IUCr electron crystallography school

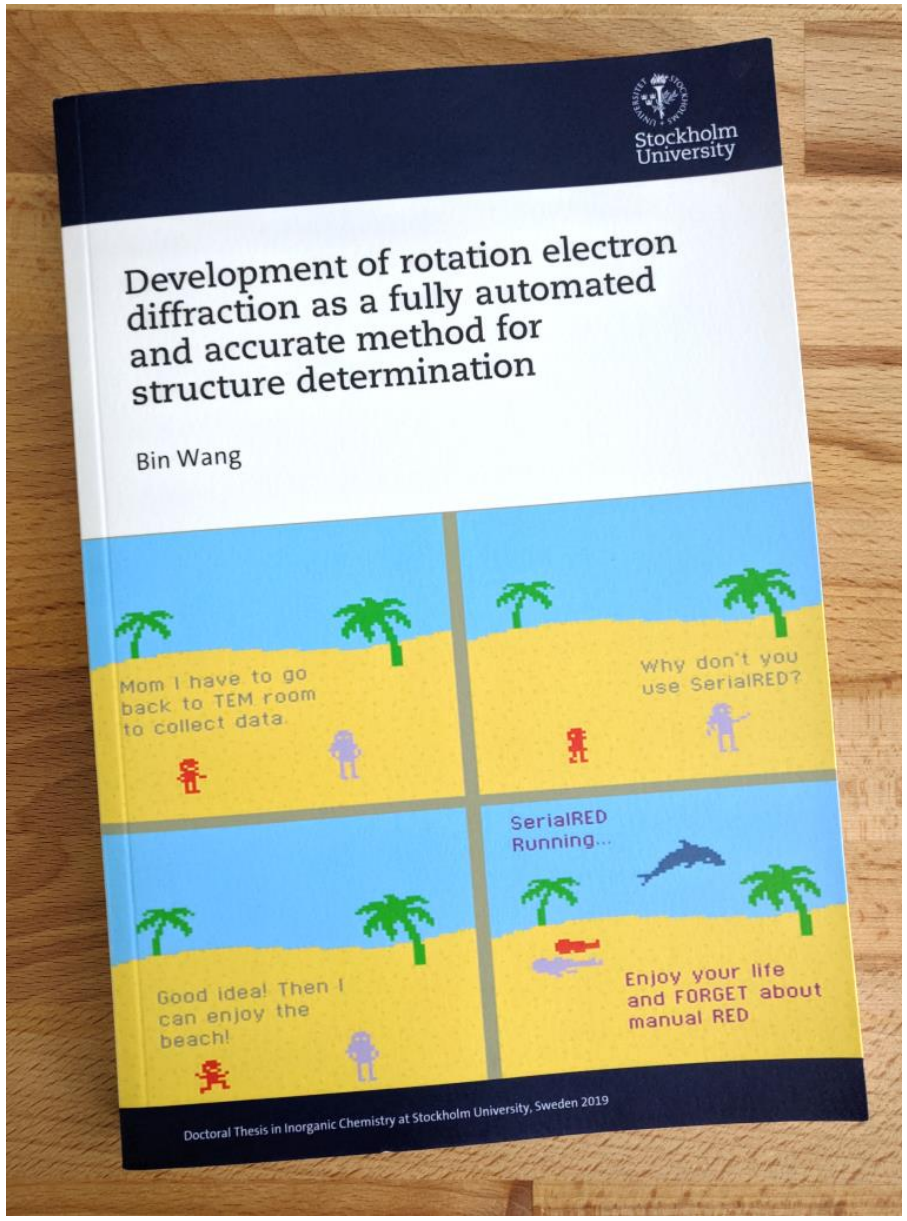
11-14 August 2021

Serial Rotation Electron Diffraction (SerialRED)

Stef Smeets

s.smeets@esciencecenter.nl

netherlands eScience center



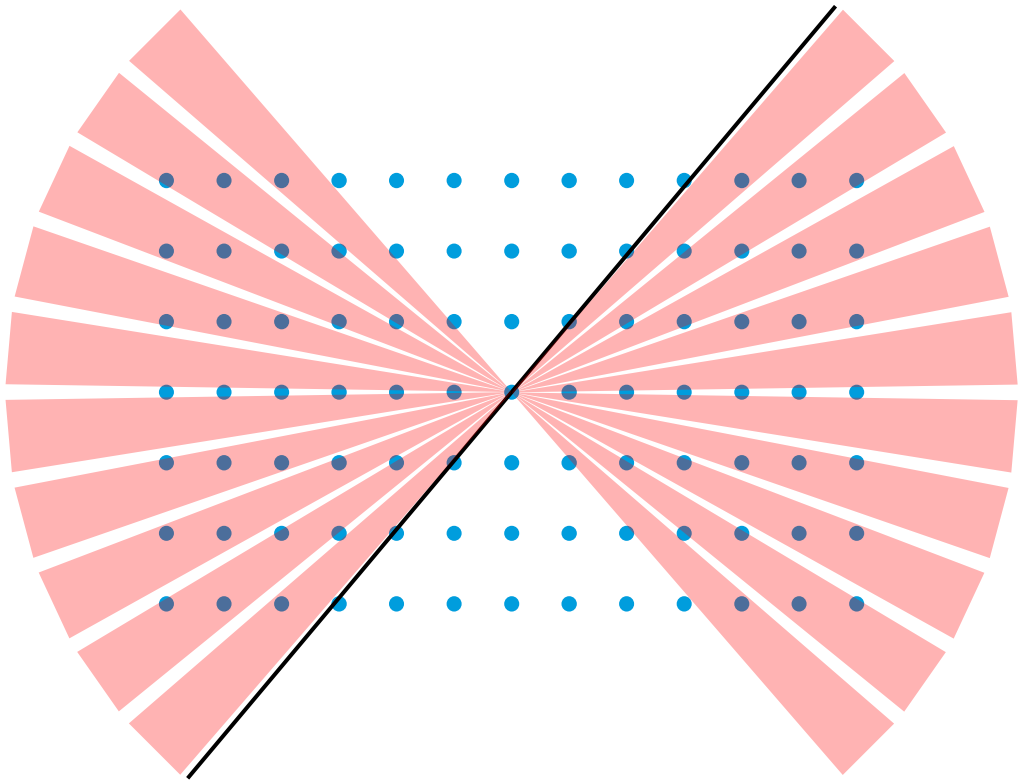
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-168157

Outline

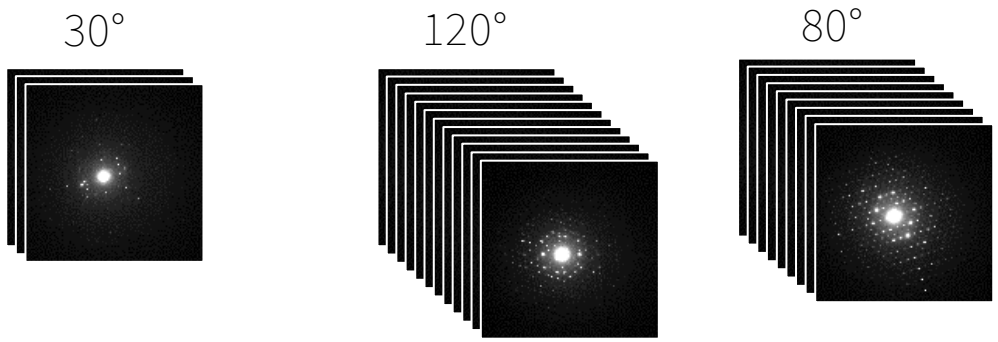
- What is SerialRED?
  - Crystal screening techniques
  - Automated data collection
  - Ensemble data processing
  - Data processing demo

Why SerialRED?

Manual data collection



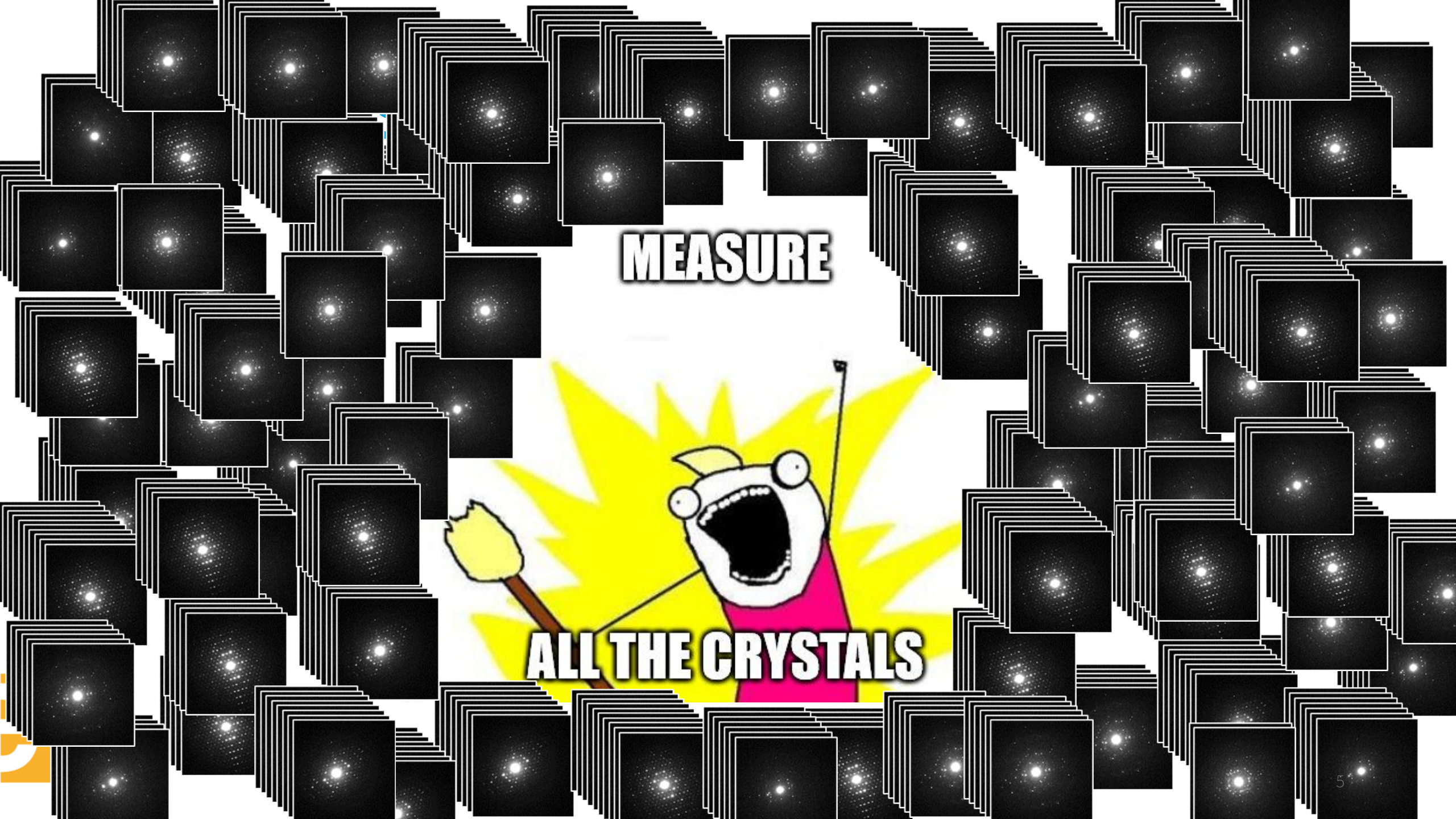
3D-ED / microED



This is great, but…

- You just collected data on 3 crystals from 10000s available 🤔
- Your crystal selection may be biased
- The experiment is probably not reproducible


CRYSTALS
LIGHTYEAR
CRYSTALS EVERYWHERE
imgflip.com


MEASURE
ALL THE CRYSTALS

When I started (~2014)

1. No open-source software
2. Manual crystal search
3. Semi-automated data collection
4. Not reproducible, ad-hoc protocols
5. Lack of experimental metadata
6. Time-consuming data processing

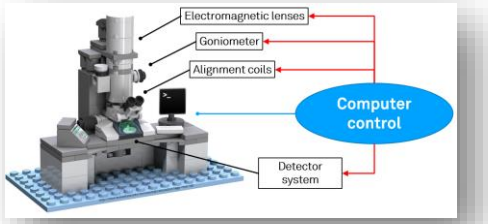

GATAN
This is brilliant.
python™
But I like this.
imgflip.com

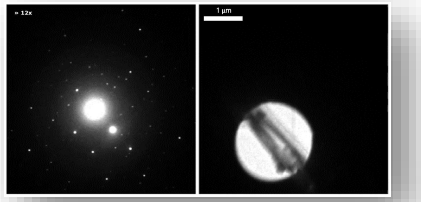
When I started (~2014)

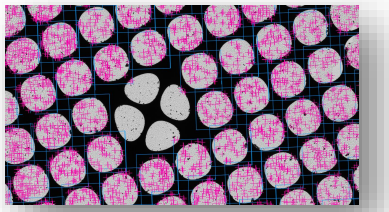
1. No open-source software
2. Manual crystal search
3. Semi-automated data collection
4. Not reproducible, ad-hoc protocols
5. Lack of experimental metadata
6. Time-consuming data processing

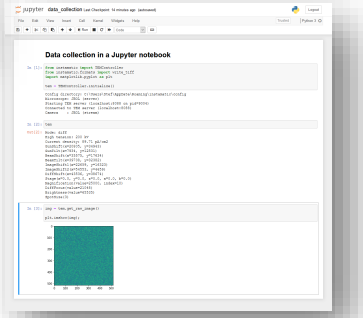
Instamatic (2017)

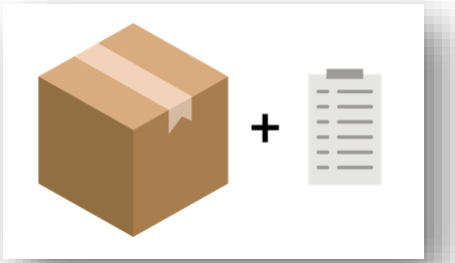
http://github.com/instamatic-dev/instamatic
Automated crystal screening
Automated data collection with crystal tracking
Reproducible, standardized experiments
Consistent metadata and logging
Data processing pipeline

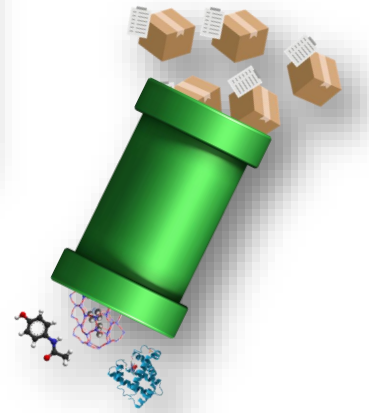




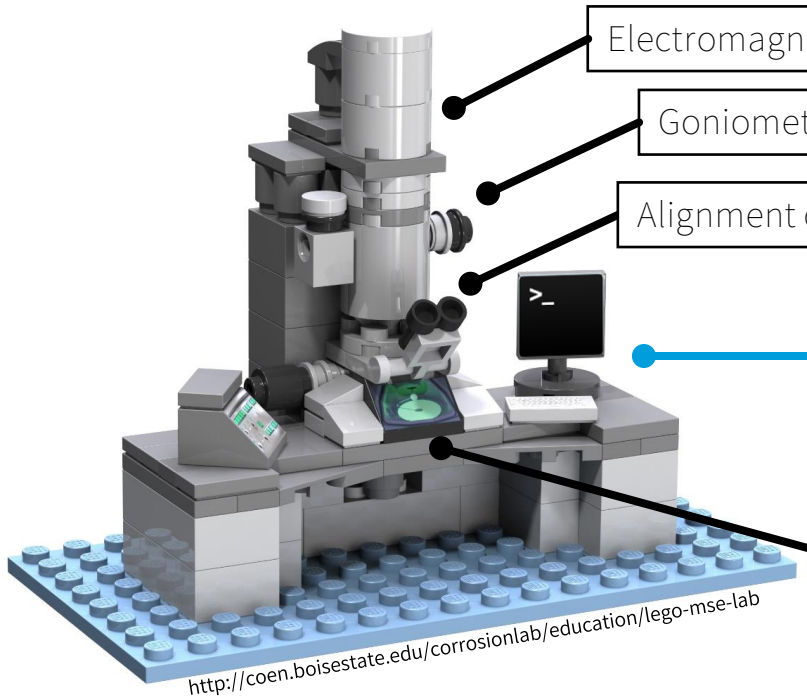








Automation toolkit: *Instamatic*



Electromagnetic lenses

Goniometer

Alignment coils

Detector system

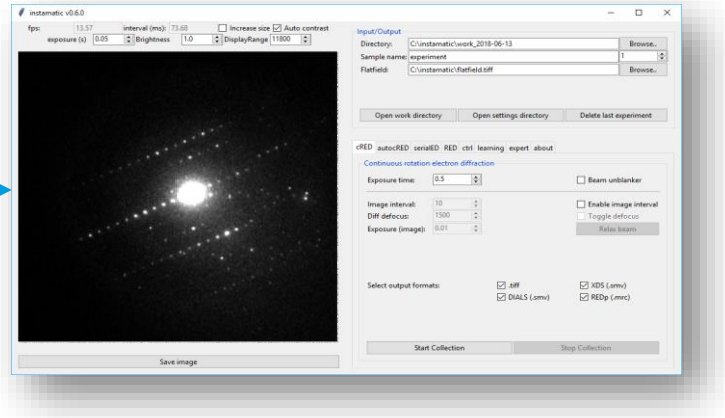
Instamatic
(Python 3.6+)

Features

- Object oriented Python API
- Grid montaging
- Image segmentation
- Crystal tracking
- Automated data processing
- Calibrations/alignments

Automated experiments

- Screening
- High-throughput 3D ED
- Serial crystallography
- Unsupervised data collection



Microscopes

TFS Titan/Themis Z

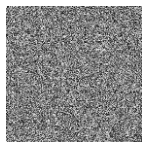
JEOL 1400/2100/
3200/ARM200

Simulated

Cameras

ASI Cheetah

Gatan



Simulated

TVIPS (X)F416

Source code:
http://github.com/instamatic-dev/instamatic

Instamatic in a Jupyter Notebook

jupyter data_collection Last Checkpoint: 14 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Data collection in a Jupyter notebook

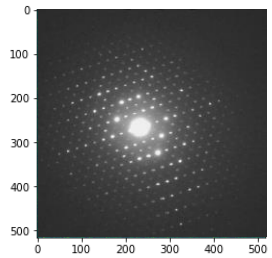
In [1]:
```
from instamatic import TEMController
from instamatic.formats import write_tiff
import matplotlib.pyplot as plt

tem = TEMController.initialize()
```

```
Config directory: C:\Users\Stef\AppData\Roaming\instamatic\config
Microscope: JEOL (server)
Starting TEM server (localhost:8088 on pid=9004)
Connected to TEM server (localhost:8088)
Camera    : JEOL (stream)
```

In [2]:
```
tem
```

Out[2]:
```
Mode: diff
High tension: 200 kV
Current density: 89.71 pA/cm2
GunShift(x=20805, y=34943)
GunTilt(x=7934, y=12531)
BeamShift(x=35575, y=17634)
BeamTilt(x=39738, y=32382)
ImageShift1(x=22659, y=16323)
ImageShift2(x=54553, y=4659)
DiffShift(x=43536, y=38674)
Stage(x=0.0, y=0.0, z=0.0, a=0.0, b=0.0)
Magnification(value=25000, index=10)
DiffFocus(value=21048)
Brightness(value=65535)
SpotSize(3)
```

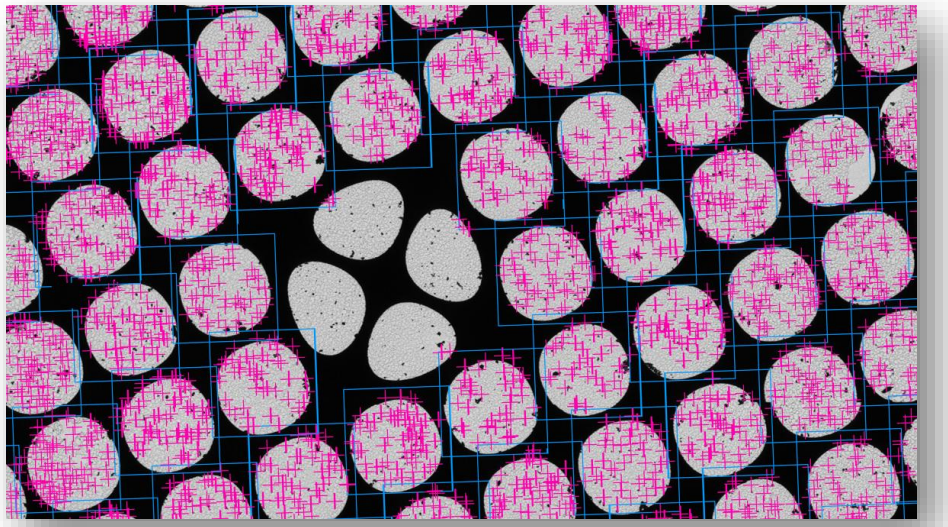
In [3]:
```
img = tem.get_raw_image()

plt.imshow(img);
```



netherlands eScience center

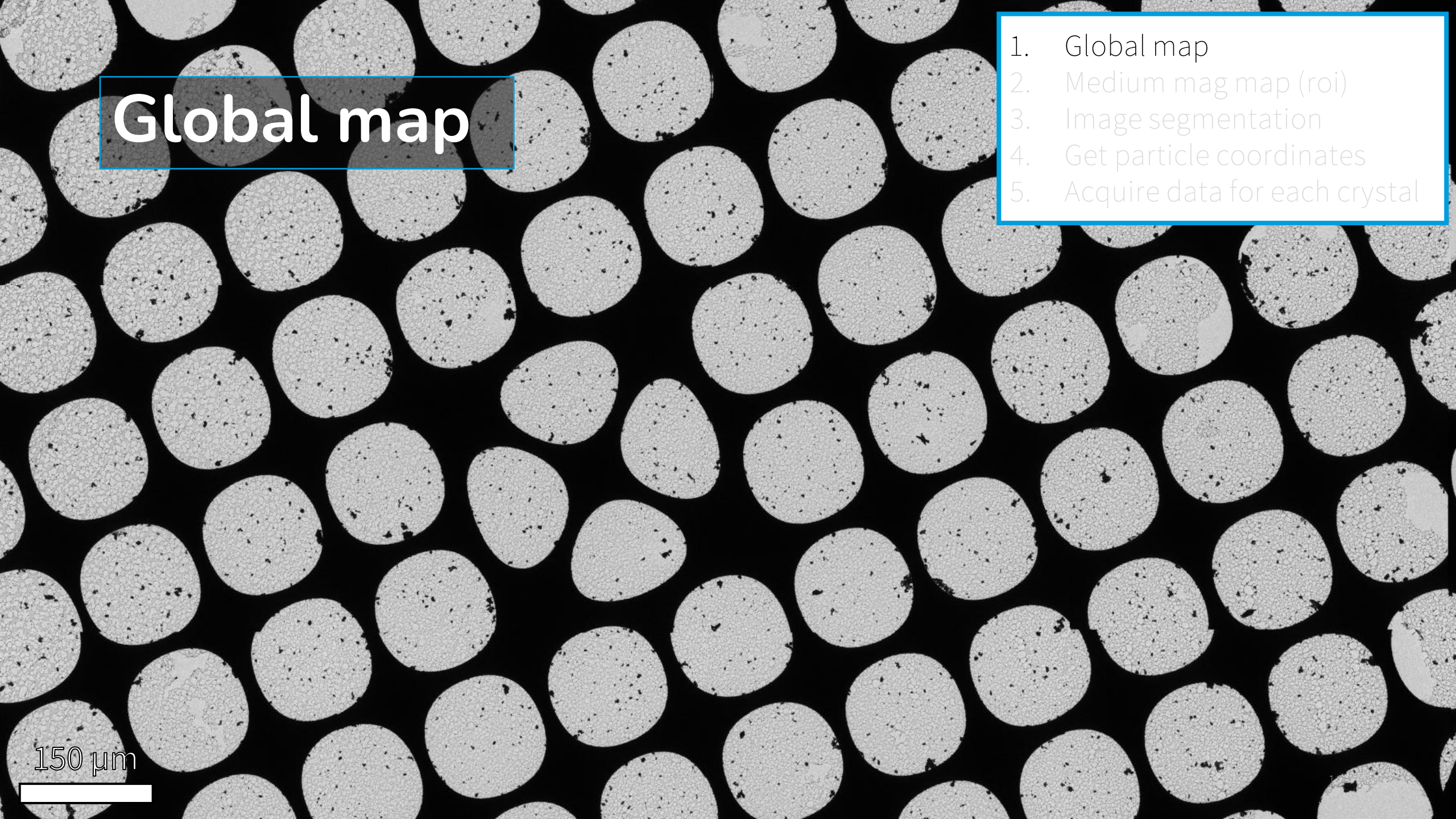
Outline

- What is SerialRED?
- Crystal screening techniques
- Automated data collection
- Ensemble data processing
- Data processing demo



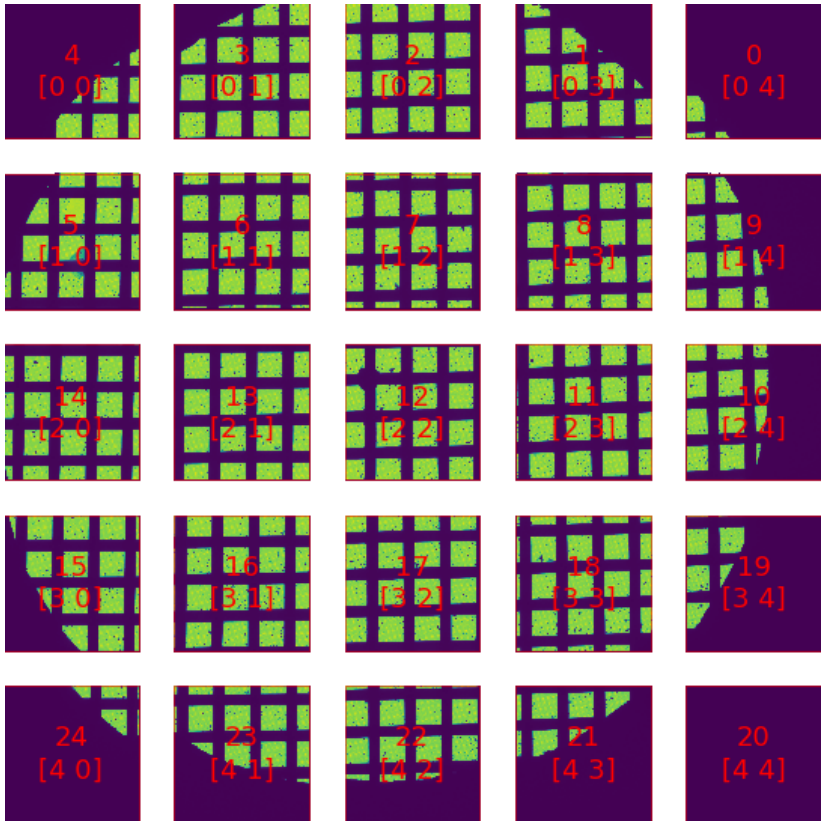
Global map

1. Global map
2. Medium mag map (roi)
3. Image segmentation
4. Get particle coordinates
5. Acquire data for each crystal

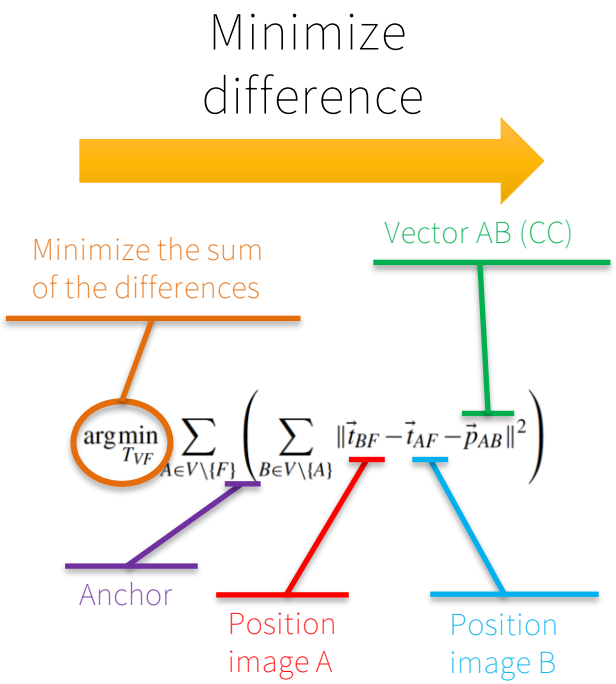


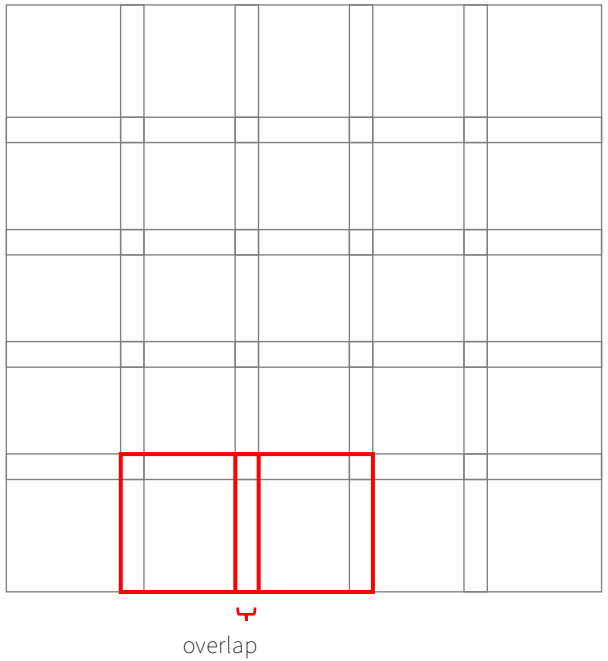
Montaging/stitching in *Instamatic*

Algorithm: Preibisch et al. (2009), Bioinformatics, 25(11):1463-1465
Implementation: https://github.com/instamatic-dev/pyserialem



5x5 grid
collected at low mag
10% overlap

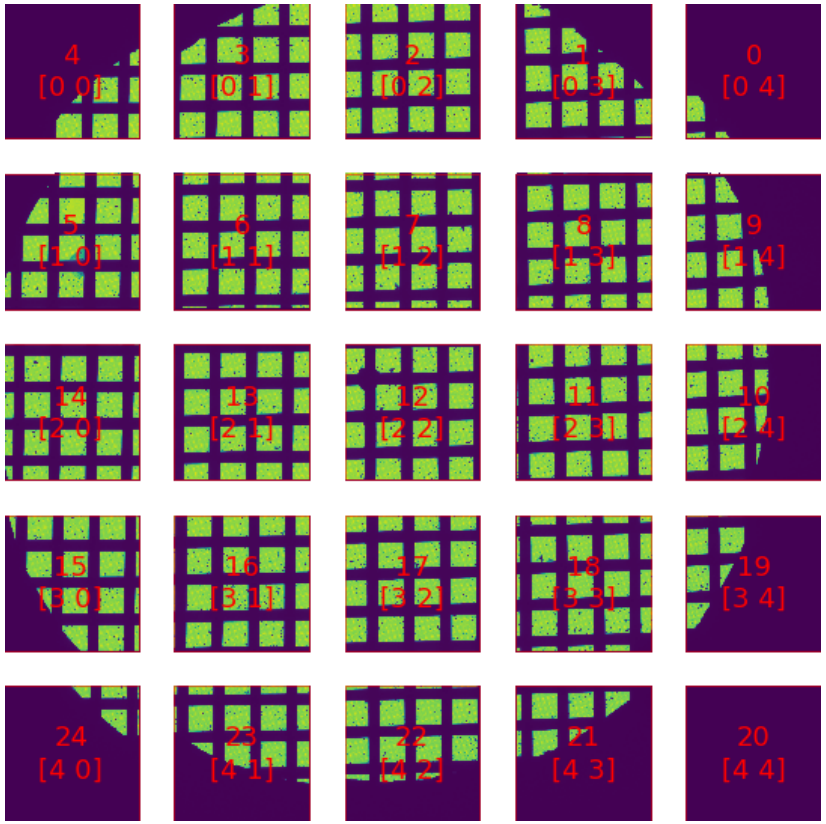




Montaging/stitching in *Instamatic*

Algorithm: Preibisch et al. (2009), Bioinformatics, 25(11):1463-1465
Implementation: https://github.com/instamatic-dev/pyserialem

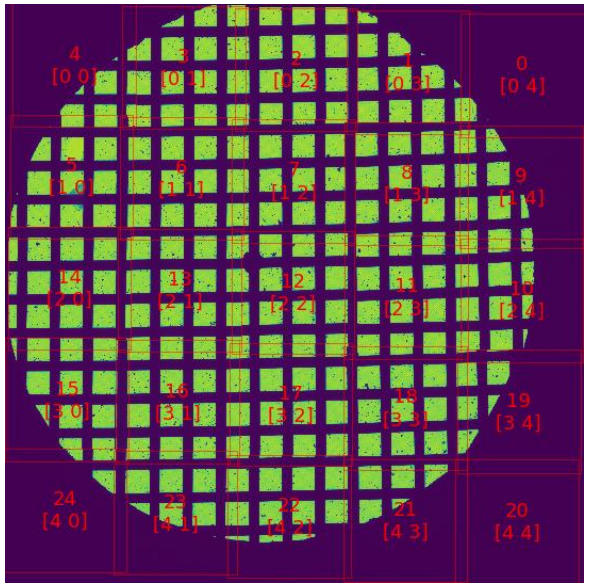
Alternative software: SerialEM
https://bio3d.colorado.edu/SerialEM/



5x5 grid
collected at low mag
10% overlap

Minimize difference

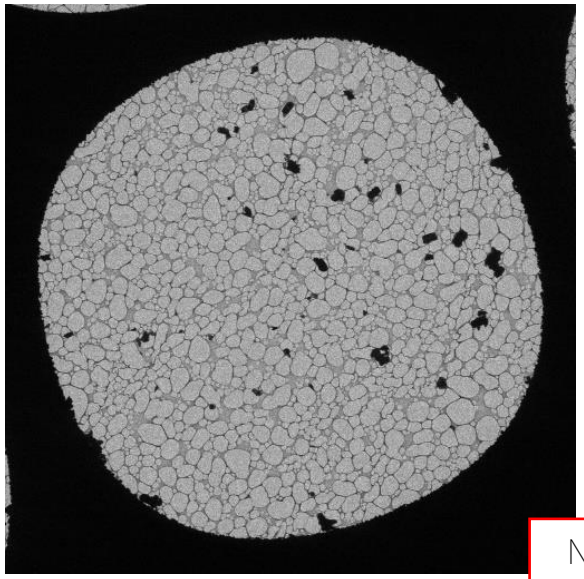
$$\underset{T_{VF}}{\arg\min} \sum_{A \in V \setminus \{F\}} \left(\sum_{B \in V \setminus \{A\}} \|\vec{t}_{BF} - \vec{t}_{AF} - \vec{p}_{AB}\|^2 \right)$$

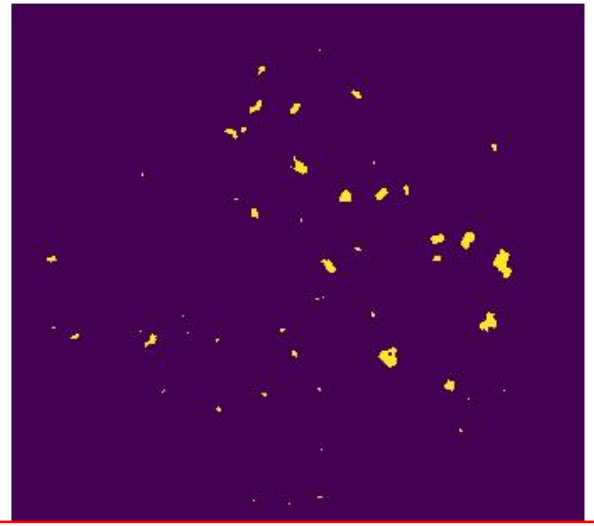


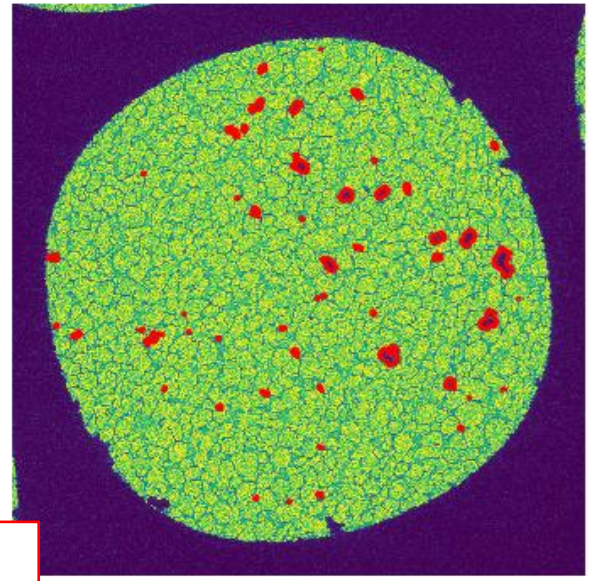
Stitched image

Global map

1. Global map
2. Medium mag map (roi)
3. Image segmentation
4. Get particle coordinates
5. Acquire data for each crystal

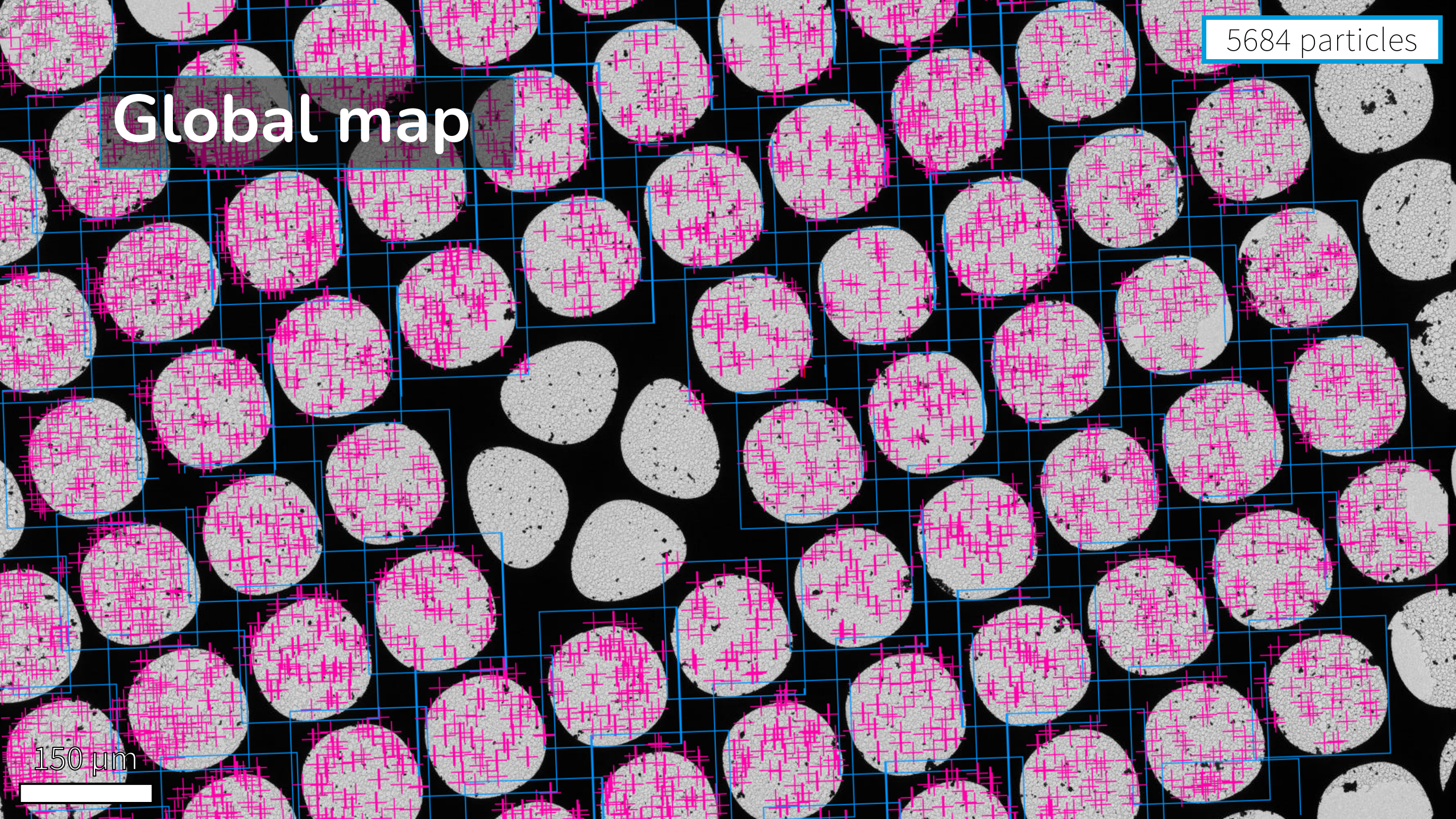




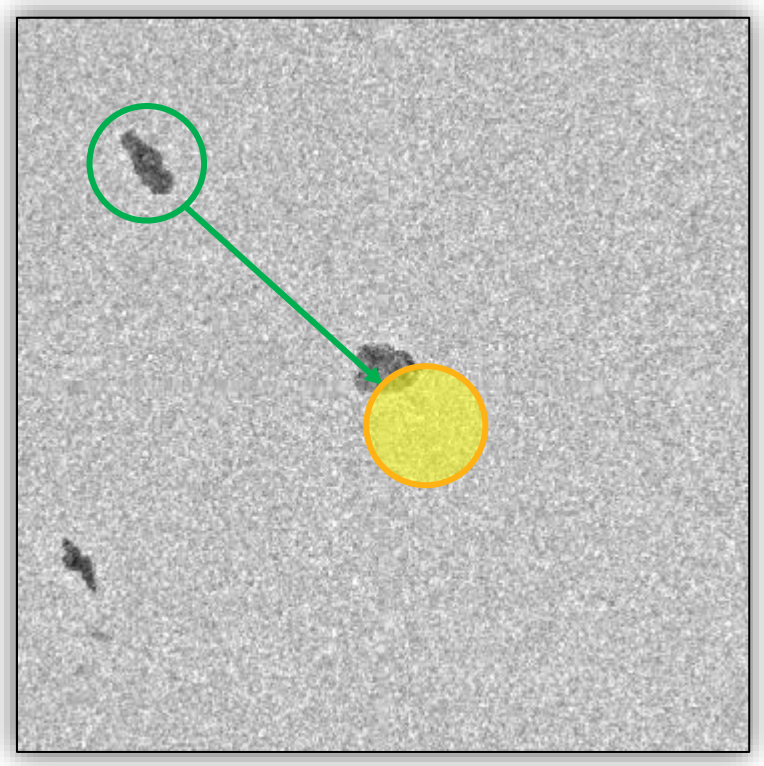


Neural network based segmentation/classification using Ilastik
(https://gitlab.tudelft.nl/aj-lab/predicrystal)

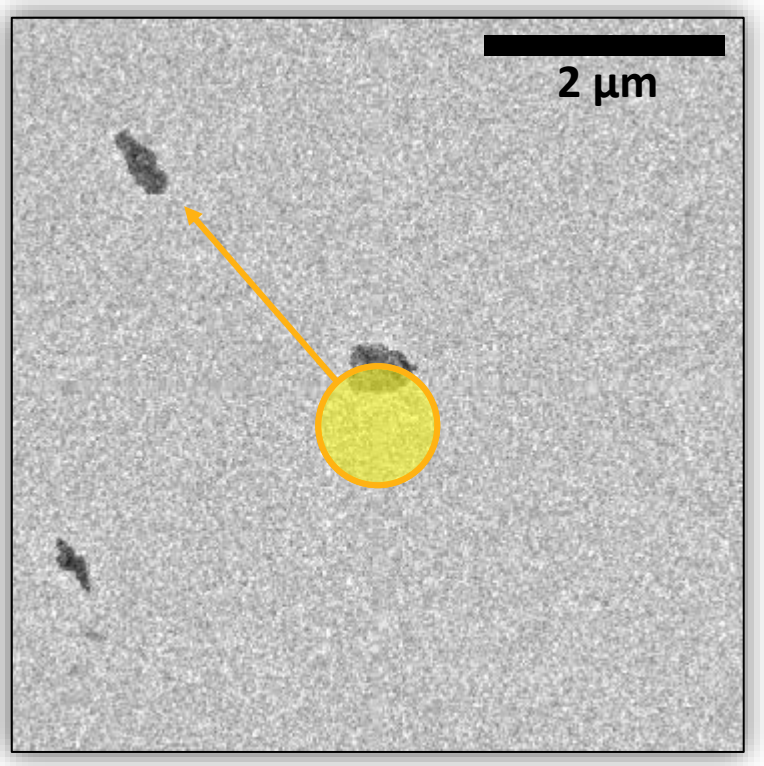
150 µm


Global map
5684 particles
150 µm

Particle centering



Crystal to beam
Uses stage movement
Works with global (stage) coords



Beam to crystal
Uses beam deflectors
Works with local (image) coords

Particle centering pros/cons

Crystal to beam

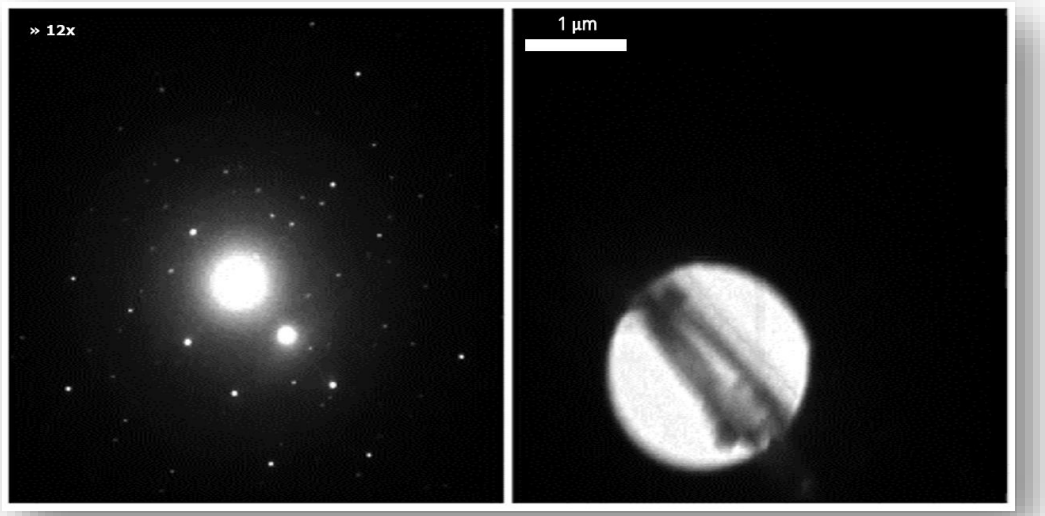
- Full control over crystal selection
- No mode switching
- Allows pre-screening
- Stage movement is slow
- Backlash correction necessary
- Requires very precise stage

Beam to crystal

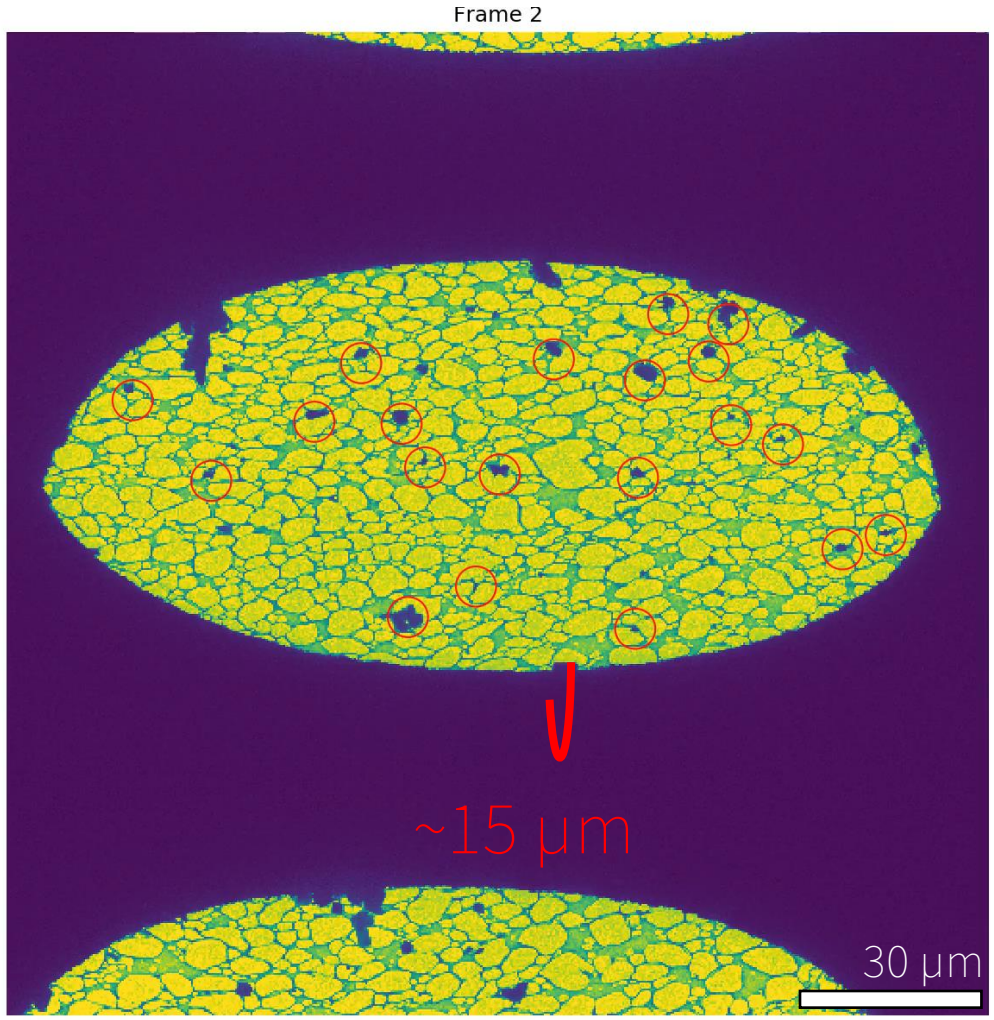
- Little control over crystal selection
- Constant mode switching
- Set-and-forget
- Deflector movement is almost instant
- Precise control over beam position
- Limited lateral range of beam

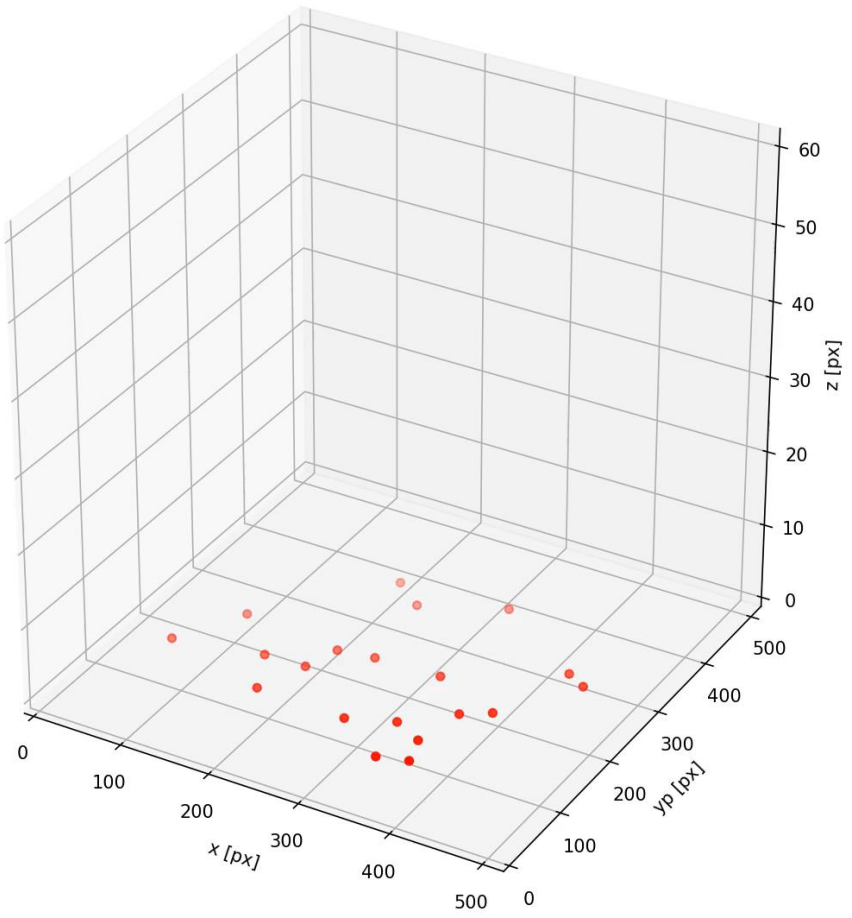
Outline

- What is SerialRED?
- Crystal screening techniques
- Automated data collection
- Ensemble data processing
- Data processing demo



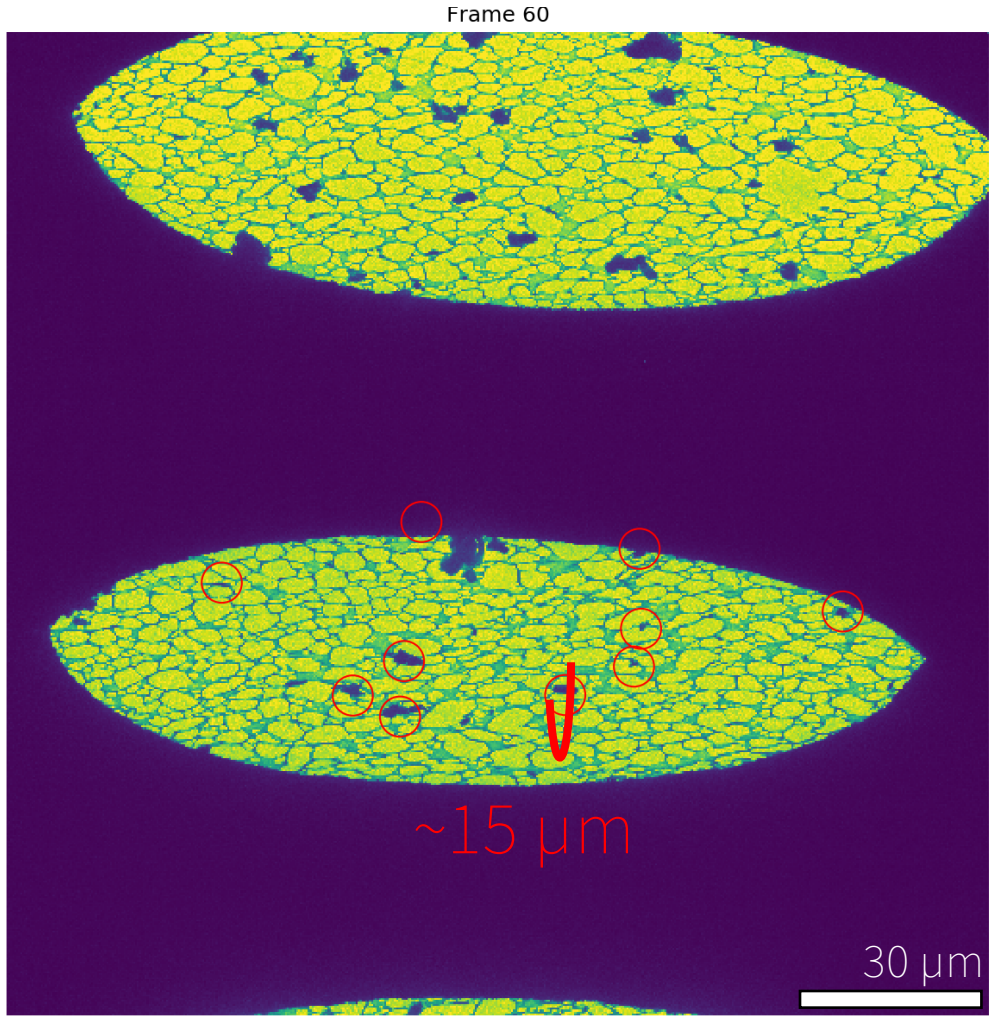
The crystal tracking problem

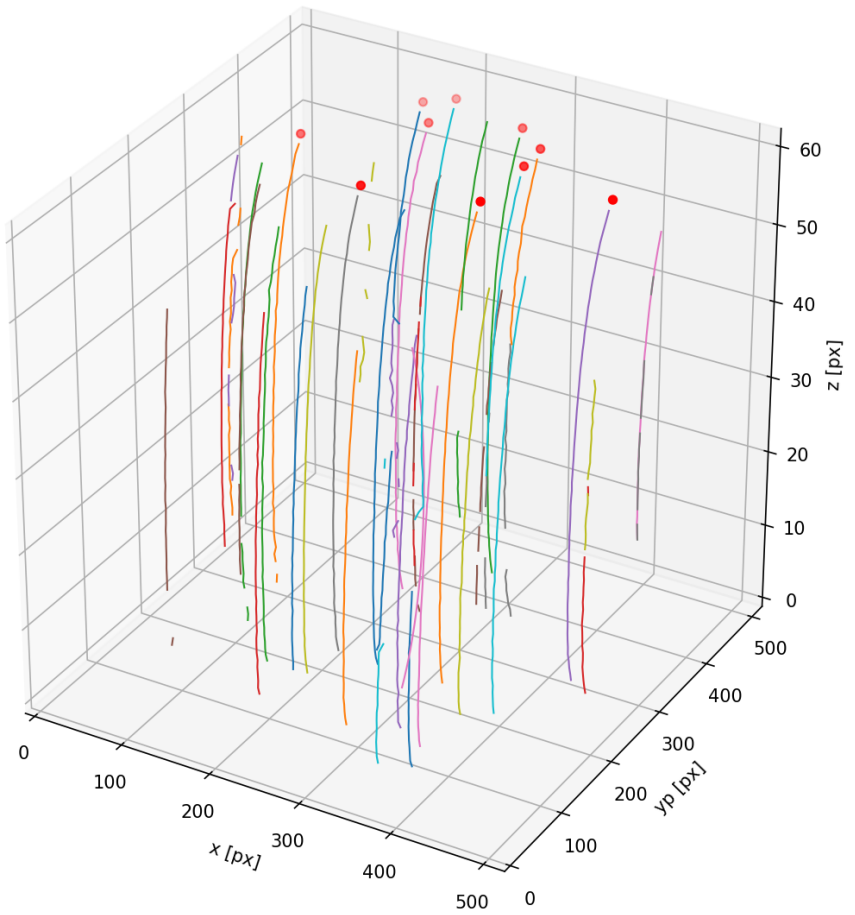




Crystal movement

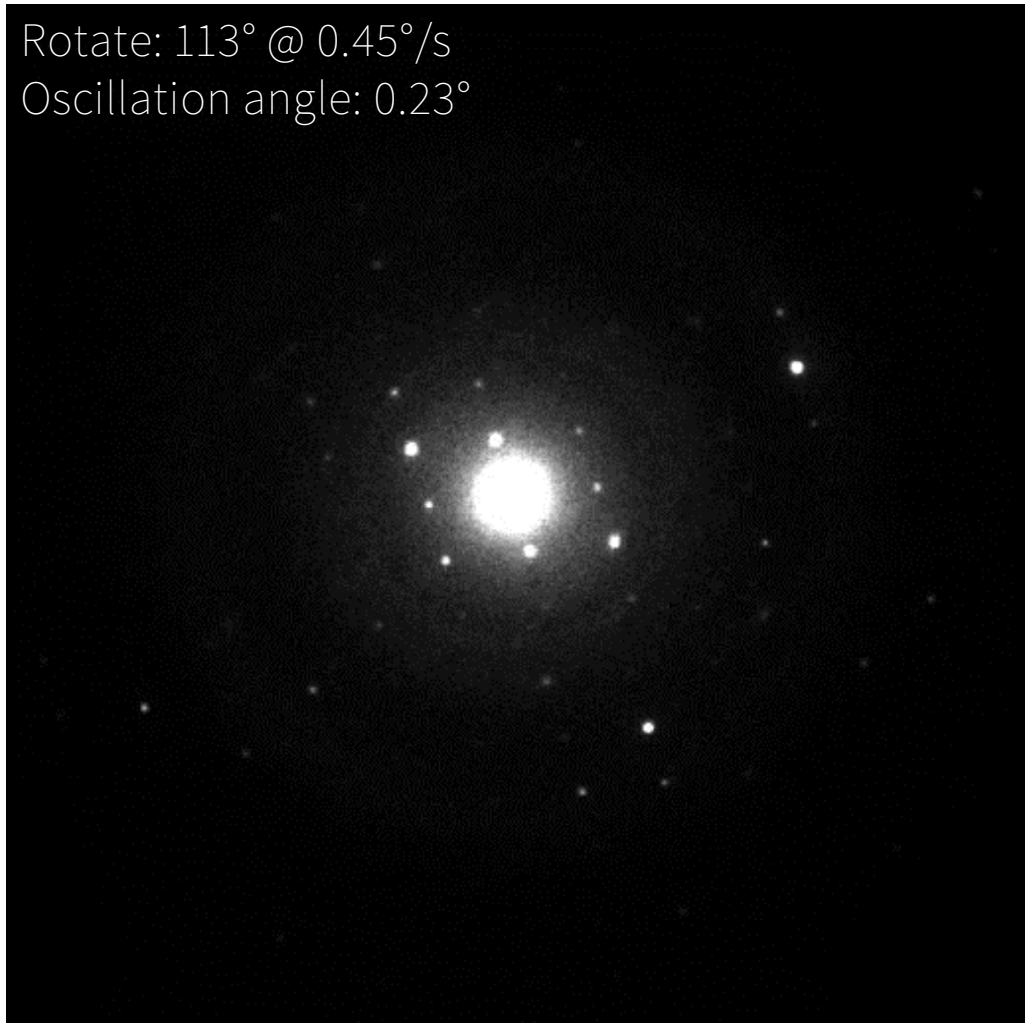
The crystal tracking problem





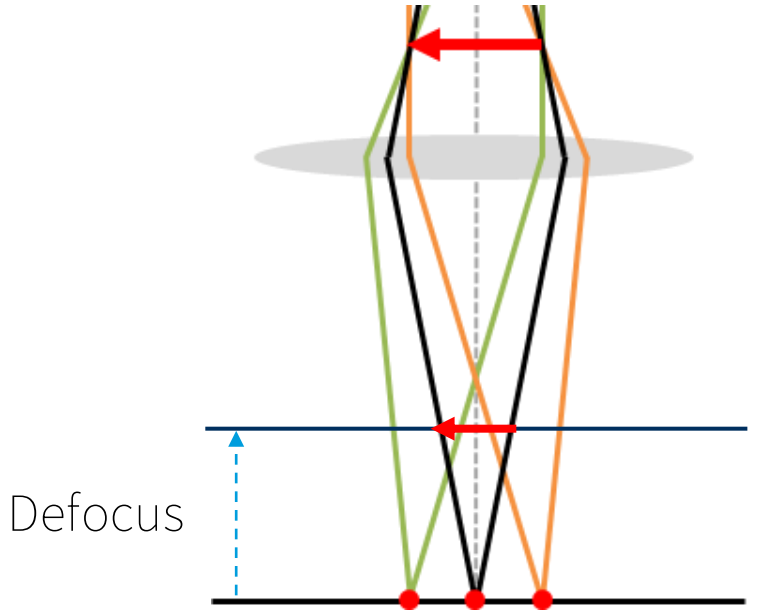
Crystal movement

Using defocus for tracking



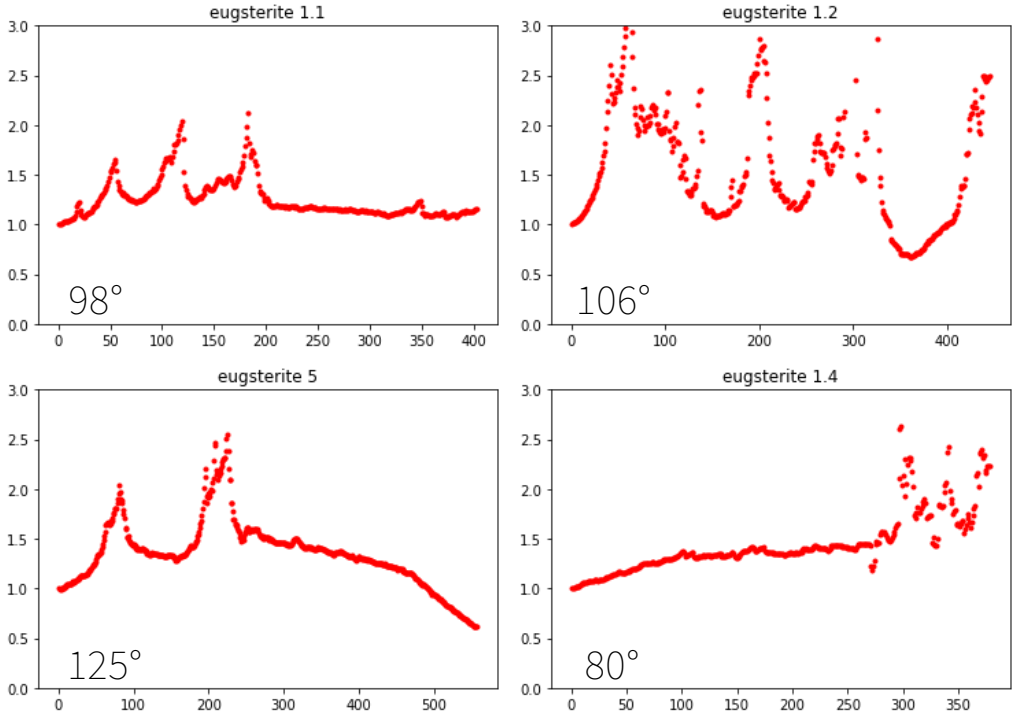
Tracking

- Defocus every 10th image (IL1)
- Manually control stage position

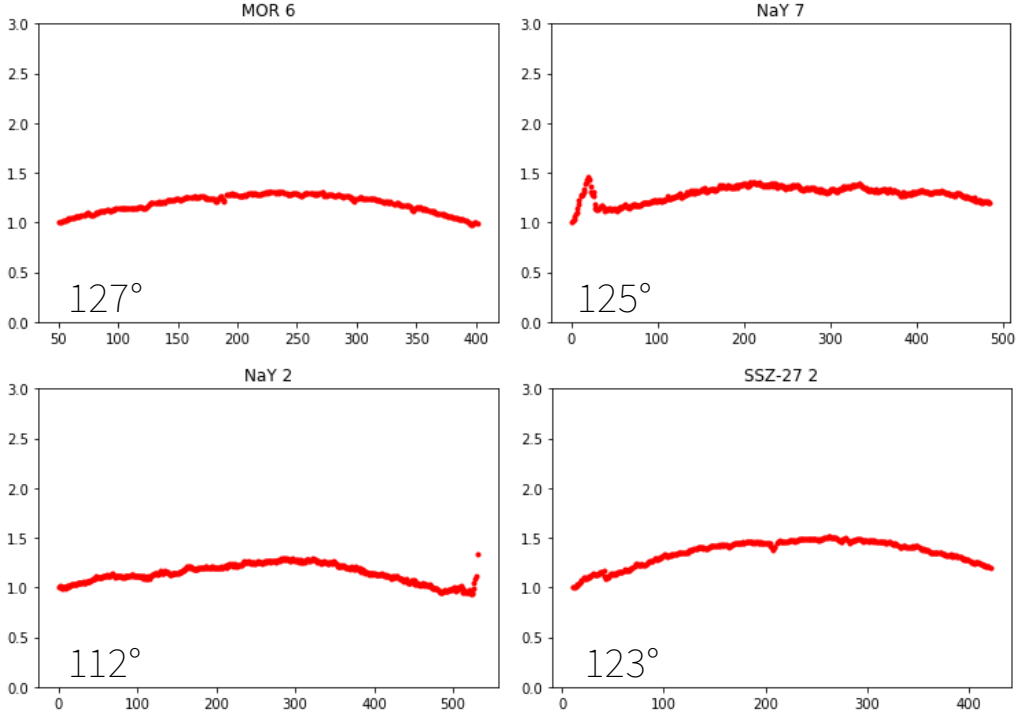


Cichocka *et al.*, *J. Appl. Crystallogr.* 51 (2018): 1652–61

Effect on image scale

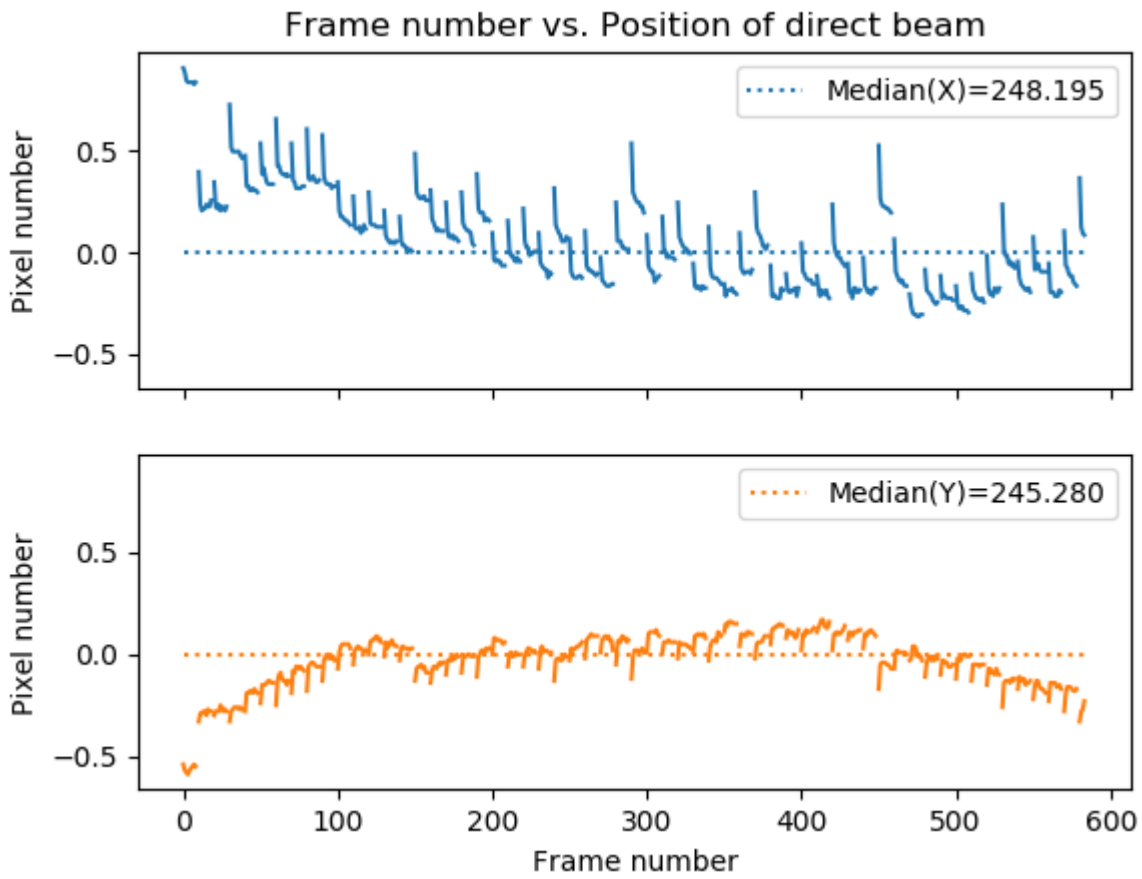


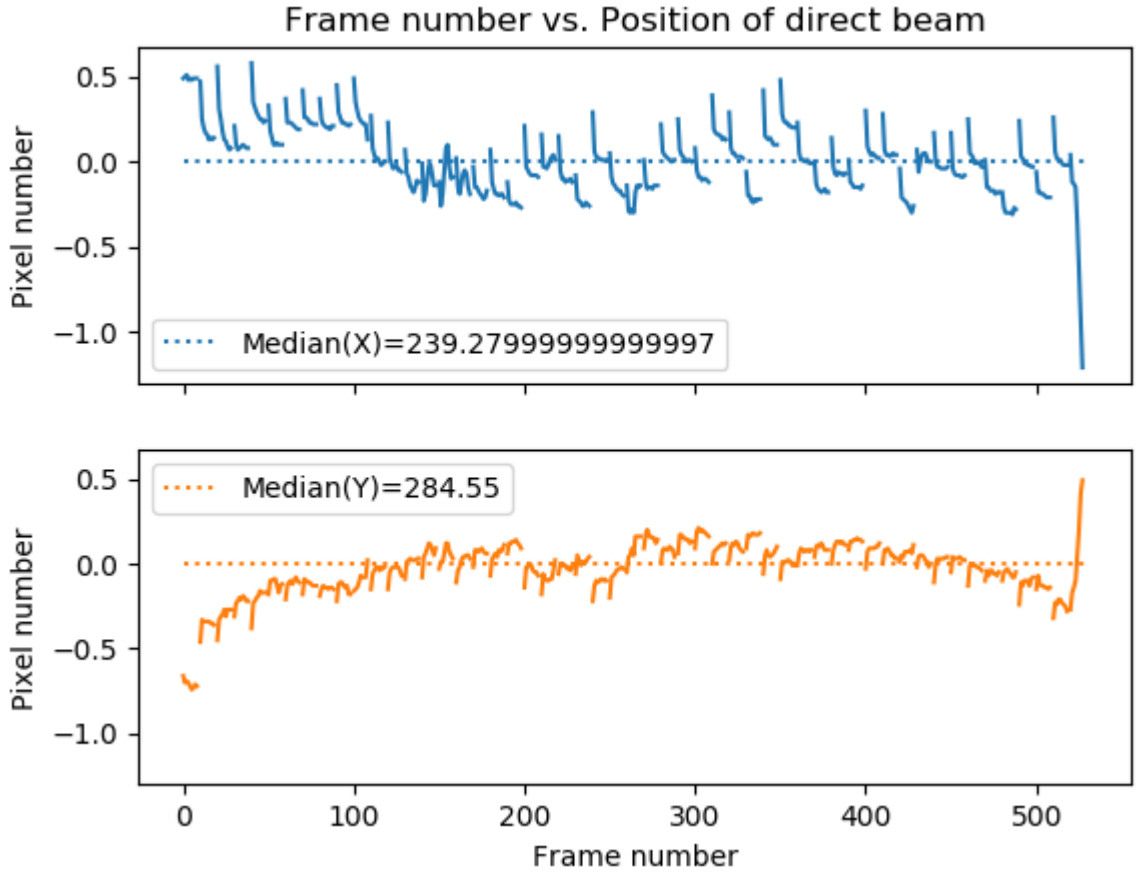
Before



After

Effect on beam position





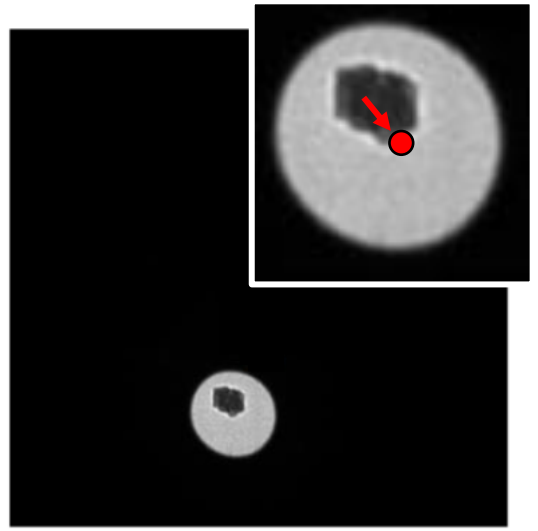
Exposure time: 0.5 s, beam relaxed before experiment

Automated crystal tracking strategy

Frame 1 (500 ms)

Frame 9

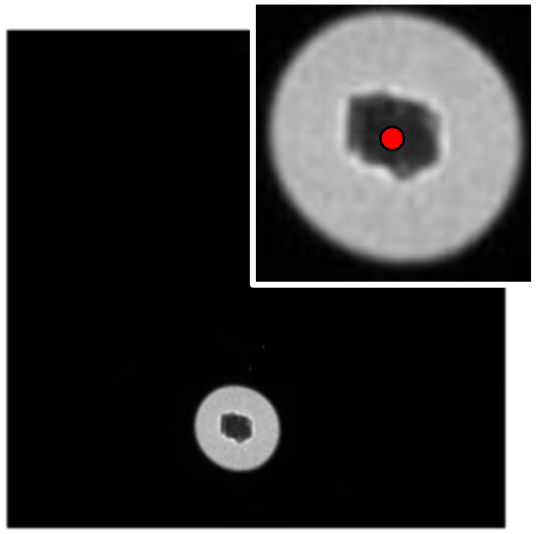
Defocus DP



Frame 10 (20 ms)

Calculate $d(x, y)$ via image analysis

Center crystal via deflectors



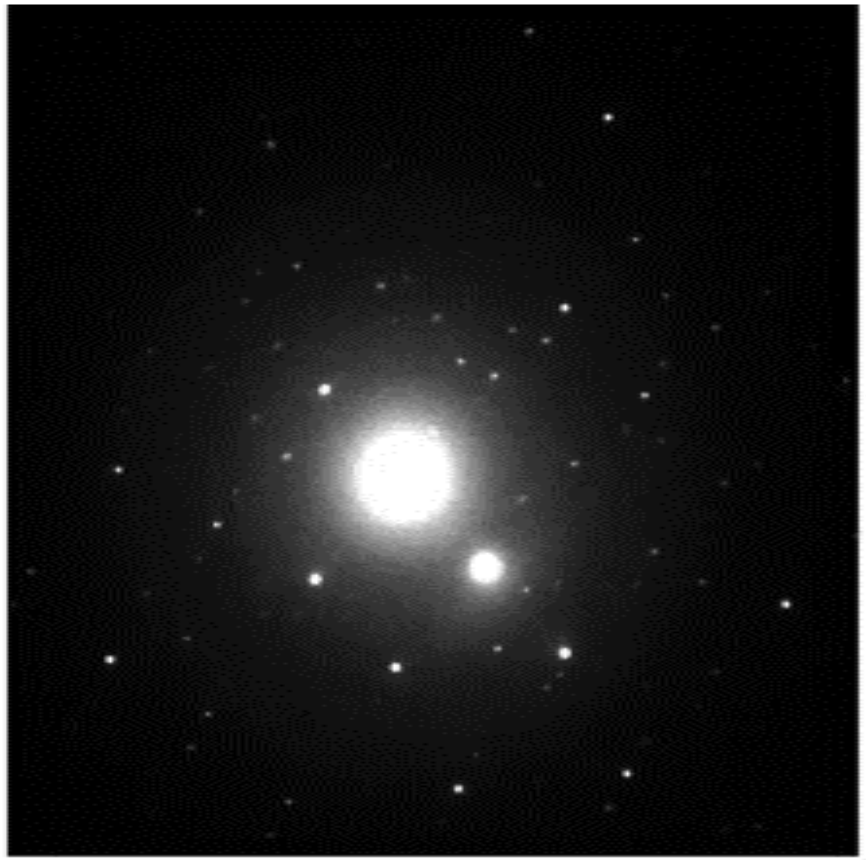
Refocus DP + wait until 500 ms passed

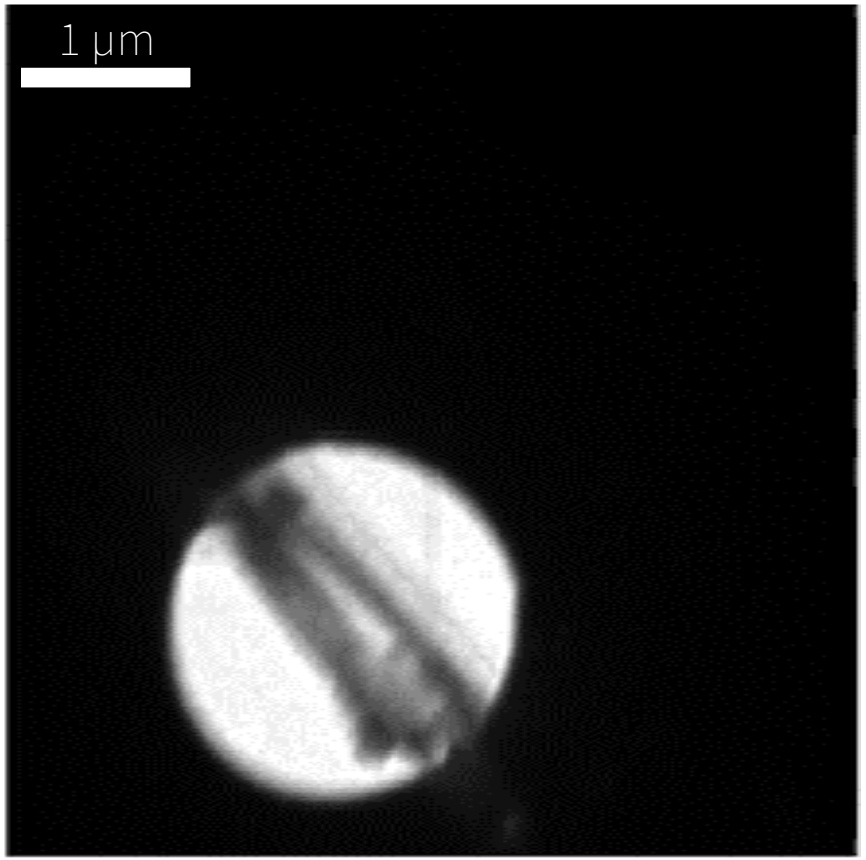
Continue data collection

Frame 11

Wang *et al.*, *IUCrJ* 6 (2019), 854-867

Automated crystal tracking strategy



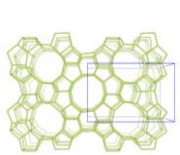
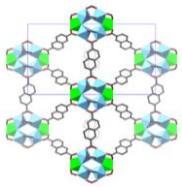
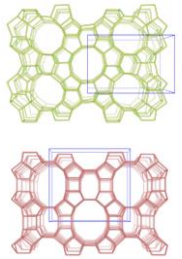
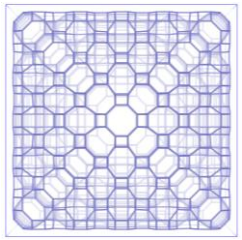
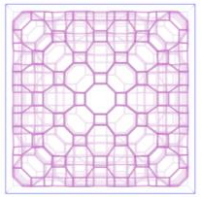


JEOL 2100-LaB_6 @ 200 kV (Timepix)
Rotation: -44.0 to 47.4° @ 0.76°/s (91.4°)
Exposure: 0.5 s, oscillation angle: 0.39°

Wang *et al.*, *IUCrJ* 6 (2019), 854-867

How to deal with all the data?!

Sample	ZSM-5	PCN-416	ZSM-5 + mordenite	PST-20 + ZSM-25
Data collection time	6 h	2 h	2 h	4 h
Rotation (mean) /°	11.9	4.0	16.3	16.1
Rotation (max) /°	76.2	44.4	73.6	78.5
# crystals	250	139	123	148
# data sets > 5°	126	66	89	99
# data sets > 20°	43	15	33	42

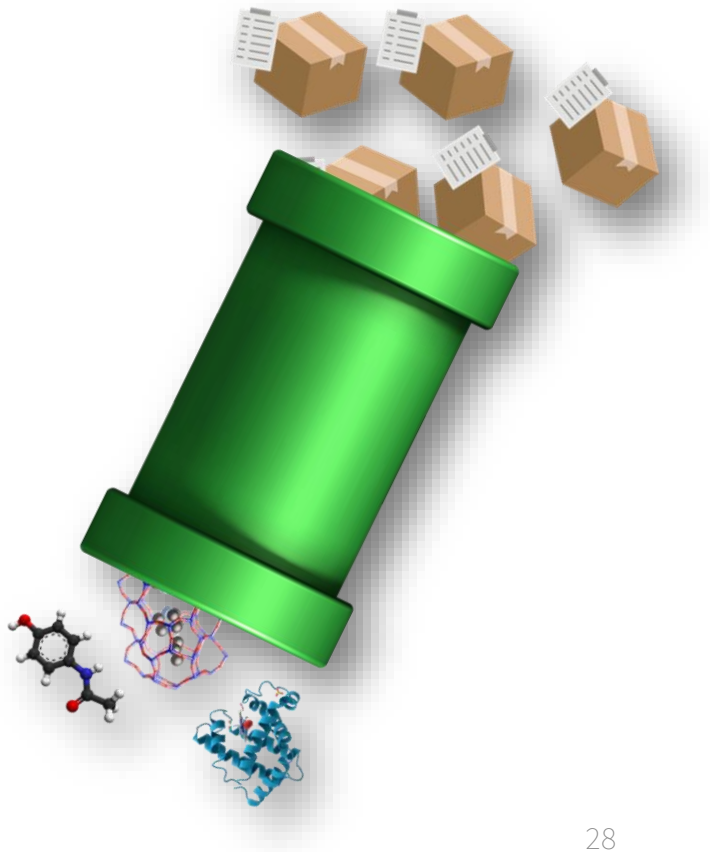






Wang *et al.*, *IUCrJ* 6 (2019), 854-867

netherlands eScience center

Outline

- What is SerialRED?
- Crystal screening techniques
- Automated data collection
- Ensemble data processing
- Data processing demo



Python data reduction pipeline

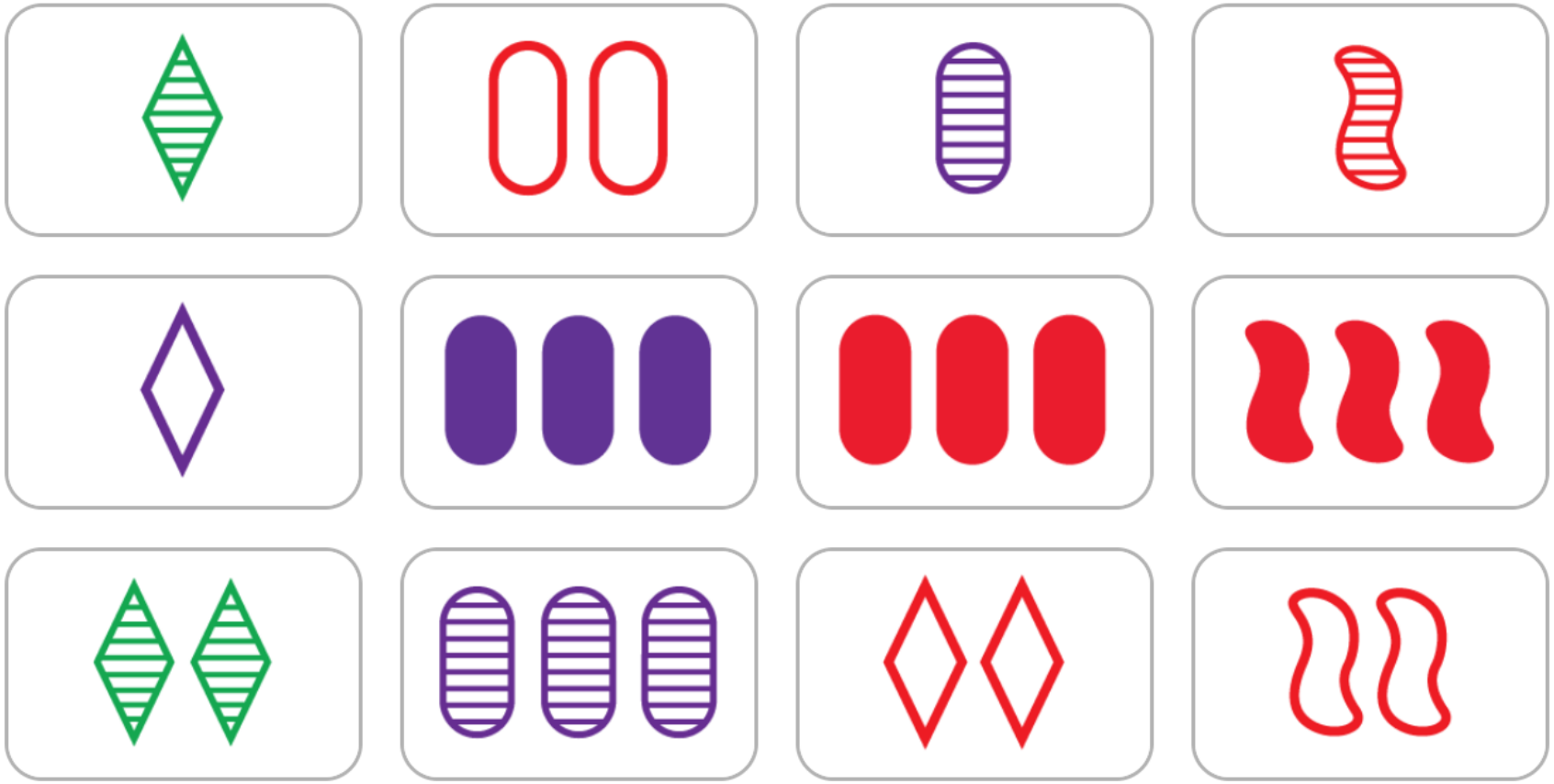
https://github.com/instamatic-dev/edtools

⬅ Data + metadata

1. Data conversion (image format, software input)
2. Data reduction (XDS)
3. Data scaling (XSCALE)
4. Cluster analysis (filter outliers, grouping)
5. Data merging (XSCALE)
6. Structure determination + refinement (SHELXS/SHELXL)

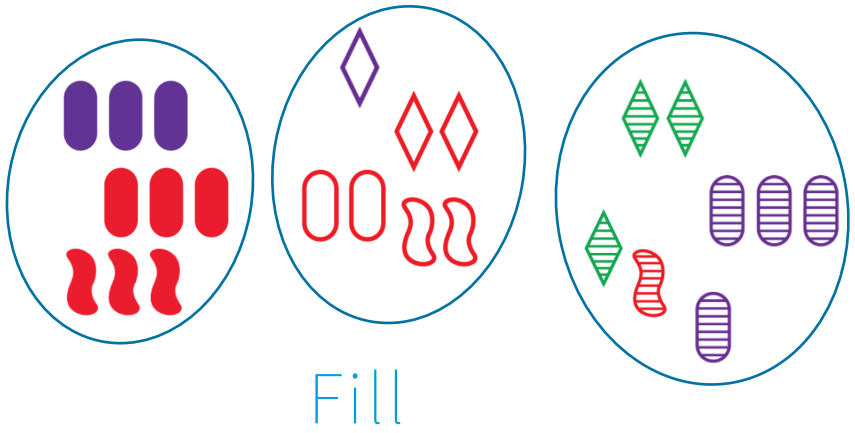
➔ Atomic model (higher precision)

What is the natural grouping of these?

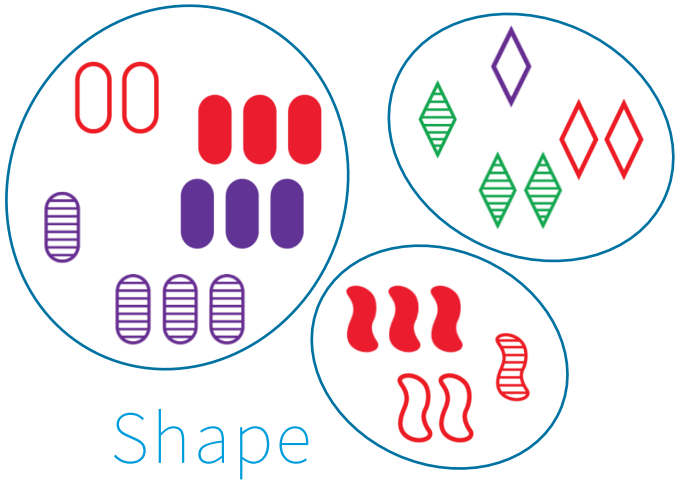


www.setgame.com, daily puzzle 2021-07-27

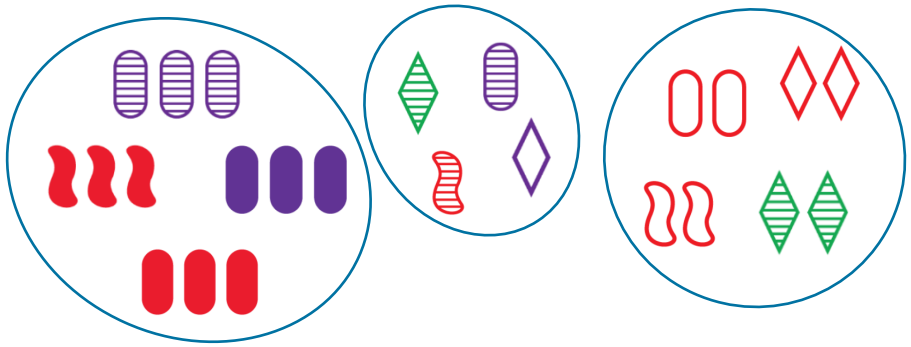
Grouping is subjective



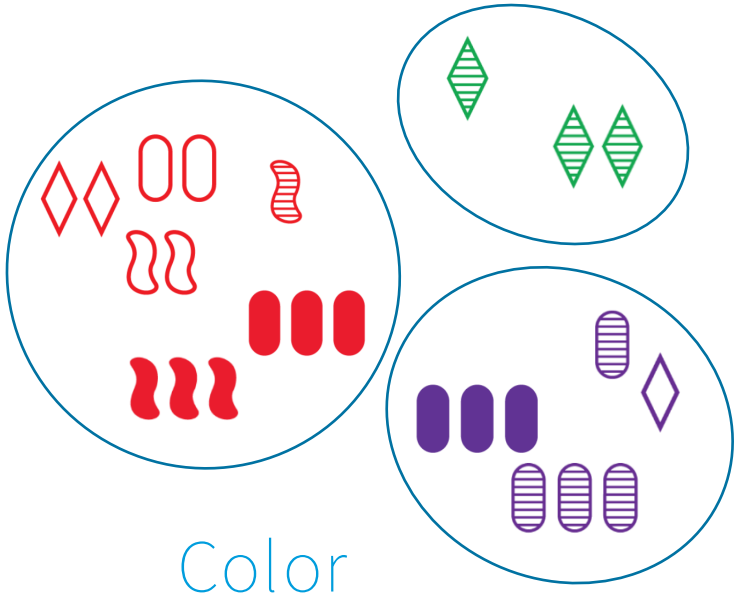
Fill



Shape

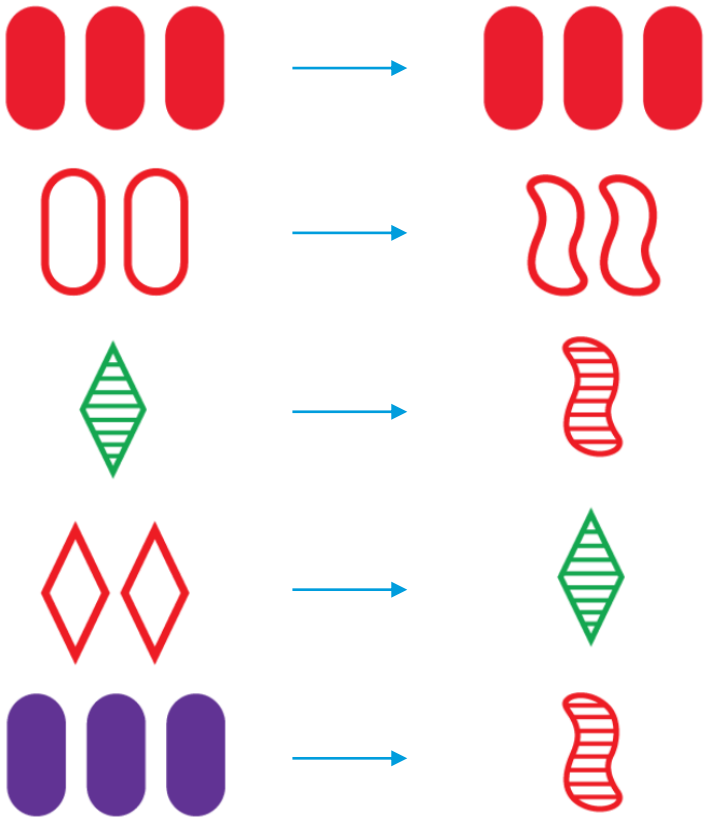


Number



Color

Similarity is defined as a distance



Distance = 0 (same!)

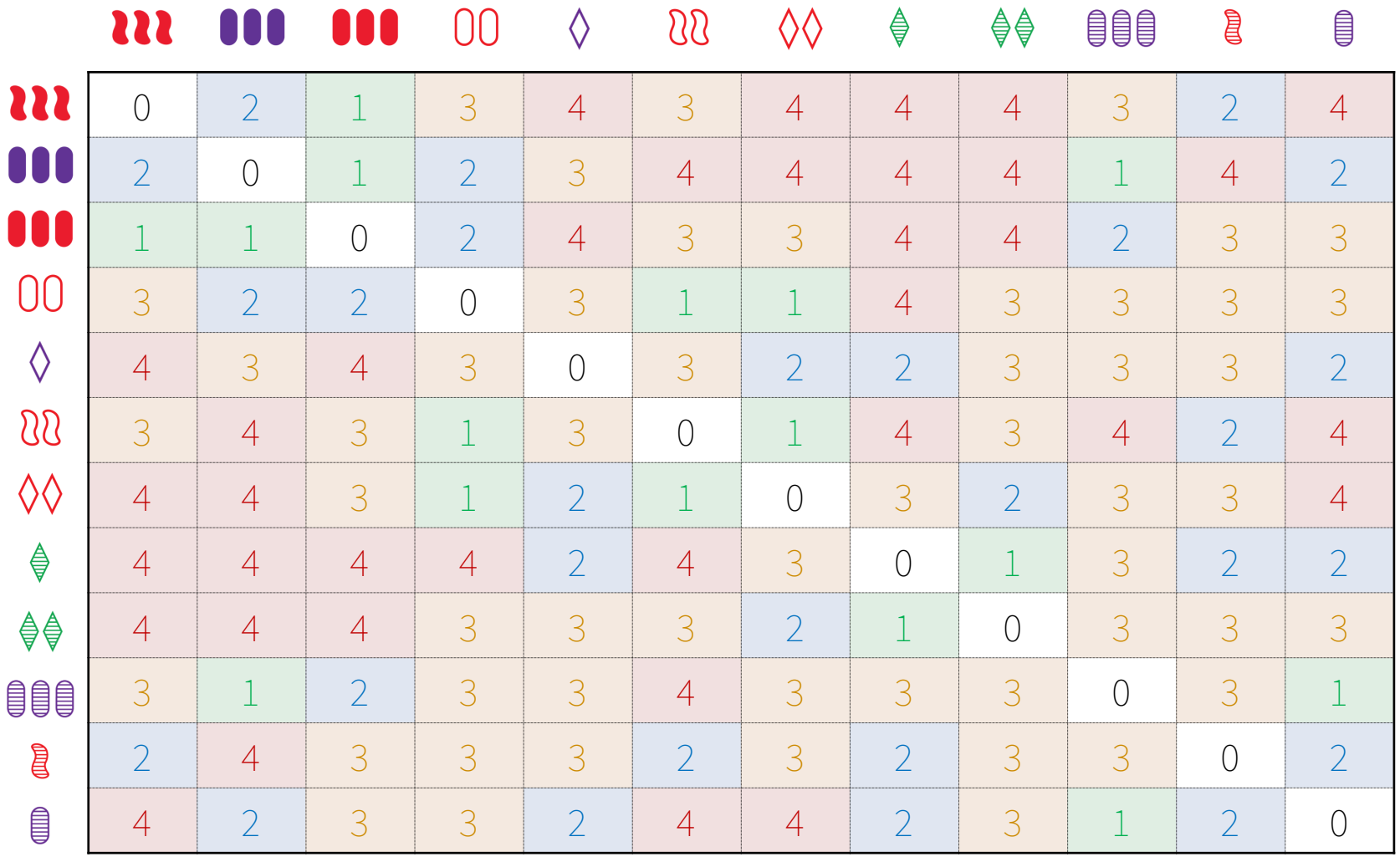
1 (shape)

2 (color, shape)

3 (number, fill, color)

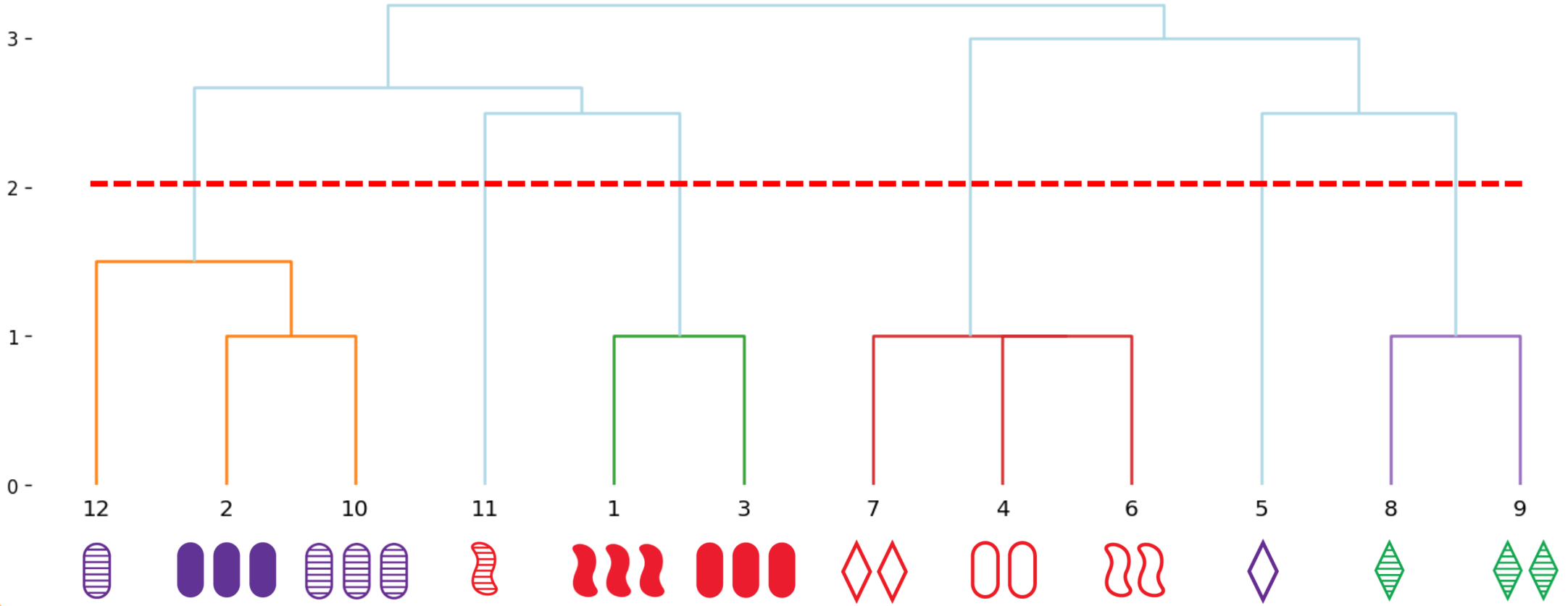
4 (number, shape, fill, color)

Similarity matrix



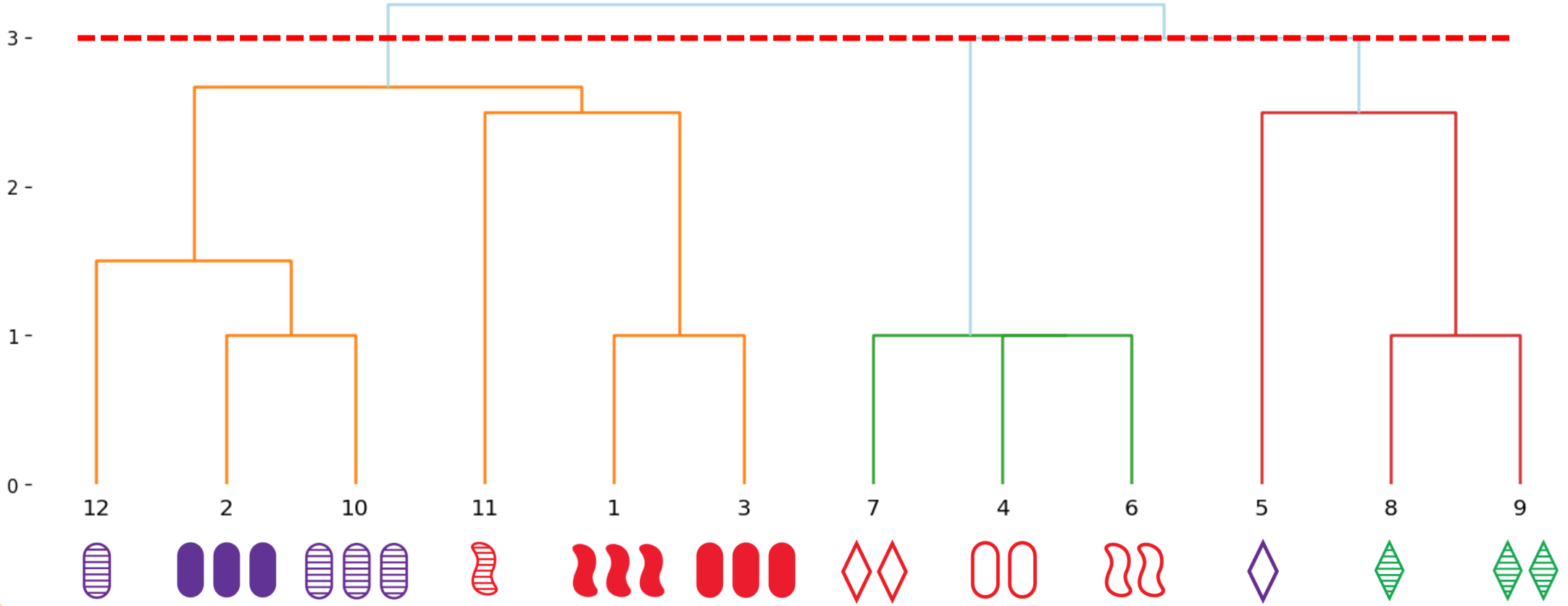
	0	2	1	3	4	3	4	4	4	3	2	4
	2	0	1	2	3	4	4	4	4	1	4	2
	1	1	0	2	4	3	3	4	4	2	3	3
	3	2	2	0	3	1	1	4	3	3	3	3
	4	3	4	3	0	3	2	2	3	3	3	2
	3	4	3	1	3	0	1	4	3	4	2	4
	4	4	3	1	2	1	0	3	2	3	3	4
	4	4	4	4	2	4	3	0	1	3	2	2
	4	4	4	3	3	3	2	1	0	3	3	3
	3	1	2	3	3	4	3	3	3	0	3	1
	2	4	3	3	3	2	3	2	3	3	0	2
	4	2	3	3	2	4	4	2	3	1	2	0

Dendrogram



scipy.cluster.hierarchy.dendrogram

Dendrogram



scipy.cluster.hierarchy.dendrogram

Cluster analysis in crystallography

- Lattice-based
  - e.g. *shape*

  - Distance:
    - Volume
    - Cell parameters
    - Linear Cell Variation (LCV)

- $d(i,j) = |V(i) - V(j)|$
- $d(i,j) = \sum_{x \in \{a,b,c,\alpha,\beta,\gamma\}} |x(i) - x(j)|$
- $D_{ab} = [a^2 + b^2 - 2ab\sin(180 - \gamma)]^{1/2}$
  $D_{ac} = [a^2 + c^2 - 2ac\sin(180 - \beta)]^{1/2}$
  $D_{bc} = [b^2 + c^2 - 2bc\sin(180 - \alpha)]^{1/2}$

  $M_{ab}(i,j) = |D_{ab}(i) - D_{ab}(j)|/\min[D_{ab}(i), D_{ab}(j)]$
  $M_{ac}(i,j) = |D_{ac}(i) - D_{ac}(j)|/\min[D_{ac}(i), D_{ac}(j)]$
  $M_{bc}(i,j) = |D_{bc}(i) - D_{bc}(j)|/\min[D_{bc}(i), D_{bc}(j)]$

  $\mathrm{LCV} = \max(M_{ab}, M_{ac}, M_{bc})$

Foadi *et al.* (2012) *Acta Cryst. D* 69: 1617–32.

Cluster analysis in crystallography

- Lattice-based
  - e.g. *shape* 
  - Distance:
    - Volume
    - Cell parameters
    - Linear Cell Variation (LCV)

- Reflection-based
  - e.g. *color* or *fill* 
  - Distance:

$$d(i,j) = [1 - CC_I^2(i,j)^{\frac{1}{2}}$$

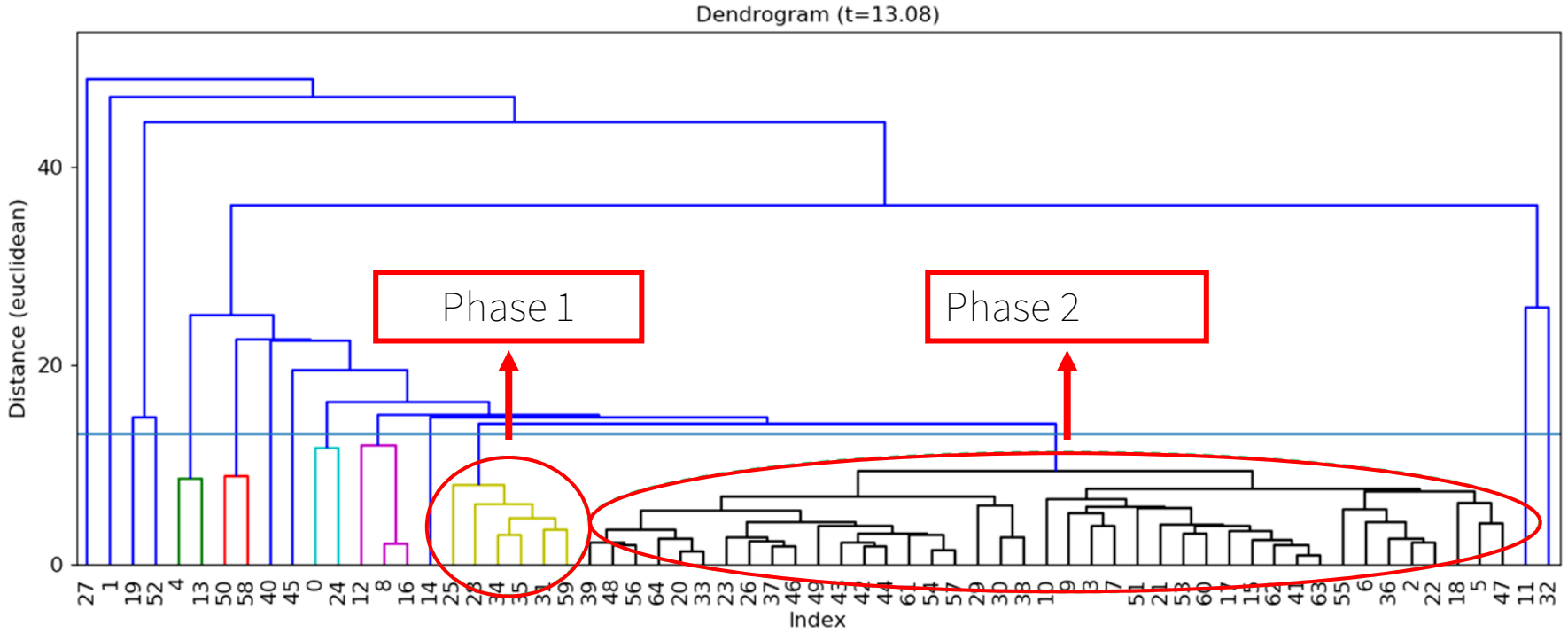
  - CC_I: correlation coefficient between common reflections

Foadi *et al.* (2012) *Acta Cryst. D* 69: 1617–32.

Giordano *et al.* (2013) *Acta Cryst. D* 68: 649–58.

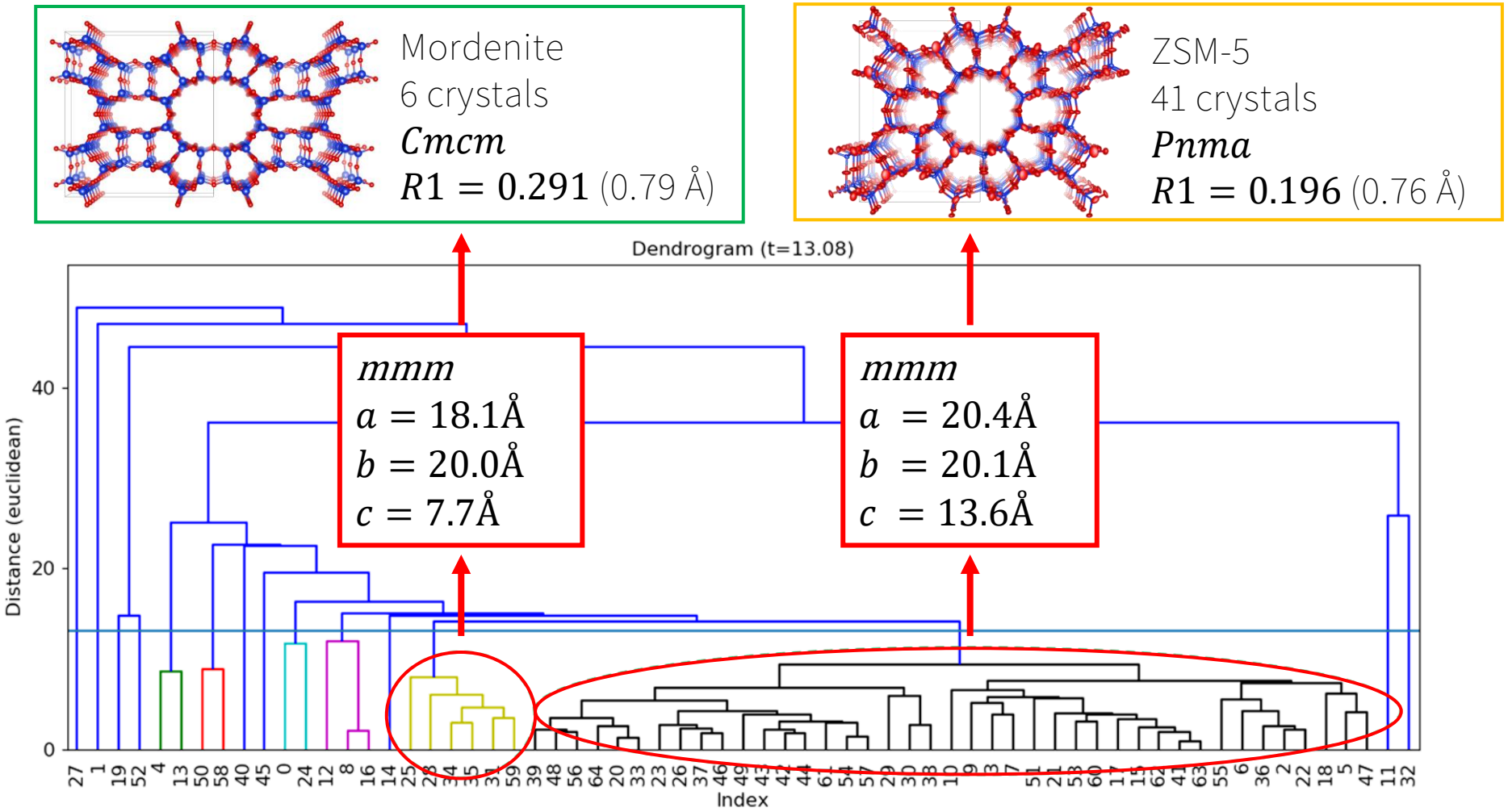
Lattice-based clustering

- Polycrystalline mixture of 2 phases
- SerialRED data from 89 crystals indexed using XDS



Wang *et al.*, *IUCrJ* 6 (2019), 854-867

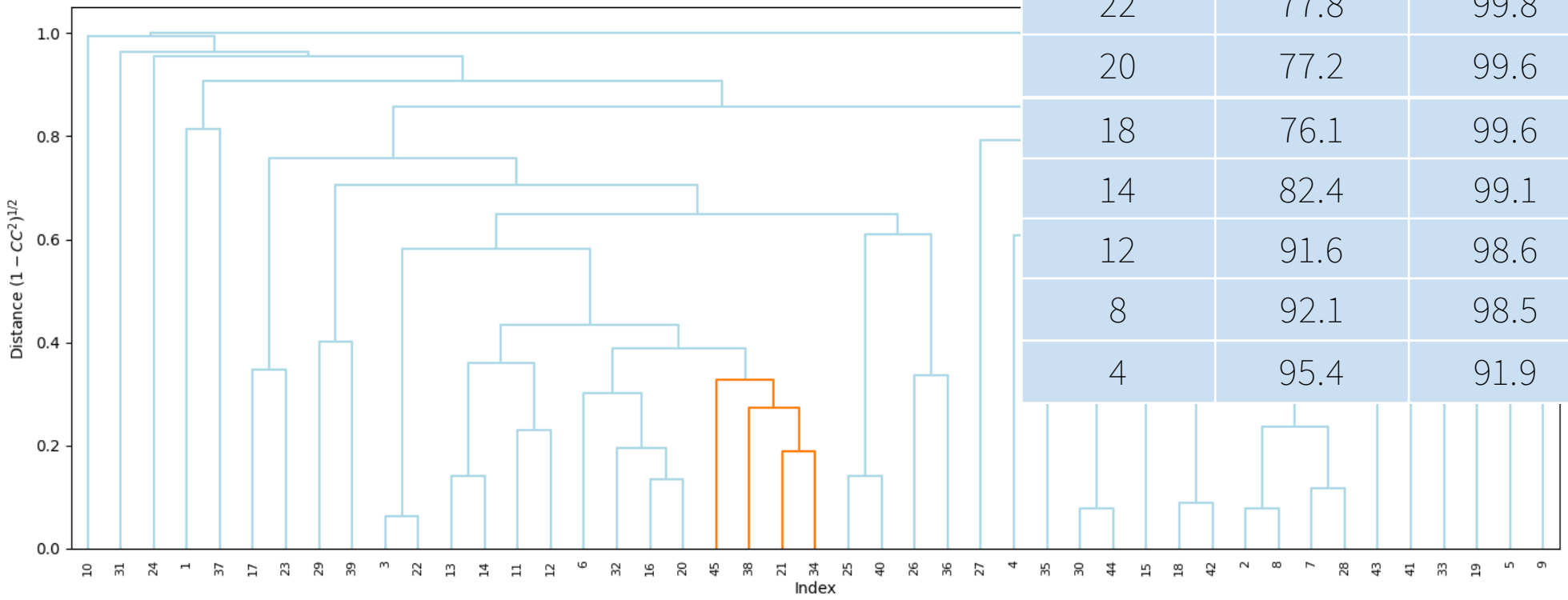
Lattice-based clustering



Wang *et al.*, *IUCrJ* 6 (2019), 854-867

Reflection-based clustering

- SerialRED data from 45 crystals (single phase)
- Processed using XDS



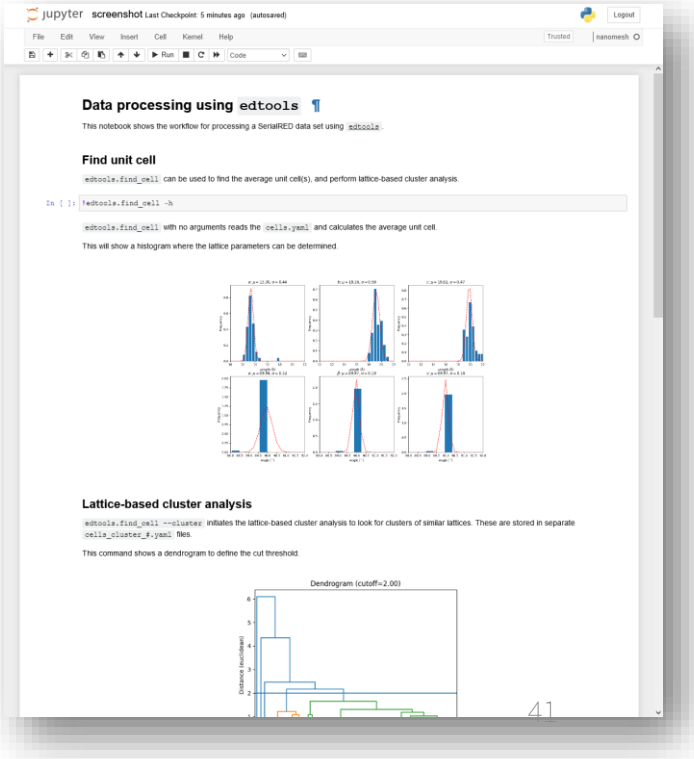
Crystals	$CC_{1/2}$	Compl. %	R_{meas}
45	79.3	99.9	0.83
34	79.3	99.9	0.77
22	77.8	99.8	0.66
20	77.2	99.6	0.66
18	76.1	99.6	0.65
14	82.4	99.1	0.63
12	91.6	98.6	0.43
8	92.1	98.5	0.41
4	95.4	91.9	0.33

Wang *et al.*, *IUCrJ* 6 (2019), 854-867

netherlands eScience center

Outline

- What is SerialRED?
- Crystal screening techniques
- Automated data collection
- Ensemble data processing
- Data processing demo



Data processing demo

- edtools: https://github.com/instamatic-dev/edtools
  - `pip install edtools`

- Data/notebooks: https://zenodo.org/record/5176268

netherlands eScience center

Let’s stay in touch

www.eScienceCenter.nl

info@esciencecenter.com

+31 (0)20 460 4770