Big Data in analytical TEM

C. Hébert EPFL, Lausanne

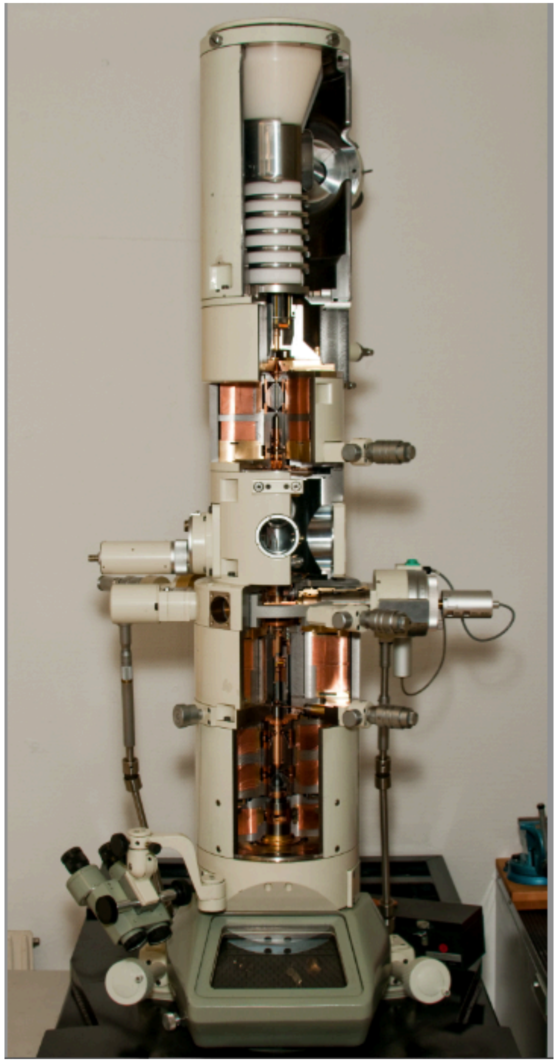
Outline

- Quick presentation of Transmission electron microscopy
- Lots of data, but is it big data ?
- (Long) way to go to be FAIR
- Do we want to move thinks ? Why ? How ?

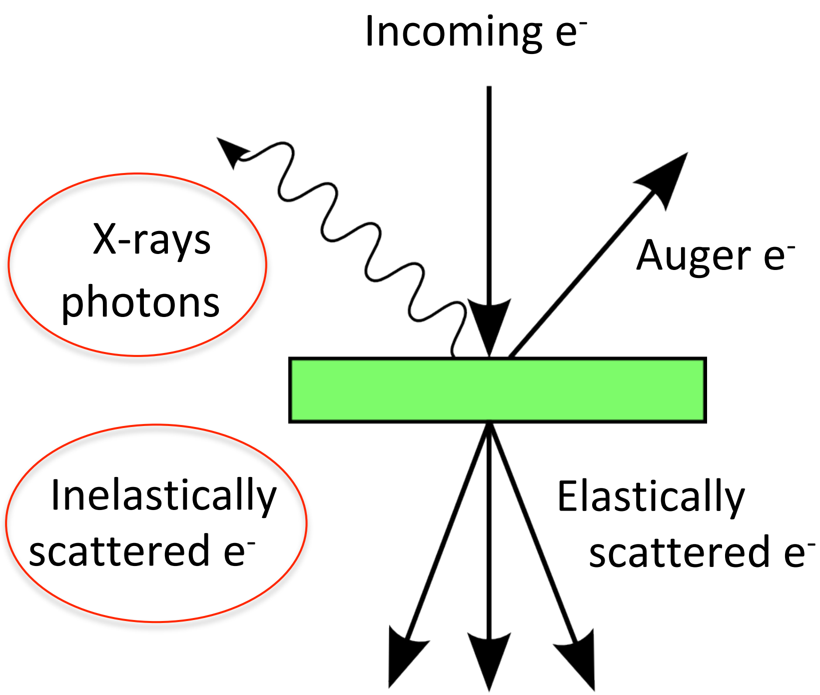
A brief history of TEM

- 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

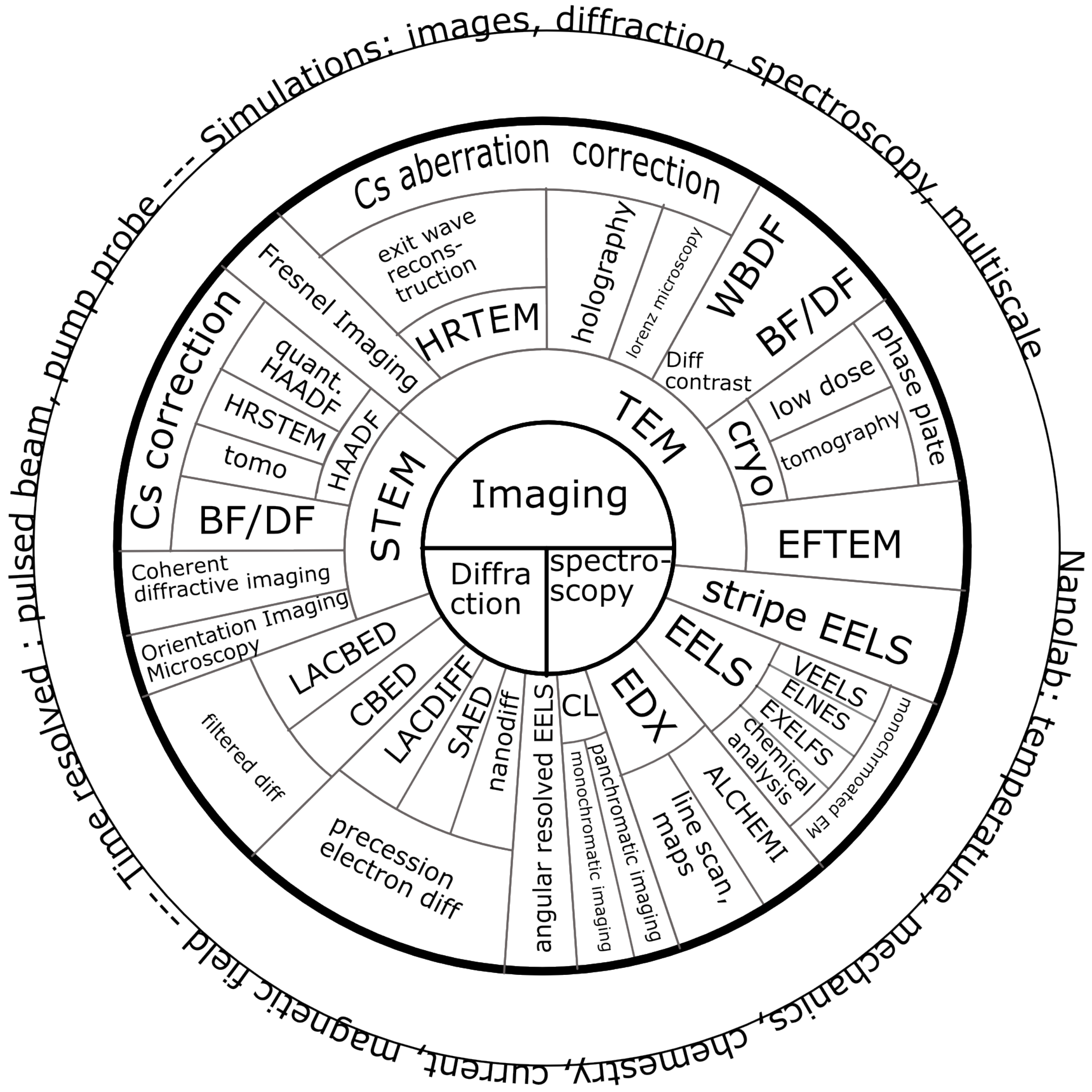
Principle of (analytical) TEM



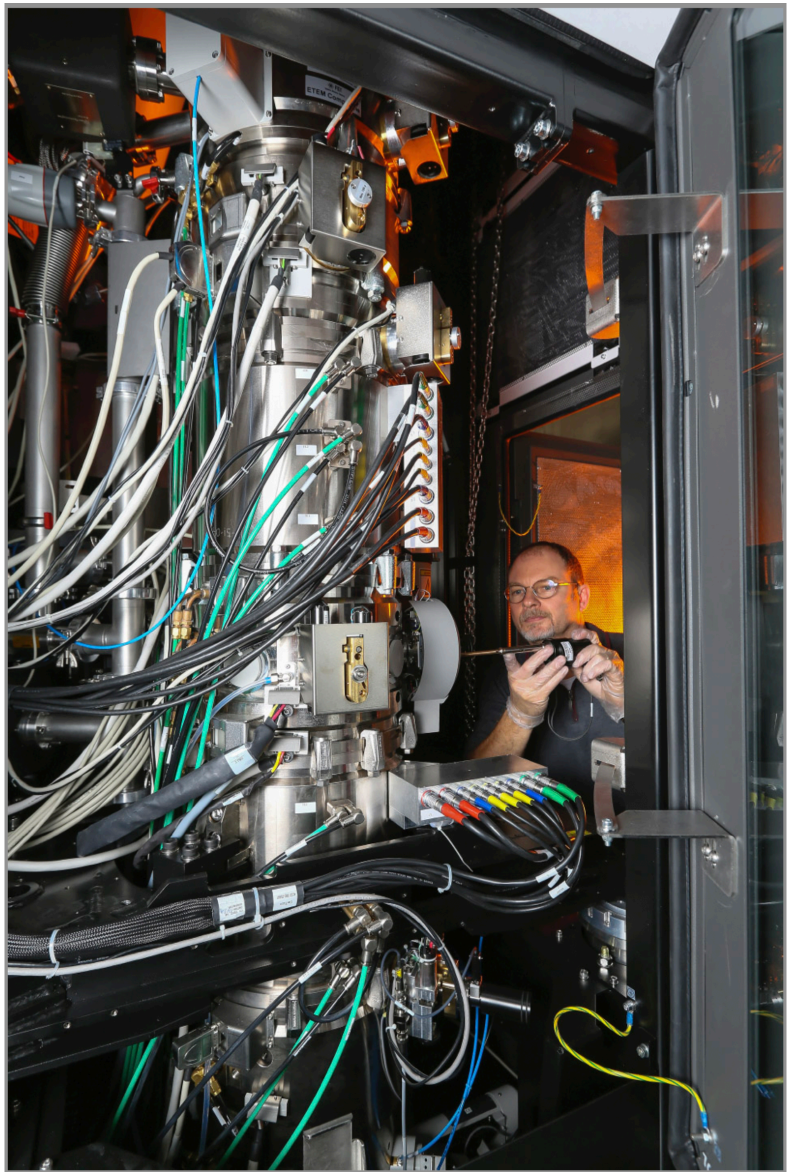
Probe = electrons
100-300 kV
(80…60…30…20…)
Velocity: 0.55-0.77 c



TEM techniques



Evolution of instruments

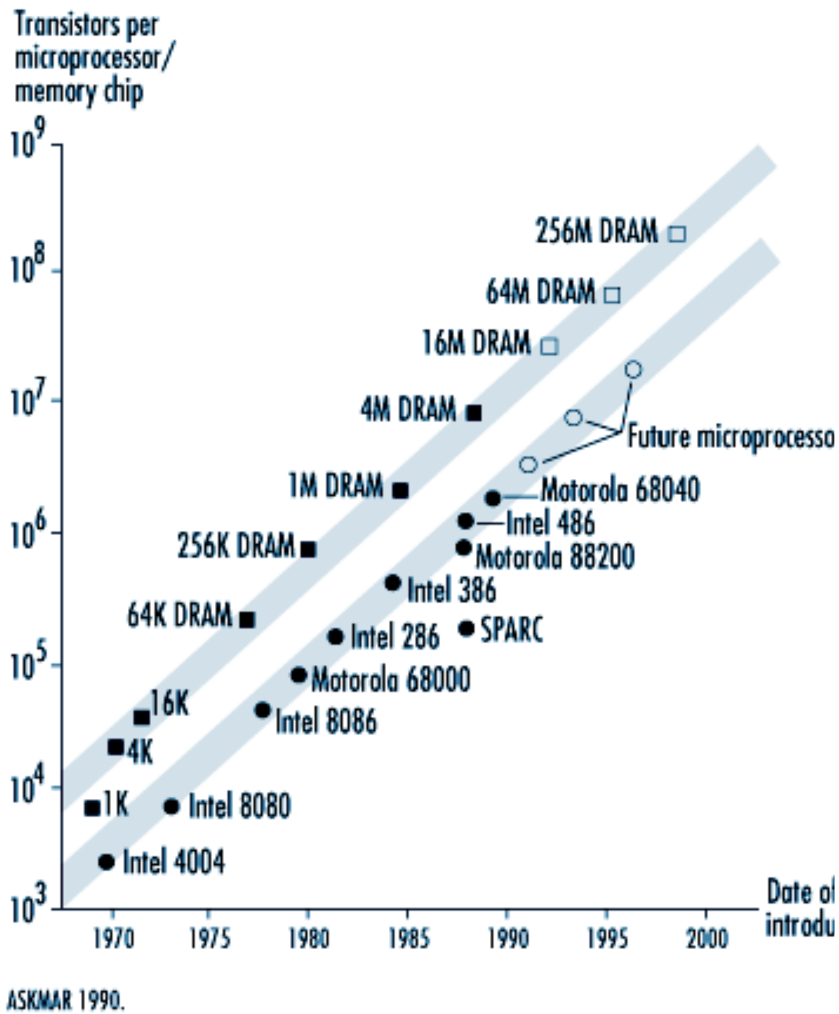


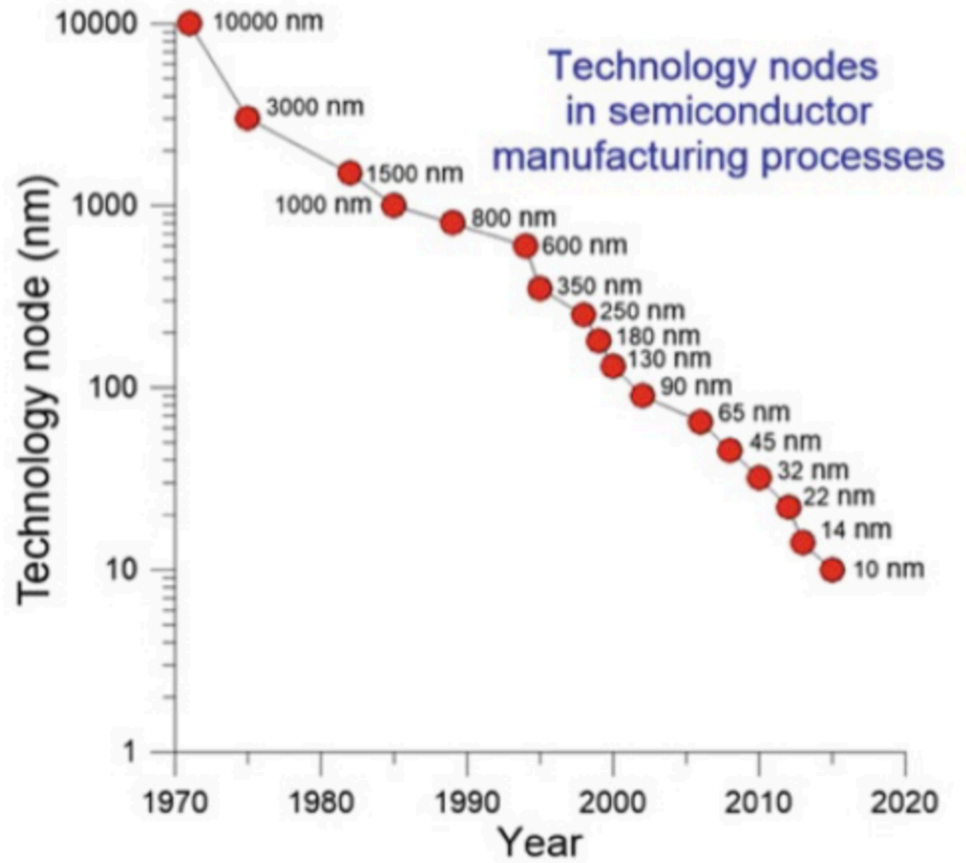
Evolution of instruments

Evolution of instruments

Evolution of instruments

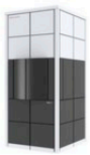
The big market for TEM manufacturers





Thermo Scientific™

Metrios™ TEM for Semiconductors

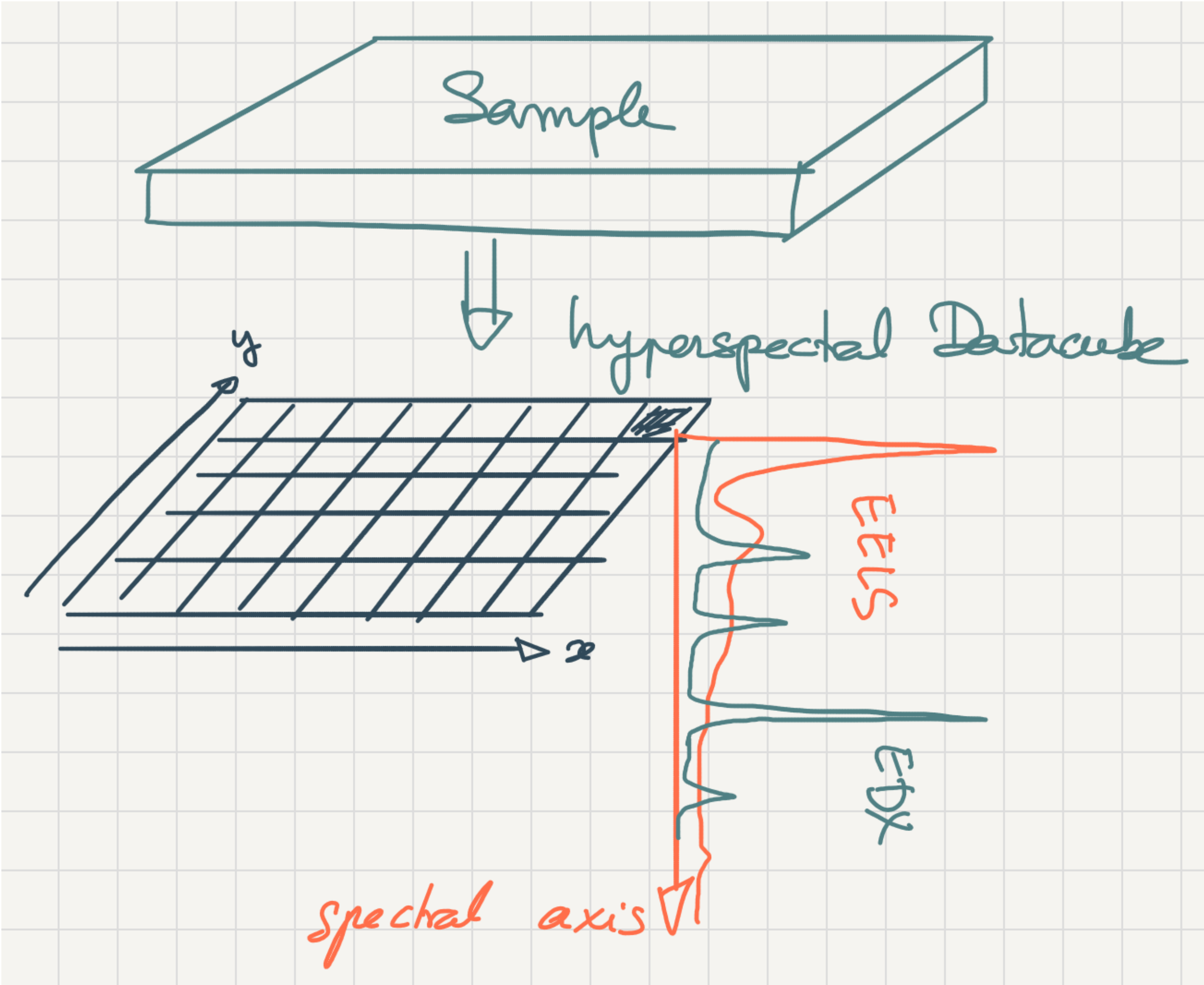


Advanced logic and memory manufacturing processes are becoming more reliant on fast turnaround of precise structural and analytical data to be able to quickly calibrate tool sets, diagnose yield excursions, and optimize process yields. At technology nodes below 28nm, especially in cases where non-planar 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.

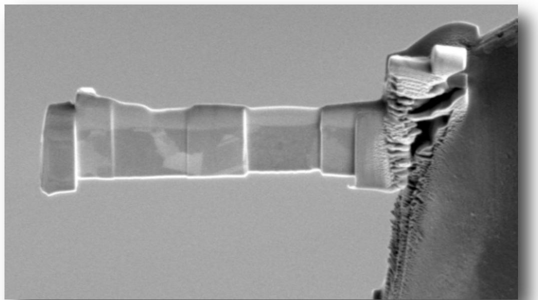
Analytical STEM: EELS & EDX



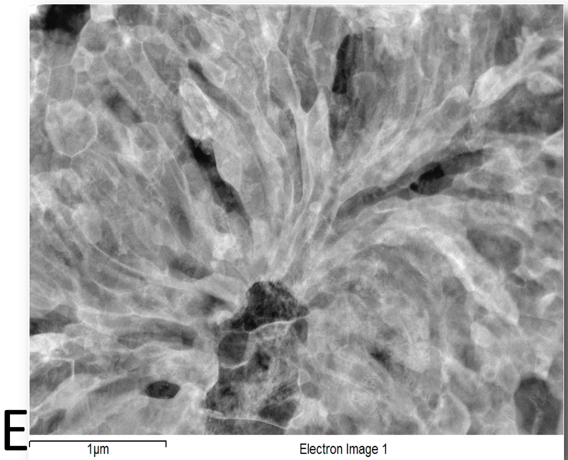
EPFL: 1990s - ~2010

- "Old tech": CM300 FEG, Si-Li detector; small solid angle of collection =>
  - Slow (100s counts/s)
  - Artefacts (signals from pole piece)
  - Poor light element sensitivity
- Applications example: Nb_3Sn superconducting cables in Cu-Sn matrix

FIB-prepared TEM lamella



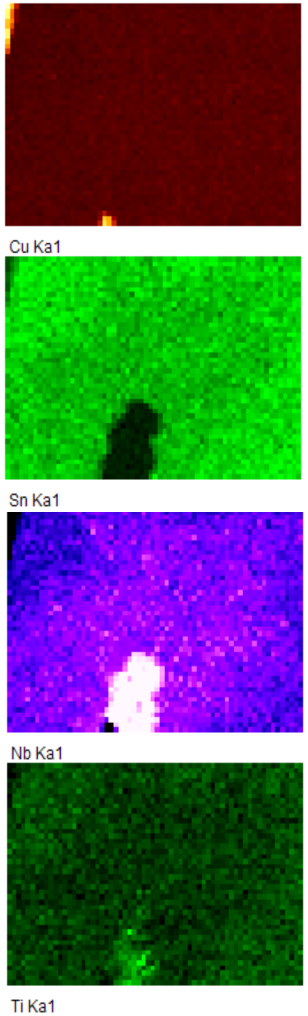
Dark-field STEM image



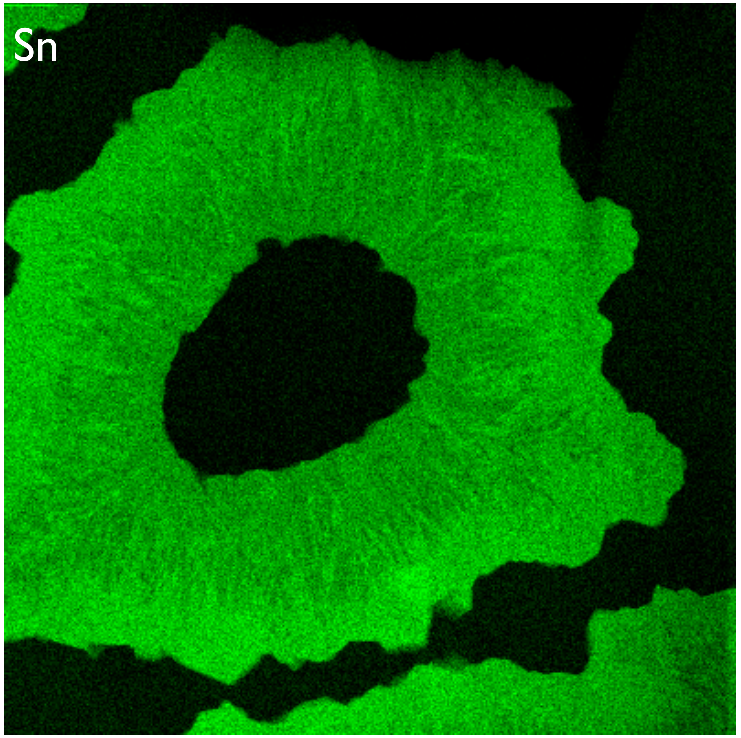
Marco Cantoni, CIME

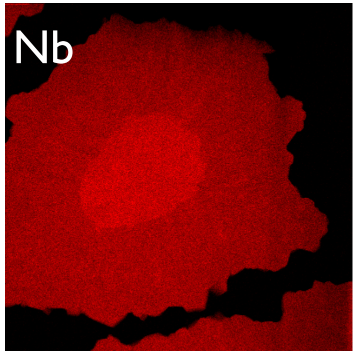
60min. 128 x 96 pixels

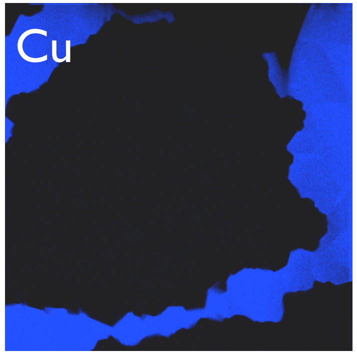
EDS, element maps

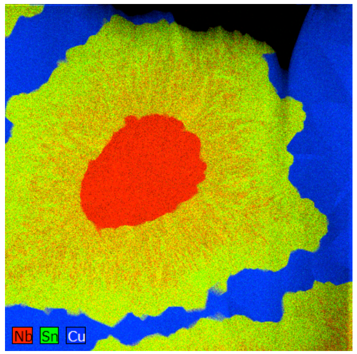


EPFL, now







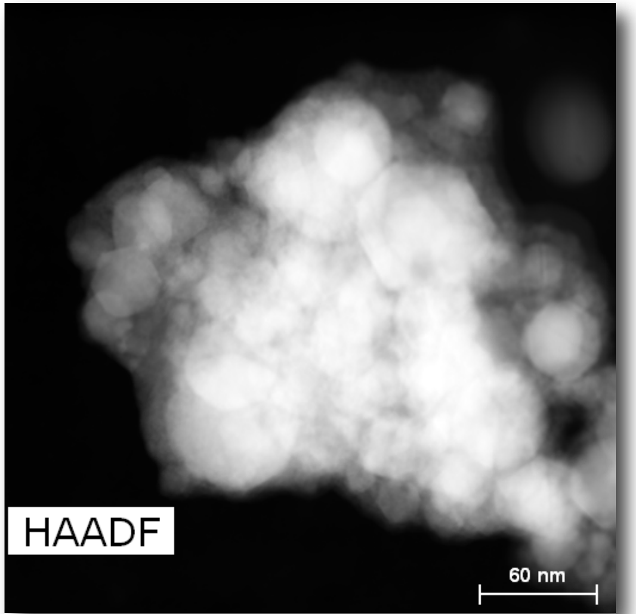


400 x 400 pixels (5 µm x 5 µm)
160’000 spectra
10 minutes total

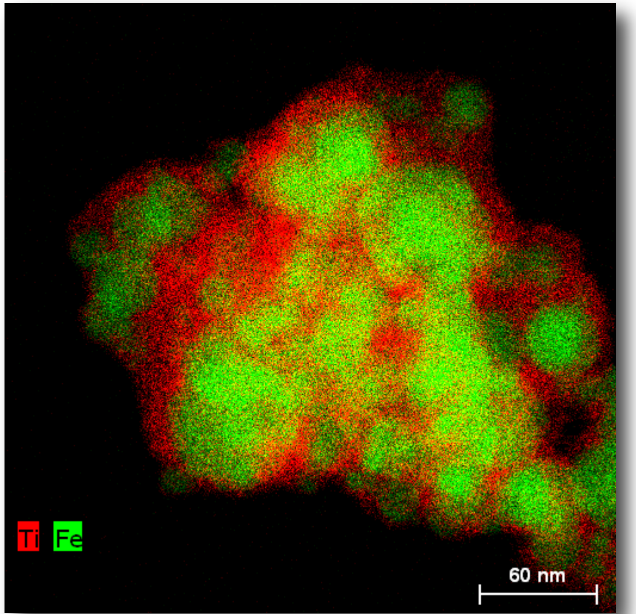
STEM-EDX as regular service tool

- With fast mapping we can use STEM-EDX as regular feedback tool for materials synthesis or deposition
- Results from a typical couple of hours session
- Example : proof of TiO_2-coated Fe_2O_3 core-shell nanoparticles

HAADF image



Map of Ti and Fe

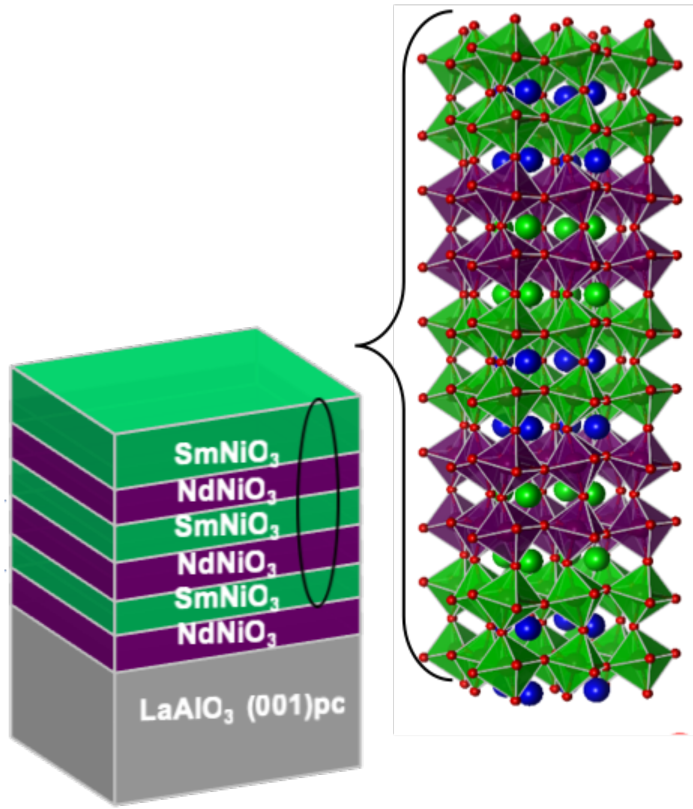


Tommaso Nardi and Yves Leterrier LTC, EPFL and Duncan Alexander, CIME

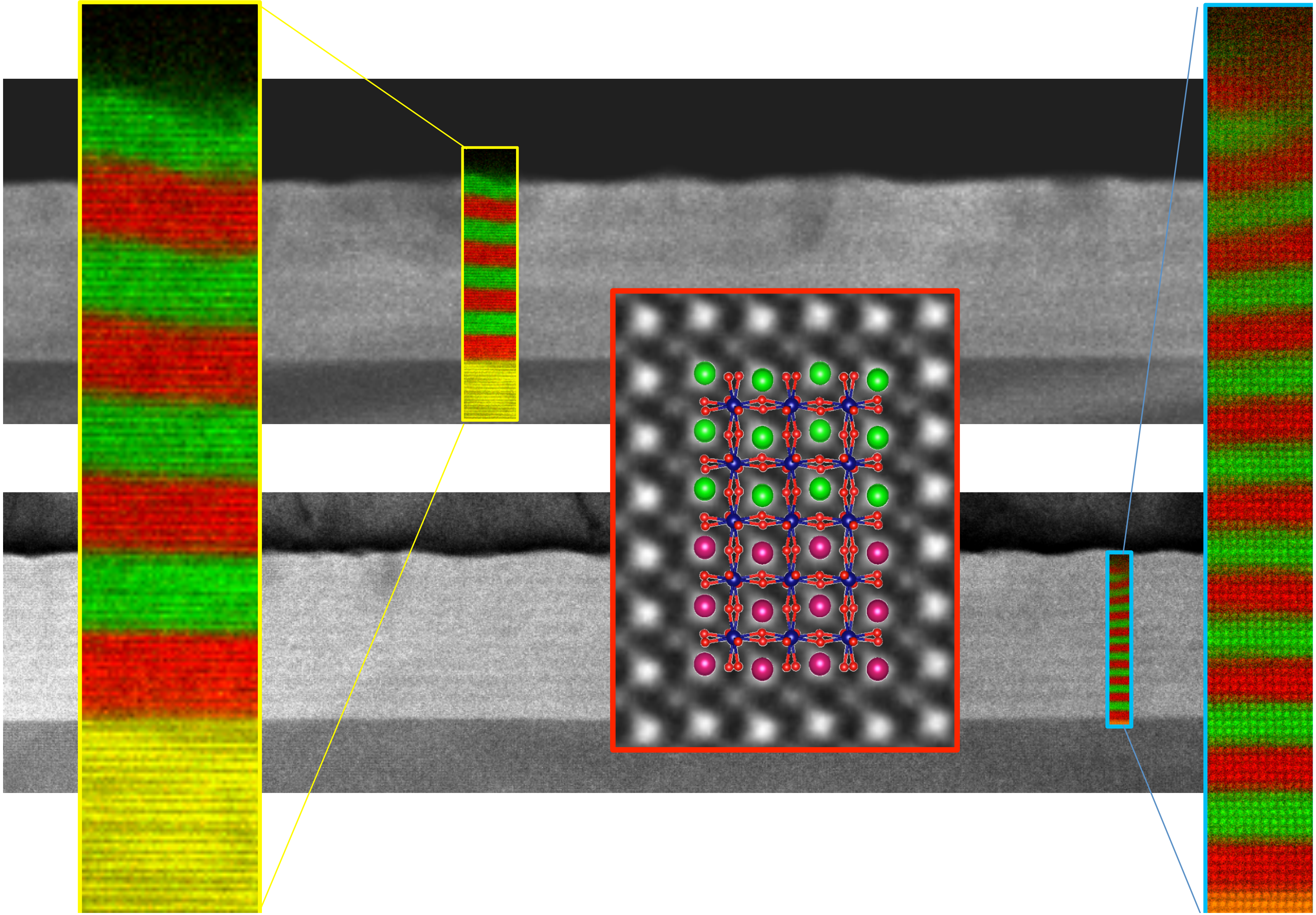
Example 2

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



Available data: HAADF-STEM, HRSTEM; EELS (not same region)



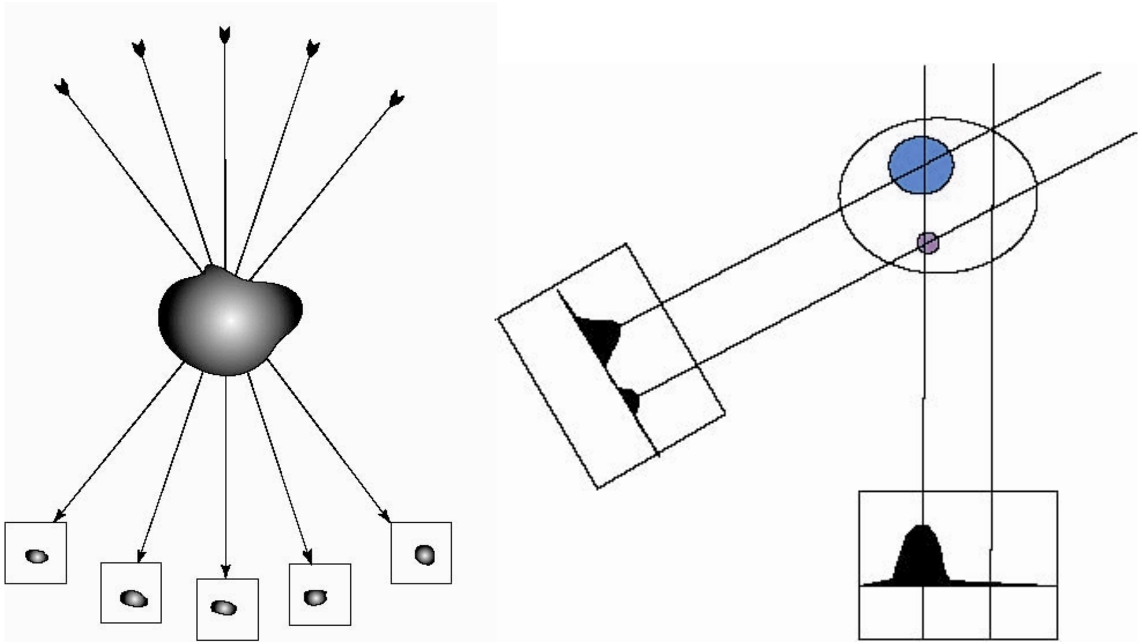
Data volume

- 1 datacube ~1 Gb
- ~5-10 datacubes / day 200 day/year
- 1-2 Tb per analytical TEM and year
- Small compared to:
  - Analytical Tomography
  - 4D STEM (can reach 100 Gb/data-hypercube)
  - In-situ observation (1000 fps at 1M pixel: 1Gb/s)

- Hence the statement of a TEM center able to generate 1 petabyte / year

TEM Tomography

Basic idea: recover the 3D structure of the object investigated
Developed for life sciences
Useful for materials sciences (but additional difficulties)



acquisition projection

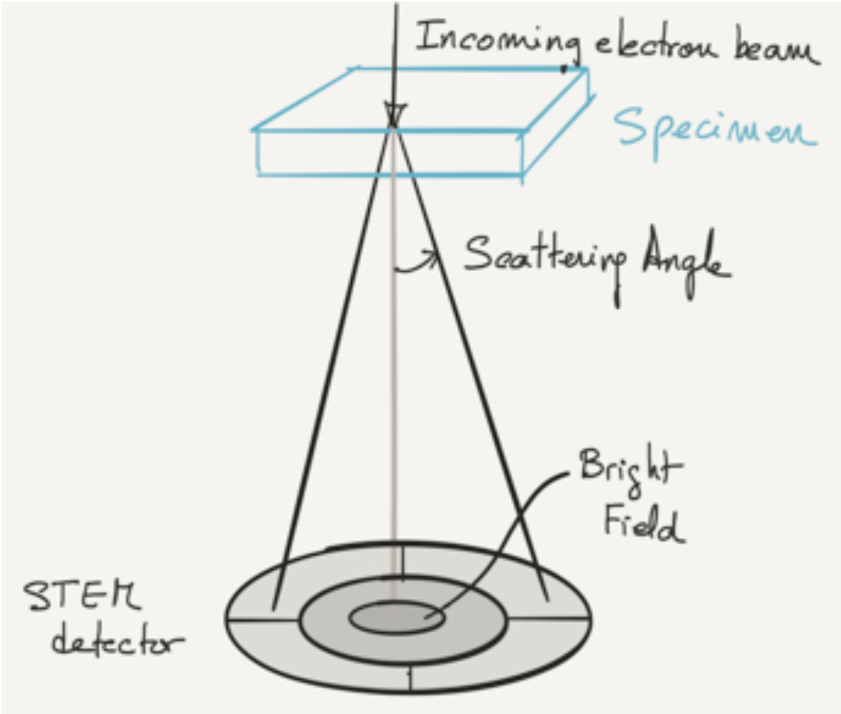
Different algorithms for backprojection

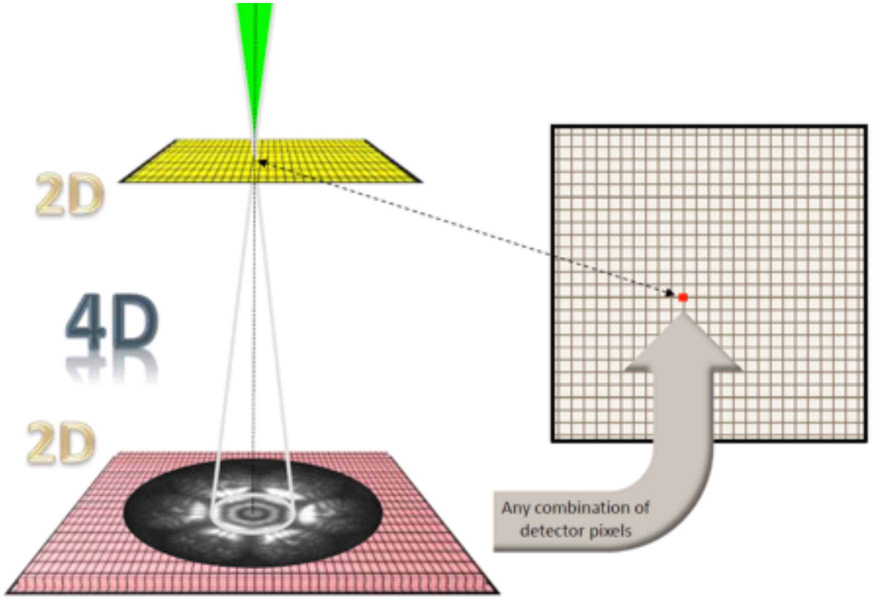
~100-150 images/ tomogramm

Can be applied to spectroscopy.

- Sparse (noisy) data
- 10-50 Gb/ hypercube

4D STEM





New technique, field of new, open developments
LiberTEM (Jülich research center, D Weber
Pycroscopy: oak ridge national lab
Dectris detector entering soon ?

Big data ??

One TEM sample : ←1 µm→ × 100 nm $(10^{-6})^2 \cdot 10^{-7} = 10^{-19}\ m^3$

Average TEM-day : 5 samples }
Average day / year 200 }   1000 samples / y

$10^{-16}\ m^3$ / year & TEM

EPFL 5 TEM 10'000 students   TU Graz 16'000
4 TEM

EMAT 6 TEM 20'000 students   ...

Take 5 TEM per 20'000 students

~ $200 \cdot 10^6$ students worldwide → 10'000 TEM worldwide

⇒ $10^{-12}\ m^3$ matter / year   ~ 80y

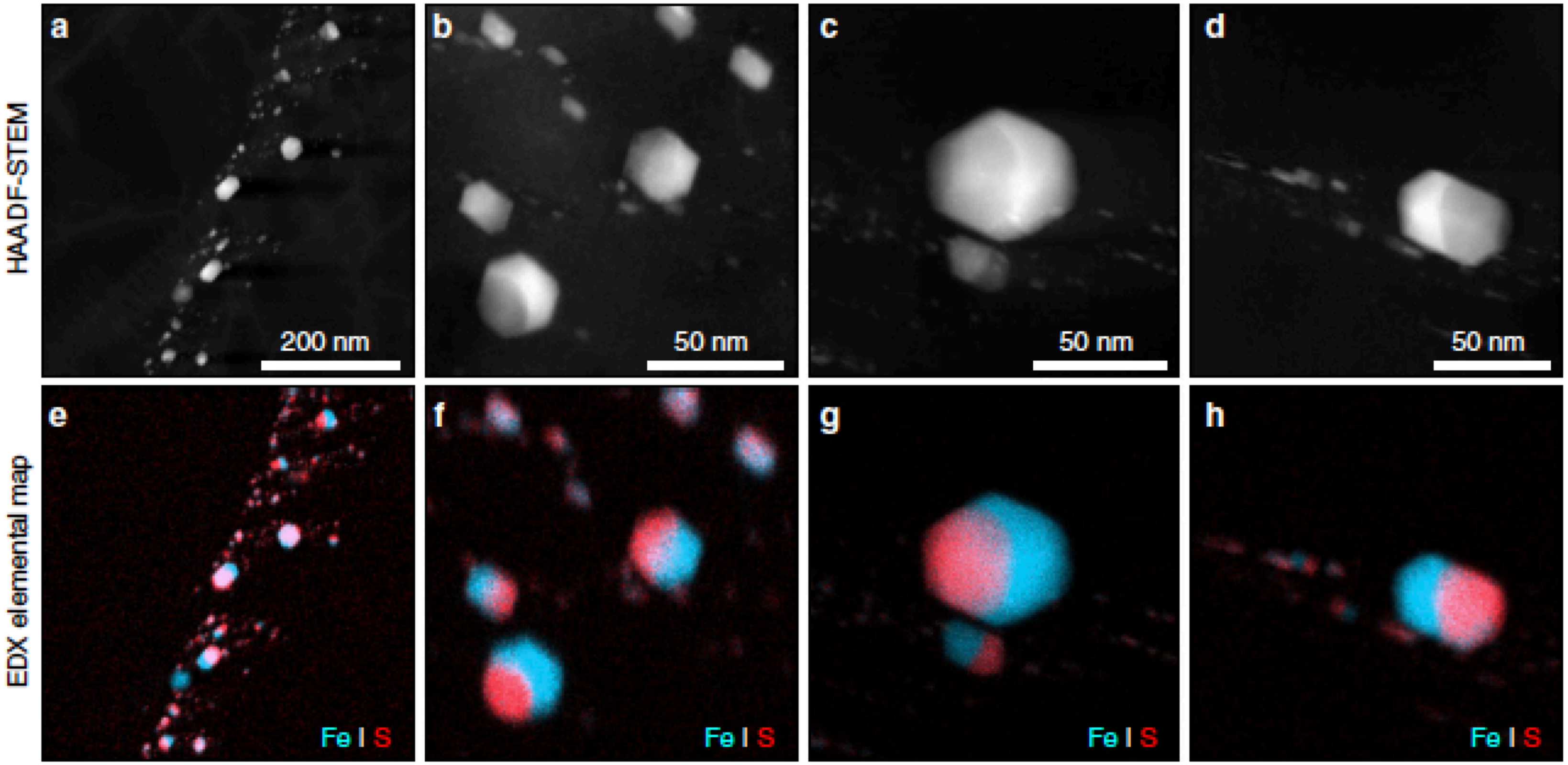
linear increase
⇒ ~50y

$5 \cdot 10^{-11}\ m^3$ matter

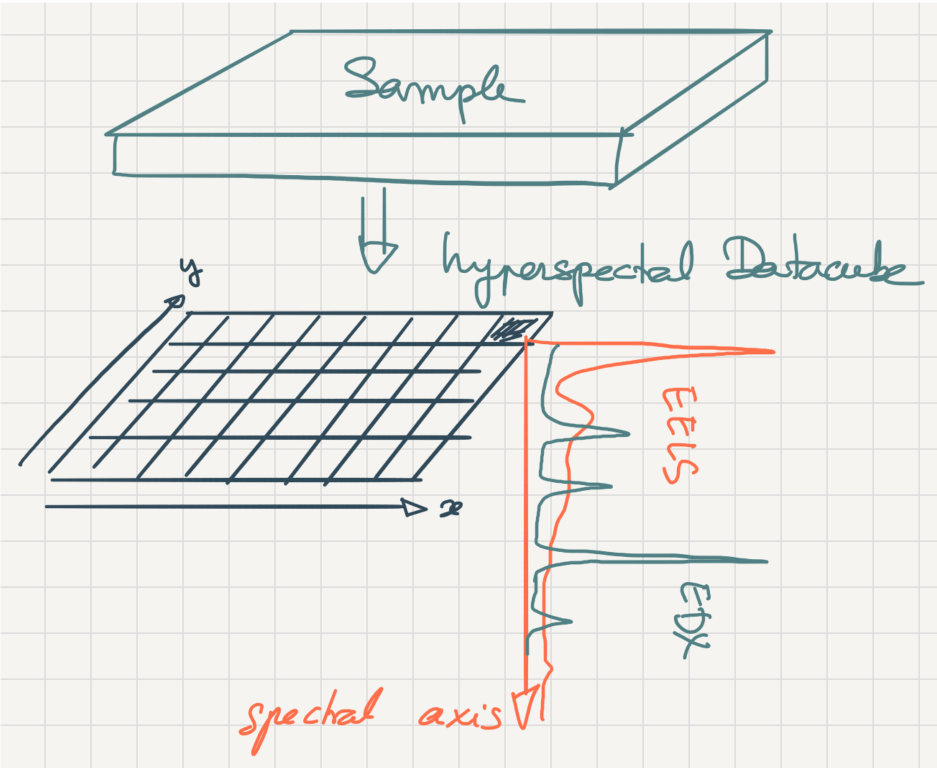
0,05 mm³

Big data ??

A large planetary body inferred from diamond inclusions in a ureilite meteorite
Nabiei & Al. Nature communications, 9 (2018)



Back to Analytical TEM:EELS & EDX



At EPFL:
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

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”

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. They do not really view value in an EDX datacube
- Microscopists (1%) needs to know *both* (manufacturer softs used at the TEM)
- Not much real hard money

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.)

Result: “Open data” now

- 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

Shall we be satisfied with this situation or shall we seek to change it ?

Why change ?

- With common, documented file format we can create databases collecting the measurements on materials.

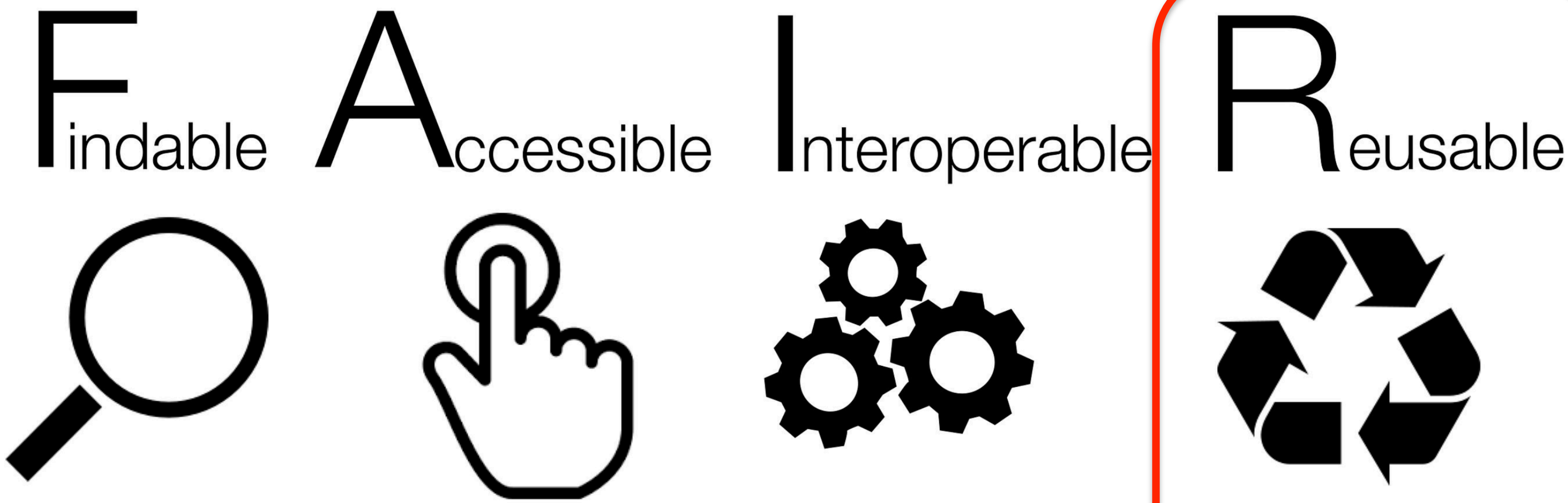
- This is requested   by funding agencies

- Sharing datasets will allow groups developing methods to make methods more robust tested against many datasets (content and variability) reproducibility of experiments.

- Possibility to use the power of a large community (kaggle)

Challenges/Discussion:

How to make the data linked to those projects FAIR ?



Database
Metadata

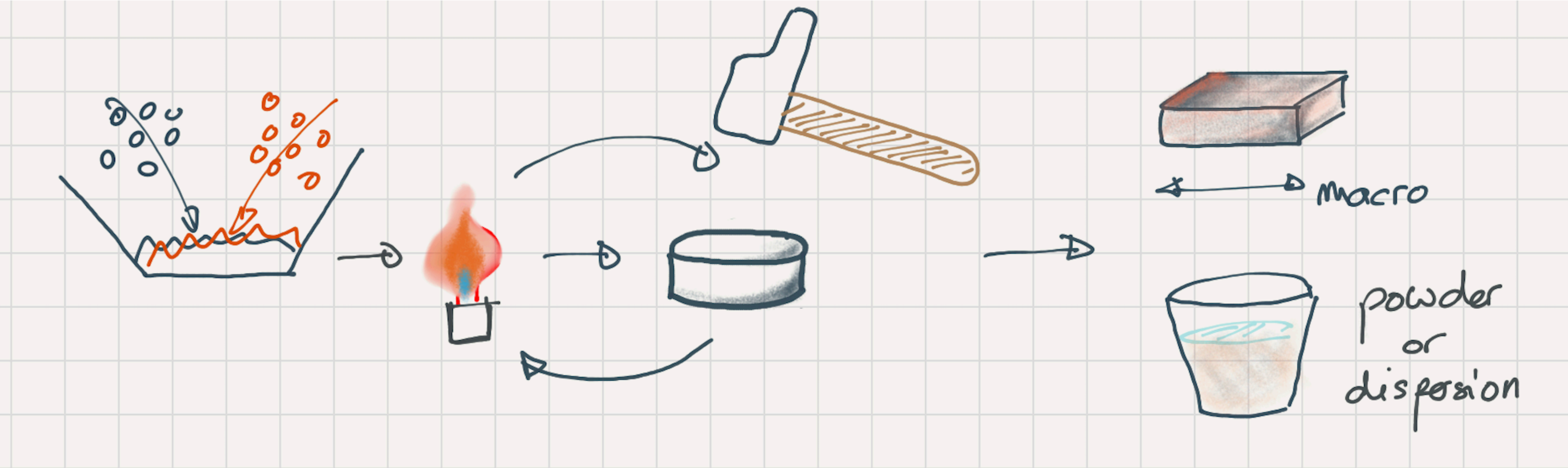
Open formats

??????

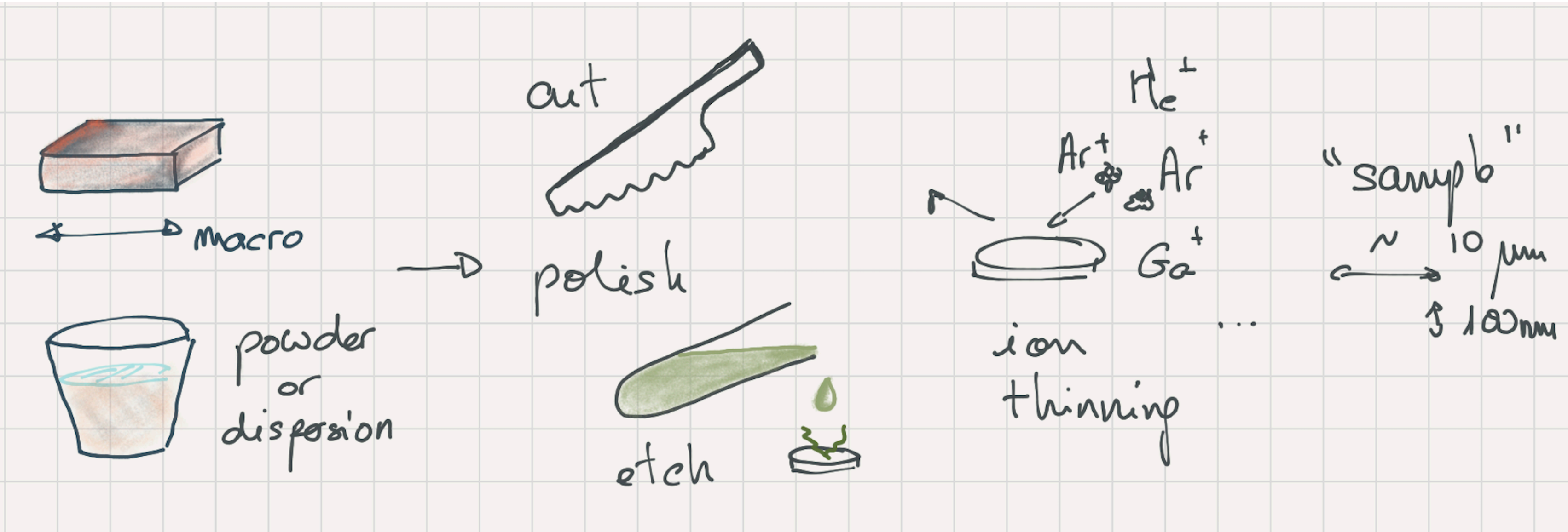
Proper documentation
metadata

Metadatas, what do we need?

How was the sample fabricated?

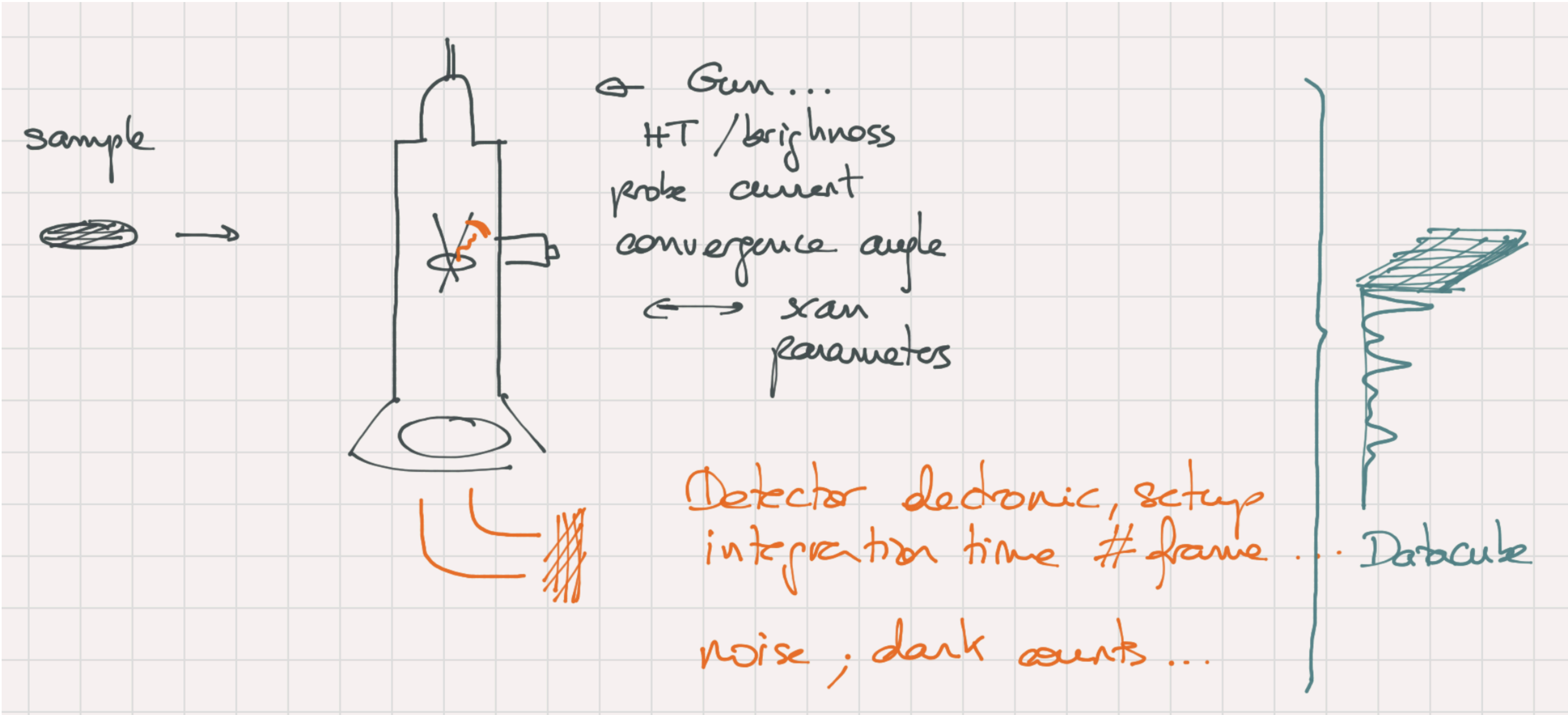


How was the TEM lamella obtained?



Metadatas, what do we need?

How was the microscope set up ? The detector configured ?



So, let’s start!

- **Open Post doc position at LSME EPFL**
-
- Advanced machine learning tools and FAIR data in TEM.
-
- The subject of this post doc will be twofold :
- Explore new approaches for hyperspectral STEM-EDX data decomposition, using contextual prior knowledge (number of phases in presence, existence of regions of pure phase, specificities of the composition of phases, etc) into a guided/supervised learning process, rather than a full unsupervised learning.
- Explore the necessary actions to put in place in order to make the data comply with the requirement of FAIR data. Especially working on the required metadata that need to be attached to the data for a correct documentation. This last part will be carried out in close collaboration with international groups (in France, Germany, Ireland, US… interested in the same subject)