Analytical TEM faces the challenge of FAIRness

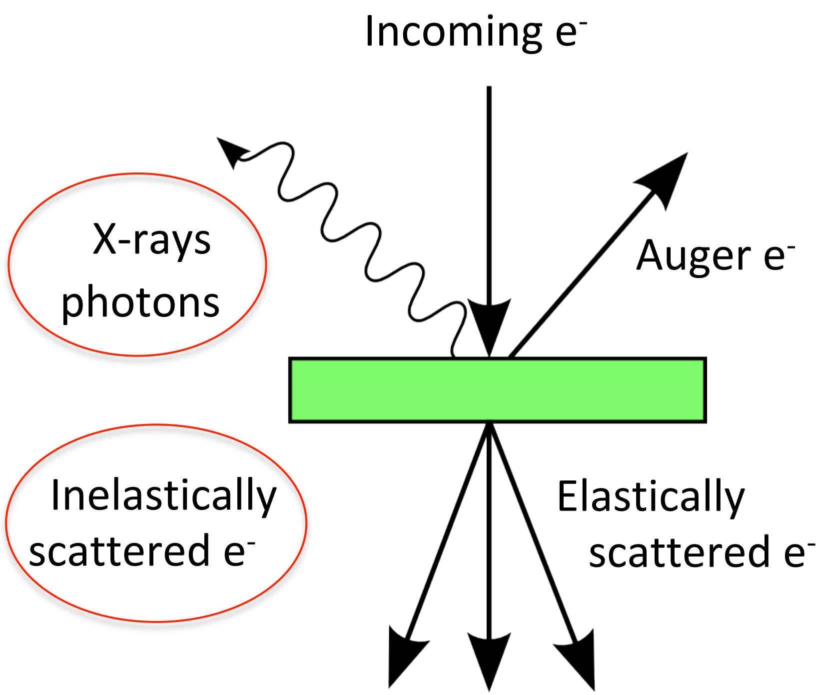
C. Hébert EPFL, Lausanne

A brief history of TEM

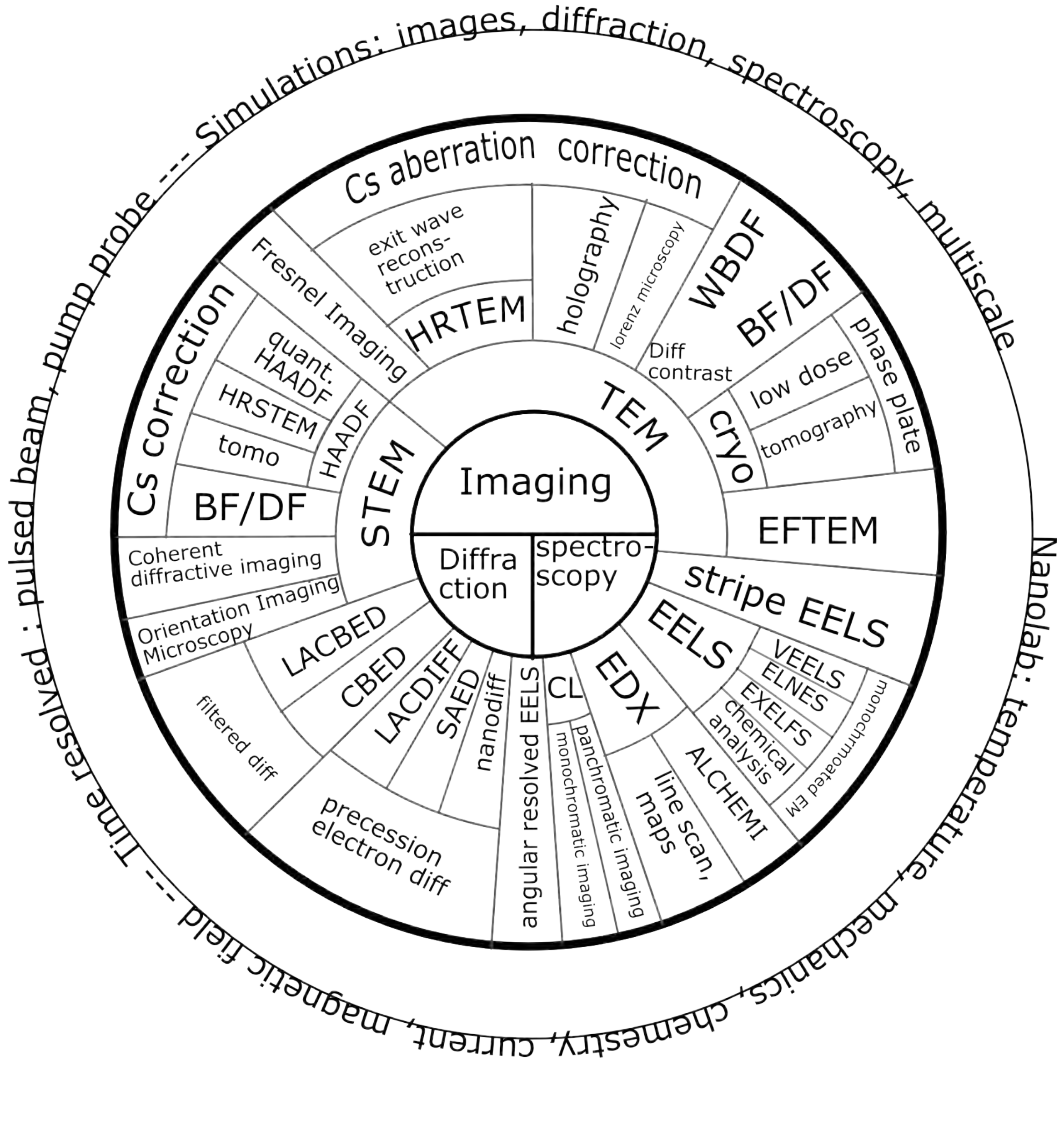
- 1933 First TEM Knoll and Ruska, resolutions overcomes that of Light microscopy
- 1939: First commercial electron microscopes are delivered
- 1941: The first EELS measurement recorded in TEM, by Ruthemann
- 1951: First microanalyzer of X-ray by Castaing
- 1965: Crewe describes the first STEM built at ANL
- 1979: Field emission Gun
- 1988: CCD in TEM using scintilators
- 1998: Cs corrector installed on a TEM by Haider et al, 1.3 Å
- 1999: Cs corrector installed on a STEM by Krivanek
- 2003: first commercial TEM with Cs correction

Principle of (analytical) TEM

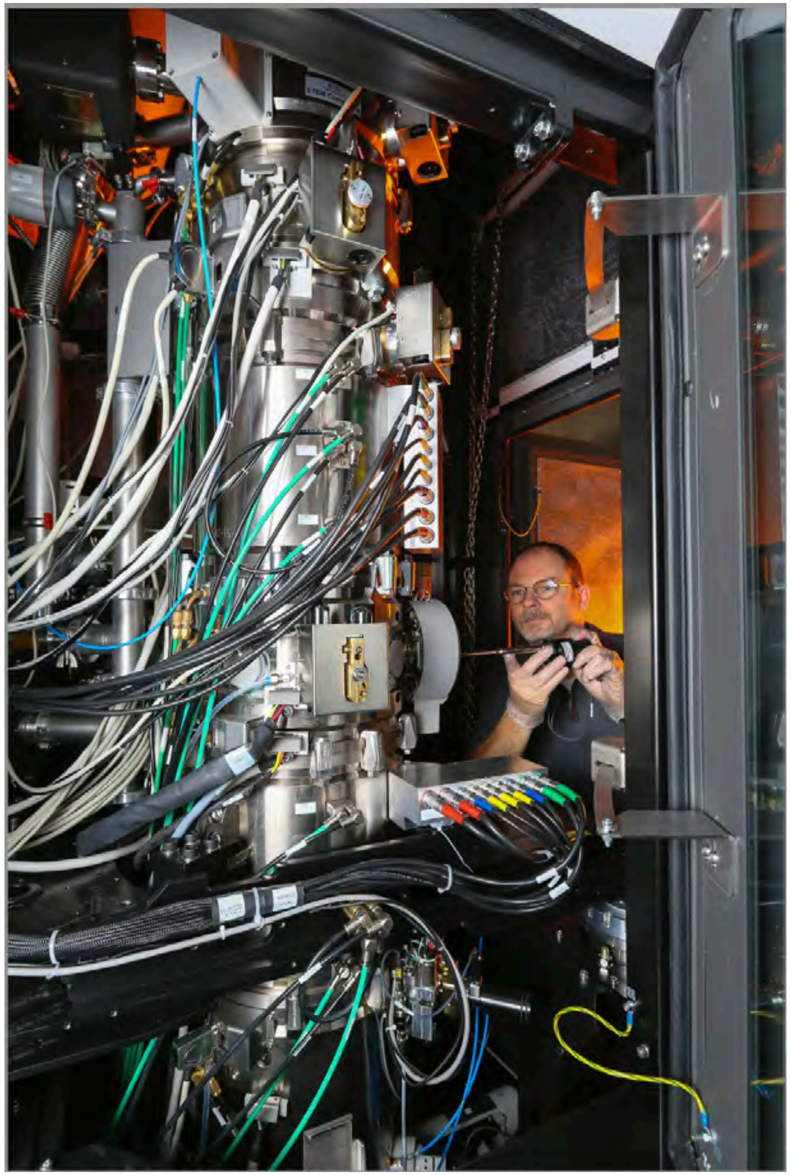
Probe = electrons
100-300 kV
(80…60…30…20…)
Velocity: 0.55-0.77 c



TEM techniques



Evolution of instruments

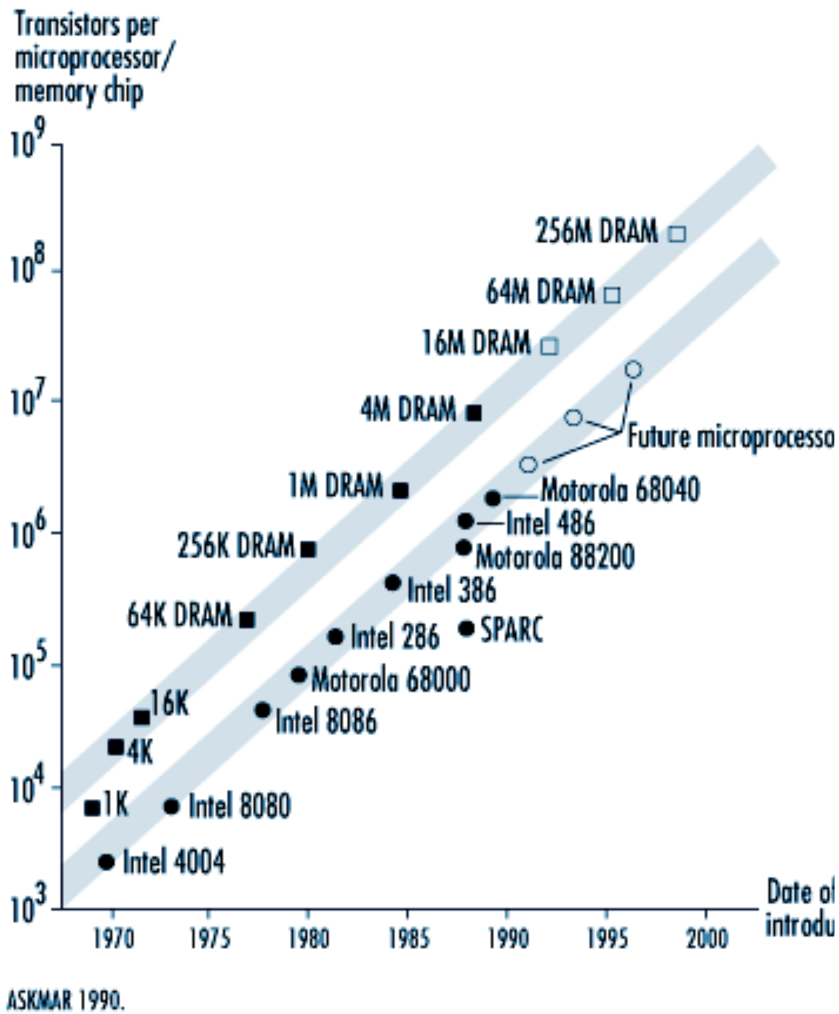


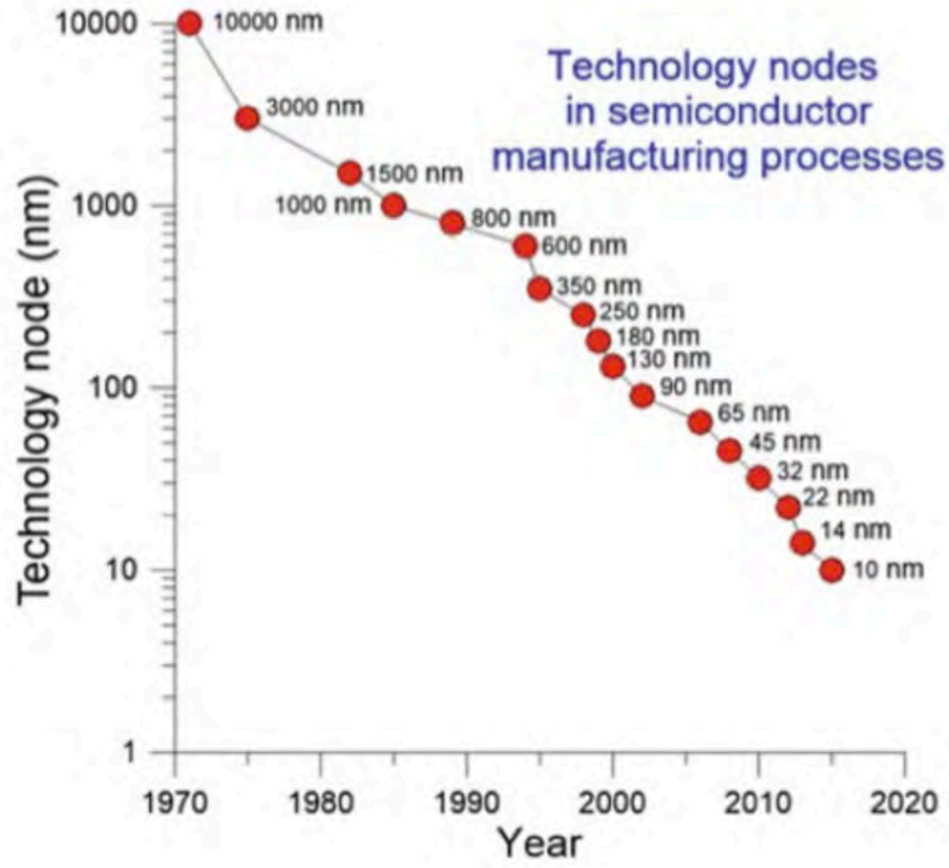
Evolution of instruments

Evolution of instruments

Evolution of instruments

Evolution of instruments





Thermo Scientific™

Metrios™ TEM for Semiconductors

Advanced logic and memory manufacturing processes are becoming more reliant on fast turnaround of precise structural and analytical data to be able to quickly calibrate tool sets, diagnose yield excursions, and optimize process yields. At technology nodes below 28nm, especially in cases where non-planar device designs are being implemented, conventional SEM or optical-based analysis and inspection tools cannot provide useful data. The Thermo Scientific™ Metrios™ transmission electron microscope (TEM) is the first TEM dedicated to providing the fast, precise measurements that semiconductor manufacturers need to develop and control their wafer fabrication processes.

High-volume TEM data, accurate and repeatable - at the lowest cost-per-sample

Metrios TEM automates the basic TEM operation and measurement procedures, minimizing the requirements for specialized operator training. Its advanced automated metrology routines deliver significantly greater precision than manual methods. The Metrios TEM is designed to provide customers with improved throughput and lower cost-per-sample than other TEMs.

Manufacturer’s statements

- *We provide export tools* (well hmh…)
- *We will not use a file format that will change from outside forcing us to adapt.* (Is there a chance to define an international standard?)
- *Readers exist anyway* (true but they are provided on a voluntary basis by scientist who have better things to do, and not always complete+take time.)

“Open data” so far

- Give access to the original file:
  Proprietary format.

- Give access to a converted file:
  Missing info and metadata

- Additional information not stored in the file
  Electronic Lab Notebook?

Hence we store them on a repository whenever demanded by a journal, but this is useless

Current situation

- Integrated microscopes, one environment to control the acquisition of all signals.

😃 linked information (one probe position, all signals)

😫 proprietary formats. No control on the metadata (but not too bad)

- Acquisition software include processing tool

😃 Intuitive usage. State of the art processing. Allows quick interactive checks on the fly during acquisition.

😫 State of the art is not cutting edge. Development very difficult

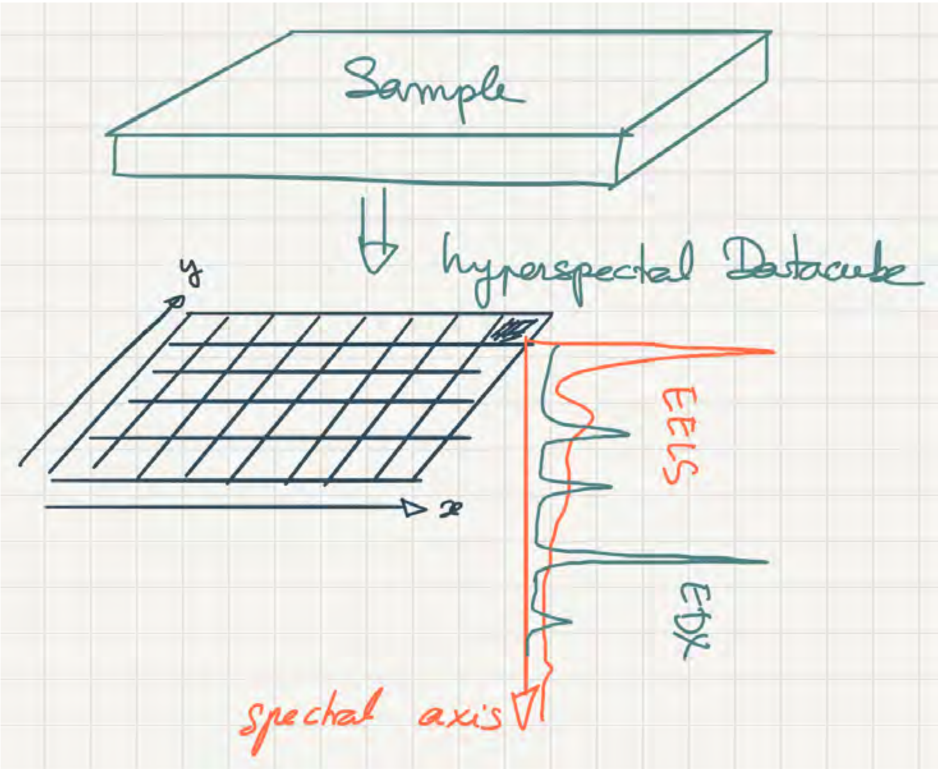
- Alternatives solution exist (hyperspy; scanning and acquisition tool LPS Orsay/ M. Tencé)

😃 Solutions **exist**

😫 conversion of metadata and format incomplete/cumbersome. Huge barrier for learning. Research “dead-time” and “dead-ressources”

Analytical (S)TEM:EELS & EDX

Life in the lab:



Acquisition on FEI Titan possible with

- TIA (FEI soft EELS & EDX)
- Velox (FEI soft EELS & EDX)
- Gatan (EELS & EDX)
- Brukker (EDX)

Same hardware, 4 file formats, 4 softs, 4 metadata scheme, all pros and cons

Difficulties in implementing good practices

- Lab culture , background of people
- Manufacturer’s tool makes it more complicated
- Open tools are *way* less intuitive to use for non programmers
- 99 % of users of a TEM facility are OK with the tools as they are and do not want to “loose more time” on the TEM investigation.
- Microscopists (1%) needs to know *both* (manufacturer softs used at the TEM)
- Not much real hard money

Light at the end of the tunnel ?

- External drive (funding agencies and institutions)
  - Not yet sufficient (data management plan does not mean anything)
- Awareness of students is increasing
  - But mainely always pushed as 2. priority
- Machine learning tools are developed in an open environment, they become a must in data analysis.
  - But are used by a minority.
- Initiatives. Hyperspy. Renku.
  - Time and energy consuming

Typical sample: example 1

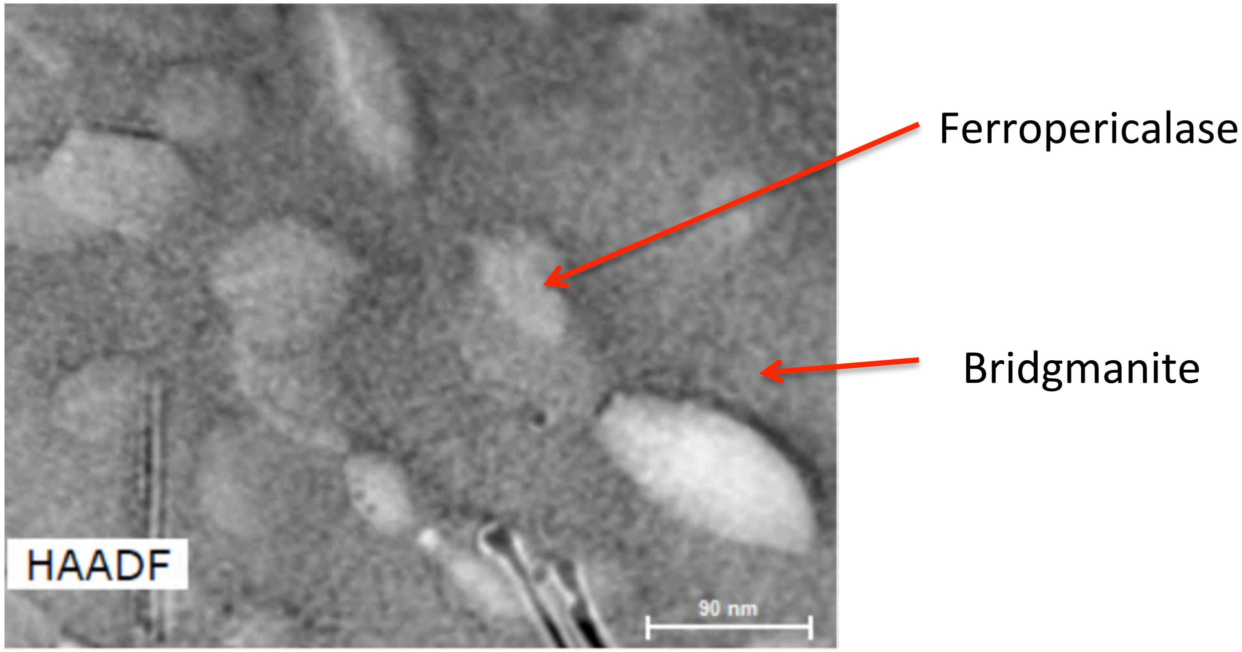
Artificial mineral synthesized by diamond anvil cell (~50GPa) + laser heating (~3000K)

Fe, Mg, Al, Si, O, Ca + (traces of Nd, Sm, U, Hf) (+Ga: FIB sample prep)

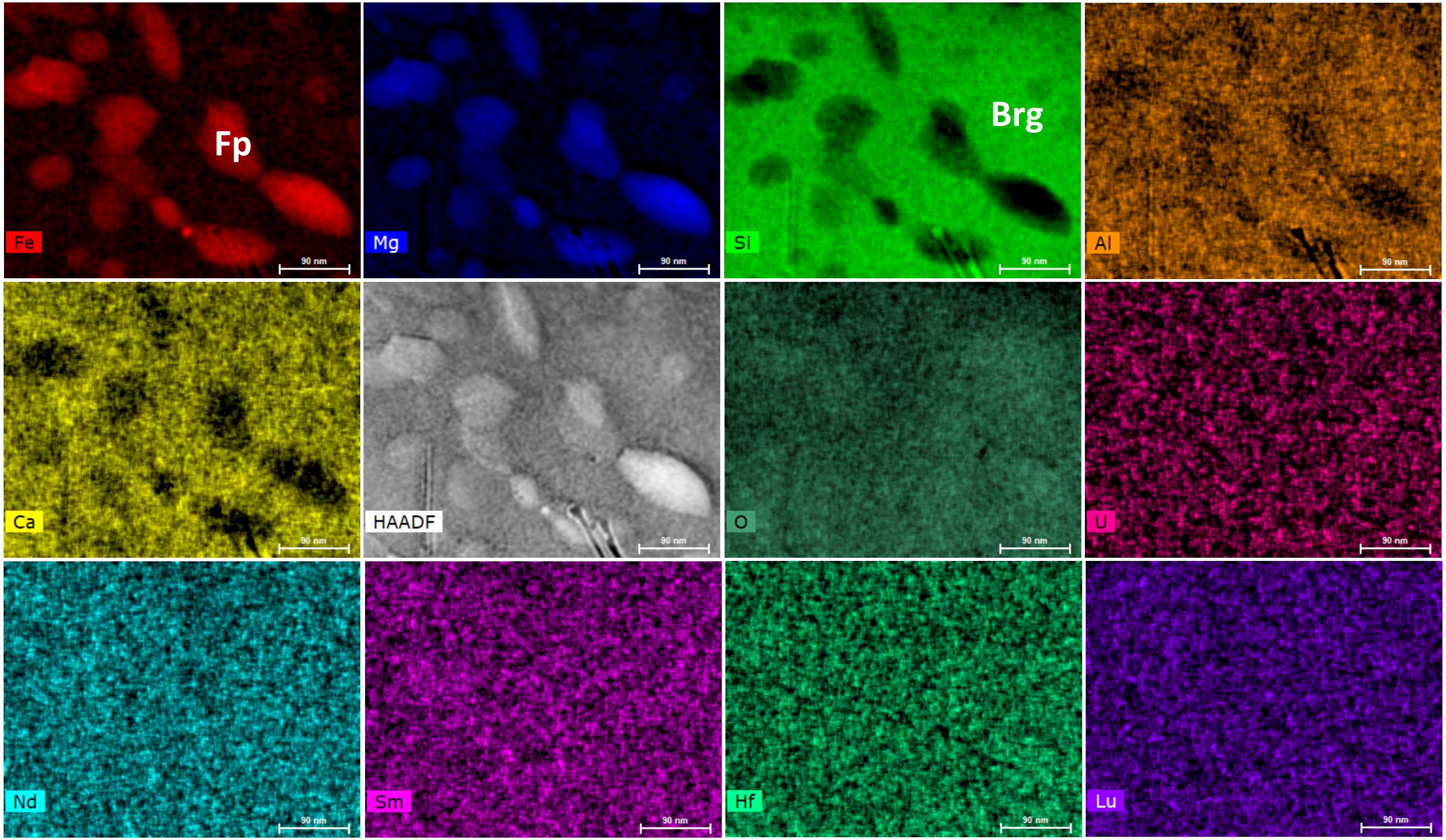
2 main phases. Aim: access the way dopants separate between phases

Hyper spectral EELS & EDX data accessible. (here EDX is shown)

Tool : machine learning, quantification methods.

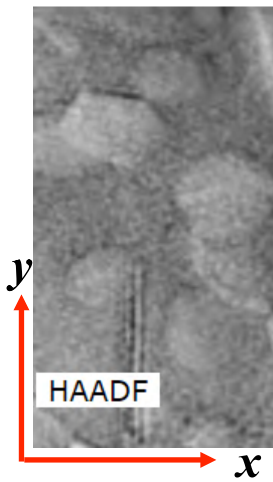


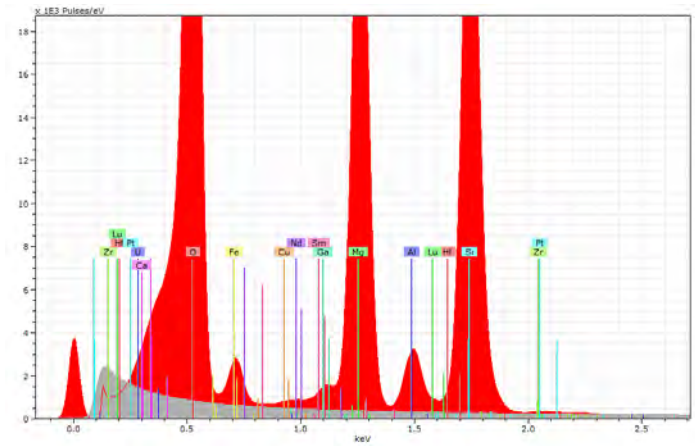
- EDS elemental intensity maps FEI tecnai Osiris (esprit)



Goto: Hyperspy

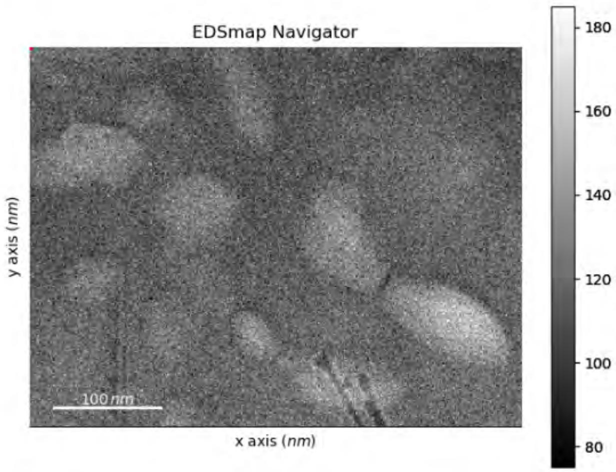
- **A representative integrated EDS spectrum**

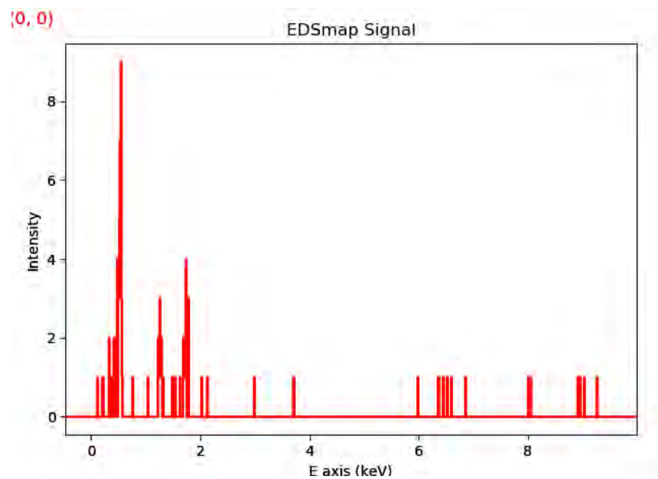




Spectrum = **background signal** + **sets of Gaussian peaks** (each specific element in the sample yields a unique set of Gaussian peaks, like a signature in ***E*** axis) + noise

- **A representative single pixel EDS spectrum**





```
In [25]: s.original_metadata
```

```
Out[25]: ├── byte-order = dont-care
         ├── data-length = 1
         ├── data-type = unsigned
         ├── date =
         ├── depth = 2048
         ├── height = 293
         ├── key = value
         ├── offset = 0
         ├── record-by = vector
         ├── signal =
         ├── time =
         ├── title =
         └── width = 379
```

```
In [26]: s.metadata
```

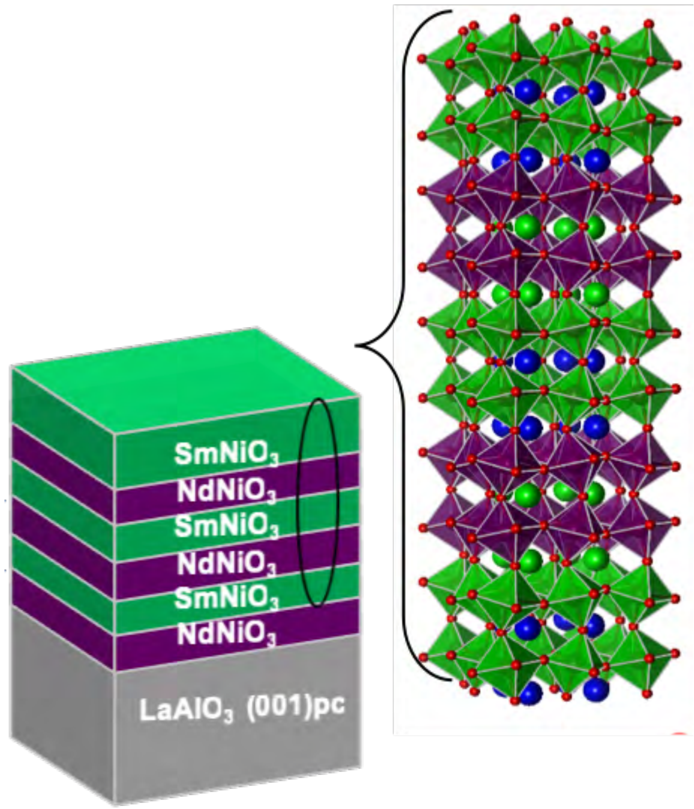
```
Out[26]: ├── Acquisition_instrument
         │   └── TEM
         │       ├── Detector
         │       │   └── EDS
         │       │       ├── azimuth_angle = 45
         │       │       ├── elevation_angle = 35.0
         │       │       └── energy_resolution_MnKa = 130.0
         │       ├── Stage
         │       │   └── tilt_alpha = 0.0
         │       └── beam_energy = 200
         ├── General
         │   ├── date =
         │   ├── original_filename = 1 selection.rpl
         │   ├── time =
         │   └── title = EDSmap
         ├── Sample
         │   ├── elements = ['Al', 'Ca', 'Fe', 'Hf', 'Lu', 'Mg', 'Nd', 'O', 'Si', 'Sm', 'U']
         │   └── xray_lines = ['Fe_Ka', 'Si_Ka', 'Mg_Ka', 'Al_Ka', 'Ca_Ka', 'Nd_La', 'Sm_La', 'Hf_La', 'Lu_La', 'U_La', 'O_Ka
         '] 
         └── Signal
             ├── binned = True
             └── signal_type = EDS_TEM
```

😩 😩 😩

Example 2

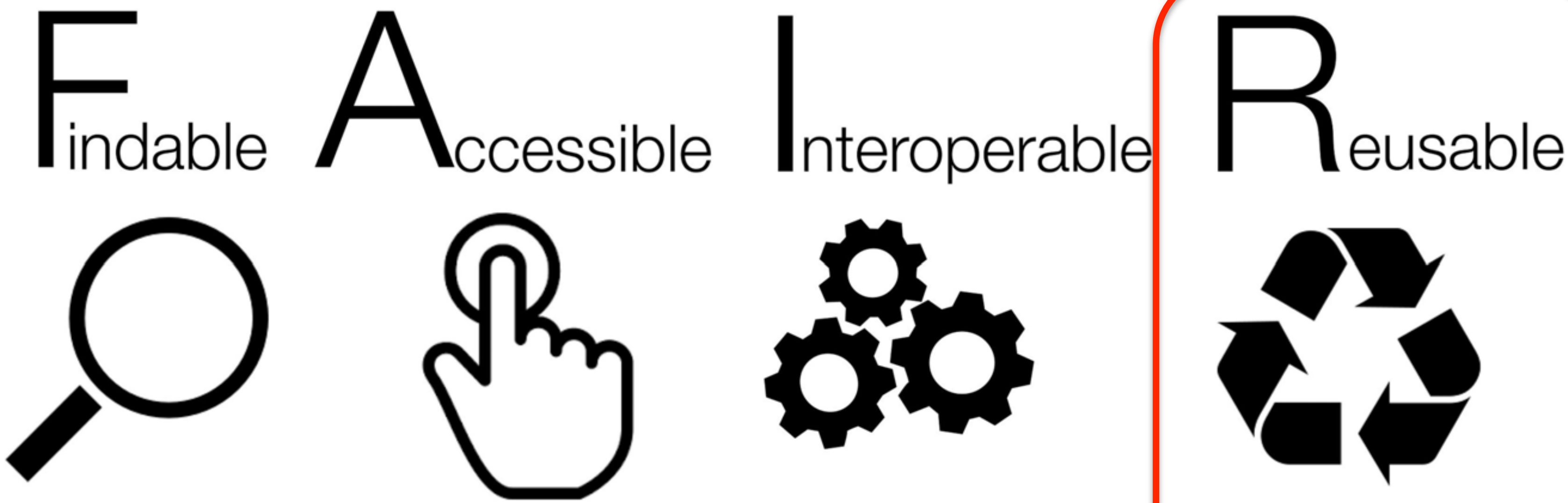
Perovskyte structure. Nickalates superlatices. Sample JM Triscone, Geneva.
TEM investigation B. Mundet, LSME, EPFL.

DFT calculations available Philippe Ghosez at Theoretical Materials Physics, Q-MAT, CESAM, University of Liège



Challenges/Discussion:

How to make the data linked to those projects FAIR ?



Database

Open formats

??????

Proper documentation
metadata

Working group 1

Cécile Hébert, Andy Stewart, Alberto Eljarrat, Rachel Nicholls, Markus Kühbach, Heike Görzig

File formats /data models

- Should we recommend a standard format? (e.g.NeXuS?), naming scheme?
- Tags/graphs vs hierarchical structure
- Table comparing key words in calculation and experiment (common language)

Motivation

- With common file format we can create databases collecting the measurements on materials
- This is requested by funding agencies
- Sharing datasets will allow groups developing methods to make methods more robust tested against many datasets (content and variability) reproducibility of experiments.
- For this purpose we need **open** and **documented** format

Data format:

- There are already several data format that could be suitable, needs investigation
- NeXuS, HDF5 EOS, FITS (astronomy) , envi/IDL, nrrd, Nion data format… ??
- Needs to define structure and namespace (start comparing existing)
- Ingredients (data, IDs, links, metadata+ terminology dictionary …)

Look at communities who had the same issues

- Synchrotron
- CryoEM
- Nomad
- ExPaNs
- RDA organisation /meetings

Organisation of format: what needs to be considered

- Storage of contents and of additional information (IDs, DOI, metadata…)
- Keep ability to link to a wider ensemble, establish logical connections between datasets, origin of sample / material etc.
- Needs to separate in different level of correlation and decide what is one file:
    - Simultaenous acquisition of signals is one experiment (high level of correlation).
    - Serial acquisition of signal of the same sample area
    - Same TEM lamella different position, same day
    - Same TEM lamella different days /microscopes
    - …
- Have a structure allowing cross-links between datasets. / can be linked to keyword discussion.
- Metadata scheme. General agreement on main categories: microscope, detector , sample. Defining fine grained structure will require large community. Shall the metadata be human readable?

We are not yet at the stage of linking experiment with theory, because this only comes when the data are processed.

As an example dielectric function can be retrieved from EELS data, but this requires some processing. Cross link will arrives at the final step. Processed data can be injected in the same database, linked to the original, including workflow.

Action steps

- Compile this into 3 pages of a paper
- Contact other other labs (Rafal: CH,)
- Find hard money
- Convince higher instances (EMS, …)
- RDA community / workshop https://www.rd-alliance.org/plenaries/rdas-14th-plenary-helsinki-finland
- m.kuehbach@mpie.de
- Chose a practical test case