AI for automated data collection & analysis - an ESRF perspective

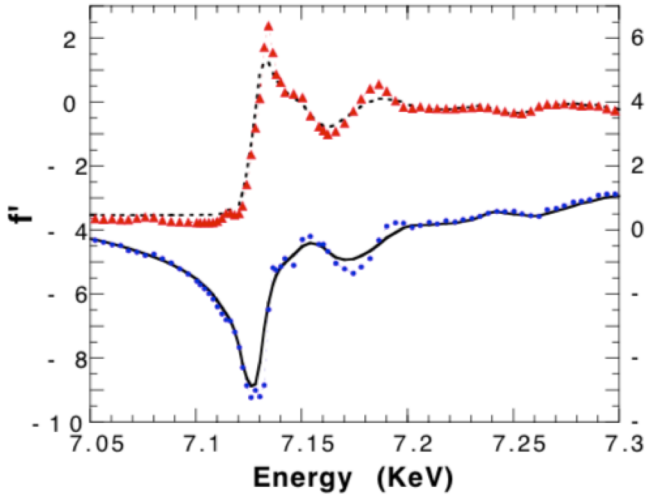
ESRF | The European Synchrotron

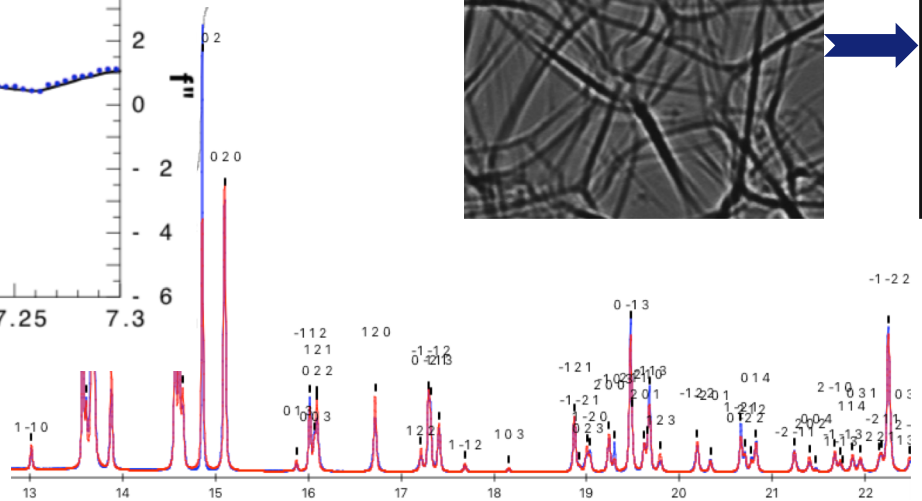
Vincent Favre-Nicolin
Algorithms & scientific Data Analysis
ESRF / Experiments division

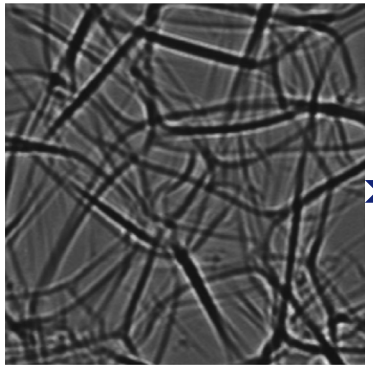
STREAMLINE

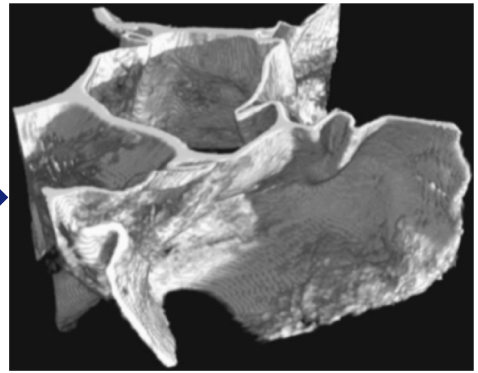
STREAMLINE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870313

SYNCHROTRON DATA ANALYSIS, CIRCA 2000





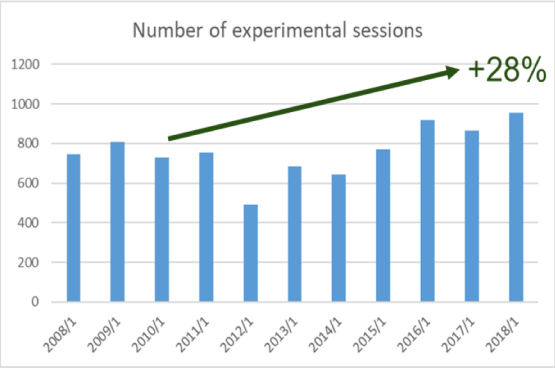


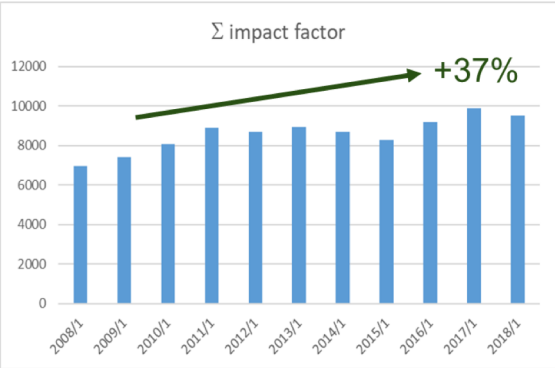


20 years ago analysing data was simpler:

1. Collect data (absorption spectrum, powder pattern, image..)
2. Fit data & determine sample structure, reconstruct 3d image (explicit modelling, local or global optimisation)
3. Publish !

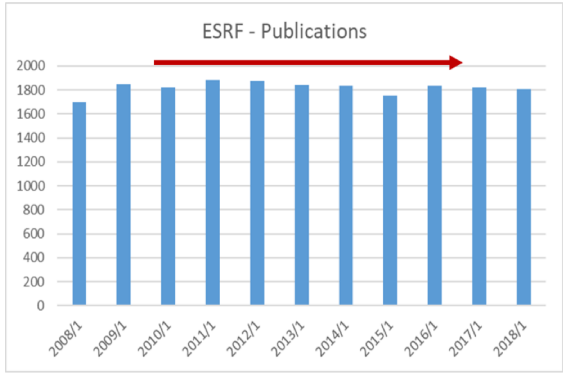
FROM DATA COLLECTION TO DATA EXPLOITATION





Beamtime usage and data exploitation over the 2008-2018 period

2/3 of the beamline portfolio are offering " expert" emerging techniques: (ptychography, coherent diffraction imaging, serial crystallography, diffraction-contrast tomography, etc.)



Dramatic increase of raw data produced by most beamlines

- #publications is stalling despite the increase in the #experiments !
- Delay experiment-publication=3years?
- Too much data ?
- No time to analyse ?
- Lack of good decisions during experiments ?

TOWARDS BIG DATA

Conservative estimate:

2020 – **6 PB**

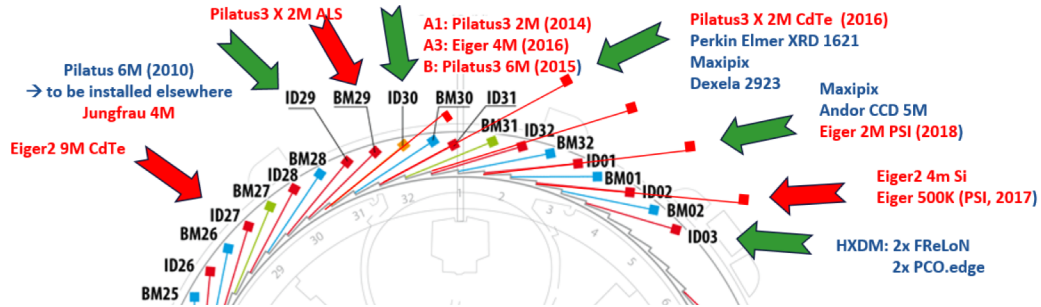
2021 – **20 PB**

2022 – **30 PB**

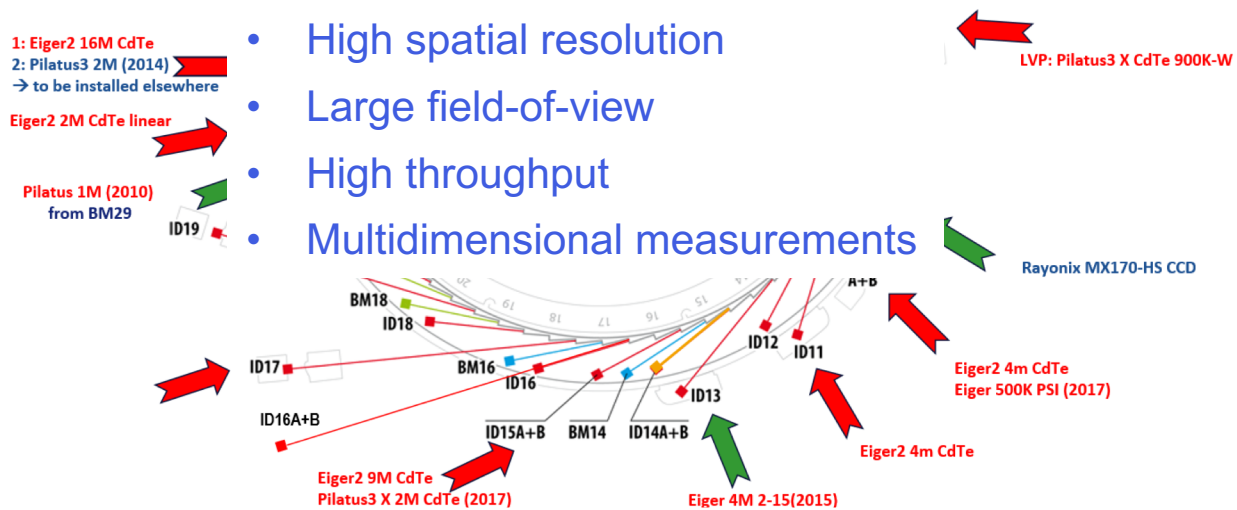
2023 – **45 PB**

Beyond imaging:
Ability to measure thousands, millions of spectra, powder patterns,…

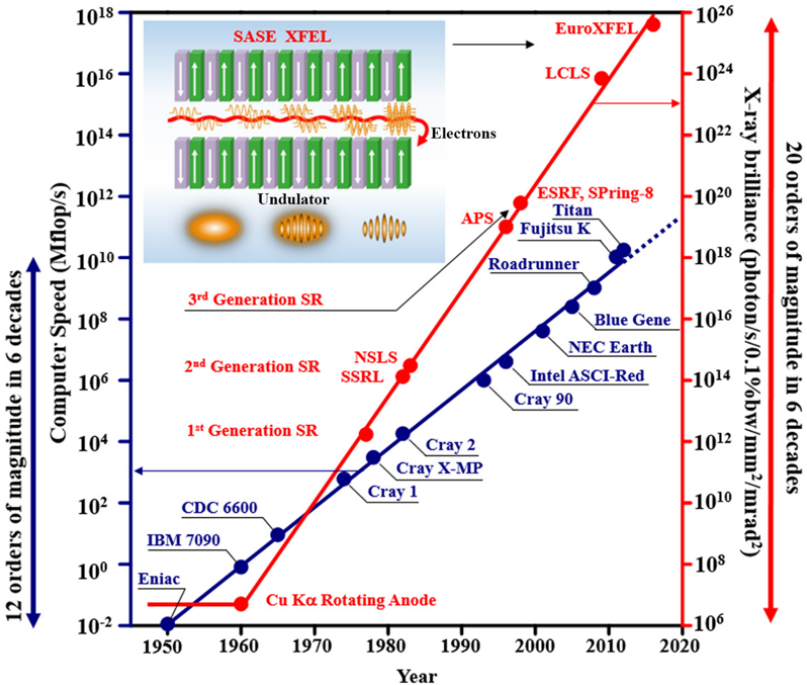
Detector portfolio for the scattering beamlines



- Time-resolved experiments
- High spatial resolution
- Large field-of-view
- High throughput
- Multidimensional measurements



CHALLENGES FOR MODERN SYNCHROTRON DATA ANALYSIS



- nb neurons in human brain: 10^{12}
- nb synapses 10^{14}-10^{15}

More photons:

- 10^{12} photons/s (10^{17}-10^{18} photons/experiment)

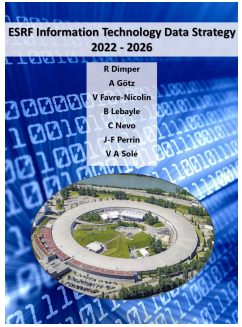
Need to:

- Process data
- Manage radiation damage
- Interpret individual datasets
- Explore encyclopaedia of data

Sample → Data collection → Data processing → Analysis (interpretation: segmentation, statistical analysis…)

THE DATA CHALLENGE REQUIRES A HOLISTIC APPROACH

- Explosion of the data volume
- Significant increase of complexity of the data sets
- limited dedicated resources



DATA EXPLOITATION

DATA reduction, pre-processing and on-line analysis activities, in collaboration with other synchrotron centres

Establishing pipelines enabling ESRF user to benefit from National and European High Power Computing Centres for DATA analysis and modelling

DATA access and handling in collaboration with many other partners.

ESRF
IT Data Strategy

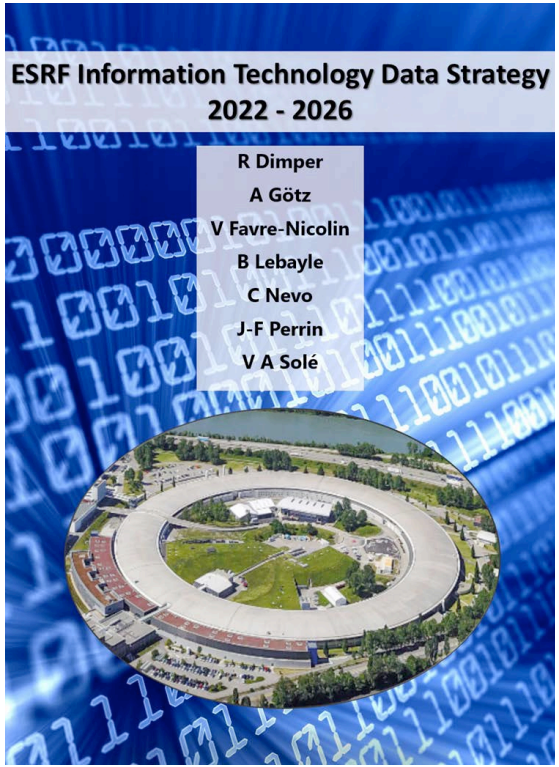
DATA POLICY

BEAMTIME USAGE

Hardware infrastructure providing DATA storage Capacity, and CPU and GPU power

Beamline and experiment control and automation. On-line data analysis

ESRF IT DATA STRATEGY



Agenda Item # 13

https://www.esrf.fr/cms/live/live/en/sites/www/home/about/documentation.html

EXPD: ALGORITHMS & SCIENTIFIC DATA ANALYSIS (ADA)

1) **Provide software for data analysis**, during or after experiments, on ESRF or external hardware
2) **Develop & improve algorithms** to handle high-throughput and big data using High Performance Computing techniques
3) **Coordinate software development** with the relevant beamlines, the ISDD software group and TID systems & communication, notably for Online Data Analysis solutions

Personnel: 6 permanent + 3 time-limited

(+7 FTE to be opened)

STREAMLINE

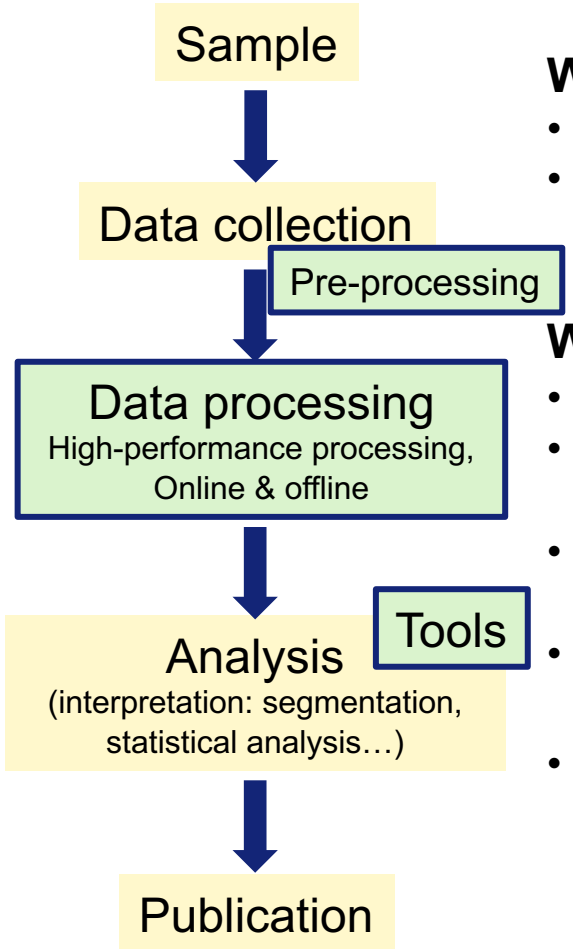
LEAPS INNOVATION

EOSC Future

Key applications (priority given to high-throughput and flagship beamlines):

- **Tomography tools** (full field, XRD-CT, XRF-CT..)
- **Scattering: fast azimuthal integration** for powder diffraction / SAXS / WAXS
- **Serial Crystallography** (ID29)
- **Coherent Imaging** techniques
- **Spectroscopy**: scanning and core spectroscopies, including machine learning

DATA ANALYSIS AS A 'SERVICE' ?



What Data Analysis As a Service <u>is not</u>:

- Doing the most of the analysis for all users
- Users with no understanding the fundamentals & limits of the techniques

What DaaS <u>is</u>:

- Handle data processing <u>when</u> it can be automatized
- Provide tools (software, computers, GPU) for **online data processing** ; enable good decisions during experiments
- Provide tools and computing resources (when possible) for **offline analysis**
- Make data available (data policy)

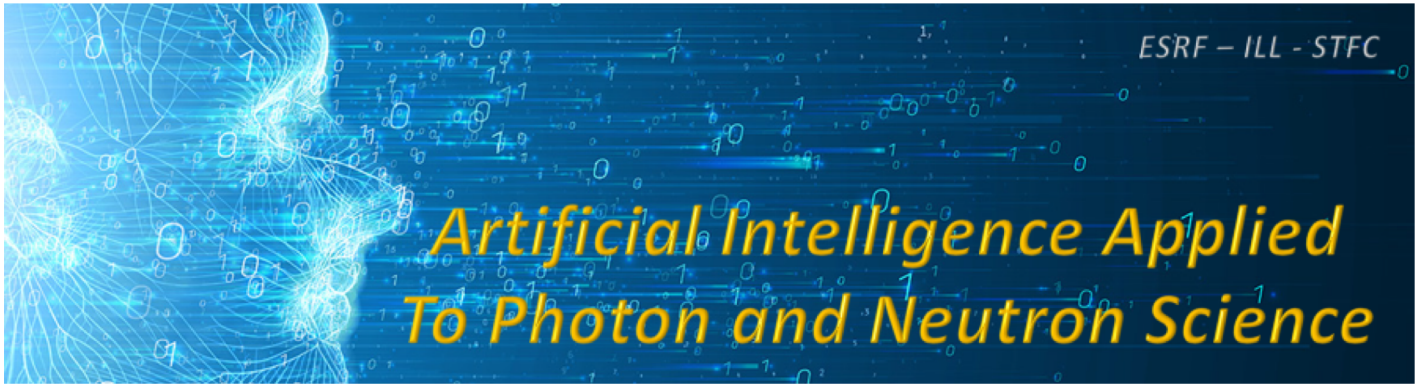
- For <u>selected</u> techniques: enable users to **focus on interpretation & scientific results** and not be overwhelmed by data handling issues

MACHINE LEARNING FOR SYNCHROTRON ?

Faster data processing

Better processing

Better interpretation



Design better experiments

Automated data collection

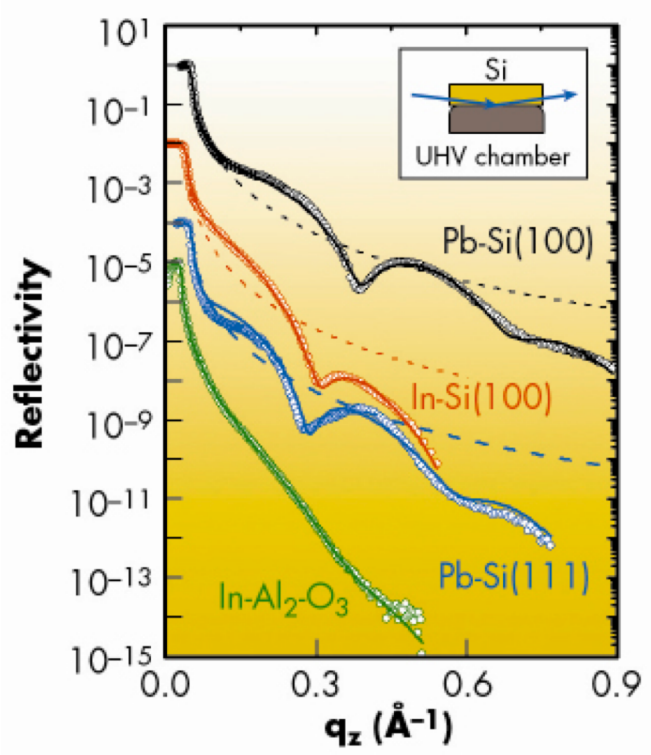
1. **Faster data processing**
2. Better processing
3. Data Analysis
4. Instrument configuration
5. Automated data collection
6. Framework for Online Data Analysis
7. Open Data

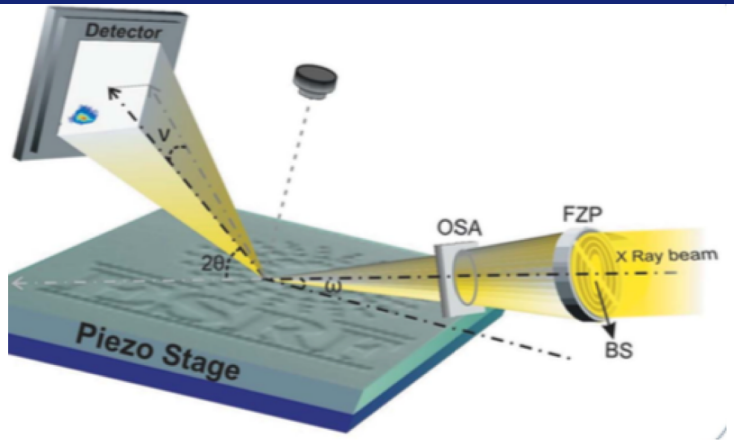
WE HAVE A PROBLEM …

Example of ESRF / ID16B
Fast holo-tomography:
In-situ experiments can be done at lower resolution every 7s for up to two hours, i.e. 2000 scans, and this can be repeated three to five times a day, possibly up to eight times with improved alignment procedures.
1 scan currently needs ~360 CPU cores.day

Assuming 3 000 scans per day, this would require a peak demand of 250 000 CPU cores

X-RAY REFLECTIVITY





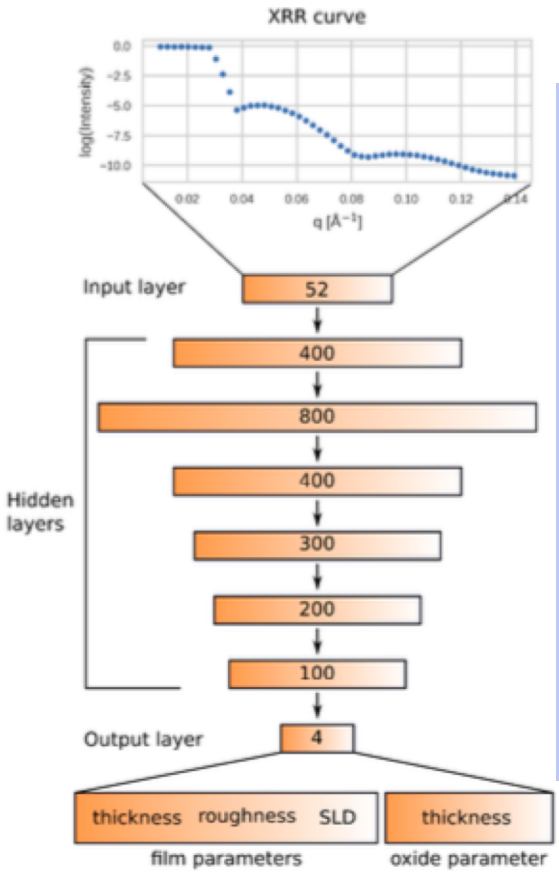
X-ray reflectivity: measure the (specular) reflected intensity allows retrieving the electronic density ρ(z) as a function of depth.

This can be **fitted** (Monte-Carlo, least squares, genetic algorithms) to retrieve the thickness and roughness of layers.

Fast (continuous) scanning allows with a bright beam allows to scan surfaces up to 10^3-10^4 positions/s, allowing to acquire reflectivity curves for thousands-millions of points.

Phys. Rev. Lett. 98, 116101 (2007)
J. Appl. Cryst. **47**, 762–769 (2014)

X-RAY REFLECTIVITY FITTING WITH NEURAL NETWORK

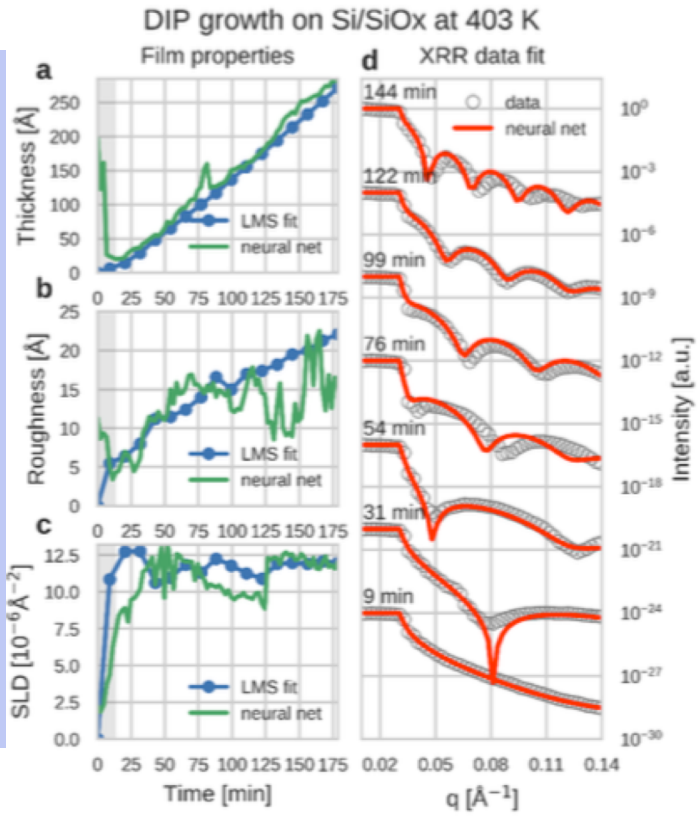


Training a neural network with a range of film parameters:

- Thickness
- Roughness
- Scattering length density
- Substrate thickness

The resulting NN allows to predict the layers parameters with a **speed of 0.03-77 milliseconds per curve** – fast enough for live analysis.

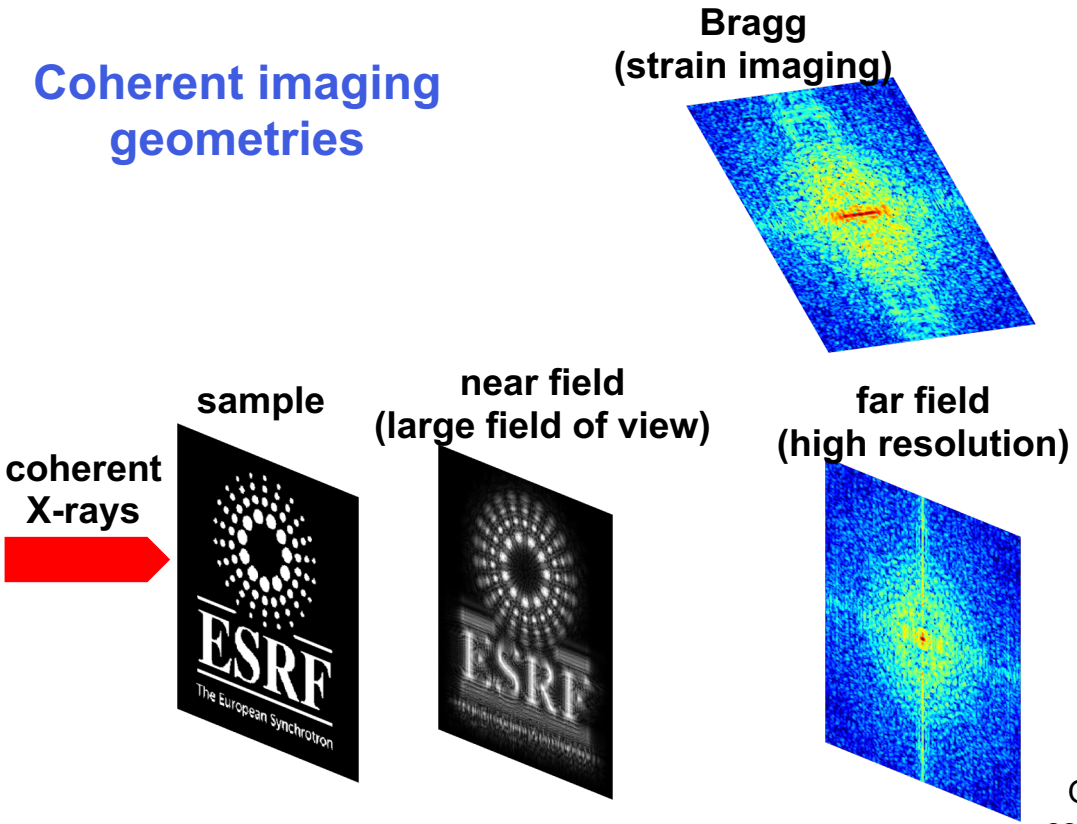
Important: predictions are limited to the range of the simulated training set.



https://arxiv.org/abs/1910.02898

COHERENT IMAGING

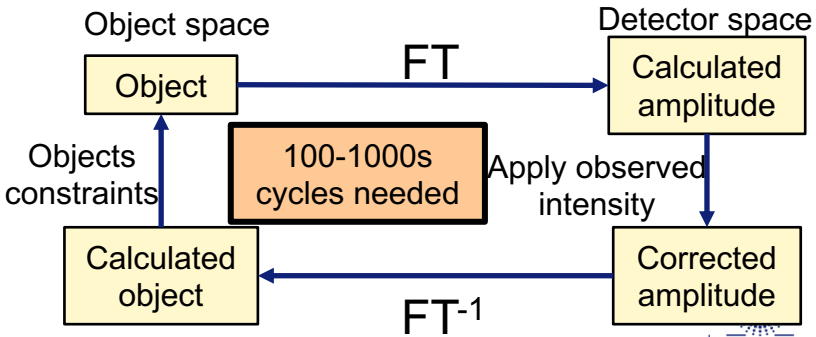
Coherent imaging geometries



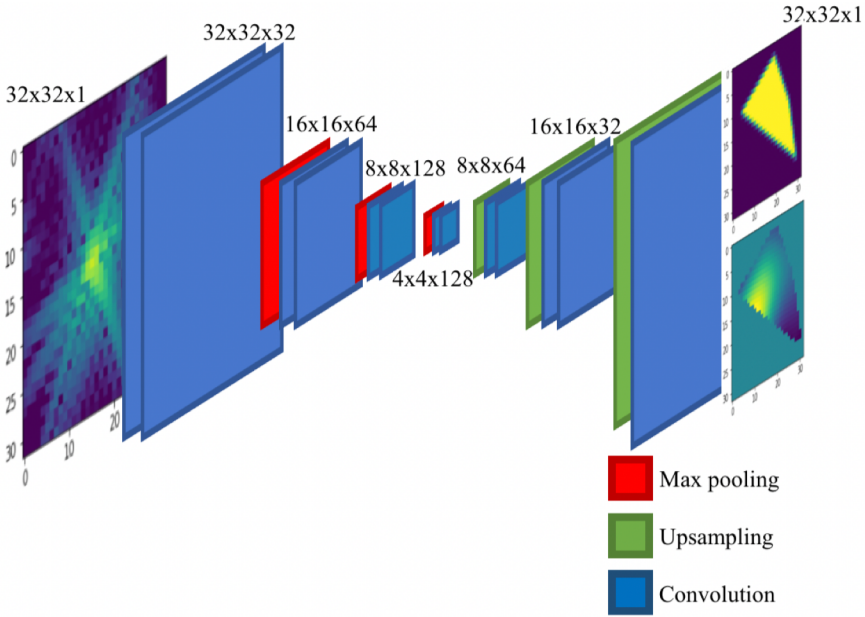
In coherent imaging techniques, either the **near-field** or **far-field projection** is recorded, being related by Fourier transform(s) to the object density (refraction index, strain).

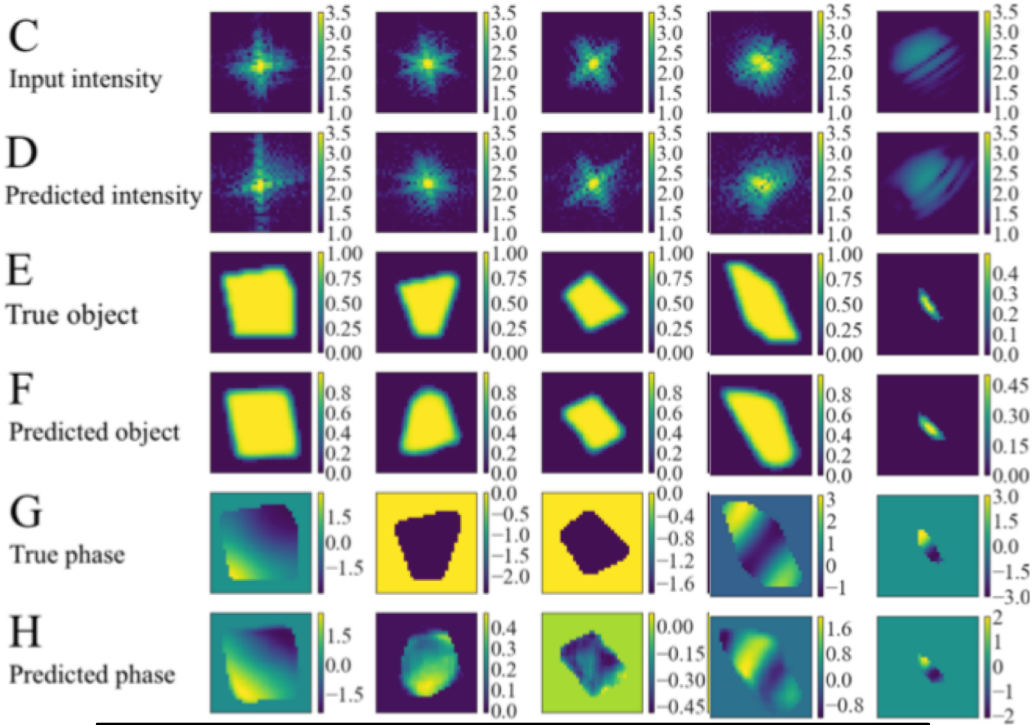
These techniques are very important for **high-resolution two and three-dimensional imaging** (resolution down to less than 10 nm).

The **phase of the scattered intensity is lost**, preventing a direct reconstruction → need **phase retrieval algorithms**.



COHERENT DIFFRACTION IMAGING





Train a neural network on a set of 180k example 2D objects (amplitude and phases).

Training time (2 GPU): ~1h

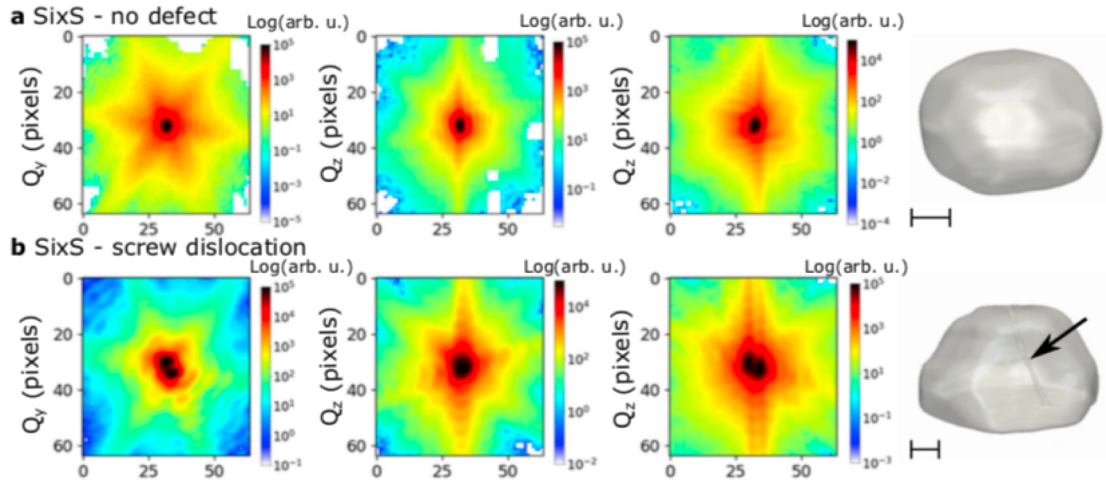
Use the NN to predict the object from diffracted intensities only. Prediction time (CPU): 2.7 ms

Caveats:

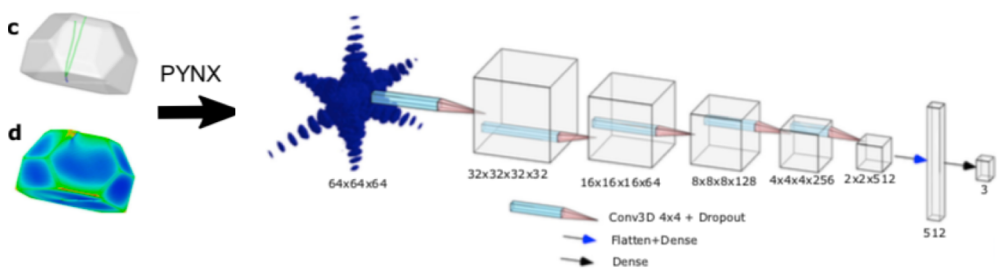
- **Real data sets often 3D, usually 200^3 to 500^3 pixels large**
- **Fine tuning of object density and phase can take longer than initial assessment**

Sci. Rep. 8 (2018), 16520

CDI– NN DEFECTS IDENTIFICATION



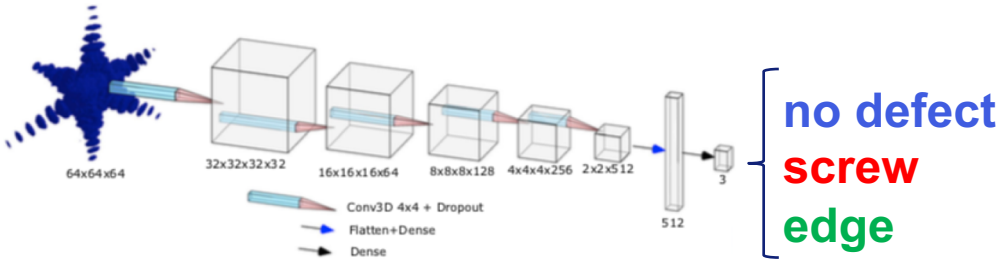
In CDI, dislocations can be identified by a split-peak in the diffraction pattern

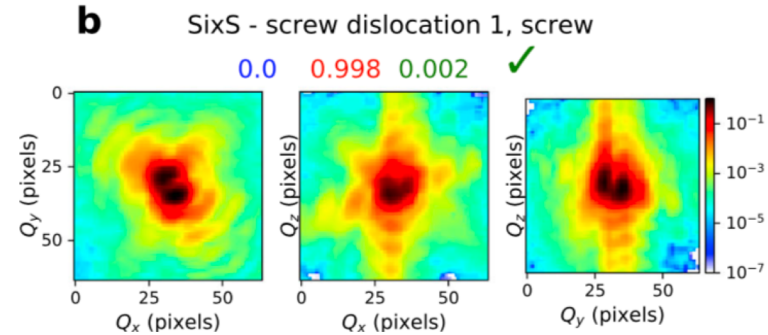


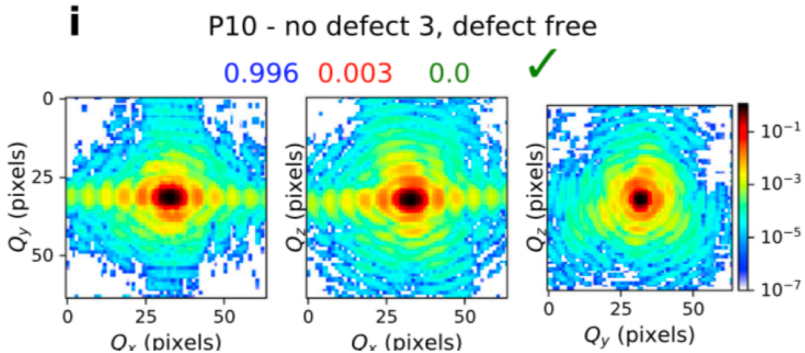
Train a NN on a variety of nano-objects shapes and for different types of dislocations (none, screw, edge).

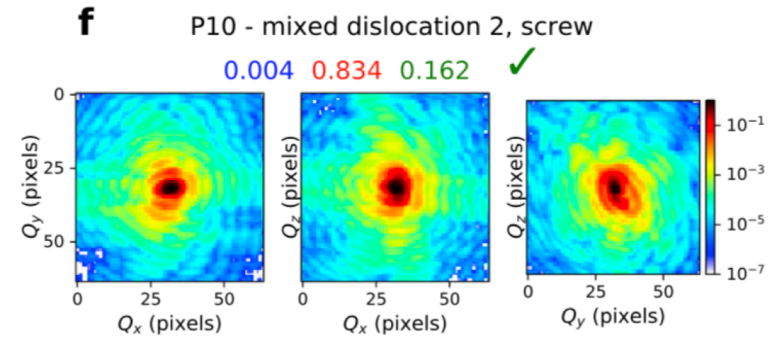
Npj Comput. Mater. 7 (2021), 115

CDI– NN DEFECTS IDENTIFICATION









The NN allows to quickly identify the presence and type of dislocation.

This could be used for quick/automated identification when scanning many particles

Npj Comput. Mater. 7 (2021), 115

ML-BASED DATA ANALYSIS LIMITS ?

Benefits of Machine Learning for data analysis are clear:

- **Fast results**
- **Nonlinear modelling**

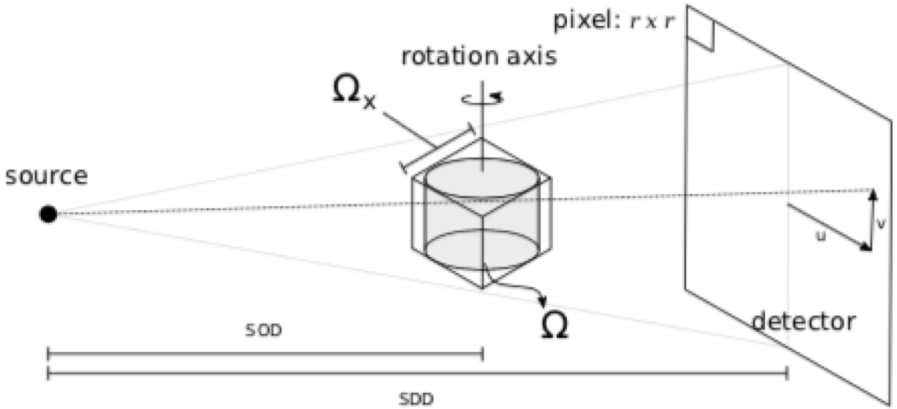
However:

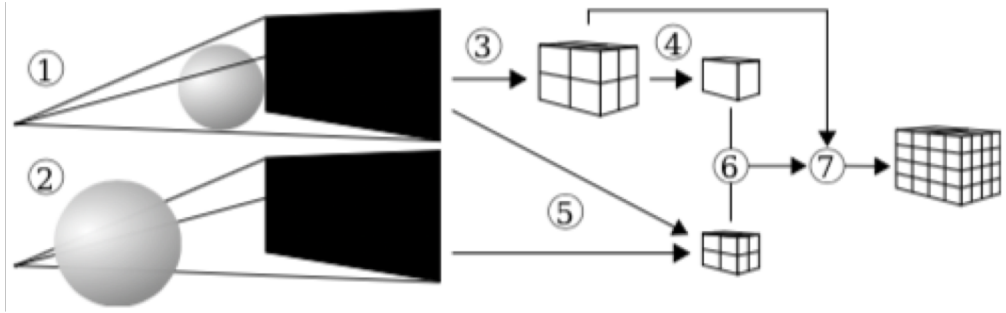
- **Results may not be better than traditional fitting..**
- **So a double approach may be required, with a fast ML step followed by a standard refinement**
- **The resources needed (human and computing) are increased (unless the approximate solution is enough)**

1. Faster data processing
2. **Better processing**
3. Data Analysis
4. Instrument configuration
5. Automated data collection
6. Framework for Online Data Analysis
7. Open Data

Key idea: try to get **more information** from the same datasets (**higher signal/noise**) compared to traditional algorithms

PHASE CONTRAST IMAGING: IMPROVE RESOLUTION





In phase contrast imaging, valuable improvements include:

- **High resolution object from low resolution image**
- **Exploiting low signal/noise data**

This would allow:

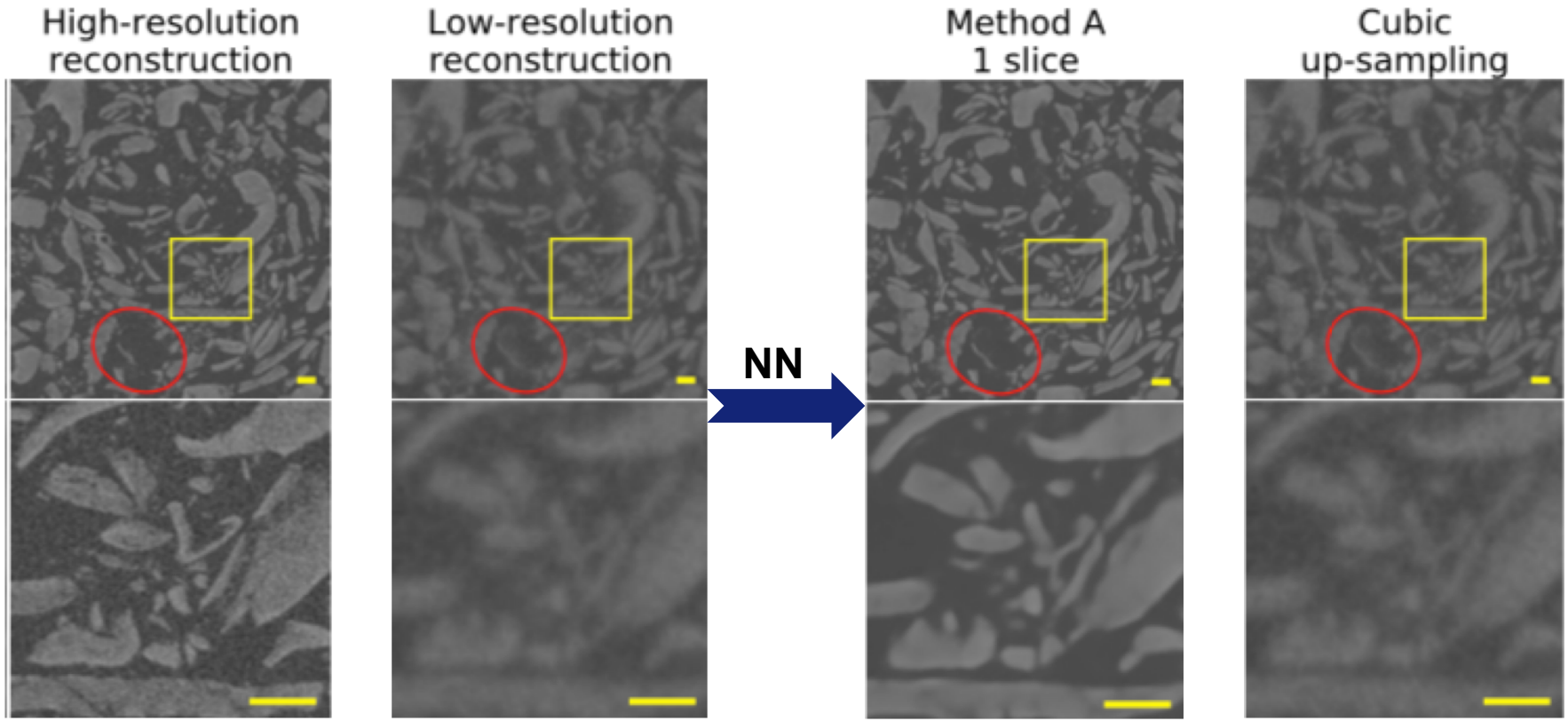
- **Faster data acquisition** (fewer projections)
- **Less radiation damage**

Method: create a neural network by acquiring a dataset at high resolution, and train it against the same data with a reduced resolution.

The NN is then tested against a different part of the same object.

Appl. Sci. 2019, 9, 2445

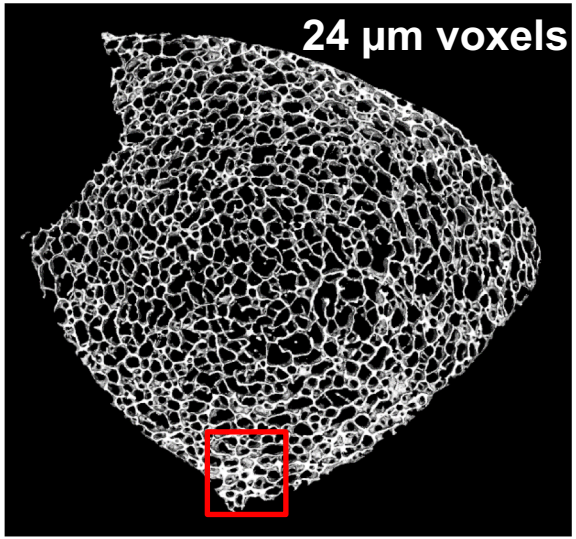
PHASE CONTRAST IMAGING: IMPROVE RESOLUTION



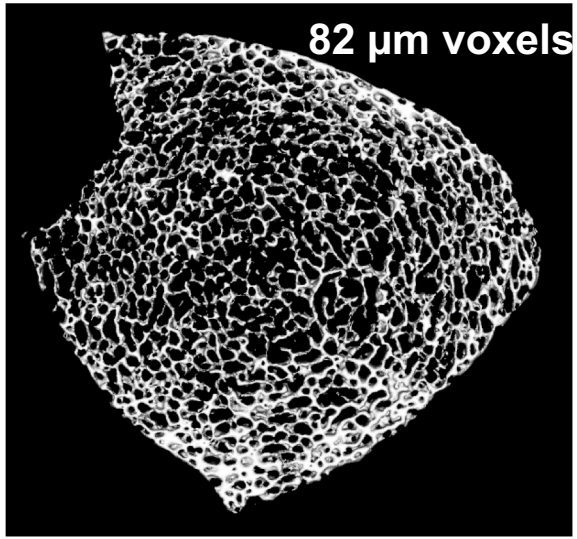
Example NN reconstruction on an oatmeal sample, training from a high resolution sub-dataset (17µm voxels) to reconstruct low-resolution ones (68µm voxels)

Appl. Sci. 2019, 9, 2445

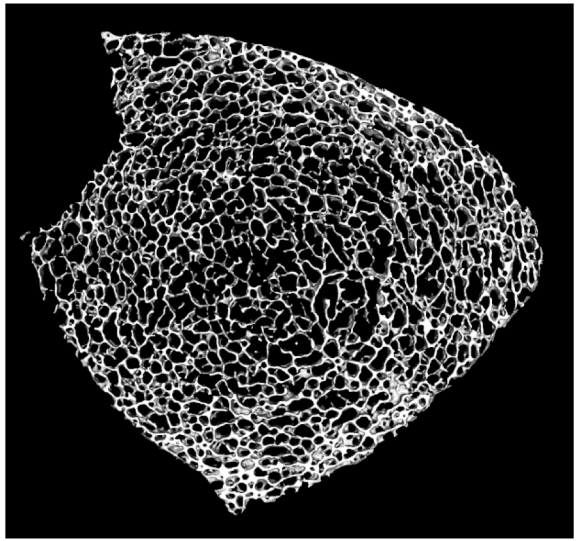
PHASE CONTRAST: IMPROVE RESOLUTION *IN VIVO*



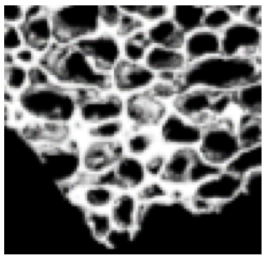
High Resolution (from µCT)

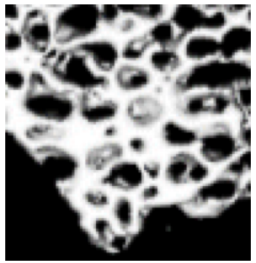


Low Resolution

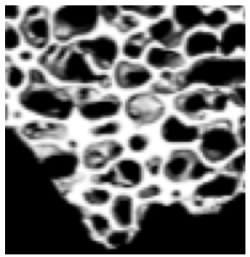


Super Resolved image





NN



Application to the study of ostheoporosis

Slide courtesy of F. Peyrin
PhD Y Li, 2019 ; B. Sixou, F. Peyrin (unpublished)

Creatis

The European Synchrotron | ESRF

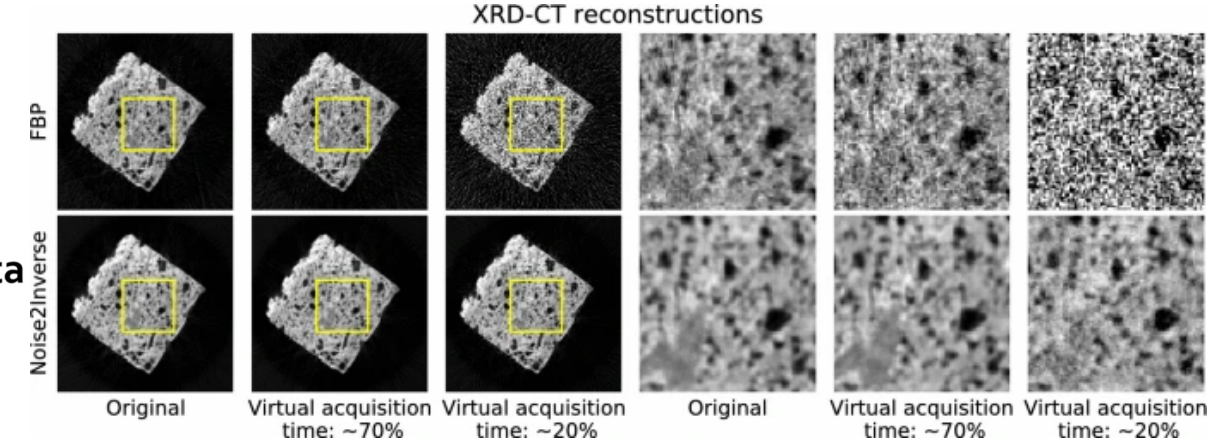
NOISE2INVERSE: HIGH RESOLUTION W/O REFERENCE DATA

Deep denoising for multi-dimensional synchrotron X-ray tomography without high-quality reference data

Allard A. Hendriksen[1], Minna Bührer[2], Laura Leone[3], Marco Merlini[3], Nicola Vigano[4], Daniël M. Pelt[1,5], Federica Marone[2], Marco di Michiel[4] & K. Joost Batenburg[1,5]

Sci. Rep. 11 (2021), 11895

Noise2Inverse: IEEE Transactions on Computational Imaging 6 (2020), 1320



X-ray diffraction tomography reconstructions of a single channel of a single slice of a ceramic fragment. The leftmost column shows the reconstruction of the originally acquired data, and the next two columns show reconstructions with synthetic noise. The rightmost three columns show magnifications of the yellow region of interest.

AI-enhanced resolution of tomography reconstructions allows:

- **Faster data collection for the same final quality**
- **Less radiation damage**

IMPROVING IMAGE RECONSTRUCTION: CHALLENGES

- Variety of imaging configurations (near field, far field)

- **Large datasets** (2048^3 to 4096^3, and much more on BM18)

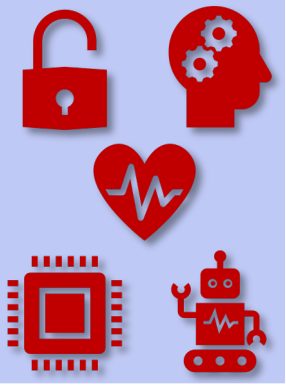
- Requires **training datasets** (not always..)
- **Transferability** of neural networks for different type of materials ?
- Would users need GPUs before their experiments for training ?

- Examples work well on samples with relatively simple density distributions (binary). What about more diverse samples ?

1. Faster data processing
2. Better processing
3. **Data Analysis**
4. Instrument configuration
5. Automated data collection
6. Framework for Online Data Analysis
7. Open Data

SEGMENTATION: MAPPING THE BRAIN CONNECTIONS

- Understand how neurons work together to interpret sensory information and to generate behavior
- Decipher the logic of neural circuits underlying learning and cognition
- Find ways to cope with neurological diseases
- Get inspired to design next generation computing architectures and improve artificial intelligence



How we go about it today

- MRI – probabilistic connections between brain regions
  (1 voxels contains tens of thousands of neurons)
- Visible light microscopy – sparse information (~ 1 in 500 000 neurons)
- EM – comprehensive mapping but we only have one *Drosophila* brain so far which took a few years of data collection

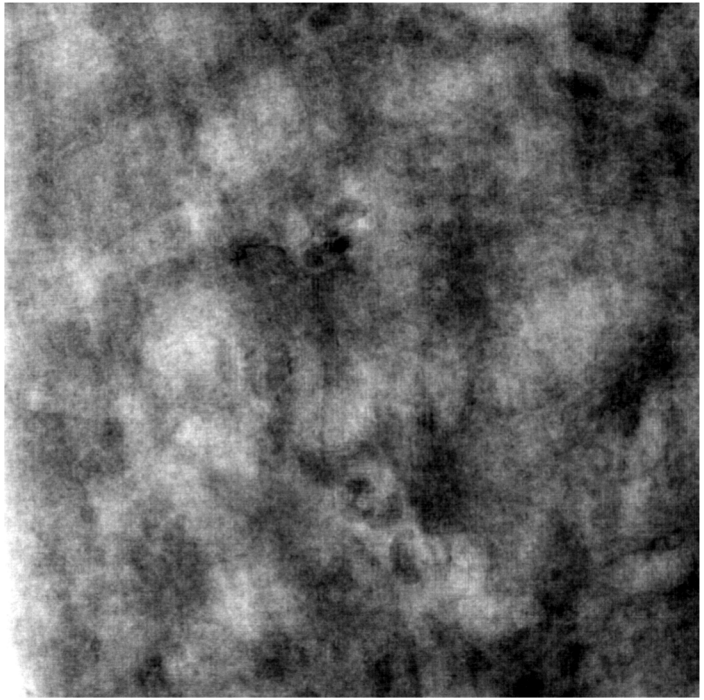
Slide courtesy of A. Pacureanu

HOLO-TOMOGRAPHY: RESOLVING NEURAL NETWORKS

Connectomics in mouse cortex – complementarity with FIB-SEM & TEM
Data collection ~4h (id16A)

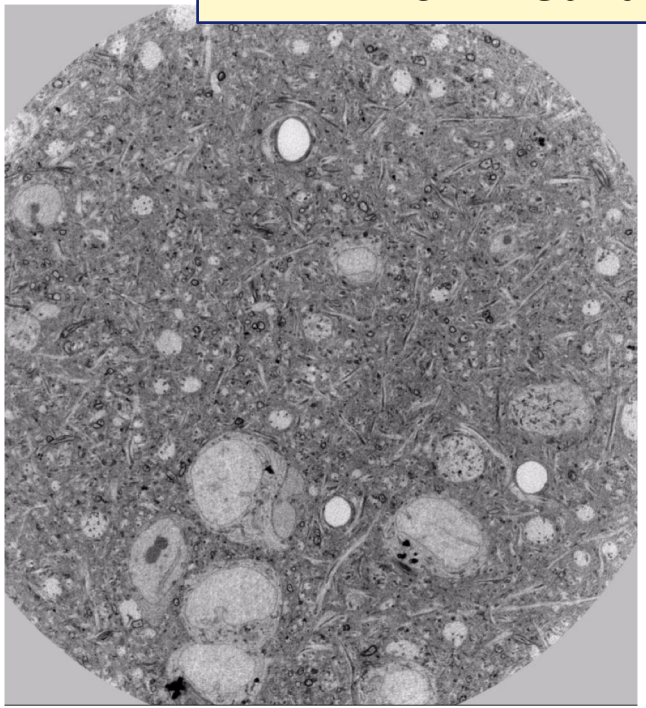
See Alexandra Joita-Pacureanu ESRF webinar

2K or 4K pixels



Phase maps
(object rotation)

Holo-tomography
reconstruction



Pixel size 40 nm

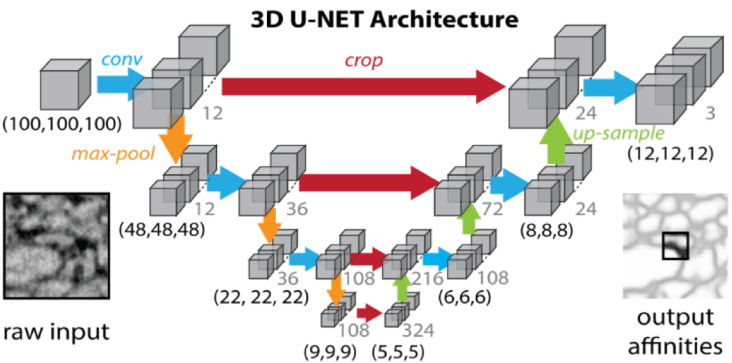
Nat. Neuroscience (2020) & bioRxiv 653188, A Pacureanu
W Lee, A Kuan, J Maniates-Selvin, Harvard Medical School

HARVARD MEDICAL SCHOOL

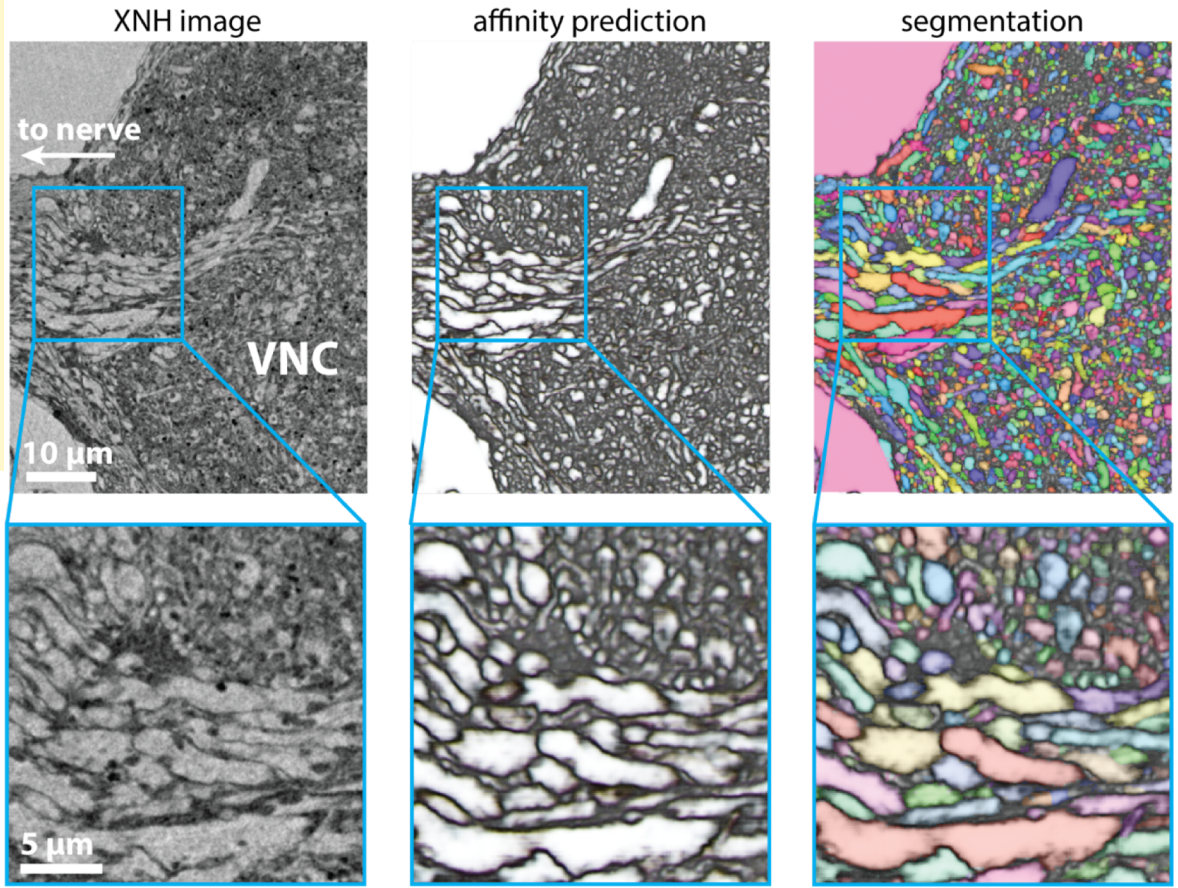
The European Synchrotron | ESRF

SEGMENTATION: MAPPING THE BRAIN CONNECTIONS

- Neuronal circuits are very complex and densely packed
- TB sized images
- Manual annotation takes tremendous resources – it took 60 human years to annotate 30% of a fruit fly brain
- Recent developments in deep learning give hope that automatic analysis is feasible



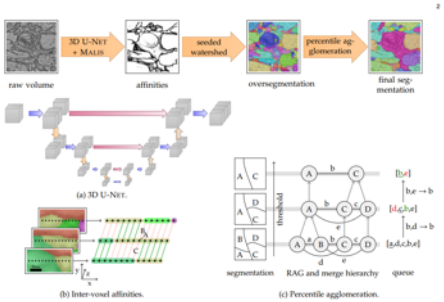
3D U-NET architecture adapted from Funke et al., TPAMI, 2018



Pacureanu et al. bioRxiv 2019

Slide courtesy of A. Pacureanu

AUTOMATIC SEGMENTATION USING DEEP LEARNING

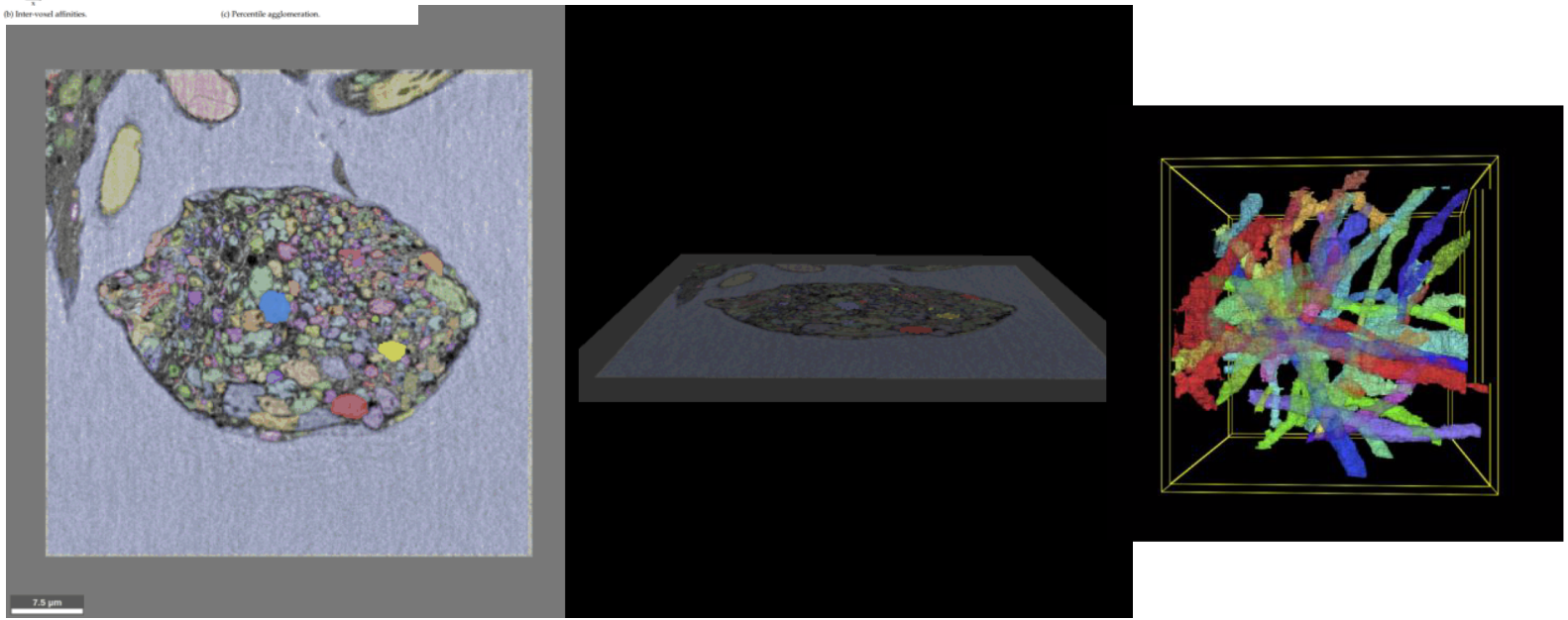


Funke et al. 2018

Deep Learning
Automatic segmentation through convolutional neural networks

VNC: ventral nerve cord

Automated segmentation of a portion of the adult *Drosophila* VNC. Right, volumetric view of selected large-diameter neurons

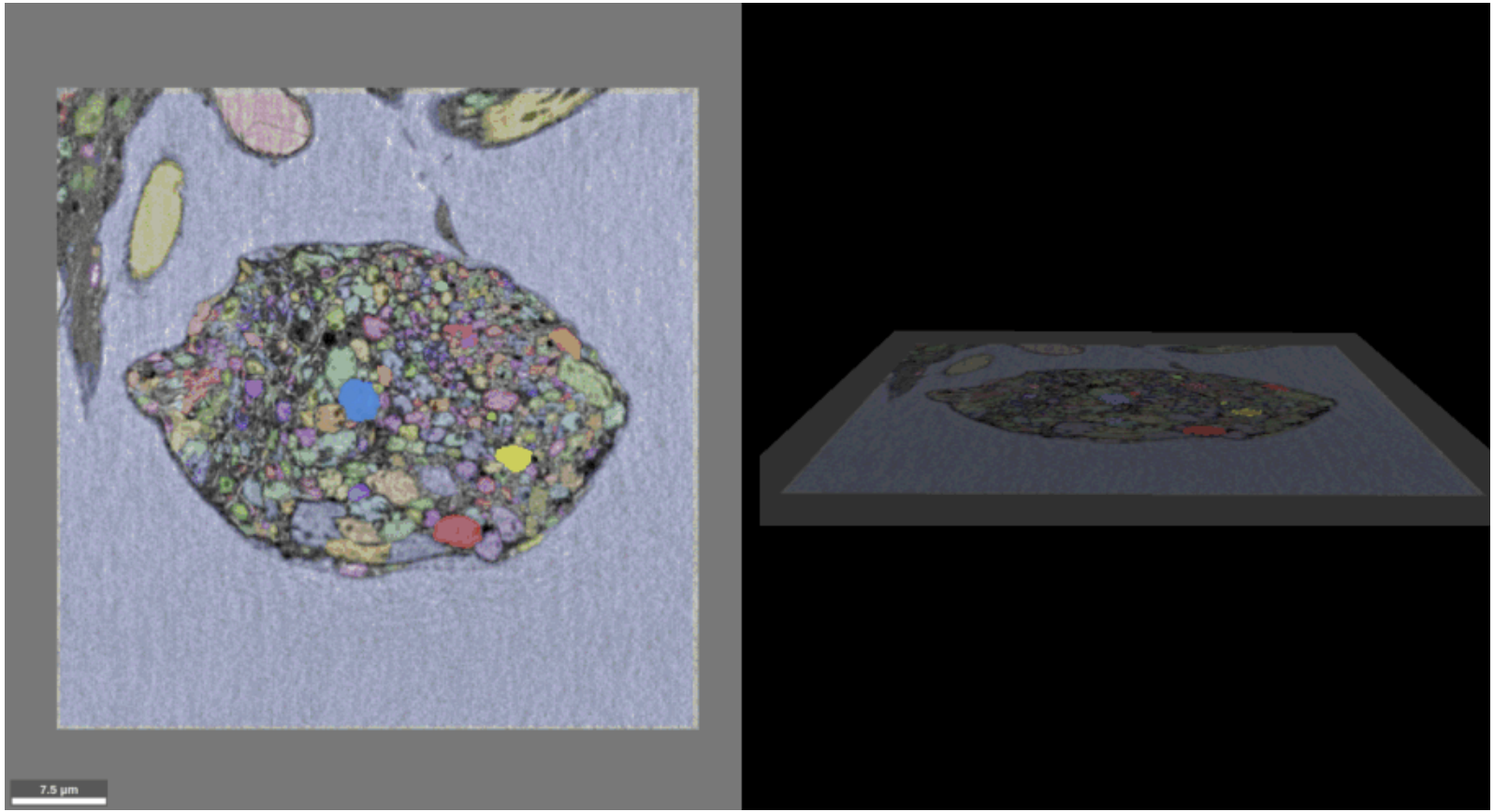


Automatic segmentation in drosophila VNC and mouse cortex

Nat. Neuroscience (2020) & bioRxiv 653188, A Pacureanu
W Lee, A Kuan, J Maniates-Selvin, Harvard Medical School

HARVARD MEDICAL SCHOOL The European Synchrotron | ESRF

SEGMENTATON IN DROSOPHILIA VENTRAL NERVE CORD



3D visualization of automatically segmented neurons
in the Drosophila Ventral nerve cord (VNC)

Pacureanu et al. bioRxiv 2019

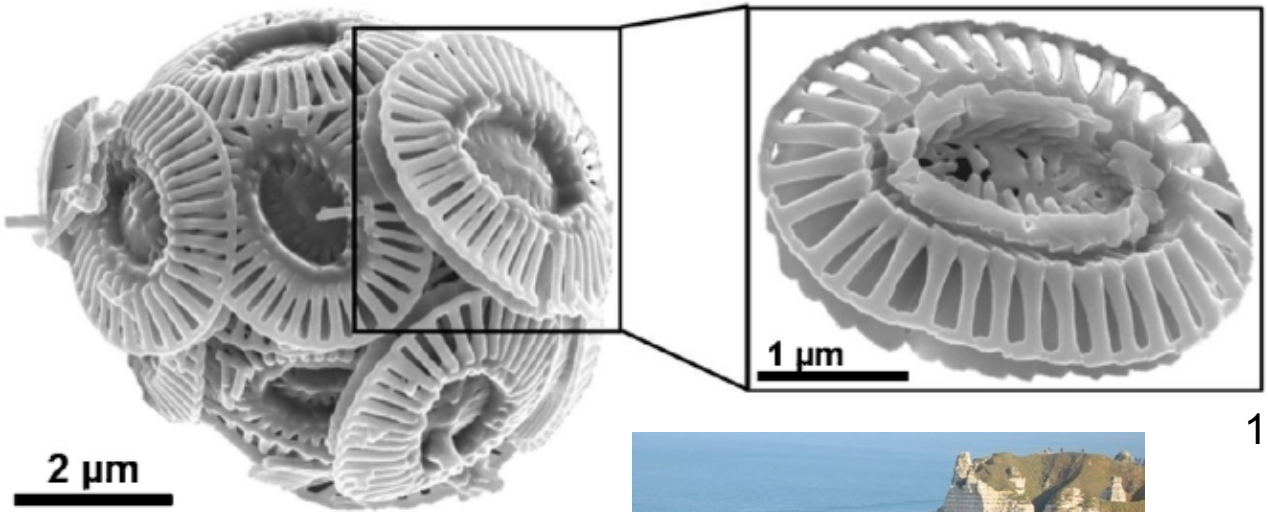
Slide courtesy of A. Pacureanu

MARINE ALGAE - COCCOLITHOPHORES

Emiliania huxleyi

Coccosphere

Coccolith



Hoffmann et al (2014)

Production is light dependent

1 per 2 hours

Calcite crystals $CaCO_3$

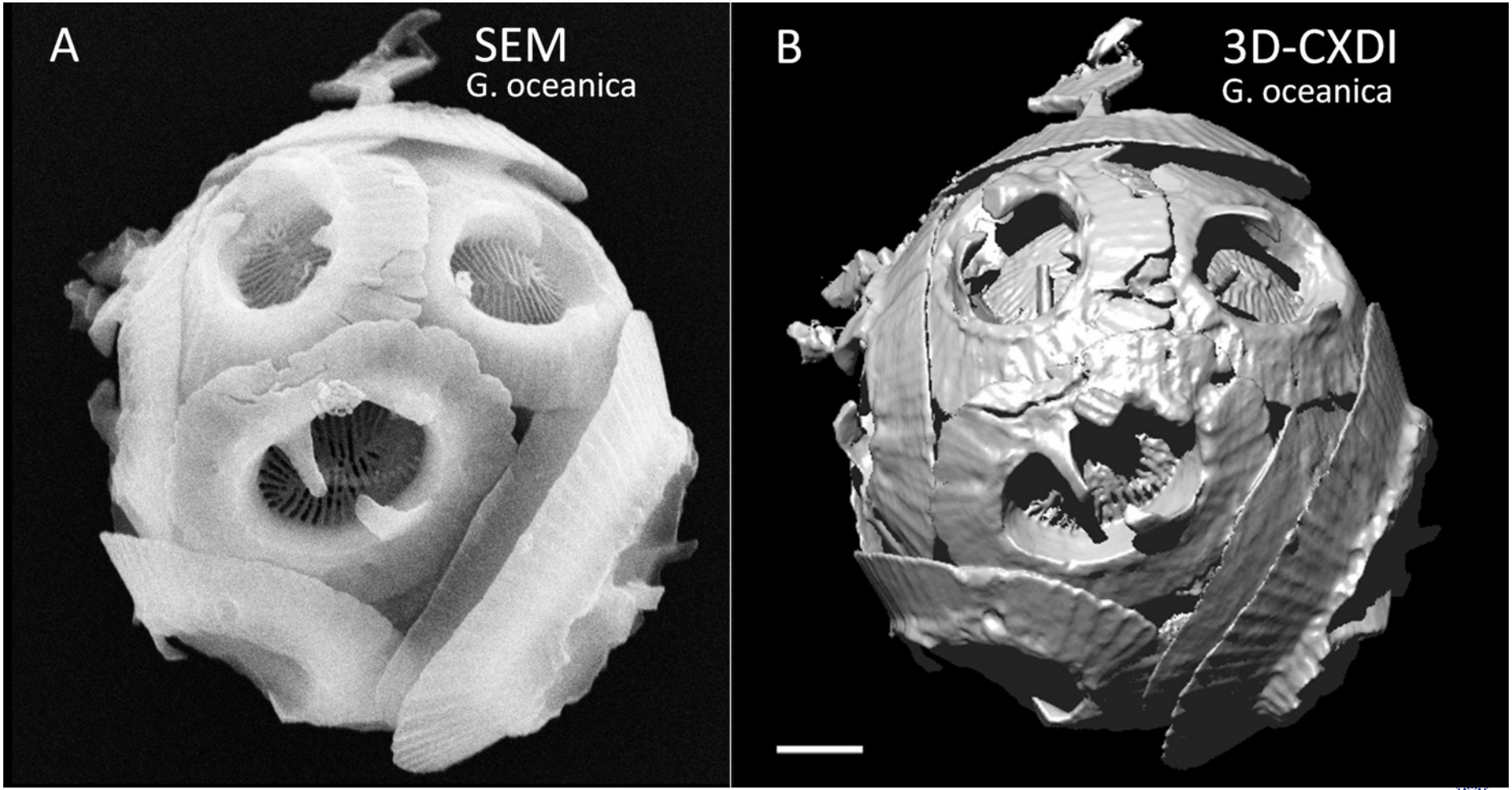
10-20 coccoliths per coccosphere

Most important calcifying organisms on Earth

Beuvier *et al.*, Nature Comm. 10, 751 (2019)
Slide: Y. Chushkin

The European Synchrotron | ESRF

3D-CXDI



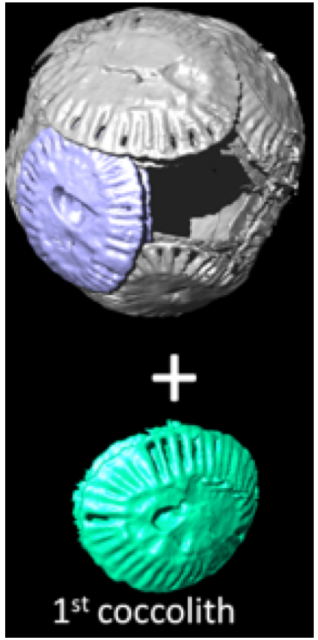
Beuvier *et al.*, Nature Comm. 10, 751 (2019)
Slide: Y. Chushkin

The European Synchrotron | ESRF

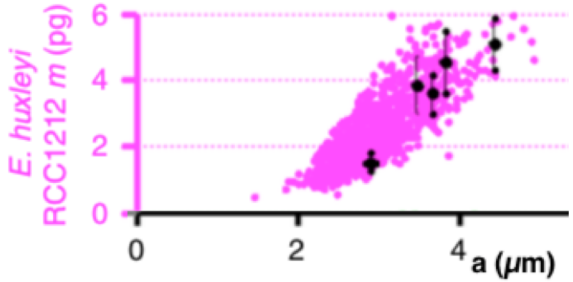
3D COHERENT DIFFRACTION IMAGING (ID10)

3D $CaCO_3$ Coccosphere

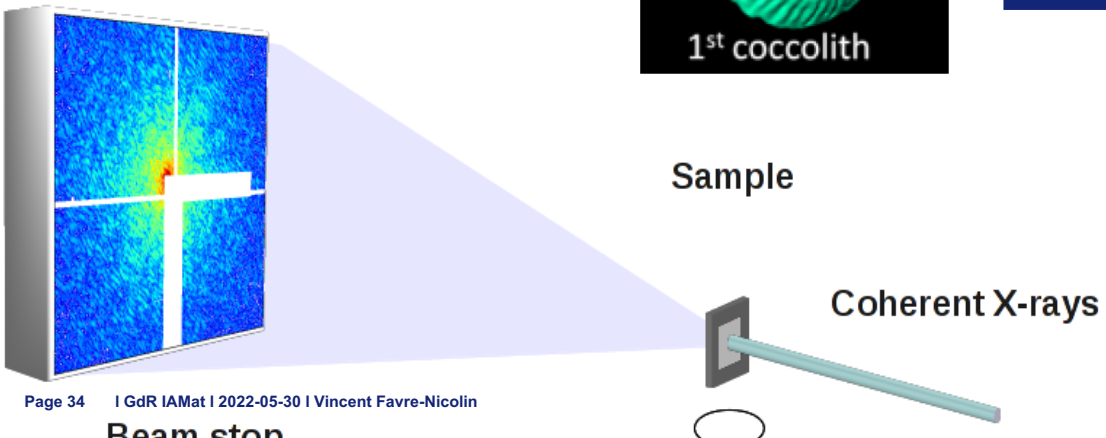
Segmentation



- Dataset: 512^3 pixels
- Pixel size: 5-20 nm
- GPU reconstruction: 32s/run
- Processing (1 GPU): 15 min
- Data collection: 1~3h
- Post-EBS: 10-100x more data

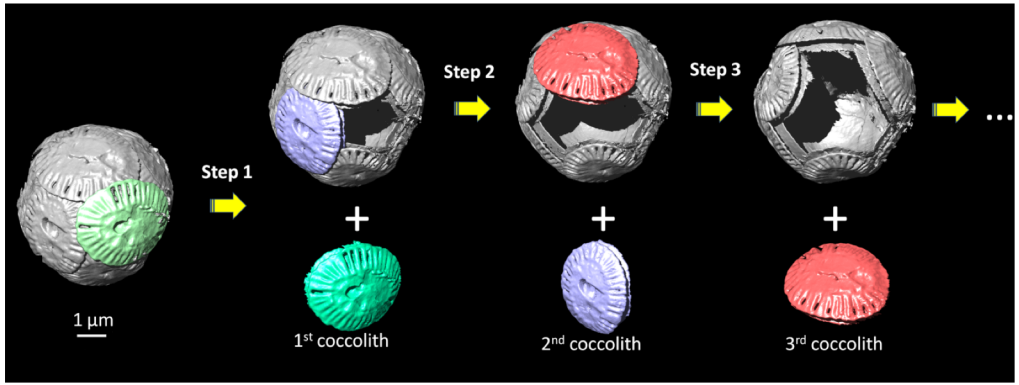


Segmentation: mass vs size of individual coccoliths



Beuvier, Nature Comm. 10 (2019), 751

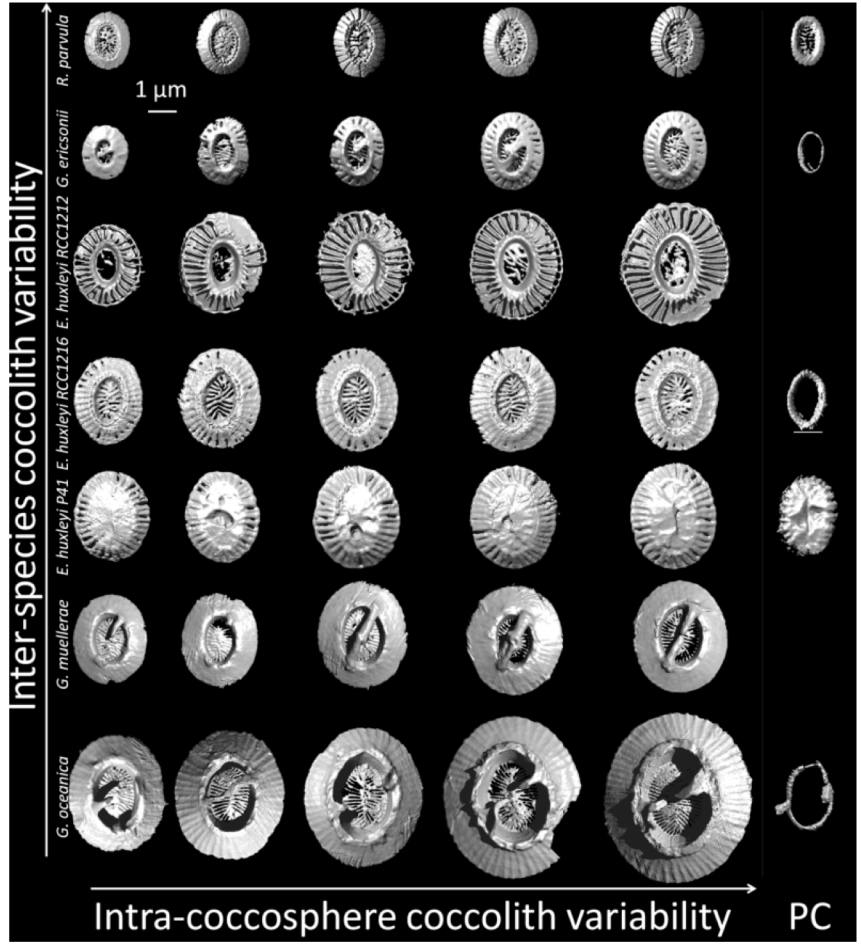
MULTIPLE DATASETS, DATABASES…



3D coherent imaging of $CaCO_3$ coccospheres (produced by phytoplankton and a large contributor to CO_2 storage) => extraction of a large collection of coccoliths

How do we process and make sense of large collections of reconstructed objects ?

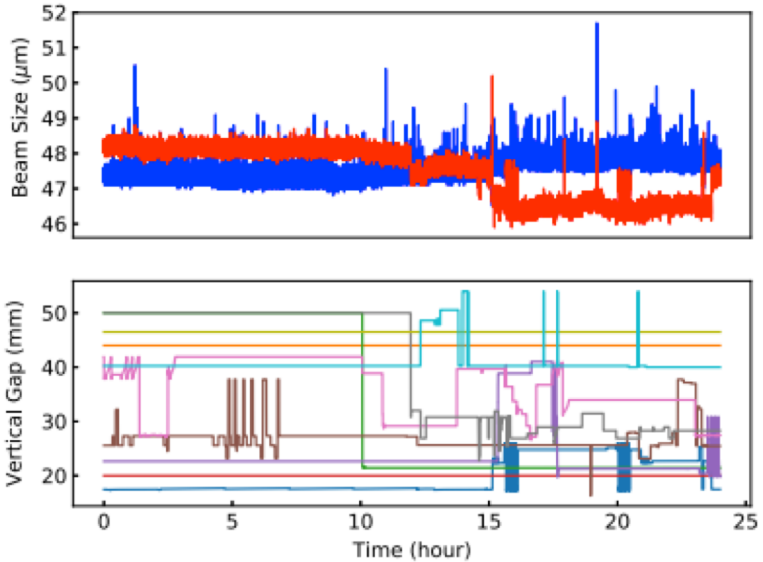
Similarities, multivariate analysis, principal component analysis… Can we go **beyond model-based statistics** ?



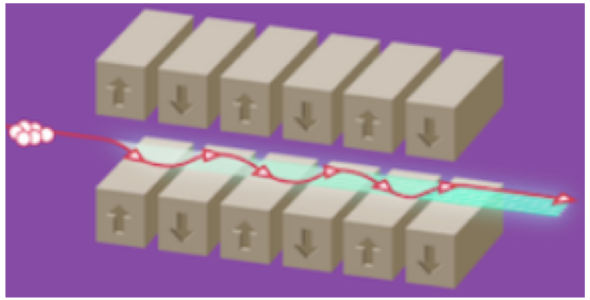
Beuvier et al, Nat. Comm. **10**, 751 (2019).

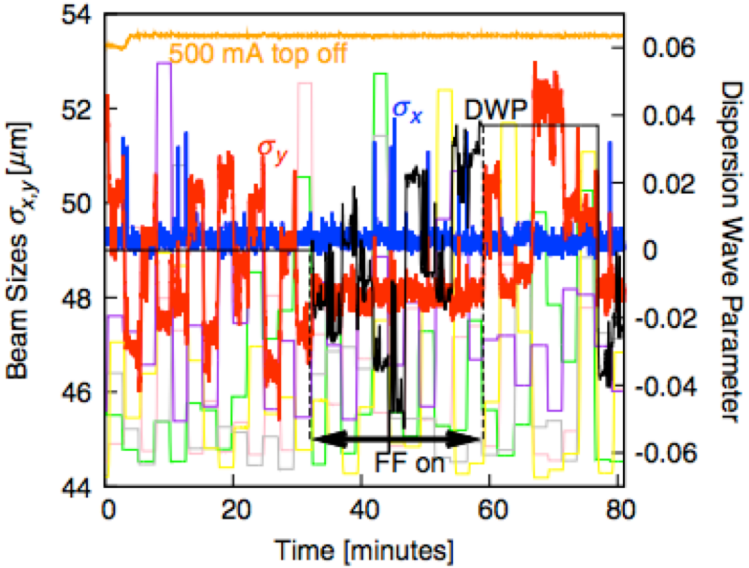
1. Faster data processing
2. Better processing
3. Data Analysis
4. **Instrument configuration**
5. Automated data collection
6. Framework for Online Data Analysis
7. Open Data

SYNCHROTRON BEAM STABILISATION



Synchrotron beam size vs time, influenced by the undulator gaps (ALS)

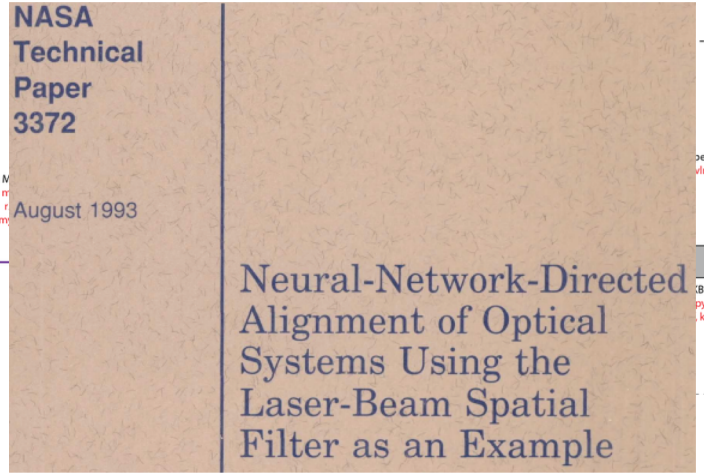




A stable beam size is essential notably for nano-focusing experiments. This is usually done using both feedback and feed-forward (FF: prediction, model based) corrections.
A machine learning FF algorithm allows to improve the beam tuning.

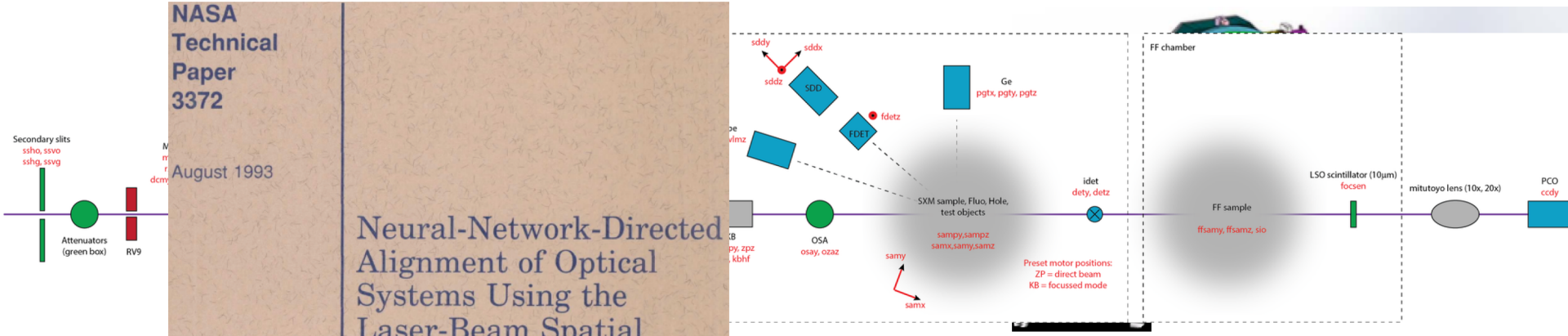
Phys. Rev. Lett. 123, 194801 (2019)

BEAMLINE OPTICS ALIGNMENT


NASA Technical Paper 3372

August 1993

Neural-Network-Directed Alignment of Optical Systems Using the Laser-Beam Spatial Filter as an Example



KB: 6 motors (+4 for slits)

From source to detector, there can be >10 optical elements, each with 2-10 motors:

- **Undulators**
- **Mirrors**
- **Slits**
- **Monochromator**
- **Nano-focusing optics**
- **Sample alignment**
- **Detectors** (distance, orientation, energy thresholds, analyser)

All alignment usually done progressively from the source, but some elements with >=4 motors (actuators), and/or non-perfect optical properties are difficult to parametrise.

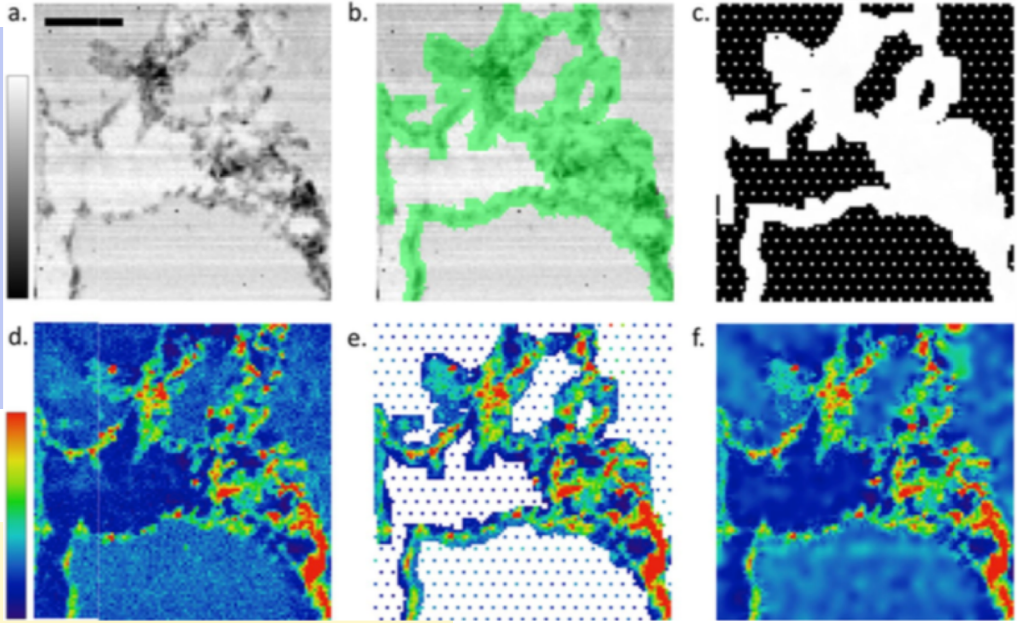
1. Faster data processing
2. Better processing
3. Data Analysis
4. Instrument configuration
5. **Automated data collection**
6. Framework for Online Data Analysis
7. Open Data

MACHINE LEARNING: AUTOMATED ACQUISITION

Accelerating Scanning X-ray microscopy

- Perform fast map (STXM)
- Identify (mask) relevant regions
- Perform high-resolution maps in the relevant areas (XRF)

Could be extended by identifying on-the-fly the relevant areas (one line to the next) ?



Sci Rep 10(2020), 9990 (Elettra)

This approach can be applied to any X-ray microscopy technique:

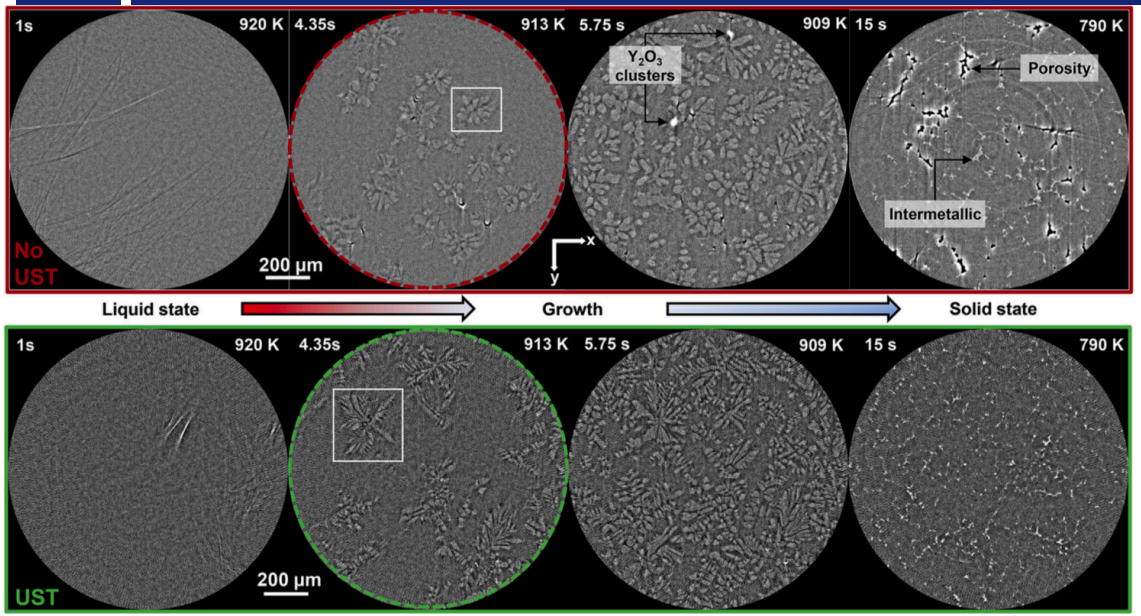
- Transmission
- SAXS/WAXS
- Powder diffraction
- Fluorescence
- XAFS/XANES (dispersive)

=> Ongoing project at ESRF, financed by STREAMLINE

STREAMLINE

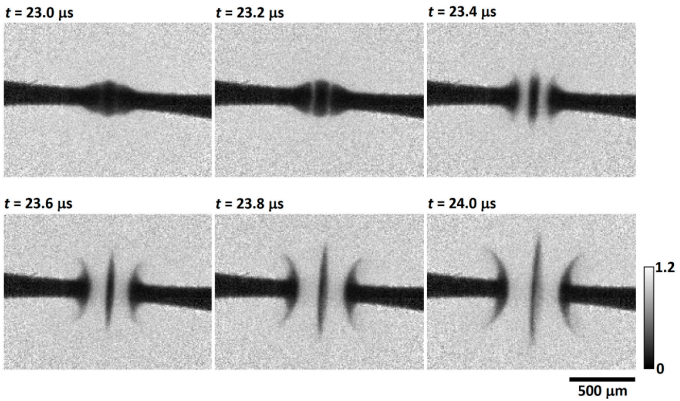
The European Synchrotron | ESRF

EVENT-BASED DATA ACQUISITION



Dendrite formation. Data collected at ~3kHz, 3x180°/s during 16s (limited by 32 GB camera memory).

Acta Materialia 129 (2017), 194



MHz radiographs of electric arc ignition

Optics Express 25 (2017),13857

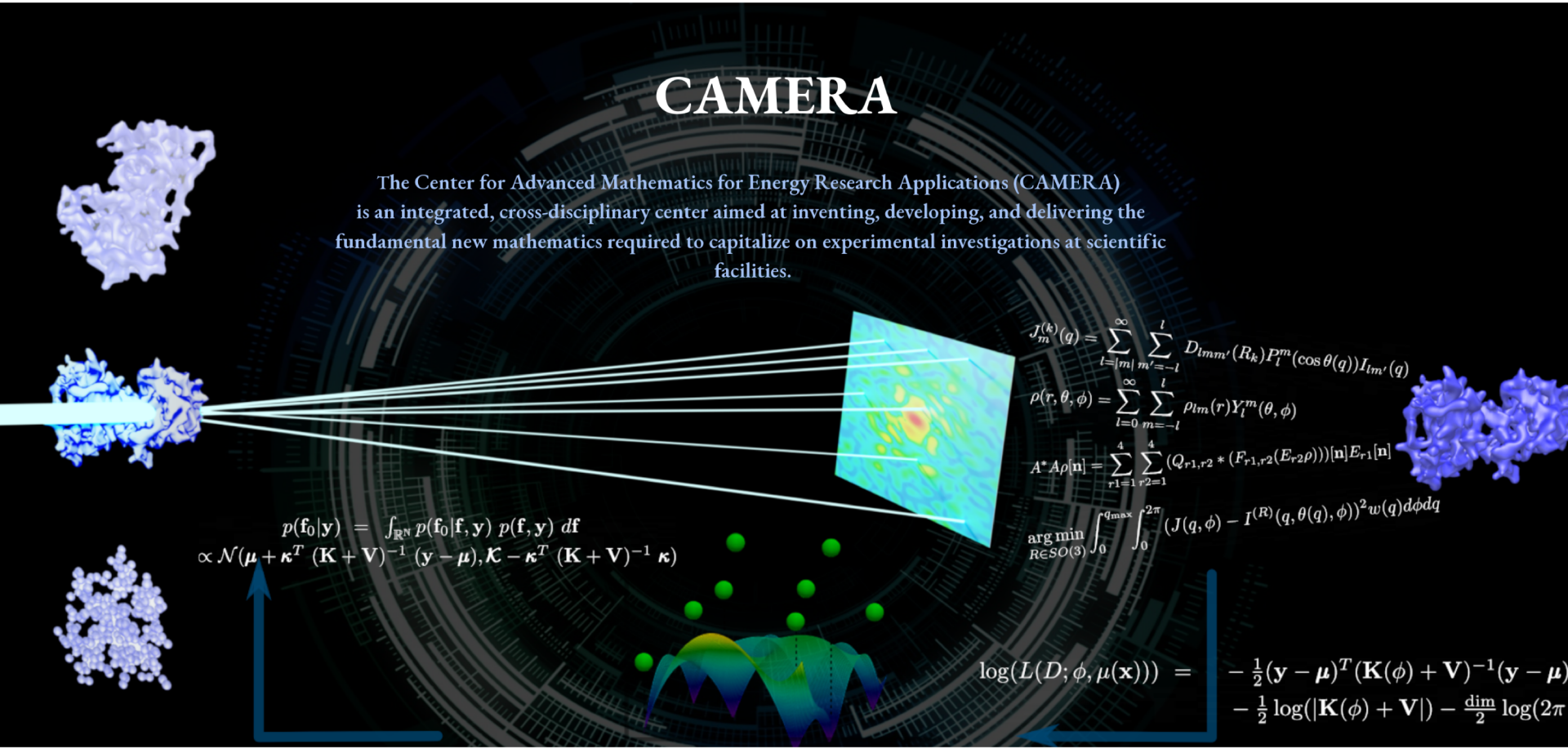
https://simap.grenoble-inp.fr/fr/equipes/high-performance-imaging

Data collected on ID19

Fast, time-resolved experiments often feature narrow relevant time frames.
Fast (on-the-fly ML) analysis would allow:

- Recording only relevant parts, or varying framerates
- Automated tuning of experimental parameters (temperature)

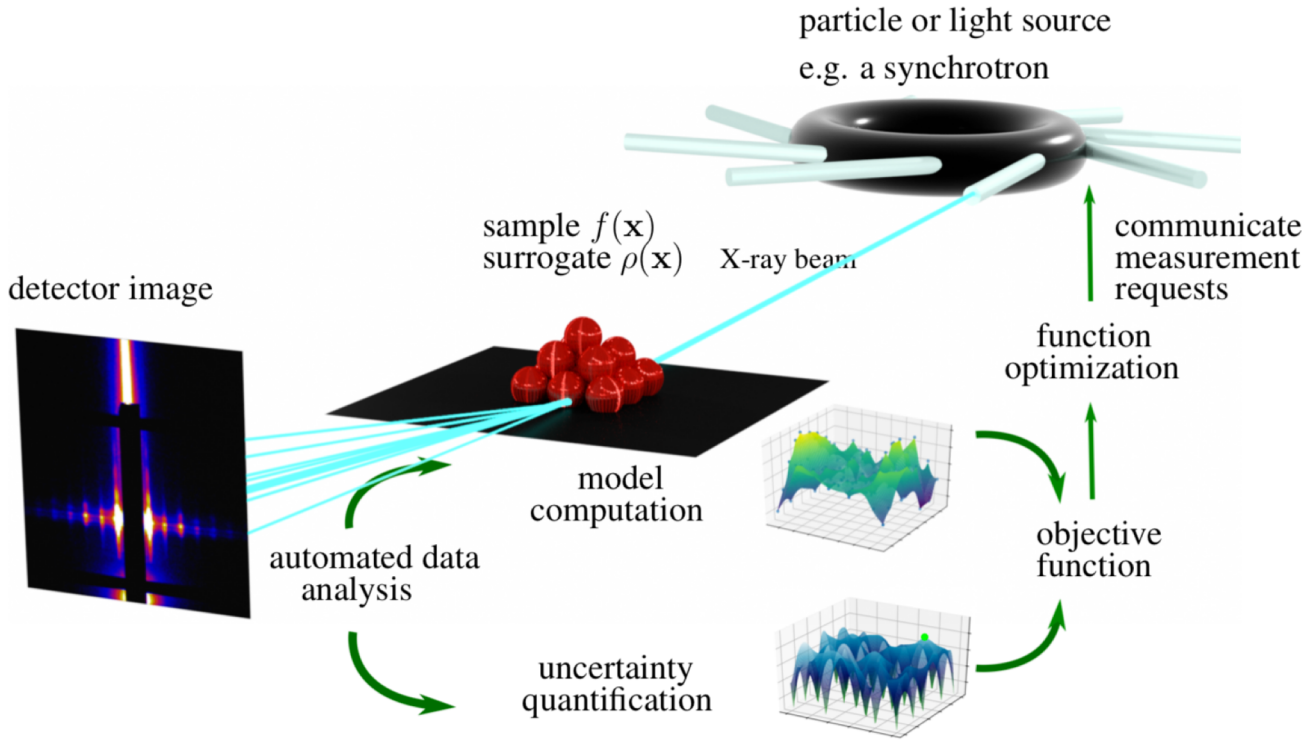
GAUSSIAN PROCESS AS A GENERIC PATHFINDER



https://camera.lbl.gov/

The European Synchrotron | ESRF

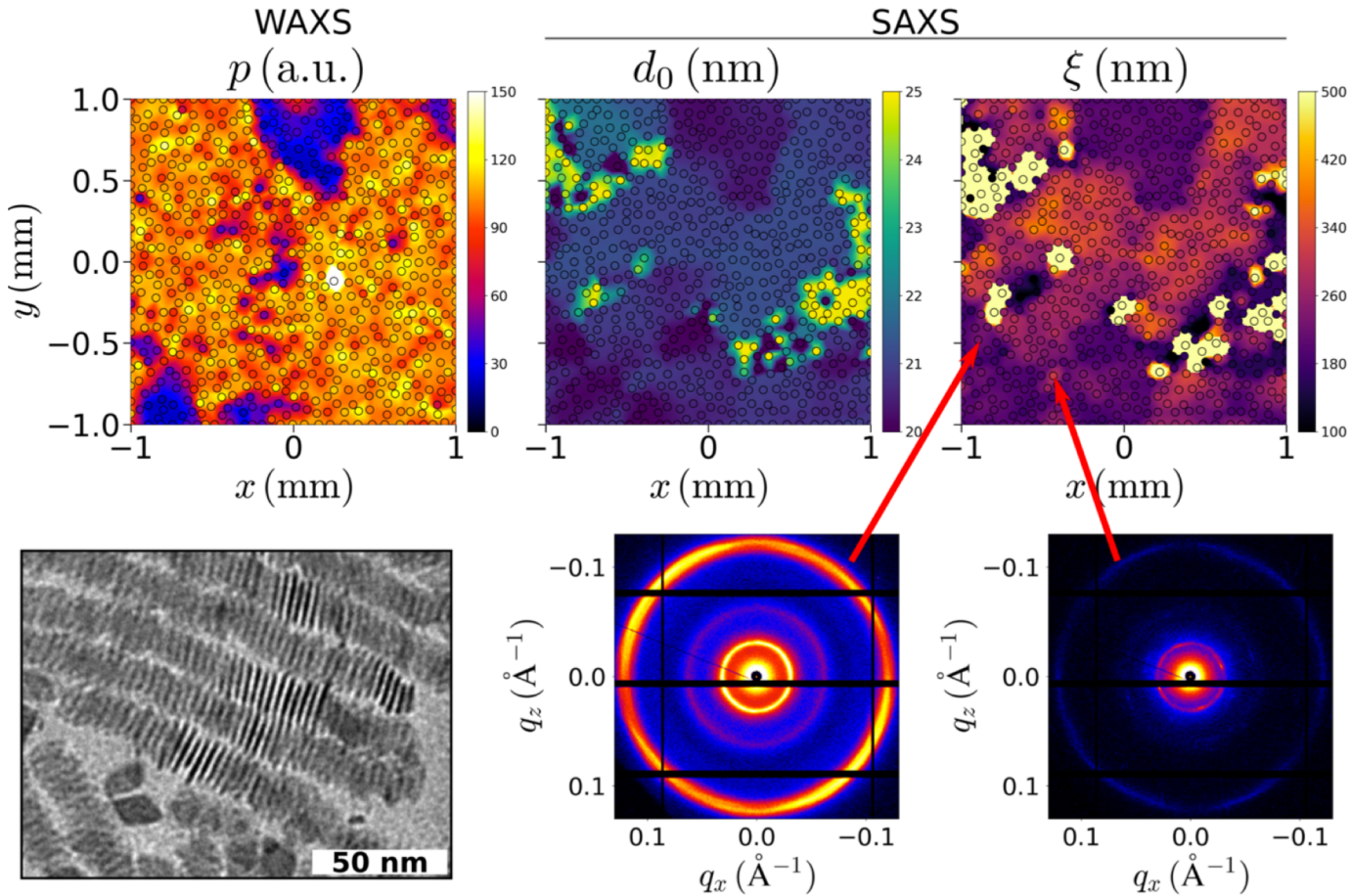
GAUSSIAN PROCESS AS A GENERIC PATHFINDER



The Gaussian process uses a Bayesian approach to evaluate the parameter space which needs to be explored. First measure a few data points, then use a model (physics-aware) to evaluate the optimal points to measure to maximise the information collected

Nature Reviews Physics, 3(10)

GAUSSIAN PROCESS AS A GENERIC PATHFINDER

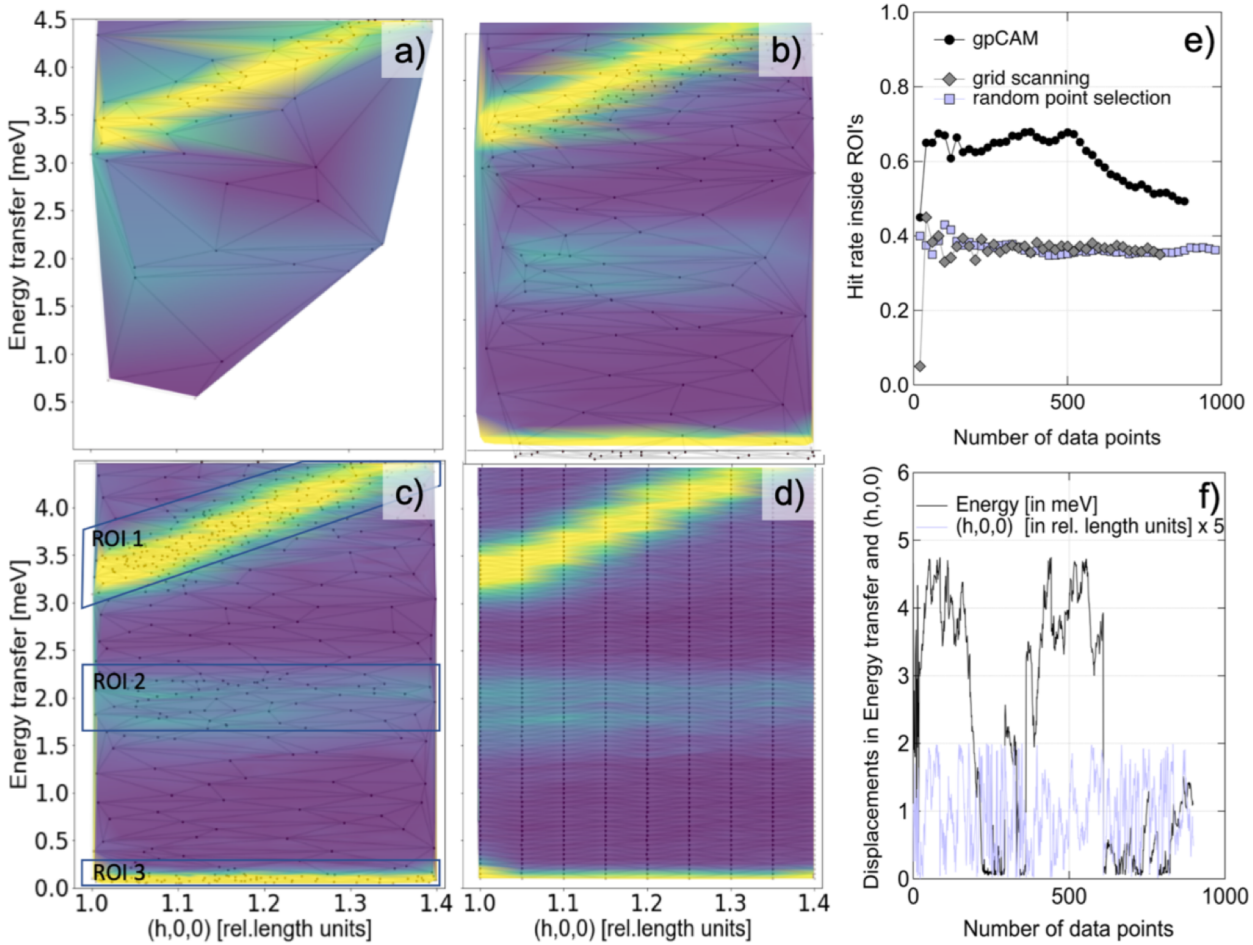


Example automated acquisition with SAXS.

The sampling points look random but are not (overlap is minimised)

Nature Reviews Physics, 3(10)

GAUSSIAN PROCESS AS A GENERIC PATHFINDER



Another example at ILL using a triple-axis spectrometer.

The gpCAM data acquisition automatically maximises the number of points measured in the relevant areas

Nature Reviews Physics, 3(10)

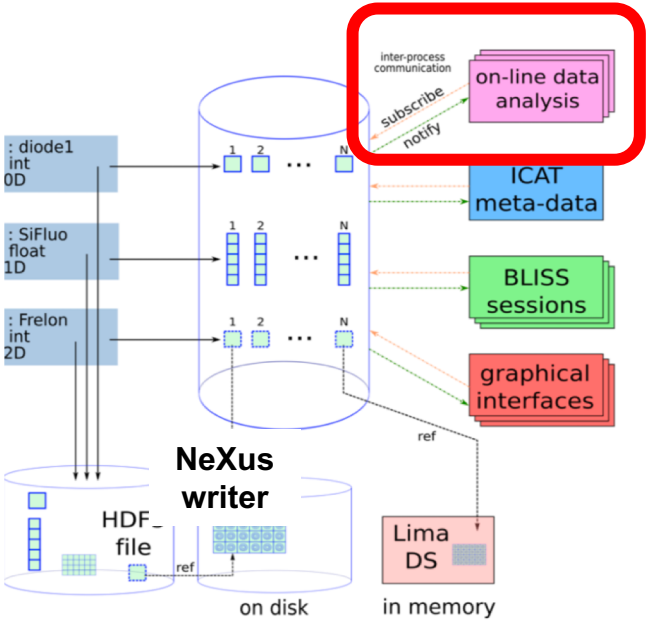
gpCAM

In [ ]:

```
from gpcam.autonomous_experimenter import AutonomousExperimenterGP
import numpy as np

def instrument(data):
    for entry in data:
        entry["value"] = np.sin(np.linalg.norm(entry["position"]))
    return data

##set up your parameter space
parameters = np.array([[3.0,45.8],
                       [4.0,47.0]])

##set up some hyperparameters, if you have no idea, set them to 1 and make the training bounds large
init_hyperparameters = np.array([1,1,1])
hyperparameter_bounds =  np.array([[0.01,100],[0.01,100.0],[0.01,100]])

##let's initialize the autonomous experimenter ...
my_ae = AutonomousExperimenterGP(parameters, init_hyperparameters,
                                 hyperparameter_bounds,instrument_func = instrument,
                                 init_dataset_size=10)
#...train...
my_ae.train()

#...and run. That's it. You successfully executed an autonomous experiment.
my_ae.go(N = 100)
```

1. Faster data processing
2. Better processing
3. Data Analysis
4. Instrument configuration
5. Automated data collection
6. **Framework for Online Data Analysis**
7. Open Data

ONLINE DATA ANALYSIS



Before:

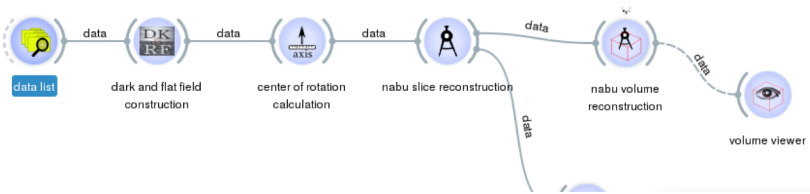
- ‘**one-way pipeline**’: acquisition, measure, display, store, analyse

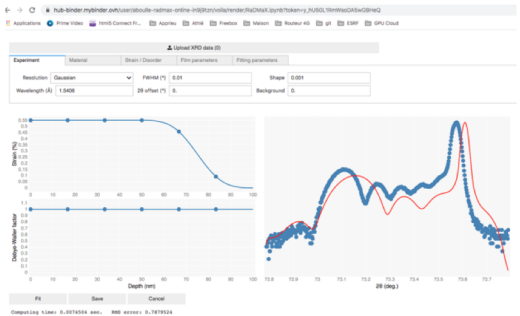
Now:

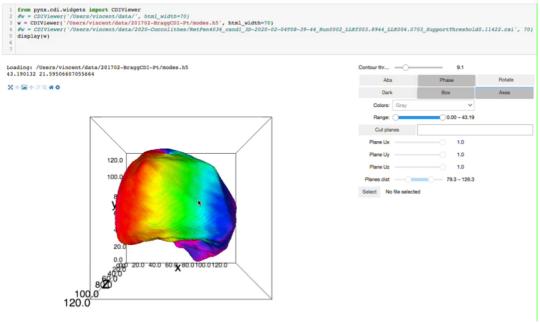
- **Need (more) online analysis** for user decision
- Automated analysis-driven acquisition
- Analysis procedures are not pre-defined except in a few cases (MX): need to have algorithm-development friendly interfaces
- **Scientist community python-educated**: provide standard access so all scientists can easily access & work on data

Development axes:

- **On-the-fly (in-memory) data access** (up to 16 GB/s) for analysis/triage/compression… (see: Memcached, ASAP:O, Bluesky’s DataBroker)
- **Workflows**
- **User interfaces** (**web**,..)







LIVE DATA ACCESS ?

- Detectors (example: Eiger 4M @ 2kHz) allow data collection > 100 Gb/s
- Default ‘fast’ network data access operates at 25 Gb/s (GPFS)

- Data access can currently be done:
  - From a REDIS database
  - From LIMA (2D data)

  } Fast (in-memory) but not standardised
  - From hdf5 files, once written/flushed to disk => slow: (de)compression + disk

For ‘live’ data processing we are working on a new standardised in-memory data access to enable fast analysis and AI-driven automated data collection

=> enable online data analysis for computational scientists

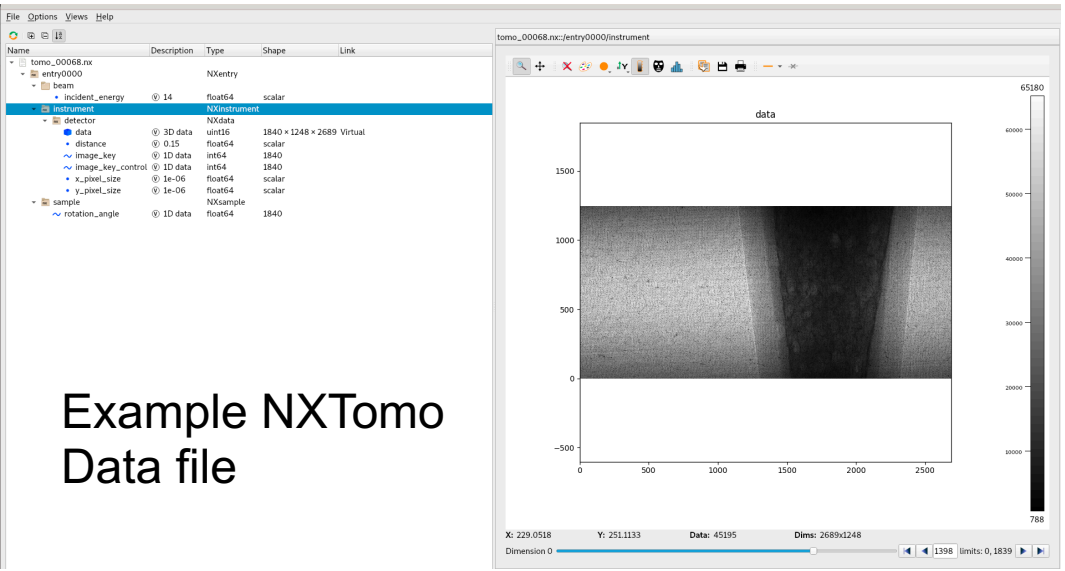
BLISS: PYTHON INTERFACE TO THE DATA ACQUISITON

```python
from bliss.scanning.scan import Scan
from bliss.scanning.chain import AcquisitionChain
from bliss.scanning.acquisition.motor import SoftwarePositionTriggerMaster
from bliss.scanning.acquisition.counter import SamplingCounterAcquisitionDevice

chain = AcquisitionChain()
chain.add(SoftwarePositionTriggerMaster(m0, 5, 10, 10),
          SamplingCounterAcquisitionDevice(i0, 0.01, npoints=10))
scan = Scan(chain)
```

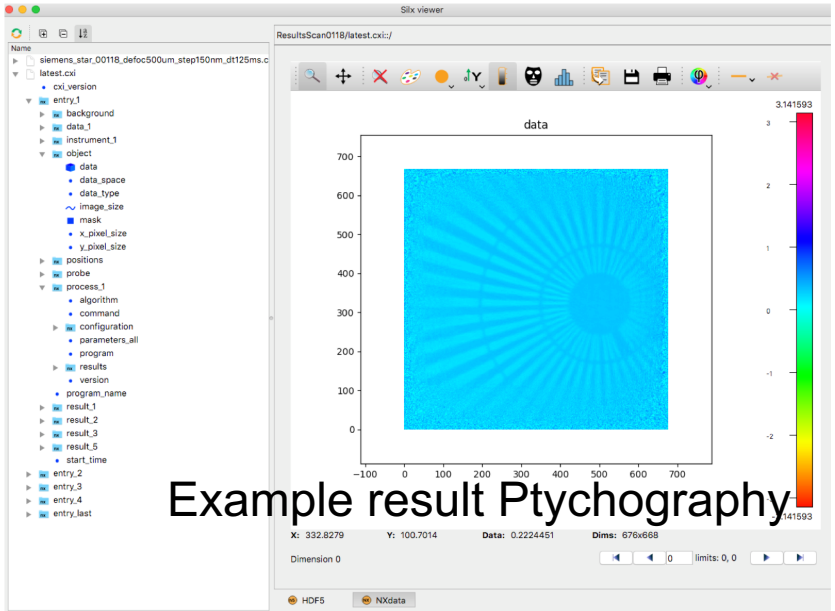
```python
def cscan(motor, start, stop, npoints, time):
    # create a new chain
    chain = AcquisitionChain(parallel_prepare=True)
    # create the monitor timer
    # npoints == 0 mean infinite
    monitor_timer = SoftwareTimerMaster(1., name="monitor_timer",
                                        npoints=0)
    # create acquisition device for the monitor diode
    diode_device = SamplingCounterAcquisitionDevice(diode1,
                                                    count_time=1.,
                                                    npoints=0)
    # Associate them in the chain
    chain.add(monitor_timer, diode_device)
    # Now the fast acquisition
    # create a motor master for a position trigger
    master = SoftwarePositionTriggerMaster(motor, start, stop, npoints,
                                           time=time)
    # The spectrum device MCA
    mca_acq = McaAcquisitionDevice(mca, npoints=npoints,
                                   trigger_mode=McaAcquisitionDevice.GATE,
                                   counters=list(mca.counters))
    chain.add(master, mca_acq)
    # The Image detector
    lima_master = LimaAcquisitionMaster(frelon,
        acq_nb_frames=npoints, acq_trigger_mode='EXTERNAL_GATE')
    lima_master.add_counter(frelon.image)
    chain.add(master, lima_master)
    # Finally build the scan and run it.
    scan = Scan(chain,name='cscan')
    scan.run()
    return scan
```

- BLISS: the ESRF's new data acquisition system
- Progressively replacing SPEC on all beamlines
- Python-based
- More friendly to interface with data analysis programs

HDF5+NEXUS: STANDARD DATA FORMAT



Example NXTomo Data file

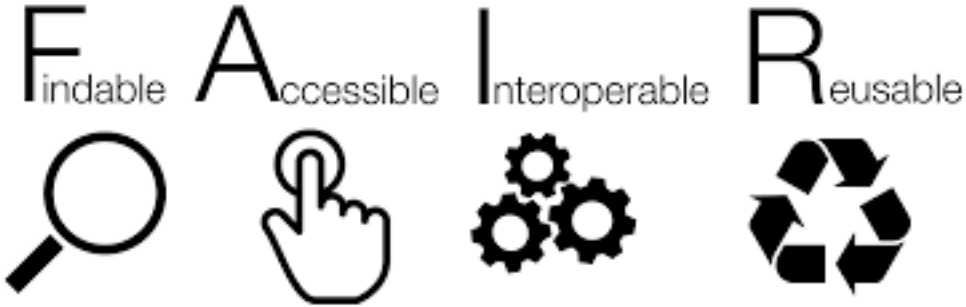


Example result Ptychography

Use of the HDF5 data format allows to store all raw data and experimental parameters for processing.
The NeXus format provides a standard organisation for the different fields.
NB: not *all* data is yet collected as hdf5, but the % is increasing

https://www.nexusformat.org/

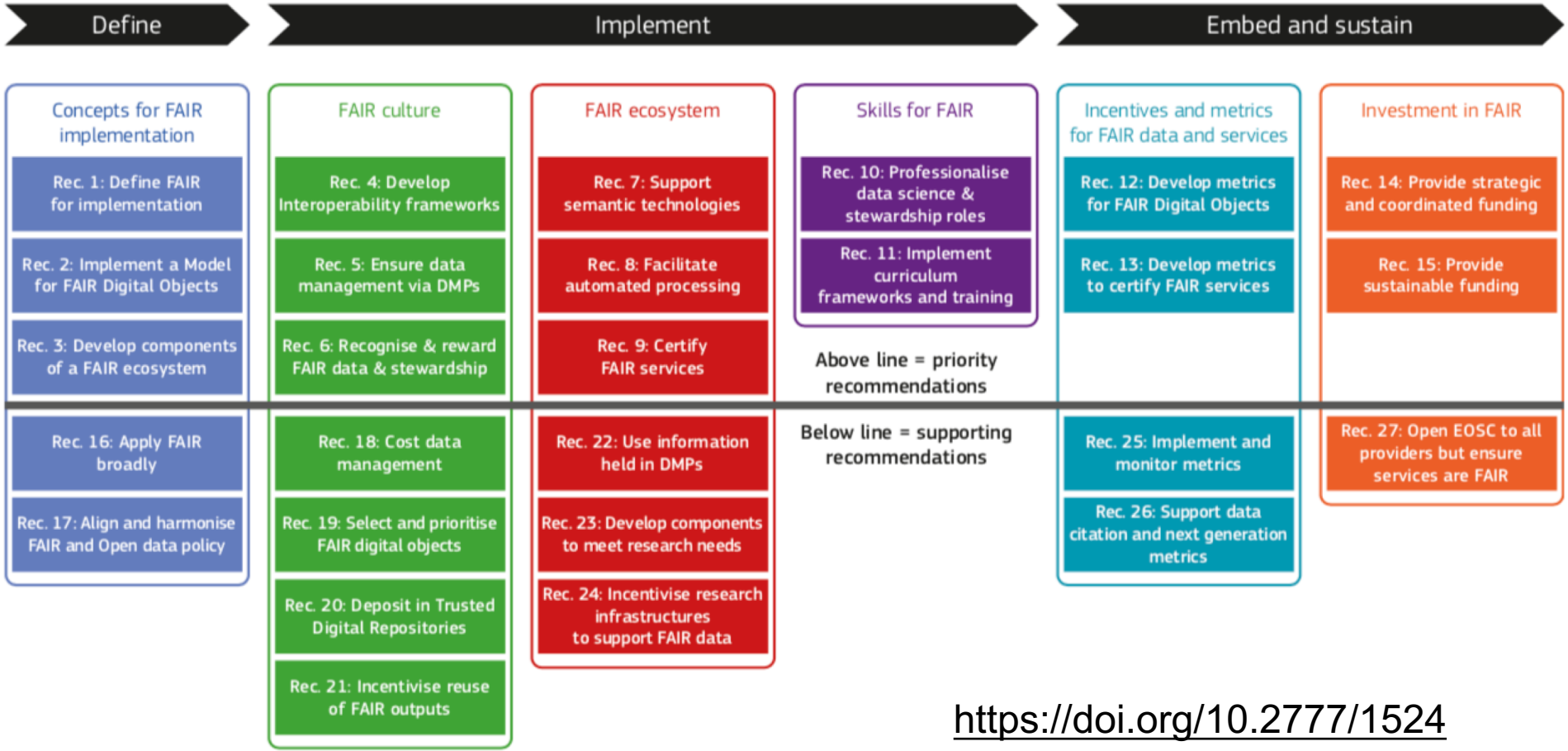
FAIR DATA



The new FAIR bible
https://doi.org/10.2777/1524

FAIR RECOMMENDATIONS

27 recommendations !



https://doi.org/10.2777/1524

Index to FAIR Action Plan recommendations

1. Faster data processing
2. Better processing
3. Data Analysis
4. Instrument configuration
5. Automated data collection
6. Framework for Online Data Analysis
7. **Open Data**

MULTIPLE DATASETS, DATABASES…

Many open databases available:

Experimental structures and properties		
ChEMBL	Bioactive molecules with drug-like properties	https://www.ebi.ac.uk/chembl
ChemSpider	Royal Society of Chemistry's structure database, featuring calculated and experimental properties from a range of sources	https://chemspider.com
Citrination	Computed and experimental properties of materials	https://citrination.com
Crystallography Open Database	Structures of organic, inorganic, metal–organic compounds and minerals	http://crystallography.net
CSD	Repository for small-molecule organic and metal–organic crystal structures	https://www.ccdc.cam.ac.uk
ICSD	Inorganic Crystal Structure Database	https://icsd.fiz-karlsruhe.de
MatNavi	Multiple databases targeting properties such as superconductivity and thermal conductance	http://mits.nims.go.jp
MatWeb	Datasheets for various engineering materials, including thermoplastics, semi-conductors and fibres	http://matweb.com
NIST Chemistry WebBook	High-accuracy gas-phase thermochemistry and spectroscopic data	https://webbook.nist.gov/chemistry
NIST Materials Data Repository	Repository to upload materials data associated with specific publications	https://materialsdata.nist.gov
PubChem	Biological activities of small molecules	https://pubchem.ncbi.nlm.nih.gov

SSHADE **S**olid **S**pectroscopy **H**osting **A**rchitecture of **D**atabases and **E**xpertise

https://www.sshade.eu

Can we make more use of open databases ?

- Cross-validation / categorisation

- Open data policy (e.g. https://www.esrf.eu/datapolicy) will provide tons of *raw* data
- … can this be combined in a more digestible, explorable form ?

K. Butler et al, Nature volume 559 (2018), 547

ESRF OPEN DATA POLICY

https://www.esrf.fr/datapolicy

ESRF EUROPEAN SYNCHROTRON RADIATION FACILITY

30 November 2015

The ESRF Data Policy

The ESRF aims to implement a Data Policy starting as soon as possible in 2016. The main elements of this policy comprise:

- **Data ownership**
- **Data curation**
- **Data archiving**
- **Open access to data**

This policy follows largely the recommendations of the PaN-data Europe Strategic Working Group laying out a common framework for scientific data management at photon and neutron facilities (Deliverable D2.1, PaN-data Europe, co-funded by the European Commission under the 7th Framework Programme)

MAIN ELEMENTS OF ESRF DATA POLICY

Data are under embargo for 3 years but can be released earlier by the experimental team

All data have a DOI assigned automatically at the session level with appropriate high-level metadata from proposal

Experimental team can register a DOI for a subset or superset of the data on provision of high-level metadata e.g. abstract

PI can request an extension to the embargo period

DOI must be cited when data are re-used

Processed data can be uploaded to data portal

Data can be downloaded from https://data.esrf.fr

After embargo period and/or on creation of bespoke DOI
Data are available under CC-BY-4.0 licence as Open Data

ESRF DATA PORTAL - HOME

ESRF | Home | My Selection 2 | Log out mchaille

Home / Investigations

Search

Name	Bea...	Title	Da...	S...	Size	Files	...	...	
MA-4353	ID16B	Investigation by 3D nano-imaging of t...					...	...	DOI 10.15151/ESRF-ES-142846529
MX-2076	BM29	BAG for the Structural Biology Group...	6	2	237 MB	186	...	...	DOI 10.15151/ESRF-ES-142840456
SC-4855	ID16A	Quantitative Localization of Dendrime...	1	1	91 KB	1	...	...	DOI 10.15151/ESRF-ES-142840375
ID21-2020	ID21	DCM					...		🔒
MA-4162	ID01	Mechanical properties of single Au n...					...	...	DOI 10.15151/ESRF-ES-142757577
MX-2147	CM01	High resolution Cryo- Electron Micros...	6	1	2 TB	1122	...	...	DOI 10.15151/ESRF-ES-142673807
IH-EV-5	ID21	Estudy of Cr speciation in soils	4	2	23 MB	15	...	...	DOI 10.15151/ESRF-ES-142481073
ES-295	ID19	ROCK-DEFORM-4D: Unravelling the ...					...	...	DOI 10.15151/ESRF-ES-141972475
IM-17	BM05	imaging of 25 samples bouillon cube...					...		🔒
IH-LS-3...	ID16B	Continuation of LS2780, radiography...					...	...	DOI 10.15151/ESRF-ES-141963373
IM-11	ID16B	Rock samples	40	10	1 TB	130720	...		🔒
IN-1102	ID19	On site industrial experiment					...		🔒
BLC-11821	ID16B	Sample stage alignment for holotomo...	8	1			...		🔒
IN-11	ID16B						...		🔒
IX-62	ID30B	Pharma studies	19	19	452 MB	114	...		🔒
MX-2073	ID30A3	London Cancer BAG comprising the I...	33	24	82 GB	1506	...	...	DOI 10.15151/ESRF-ES-140506618

ESRF DATA PORTAL - DATASET

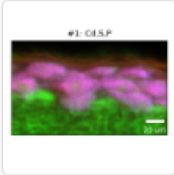
11:34 AM
October 28, 2018

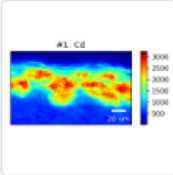
ble_Cd_aleurone

Scanning X-ray Microscope

Summary | Metadata List | Files 274 | DOI

Name	ble_Cd_aleurone
Definition	SXM
Start	11:34:00 AM
End	12:28:00 PM
Sample	ble_Cd
Description	wheat grain 16um thick Cd





/data/visitor/ev355/id21/ble_Cd/ble_Cd_aleurone

Download

CREATE A DOI FOR YOUR ESRF DATA

ESRF | Home | Open Data | ★ My Selection 11 | Log out demariaa

Home

Please fill the next form

Title *

Structural Evidence for a Role of the Multi-functional Human Glycoprotein Afamin in Wnt Transport

Abstract *

Afamin, a human plasma glycoprotein and putative transporter of hydrophobic molecules, has been shown to act as extracellular chaperone for poorly soluble, acylated Wnt proteins, forming a stable, soluble complex with functioning Wnt proteins. The 2.1-Å crystal structure of glycosylated human afamin reveals an almost exclusively hydrophobic binding cleft capable of harboring large hydrophobic moieties. Lipid analysis confirms the presence of lipids, and density in the primary binding pocket of afamin was modeled as palmitoleic acid, presenting the native O-acylation on serine 209 in human Wnt3a. The modeled complex between the experimental afamin structure and a Wnt3a homology model based on the XWnt8-Fz8-CRD fragment complex crystal structure is compelling, with favorable interactions comparable with the crystal structure complex. Afamin readily accommodates the conserved palmitoylated serine 209 of Wnt3a, providing a structural basis how afamin solubilizes hydrophobic and poorly soluble Wnt proteins.

Authors

Name Bernhard **Surname** Rupp Add author

Delete

Name	Surname
Andreas	Naschberger
Matthew W.	Bowler
Bernhard	Rupp

Mint

Dataset List

Name	Proposal	Technique
mesh-AFAMIN-revi-B5-1_1_1719726	OPID-1	
mesh-AFAMIN-revi-B5-1_1_1719728	OPID-1	
mesh-AFAMIN-revi-B5-1_1_1719731	OPID-1	
mesh-AFAMIN-revi-B5-1_1_1719733	OPID-1	
mesh-AFAMIN-revi-B5-1_1_1719736	OPID-1	
mesh-AFAMIN-revi-B5-1_1_1719739	OPID-1	
line-AFAMIN-revi-B5-1_2_1719742	OPID-1	
line-AFAMIN-revi-B5-1_3_1719744	OPID-1	
line-AFAMIN-revi-B5-1_4_1719745	OPID-1	
ref-AFAMIN-revi-B5-1_4_1719746	OPID-1	
AFAMIN-revi-B5-1_1_1719747	OPID-1	

Create your own DOI and cite them in your publications!

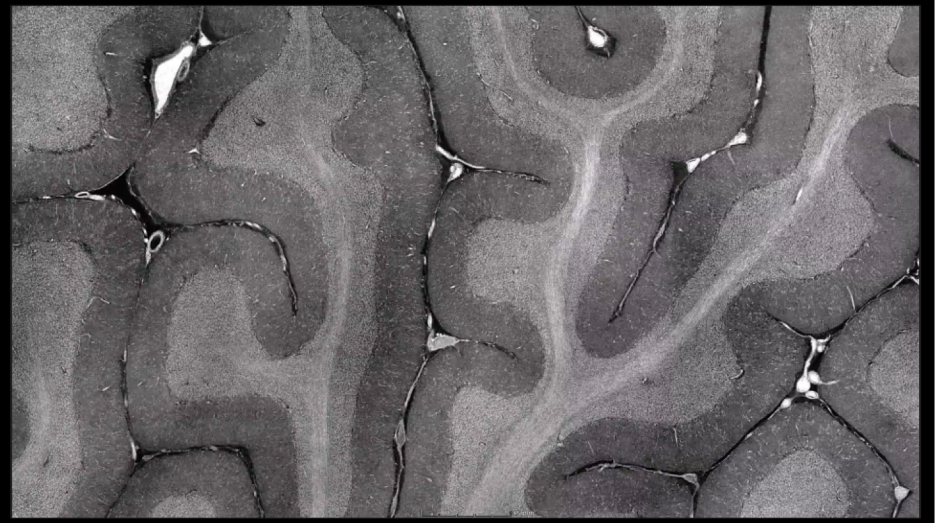
European Synchrotron Radiation Facility

Welcome to the Human Organ Atlas

The Human Organ Atlas uses **Hierarchical Phase-Contrast Tomography** to span a previously poorly explored scale in our understanding of human anatomy, the micron to whole intact organ scale.

Histology using optical and electron microscopy images cells and other structures with sub-micron accuracy but only on small biopsies of tissue from an organ, while clinical CT and MRI scans can image whole organs, but with a resolution only down to just below a millimetre. HiP-CT bridges these scales in 3D, imaging intact organs with ca. 20 micron voxels, and locally down to microns.

We hope this open access Atlas, enabled by the ESRF-EBS, will act as a reference to provide new insights into our biological makeup in health and disease. To stay up to date, follow @HiP-CT



HiP-CT imaging and 3D reconstruction of a complete brain from the body donor LADAF-2020-31. More videos can be viewed on the HiP-CT YouTube channel.

Funding

This project has been made possible by funding from:

- The European Synchrotron Radiation Facility (ESRF) — funding proposal MD-1252
- The Chan Zuckerberg Initiative, a donor-advised fund of the Silicon Valley Community Foundation
- The German Registry of COVID-19 Autopsies (DeRegCOVID), supported by the German Federal Ministry of Health
- The **Royal Academy of Engineering**, UK
- The **UK Medical Research Council**
- The **Wellcome Trust**

Collaborators

- UCL, London, England: **Peter D Lee, Claire Walsh, Simon Walker-Samuel, Rebecca Shipley, Sebastian Marussi, Joseph Jacob, David Long, Daniyal Jafree, Ryo Torii, Charlotte Hagen**
- ESRF, Grenoble, France: **Paul Tafforeau, Elodie Boller**
- Medizinische Hochschule Hannover, Germany: **Danny D Jonigk, Christopher Werlein, Mark Kuehnel**
- Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Germany:**M Ackermann**
- University Hospital of Heidelberg, Germany: **Willi Wagner**
- Grenoble Alpes University, Department of Anatomy, French National Center for Scientific Research: **A Bellier**
- Diamond Light Source, Harwell, UK: **Andy Bodey, Robert C Atwood**
- Imperial College London, UK: **JL Robertus**

https://human-organ-atlas.esrf.eu/

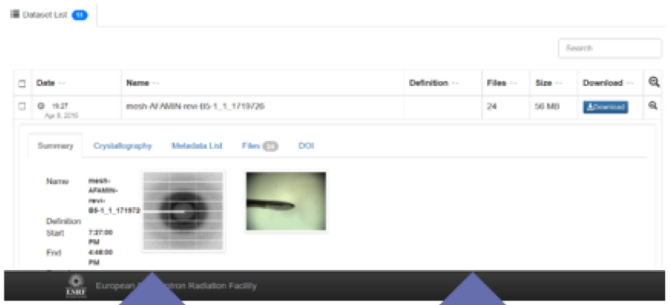
Reference

Walsh, C.L., Tafforeau, P., Wagner, W.L. *et al.* Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. *Nat Methods* (2021). https://doi.org/10.1038/s41592-021-01317-x

Aknowledgements

The development of this portal has been done as part of the PaNOSC project. PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852. The following people were involved in the development: Paul Tafforeau, Alejandro De Maria Antolinos, Axel Bocciarelli, Marjolaine Bodin and Andrew Götz from the ESRF, Jiří Majer from ELI, as well as the broader PaNOSC and ICAT communities.

OPEN DATA SEARCH



50 PB/yr
ESRF
(icat)

15 PB/yr
CERIC
(icat)

<1 PB/yr
ESS
(SciCat)

.6 PB/yr
ILL
(local)

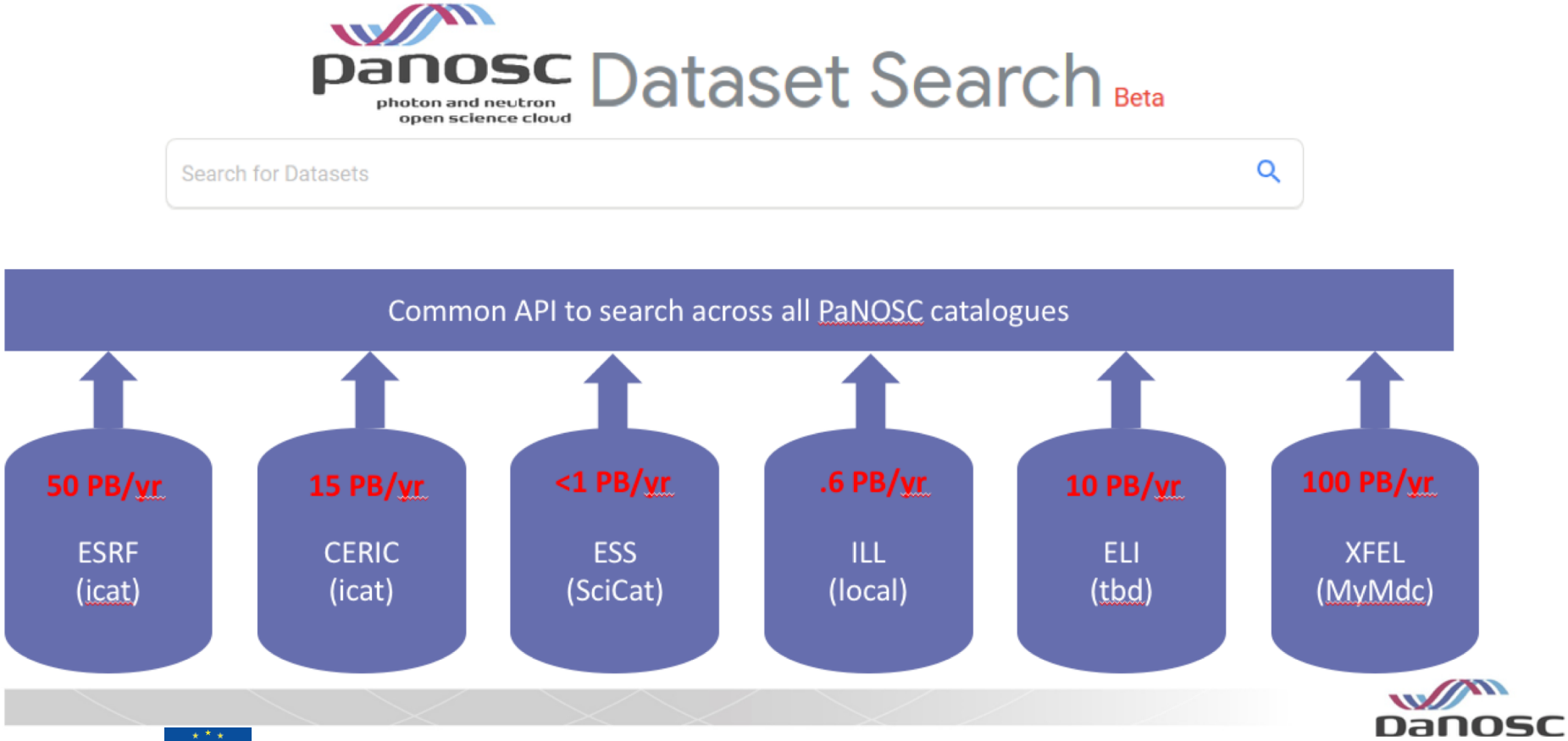
10 PB/yr
ELI
(tbd)

100 PB/yr
XFEL
(MyMdc)

https://data.esrf.fr

https://data.ill.fr

PANOSC DATASET SEARCH



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

GOOGLE – DATASET SEARCH (BETA)

datasetsearch.research.google.com/search?src=0&query=battery%20esrf%20tomography&docid=L2cvMTFqbnoyazd0Ng%3D%3D

Apprieu | Sports | Livebox | Maison | Routeur 4G | git | ESRF | GPU Cloud | Autre

Google

battery esrf tomography

Mis à jour | Format de téléchargement | Droits d'usage | Thème | Gratuits | Ensembles de données enregistrés

2 ensembles de données trouvés

In situ characterisation of the failure mechanisms of sulfur-...
doi.esrf.fr
Dernière mise à jour : 2021

Large field of view scanning, holographic nano-tomography f...
doi.esrf.fr
Dernière mise à jour : 2021

Un résultat que vous attendiez ne s'affiche pas ?
Découvrez comment ajouter de nouveaux ensembles de données à notre index

Large field of view scanning, holographic nano-tomography for connectomics and Li batteries characterization

Découvrir sur doi.esrf.fr

Identifiant unique
https://doi.org/10.15151/esrf-es-127995557

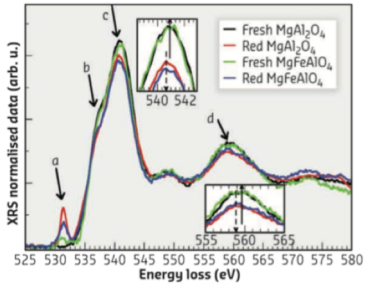
Ensemble de données mis à jour
2021

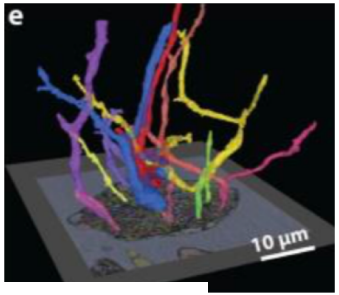
Ensemble de données fourni par
ESRF - The European Synchrotron Radiation Facility
datacite

Auteurs
Alexandra JOITA PACUREANU; Yang Yang; Peter CLOETENS; Julio Cesar Da Silva

Strategy: make ESRF data discoverable via google dataset search and other scientific search machines

PROJECTS @ ESRF & CONCLUSION





- Automated segmentation (BM18, ID16A), Connectomics
- Classification in spectroscopy
- Open data available for training
- API for fast data access & instrument control in development

- Ongoing EU projects:
  - STREAMLINE: AI-driven X-ray Microscopy
  - ENGAGE COFUND: 5 PhD @ESRF:
    - https://engage.cyi.ac.cy/?page_id=43
    - X-ray diffraction mapping applied to cultural heritage
    - Coherent X-ray imaging (near field)
    - Real-time analysis of spectroscopic data
    - AI for protein Xtallography
    - Bragg (strain) coherent diffraction imaging

- Future ? Infratech EU AI4SI project, BMBF…

- **Long-Term Projects** with user groups are highly encouraged and would be a great tool to collaborate on AI applications !

THANK YOU for listening !

Acknowledgements:

Gary Admans, Yuriy Chushkin, Andy Gotz, Alessandro Greco, Matias Guijarro, Steven Leake, Alessandro Mirone, Edward Mitchell, Wout de Nolf, Alexandra Pacureanu, Françoise Peyrin, Marius Retegan, Marie-Ingrid Richard, Mauro Rovezzi, Armando Sole, Olof Svensson, Nicola Vigano…