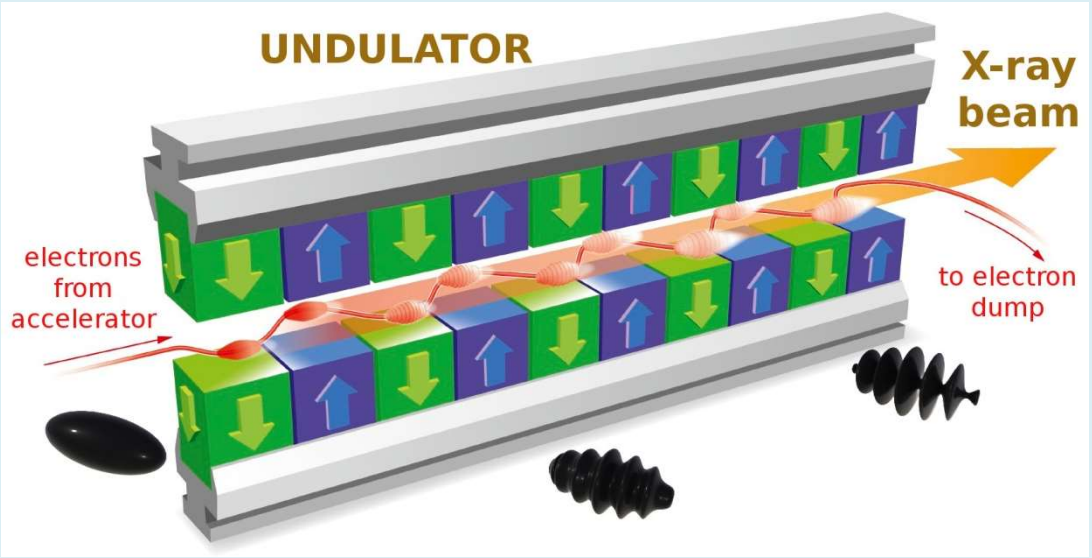
X-ray emitted by electrons interact with the electrons and modulate their velocity, creating periodically distributed electron bunches and periodic bunches of emitted radiation.

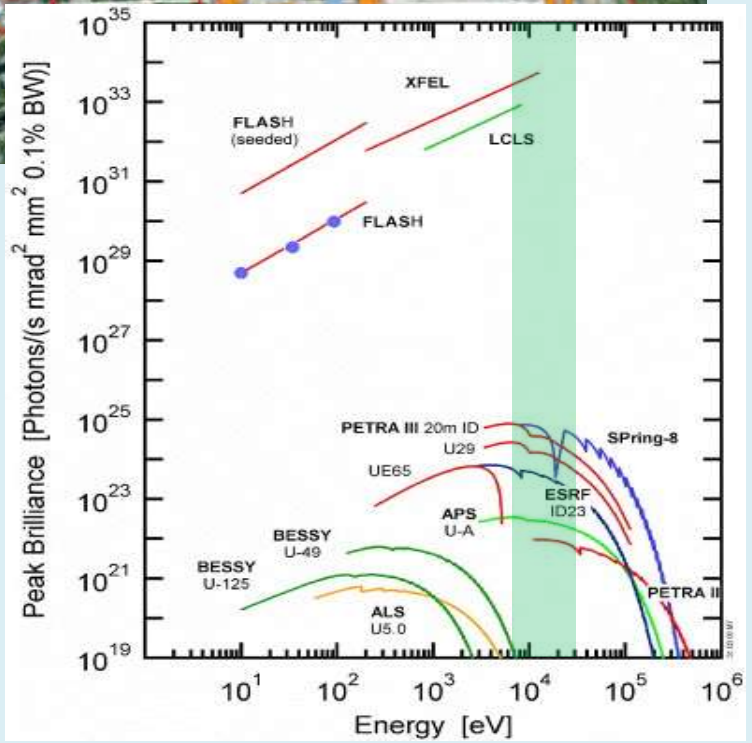
Period of bunches of radiation: 0.1 psec

X-ray Free Electron Laser (XFEL)

New X-ray source:
extremely brilliant, pulsed X-ray beams







SFX: data collection

Diffraction before destruction:

X-ray 1.0-6.0 Å
pulse = 50 fs
spot size = 1.5 μm
$3.5 \cdot 10^{10}$ ph/pulse

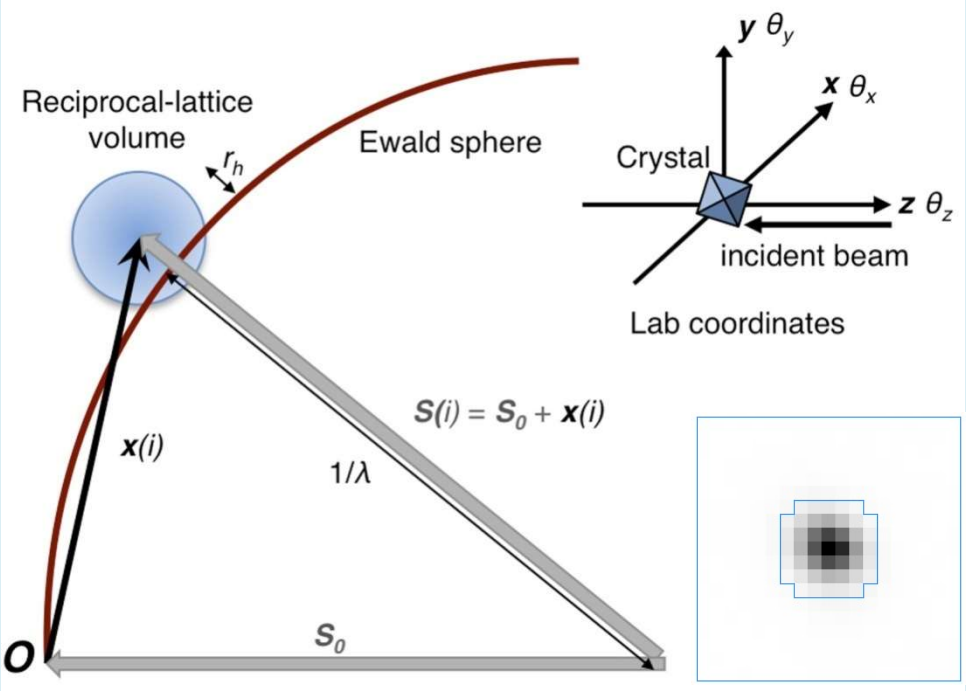
crystals 1-10 μm
25 MGy/crystals

detector pixel array
res. 1.5-2 Å

data collection
≈8000 pattern/min

Up to 5 million patterns recorded, of which
≈ 4% with diffraction
< 1% indexed

SFX: data analysis



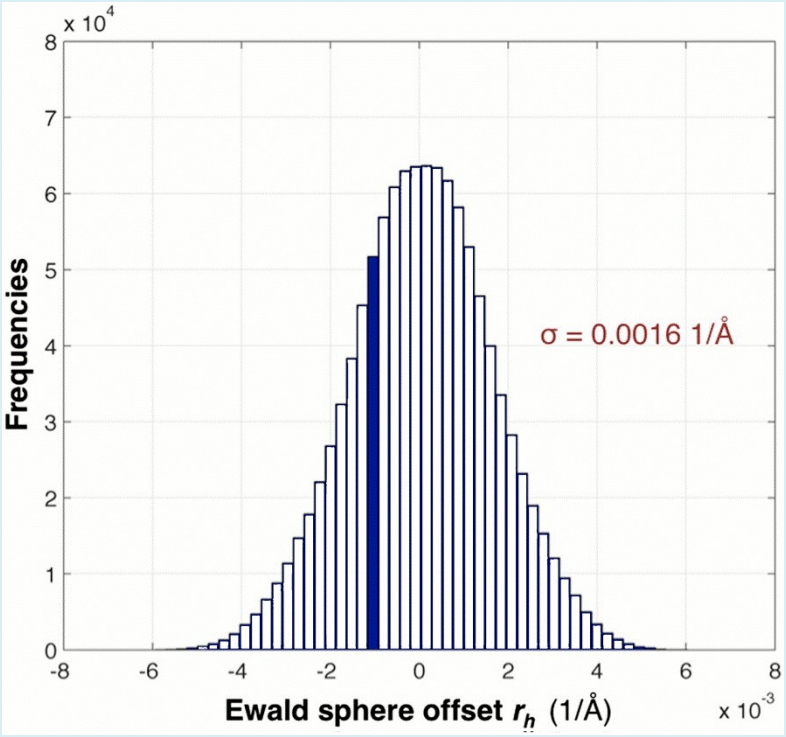
Orientation of the crystal, mosaicity, unit cell parameters have to be calculated for each frame/crystal.

Fluctuations due to X-ray source (energy, pulse, intensity) and crystals (size, quality).

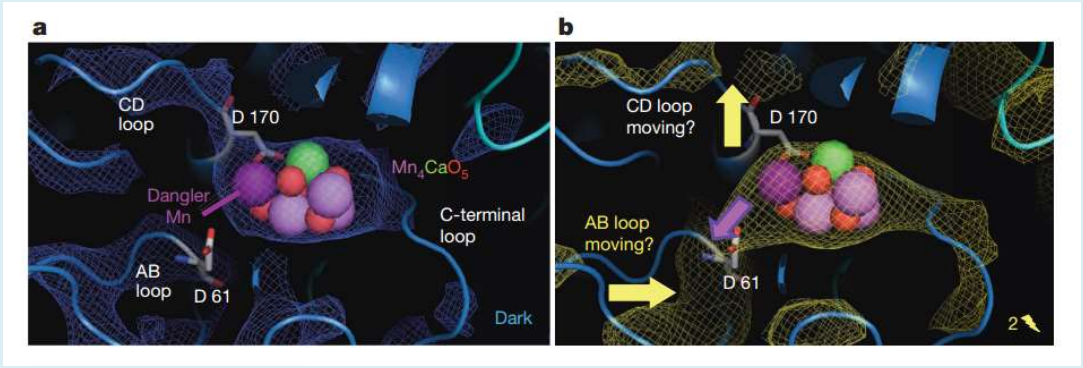
Monte Carlo methods.

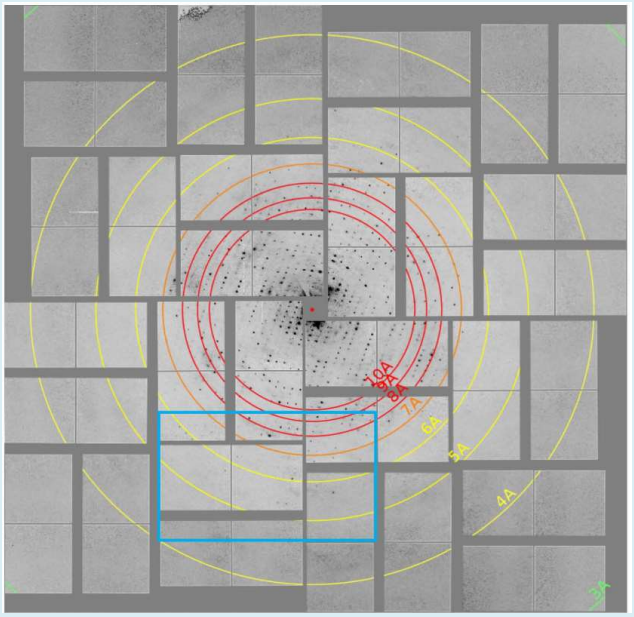
In oscillation mode (traditional X-ray crystallography), spots represent the **integrated diffraction intensity** from a node of the reciprocal lattice that enters and exits the Ewald sphere.

In SFX, a single frame yields a **partial intensity**. High values of multiplicity are required to plot and integrate the full spot intensity.

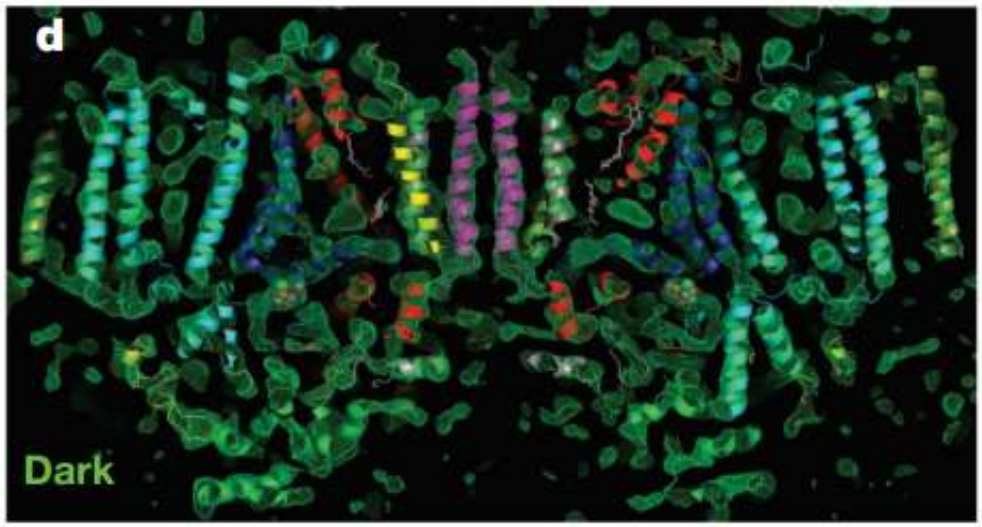


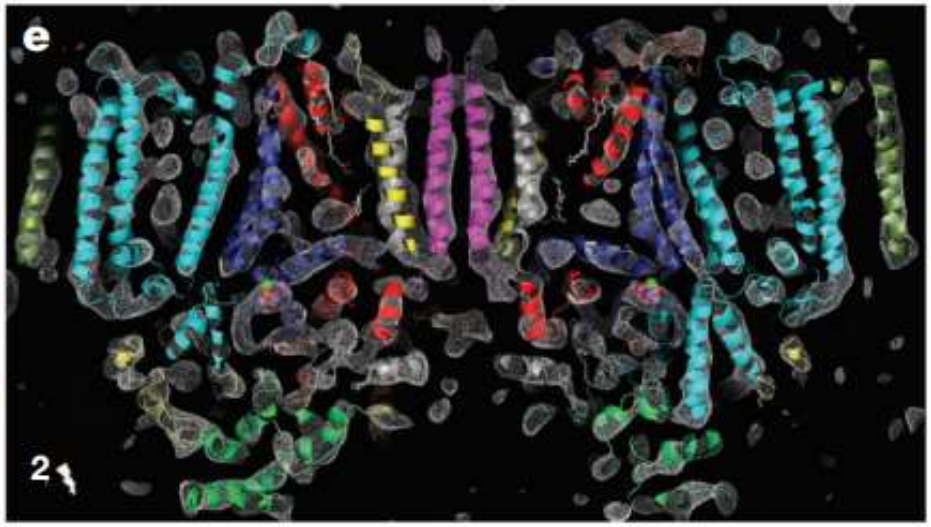
Photosystem II – Structure by SFX





SLAC@Stanford
Structures of the S_1 state (5Å resolution) and of the the S_3 state (5.5Å resolution)





Systematic absences and space group

Space group determination based on **systematic absences** (i.e. reflections whose intensity is zero due to translational symmetry elements present in the lattice). Owing to experimental errors, intensity of these reflections is actually very low/close to zero.

E.g. Data collection with Laue group *4/mmm.*

Expected reflection conditions from International Tables for Crystallography

TETRAGONAL, Laue classes $4/m$ and $4/mmm$

1. Check presence of reflections *0k0* with *k = 2n+1*

1. 2_1 screw axis along *b*

Reflection conditi							Extinction	Laue class						
								$4/m$			$4/mmm$ (			
								Point group						
hkl	$hk0$	$0kl$	hhl	$00l$	$0k0$	$hh0$	symbol	4	$\bar{4}$	$4/m$	422	4mm	$\bar{4}2m$ $\bar{4}m2$	$4/mmm$
							$P---$	$P4$ (75)	$P\bar{4}$ (81)	$P4/m$ (83)	$P422$ (89)	$P4mm$ (99)	$P\bar{4}2m$ (111)	$P4/mmm$ (123)
					k		$P-2_1-$				$P42_12$ (90			
				l			$P4_2--$	$P4_2$ (77)		$P4_2/m$ (84)	$P4_222$ (93			
				l	k		$P4_22_1-$				$P4_22_12$ (9			
				$l=4n$			$P4_1--$	$\left\{\begin{matrix}P4_1\ (76)\\P4_3\ (78)\end{matrix}\right\}^{\dagger}$			$\left\{\begin{matrix}P4_122\ (9\\P4_322\ (9\end{matrix}\right.$			
				$l=4n$	k		$P4_12_1-$				$\left\{\begin{matrix}P4_12_12\ (\\P4_32_12\ (\end{matrix}\right.$			
			l	l			$P---c$							
			l	l	k		$P-2_1c$							
		k			k		$P-b-$							
		k	l	l	k		$P-bc$							
		l		l			$P-c-$							
		l	l	l			$P-cc$							
		$k+l$		l	k		$P-n-$							
		$k+l$	l	l	k		$P-nc$							
	$h+k$				k		$Pn--$			$P4/n$ (85)				
	$h+k$			l	k		$P4_2/n--$			$P4_2/n$ (86)				
	$h+k$		l	l	k		$Pn-c$							

0	2	0	1241.84	30.44	1
0	3	0	-0.0686	0.2267	1
0	4	0	12.287	0.409	1
0	5	0	0.1013	0.3251	1
0	6	0	238.64	3.51	1
0	7	0	0.1717	0.4471	1
0	8	0	61.349	1.171	1
0	9	0	-0.9486	0.5171	1
0	10	0	556.26	8.04	1
0	11	0	-0.7931	0.6374	1
0	12	0	184.42	3.75	-1
0	13	0	-2.055	1.056	1
0	14	0	99.319	1.616	1
0	15	0	-3.781	0.923	1
0	16	0	788.01	7.14	1

Systematic absences and space group

Space group determination based on **systematic absences** (i.e. reflections whose intensity is zero due to translational symmetry elements present in the lattice). Owing to experimental errors, intensity of these reflections is actually very low/close to zero.

E.g. Data collection with Laue group *4/mmm.*

Expected reflection conditions from International Tables for Crystallography

TETRAGONAL, Laue classes $4/m$ and $4/mmm$

2. Check presence of reflections *00l* with *k = 2n+1*

Reflection conditi							Extinction	Laue class			
								$4/m$			$4/mmm$ (
								Point group			
hkl	$hk0$	$0kl$	hhl	$00l$	$0k0$	$hh0$	symbol	4	$\bar{4}$	$4/m$	422
							$P---$	$P4$ (75)	$P\bar{4}$ (81)	$P4/m$ (83)	$P422$ (8
					k		$P-2_1-$				$P42_12$ (
				l			$P4_2--$	$P4_2$ (77)		$P4_2/m$ (84)	$P4_222$ (
				l	k		$P4_22_1-$				$P4_22_12$
				$l=4n$			$P4_1--$	$\left\{\begin{matrix}P4_1\ (76)\\P4_3\ (78)\end{matrix}\right\}^{\dagger}$			$\left\{\begin{matrix}P4_122\\P4_322\end{matrix}\right.$
				$l=4n$	k		$P4_12_1-$				$\left\{\begin{matrix}P4_12_12\\P4_32_12\end{matrix}\right.$
			l	l			$P---c$				
			l	l	k		$P-2_1c$				
		k			k		$P-b-$				
		k	l	l	k		$P-bc$				
		l		l			$P-c-$				
		l	l	l			$P-cc$				
		$k+l$		l	k		$P-n-$				
		$k+l$	l	l	k		$P-nc$				
	$h+k$				k		$Pn--$			$P4/n$ (85)	
	$h+k$			l	k		$P4_2/n--$			$P4_2/n$ (86)	
	$h+k$		l	l	k		$Pn-c$				

1. 2_1 screw axis along *b*
2. 4_1 or 4_3 screw axis along *c*

0	0	6	-0.0783	0.3331	1
0	0	7	1.0234	0.3993	1
0	0	8	1720.30	34.47	1
0	0	9	1.7896	1.4854	1
0	0	10	-0.0409	0.4136	1
0	0	11	0.6631	0.4547	1
0	0	12	1076.91	21.66	1
0	0	13	1.2668	0.6038	1
0	0	14	0.3309	0.5556	1
0	0	15	0.1202	0.5820	1
0	0	16	514.85	0.6679	1
0	0	17	0.0707	0.6673	1
0	0	18	-0.7296	0.6856	1
0	0	19	-0.3101	0.7151	1
0	0	20	190.92	0.7951	1
0	0	21	0.6292	0.7126	1
0	0	22	0.0268	0.7811	1
0	0	23	0.0680	0.7936	1
0	0	24	26.634	1.132	1

TETRAGONAL, Laue classes $4/m$ and $4/mmm$

1. 2_1 screw axis along *b*
2. 4_1 or 4_3 screw axis along *c*

Reflection							Laue class						
										$4/mmm$ $(4/m\ 2/m\ 2/m)$			
hkl							oup	$\bar{4}$	$4/m$	422	4*mm*	$\bar{4}2m$ $\bar{4}m2$	$4/mmm$
							$P---$	$P4$ (75)	$P\bar{4}$ (81)	$P4/m$ (83)	$P422$ (89)	$P4mm$ (99)	$P\bar{4}2m$ (111)	$P4/mmm$ (123)
													$P\bar{4}m2$ (115)
					k		$P-2_1-$				$P42_12$ (90)		$P\bar{4}2_1m$ (113)
				l			$P4_2--$	$P4_2$ (77)		$P4_2/m$ (84)	$P4_222$ (93)		
				l	k		$P4_22_1-$				$P4_22_12$ (94)		
				$l=4n$			$P4_1--$	$\left\{\begin{matrix}P4_1\ (76)\\P4_3\ (78)\end{matrix}\right\}\dagger$			$\left\{\begin{matrix}P4_122\ (91)\\P4_322\ (95)\end{matrix}\right\}\dagger$		
				$l=4n$	k		$P4_12_1-$				$\left\{\begin{matrix}P4_12_12\ (92)\\P4_32_12\ (96)\end{matrix}\right\}\dagger$		
			l	l			$P---c$					$P4_2mc$ (105)	$P\bar{4}2c$ (112)	$P4_2/mmc$ (131)
			l	l	k		$P-2_1c$						$P\bar{4}2_1c$ (114)	
		k			k		$P-b-$					bm (100)	$P\bar{4}b2$ (117)	$P4/mbm$ (127)
		k	l	l	k		$P-bc$					bc (106)		$P4_2/mbc$ (135)
		l		l			$P-c-$					cm (101)	$P\bar{4}c2$ (116)	$P4_2/mcm$ (132)
		l	l	l			$P-cc$					cc (103)		$P4/mcc$ (124)
		$k+l$		l	k		$P-n-$					$2nm$ (102)	$P\bar{4}n2$ (118)	$P4_2/mnm$ (136)
		$k+l$	l	l	k		$P-nc$					$P4nc$ (104)		$P4/mnc$ (128)
	$h+k$				k		$Pn--$			$P4/n$ (85)				$P4/nmm$ (129)
	$h+k$			l	k		$P4_2/n--$			$P4_2/n$ (86)				
	$h+k$		l	l	k		$Pn-c$							$P4_2/nmc$ (137)

Possible space groups:
$P4_12_12$ or $P4_32_12$

It is not possible, however, to distinguish between some space groups, such as enantiomorphic space groups. In this case, both possible space groups should be tested in the following phasing step...

Scaling and merging

Scaling: Diffraction data can be used only if intensities are on the same scale, but multiple phenomena affect experimental intensities (intensity of incident beam, crystal volume crossed by X-ray beam, radiation damage, scattering of buffer/loop...)

This is a fundamental step of data processing!

Merging: Redundant observations of the same reflections are merged together. In addition, according to space group, symmetry related reflections are averaged together (often weigthing the average with experimental errors of each reflection). Friedel pairs are also merged.

If anomalous signal is required for phasing, Bijvoet mates are merged, but Friedel pairs are not.

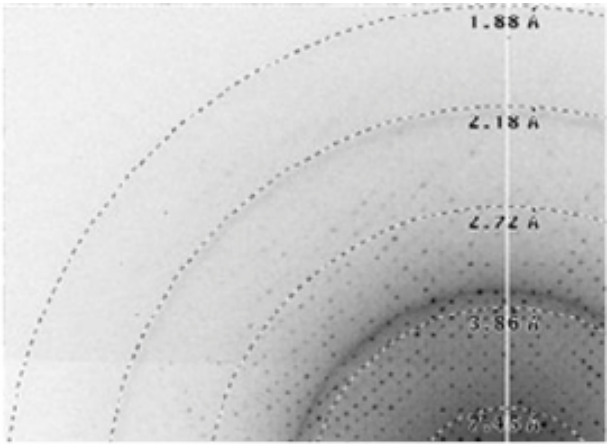
Final goal: the list of *hkl* reflections is on the same scale and not redundant.

Evaluation of data quality

Signal-to-noise ratio, in crystallography this indication is given by the average of intensities over their respective errors:

$$\langle |I|/\sigma(I) \rangle = \frac{1}{N} \sum_{\boldsymbol{h}} \frac{|I_{\boldsymbol{h}}|}{\sigma(I_{\boldsymbol{h}})}$$

with *N* number of all reflections in the dataset, or in a specific resolution shell.



Completeness: how many reflections within the unique wedge of the reciprocal space have been collected? Completeness can be evaluated for each resolution shell (each containing an equal number of reflections).

Random missing reflections cause noise, but are not problematic.
Systematically missing reflections can cause distorsions and hamper phasing.

Redundancy: average number of equivalent reflections for each unique (merged) reflection.

A higher redundancy can improve data quality and reduce errors.

Evaluation of data quality

R-values, evaluate the agreement between multiple measurements of the same intensity:

$$R_{merge} = \frac{\sum_{\boldsymbol{h}} \sum_{i=1}^{N} \left| I_{\boldsymbol{h},i} - \overline{I_{\boldsymbol{h}}} \right|}{\sum_{\boldsymbol{h}} \sum_{i=1}^{N} I_{\boldsymbol{h},i}}$$

with *N* number of equivalent reflections.

But...
increasing redundancy increases R_{merge} due to random experimental errors...

This parameter can be calculated for each resolution shell. For higher resolution shells, R-values are usually higher.

Redundancy independent R-value:

$$R_{rim} = \frac{\sum_{\boldsymbol{h}} \sqrt{\frac{N}{N-1}} \sum_{i=1}^{N} \left| I_{\boldsymbol{h},i} - \overline{I_{\boldsymbol{h}}} \right|}{\sum_{\boldsymbol{h}} \sum_{i=1}^{N} I_{\boldsymbol{h},i}}$$

Always larger than R_{merge}.

Precision indicating R-value:

$$R_{pim} = \frac{\sum_{\boldsymbol{h}} \sqrt{\frac{1}{N-1}} \sum_{i=1}^{N} \left| I_{\boldsymbol{h},i} - \overline{I_{\boldsymbol{h}}} \right|}{\sum_{\boldsymbol{h}} \sum_{i=1}^{N} I_{\boldsymbol{h},i}}$$

Decreases increasing redundancy.

Correlation coefficients, between random half datasets

$$CC_{1/2} = \frac{\sum_{i=1}^{N} [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

Ideally equal to 1, but data are significant even if $CC_{1/2}$ is lower.

Resolution

What is the resolution cutoff of the data? How far are data significant?

There is no general rule, but data are usually considered significant up to the resolution shell with a value

$$\langle |I|/\sigma(I) \rangle > 2.0$$

If the statistical distribution of intensitites is correct and the estimation of errors ($\sigma(I)$) is reliable, this value should correspond to $R_{merge} \approx 40\%$ in the same resolution shell.

However, new studies suggest that significant information is present in the diffraction intensities also beyond the resolution cutoff obtained using the previous threshold. It has been suggested to use the correlation coefficient $CC_{1/2}$ instead.

The suggested threshold varies according to authors, but a conservative criterion would suggest that $CC_{1/2} > 0.15$ in the last meaningful resolution shell.

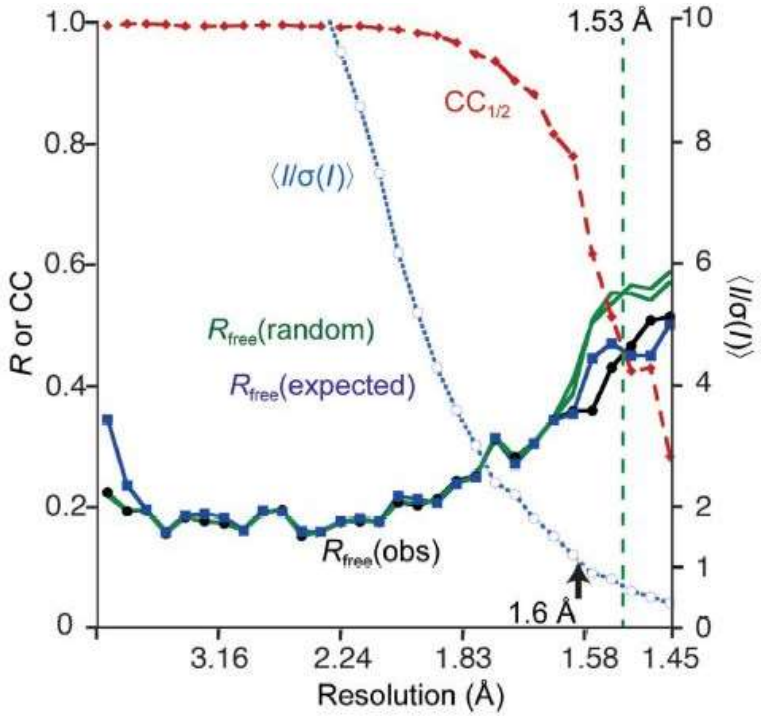


Table of data collection statistics

Table 1 Data collection and refinement statistics

	KirBac3.1(S129R)
Data collection	
Space group	$P42_12$
Cell dimensions	
$a = b$, c (Å)	106.24, 89.80
Resolution (Å)	106.24–3.05 (3.21–3.05)[a]
R_{merge}	0.262 (0.945)
$I / \sigma I$	5.5 (2.0)
Completeness (%)	99.9 (99.8)
Redundancy	6.8 (6.9)

Lattice dimensions and symmetry*

*For protein structures, inaccuracies on parameters limit precision

(Often also wavelength and source included)

Overall resolution (and maximum resolution shell)

Data quality statistics, over the whole dataset (and in the highest resolution shell)

(Sometimes total number of reflections and number of unique reflections are included)

References

- Pilatus detector: Broennimann Ch. *et al.*, “The PILATUS 1M detector.”, **J Synchrotron Radiat. 2006**, *13(Pt 2)*:120-30; Kraft P. *et al.*, “Performance of single-photon-counting PILATUS detector modules.”, **J Synchrotron Radiat. 2009**, *16(Pt 3)*:368-75.
- Microfocus beamlines: Smith J.L. *et al.*, “Micro-crystallography comes of age.”, **Curr Opin Struct Biol. 2012**, *22(5)*:602-12; Axford D. *et al.*, “In situ macromolecular crystallography using microbeams.”, **Acta Crystallogr D Biol Crystallogr. 2012**, *68(Pt 5)*:592-600; Aishima J. *et al.*, “High-speed crystal detection and characterization using a fast-readout detector.”, **Acta Crystallogr D Biol Crystallogr. 2010**, *66(Pt 9)*:1032-5; Warren A.J. *et al.*, “Exploiting Microbeams for Membrane Protein Structure Determination.”, **Adv Exp Med Biol. 2016**, *922*:105-117.
- Automation at beamlines: Nurizzo D. *et al.*, “RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments.”, **Acta Crystallogr D Struct Biol Crystallogr. 2016**, *72(Pt 8)*:966-75.
- SFX: Neutze R *et al.*, “Membrane protein structural biology using X-ray free electron lasers.”, **Curr Opin Struct Biol. 2015**, *33*:115-25; Liu W. *et al.*, “Femtosecond crystallography of membrane proteins in the lipidic cubic phase .”, **Philos Trans R Soc Lond B Biol Sci. 2014**, *369 (1647)*: 20130314.
- Evaluation of data quality: Evans P.R. and Murshudov G.N., “How good are my data and what is the resolution?”, **Acta Crystallogr D Biol Crystallogr. 2013**, *69*:1204-1219; Karplus P.A. and Diederichs K., “Linking crystallographic model and data quality”, **Science 2012**, *336(6084)*: 1030-1033; Karplus P.A. and Diederichs K., “Assessing and maximizing data quality in macromolecular crystallography”, **Curr Opin Struct Biol. 2015**, *34*: 60-68.