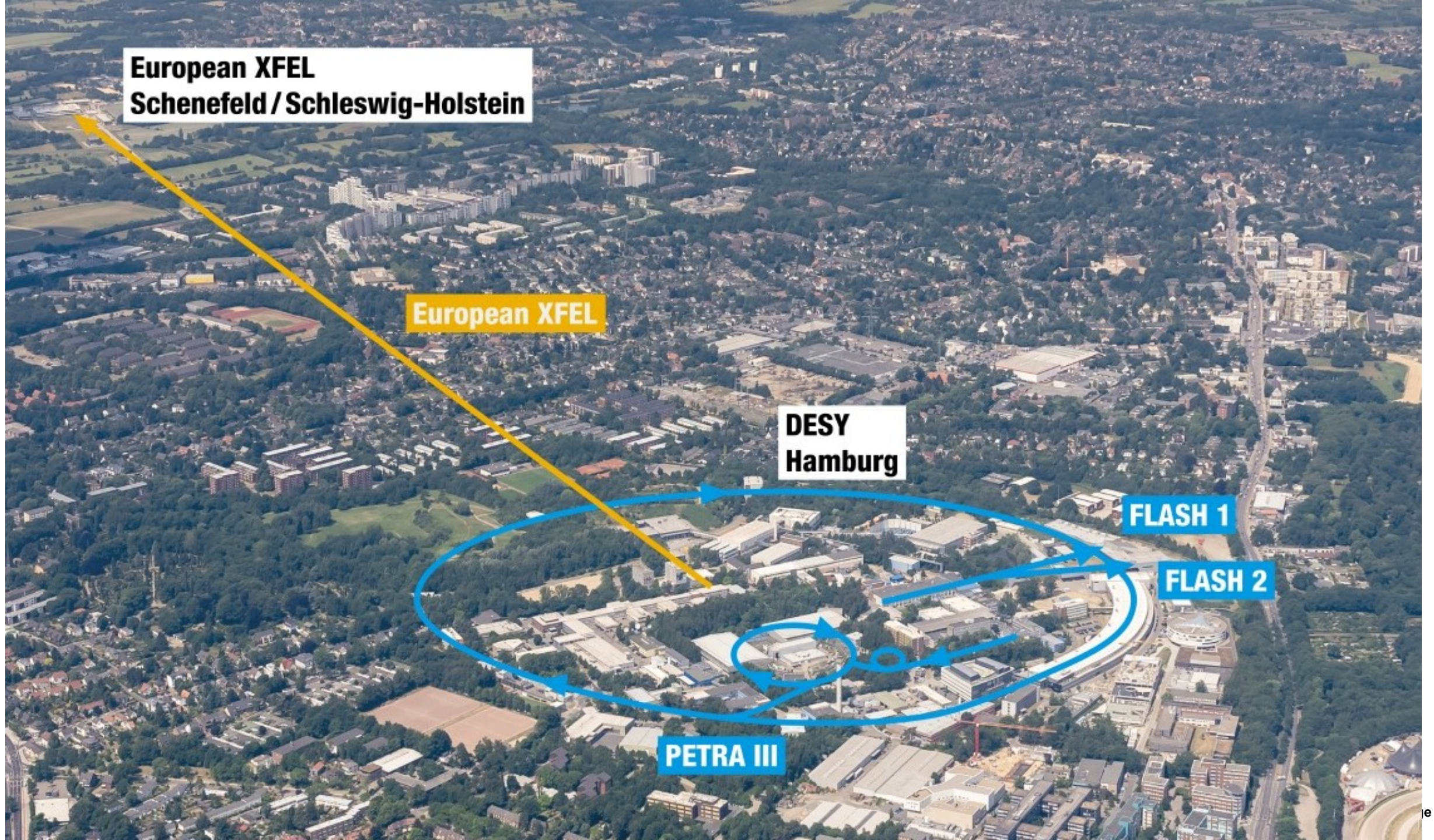
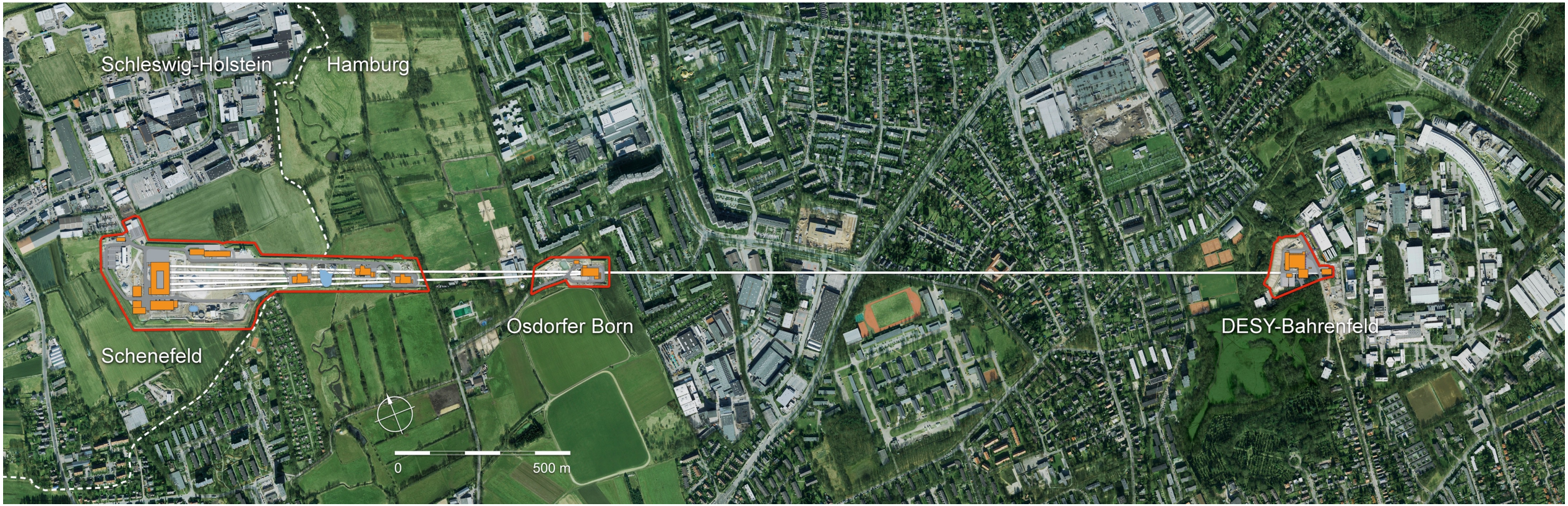
# ESS deployments @ EuXFEL

## from physics to data to physics

Martin Gasthuber for the colleagues at European XFEL and DESY – several slides are stolen from them
GPFS User Group, March 2023

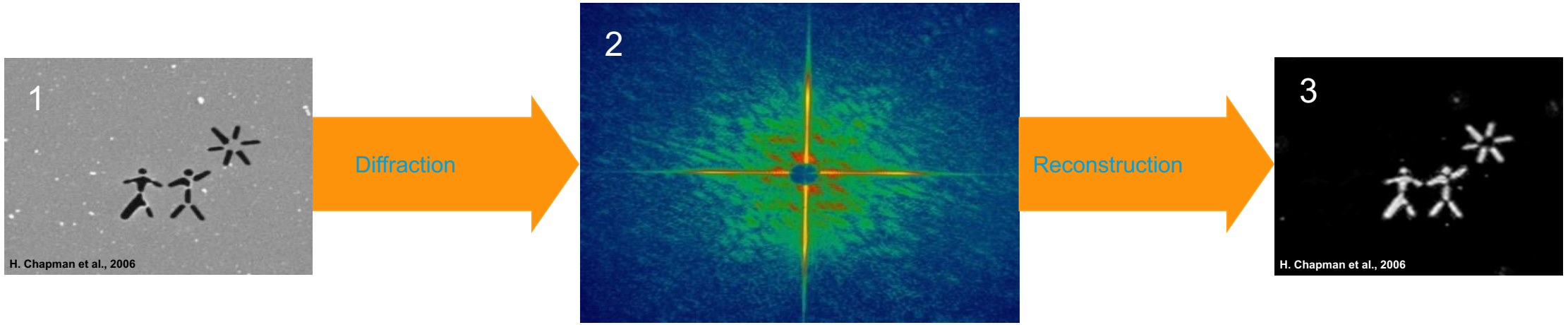
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES


European XFEL
Schenefeld / Schleswig-Holstein
European XFEL
DESY
Hamburg
FLASH 1
FLASH 2
PETRA III

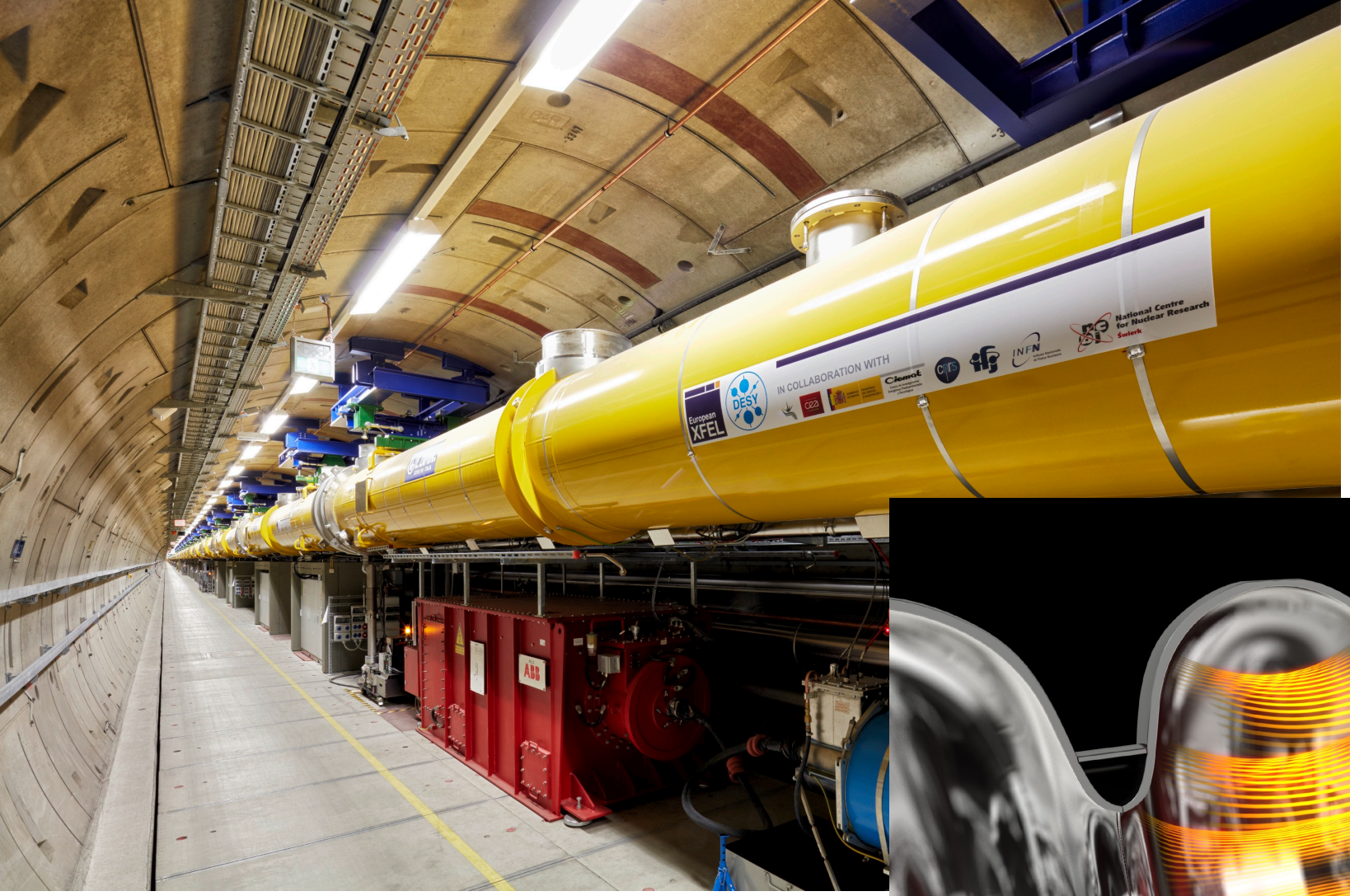

Schleswig-Holstein
Hamburg
Schenefeld
Osdorfer Born
DESY-Bahrenfeld
0
500 m

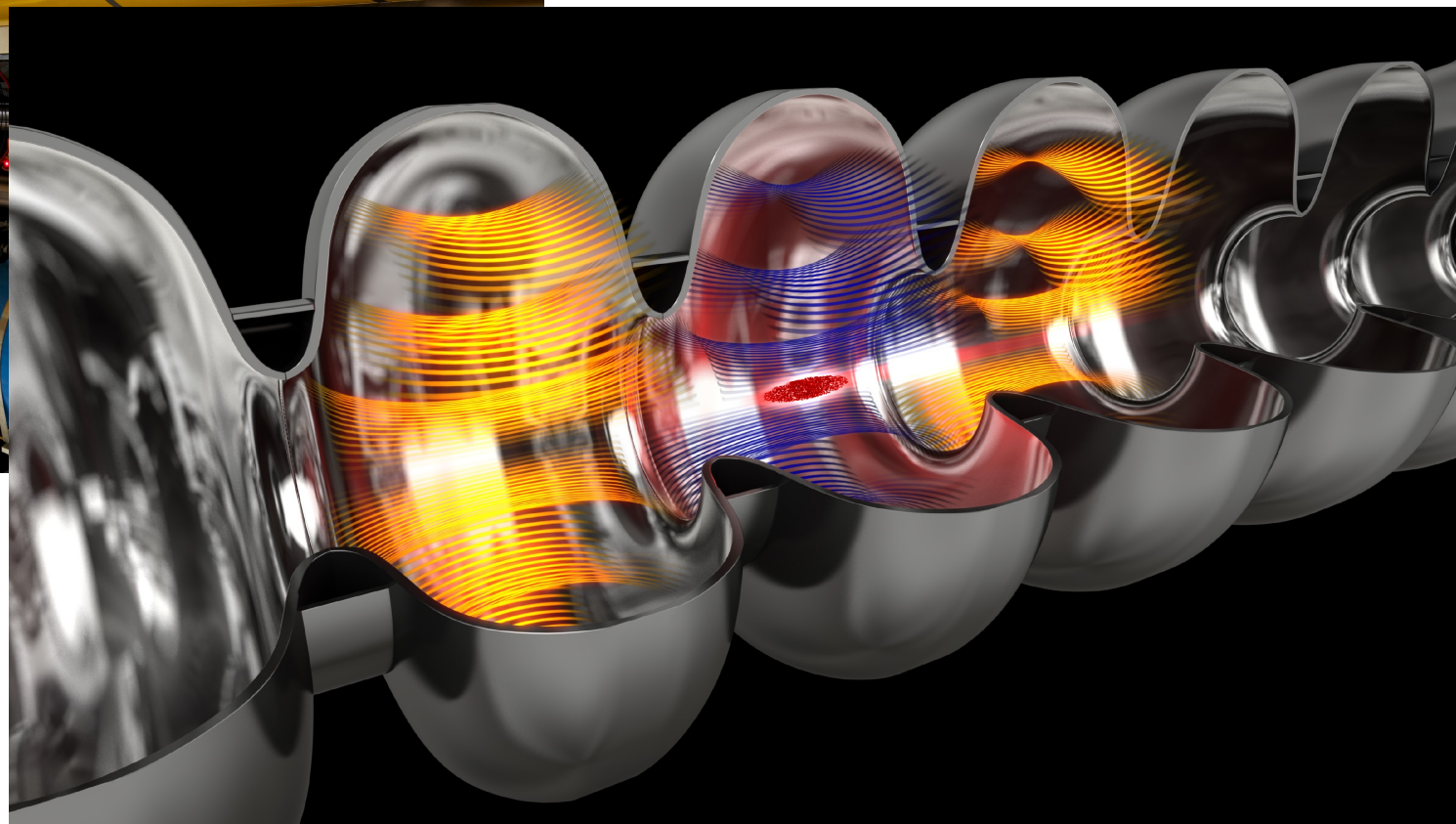
# Making pictures without a camera lens

- Crystallography method developed by Laue and Bragg, 1912–1914
- Similar method used in X-ray FELs

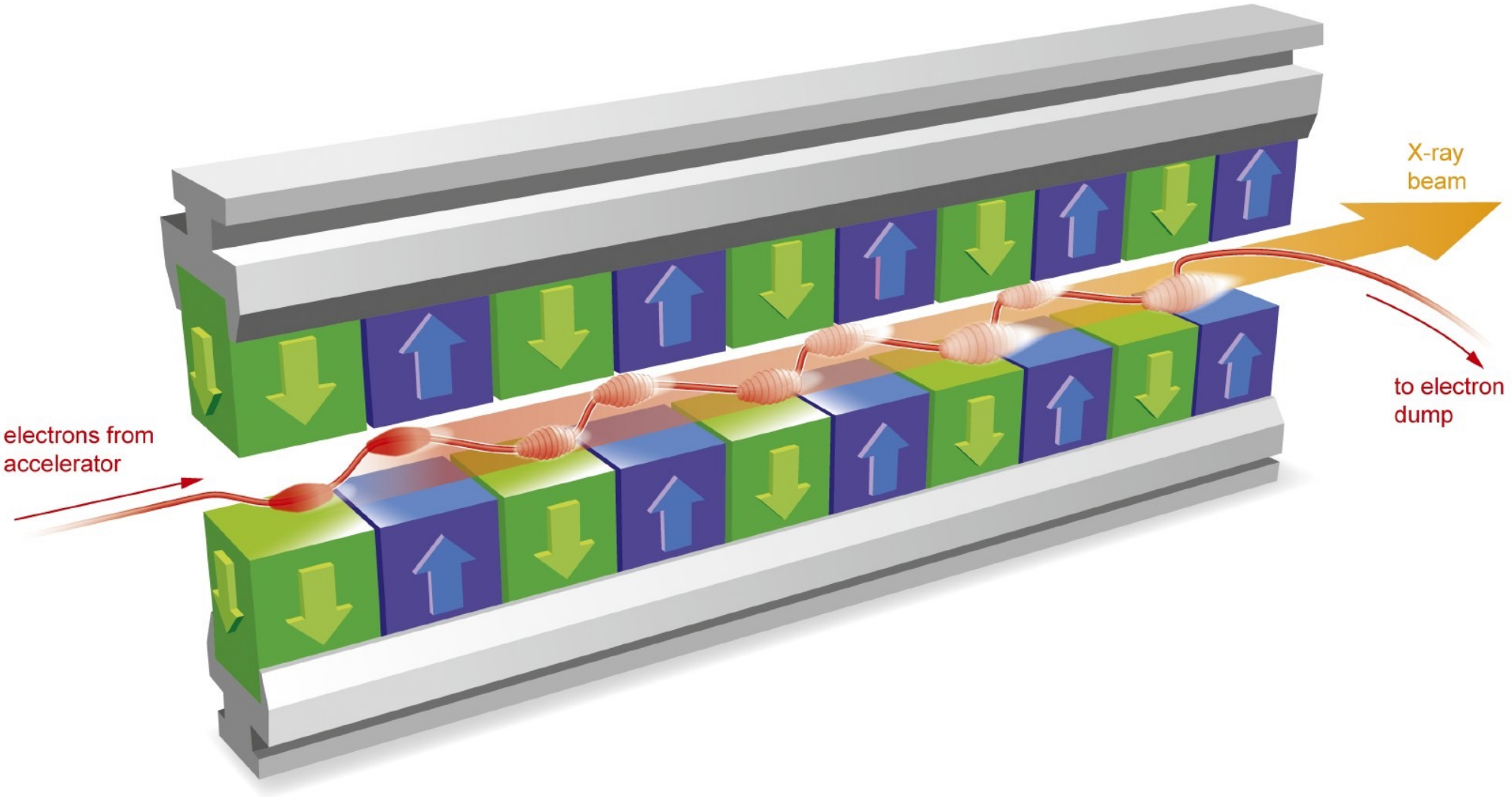



- X-rays scatter (diffract) off objects (1, microscopic shapes cut from metal)
- Detectors record the scattered X-rays (2, diffraction pattern)
- Original shapes reconstructed in high detail from detector data (3, reconstructed image)

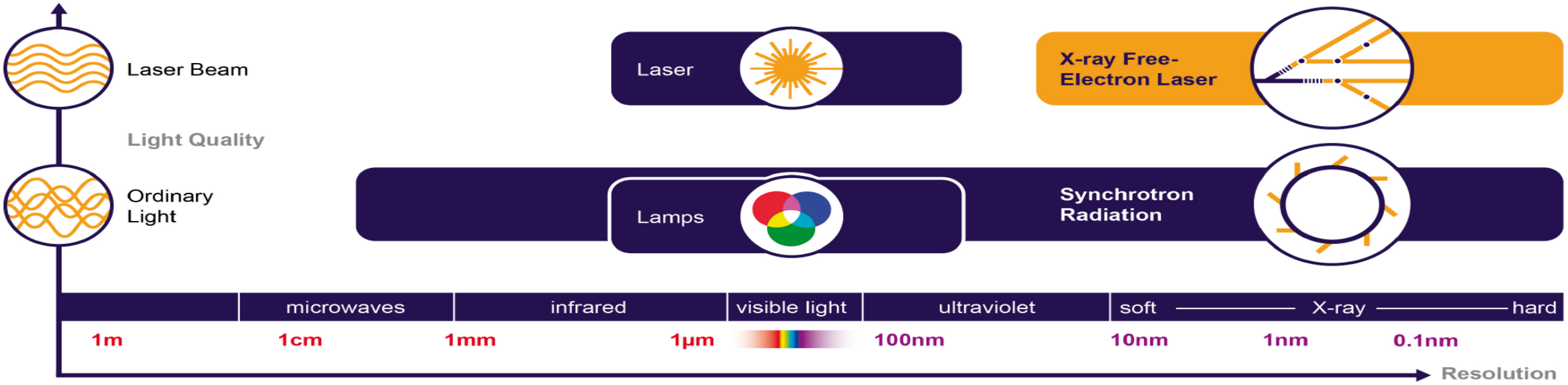

European XFEL
DESY
IN COLLABORATION WITH
Cnrs
INFN
National Centre for Nuclear Research
Świerk
ABB



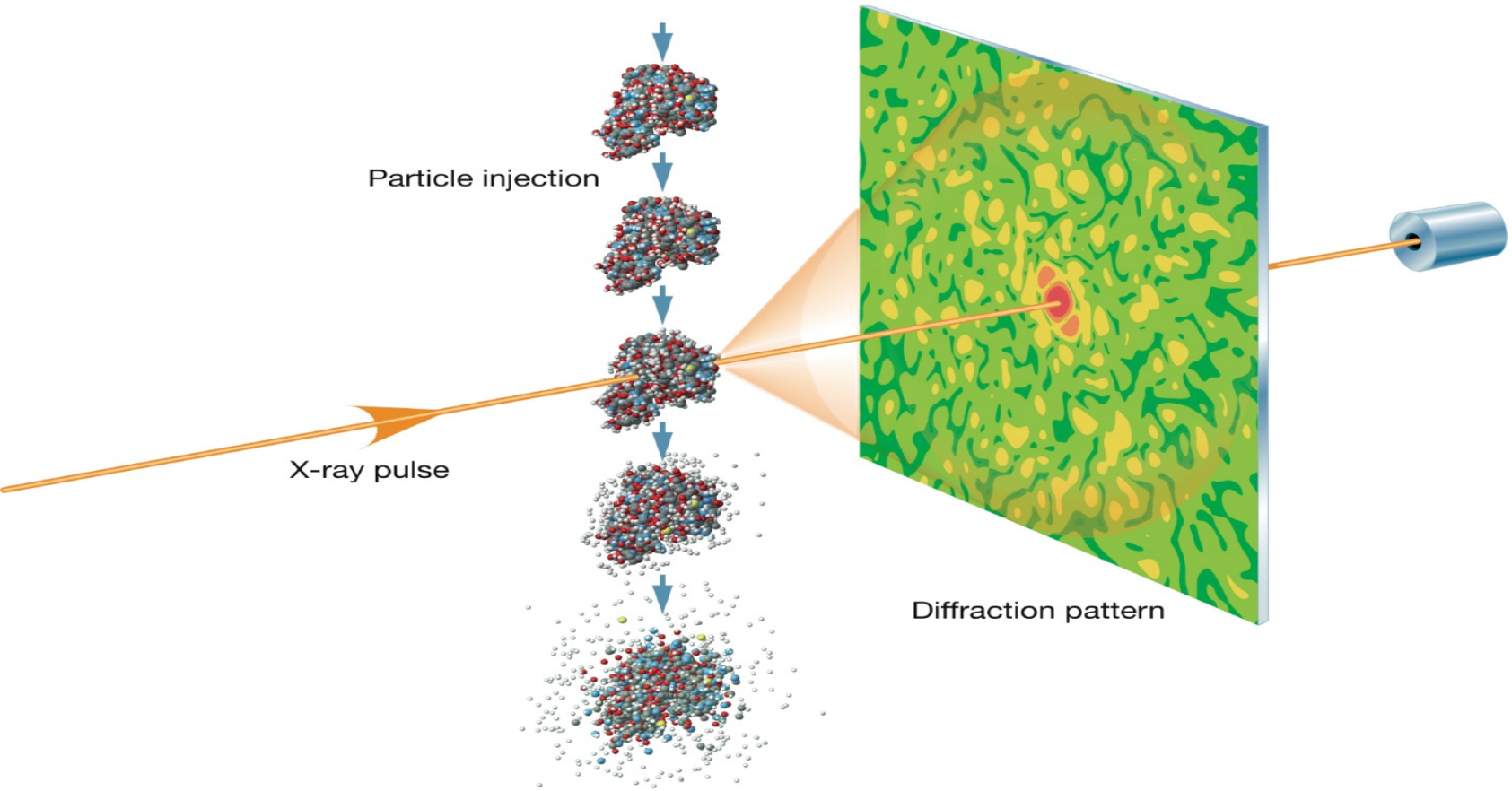
# SASE - Self-Amplified Spontaneous Emission



# Light sources and the light they generate

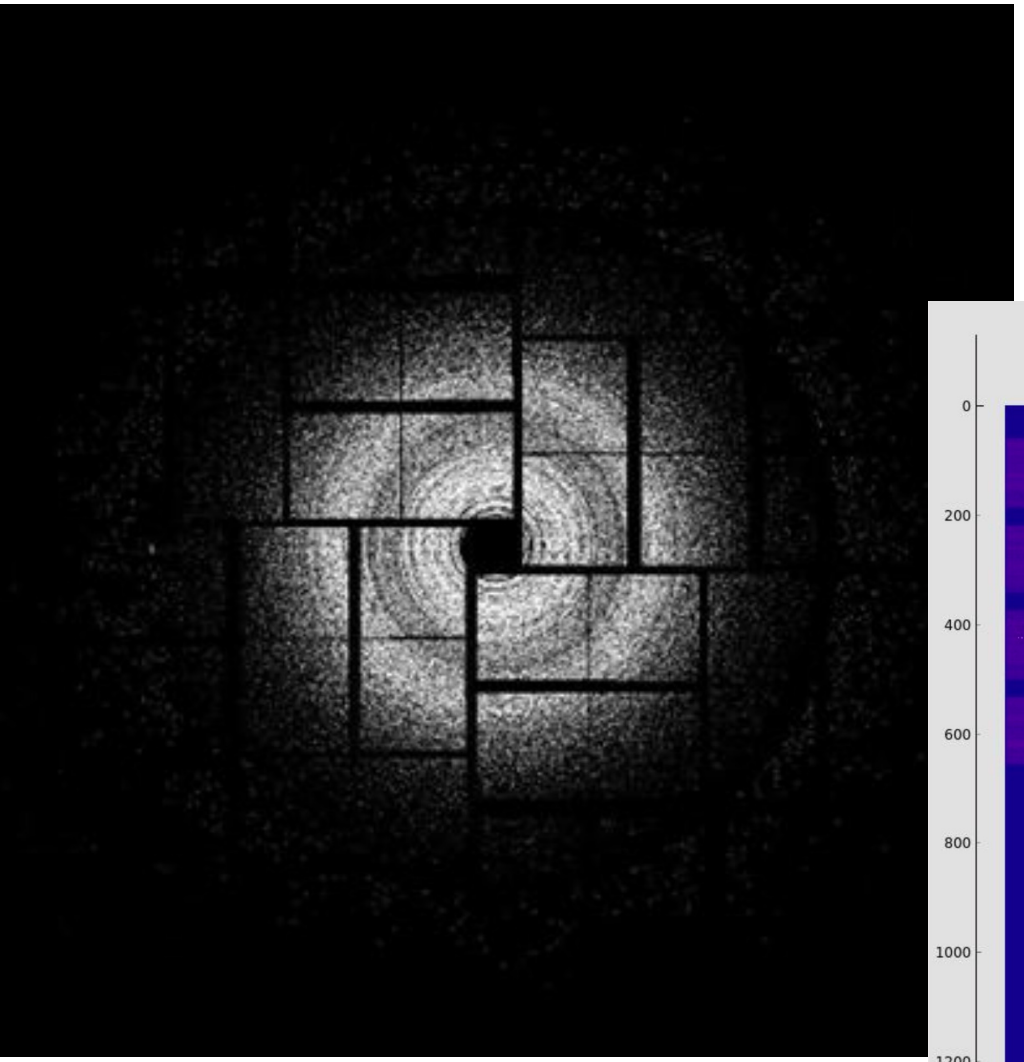


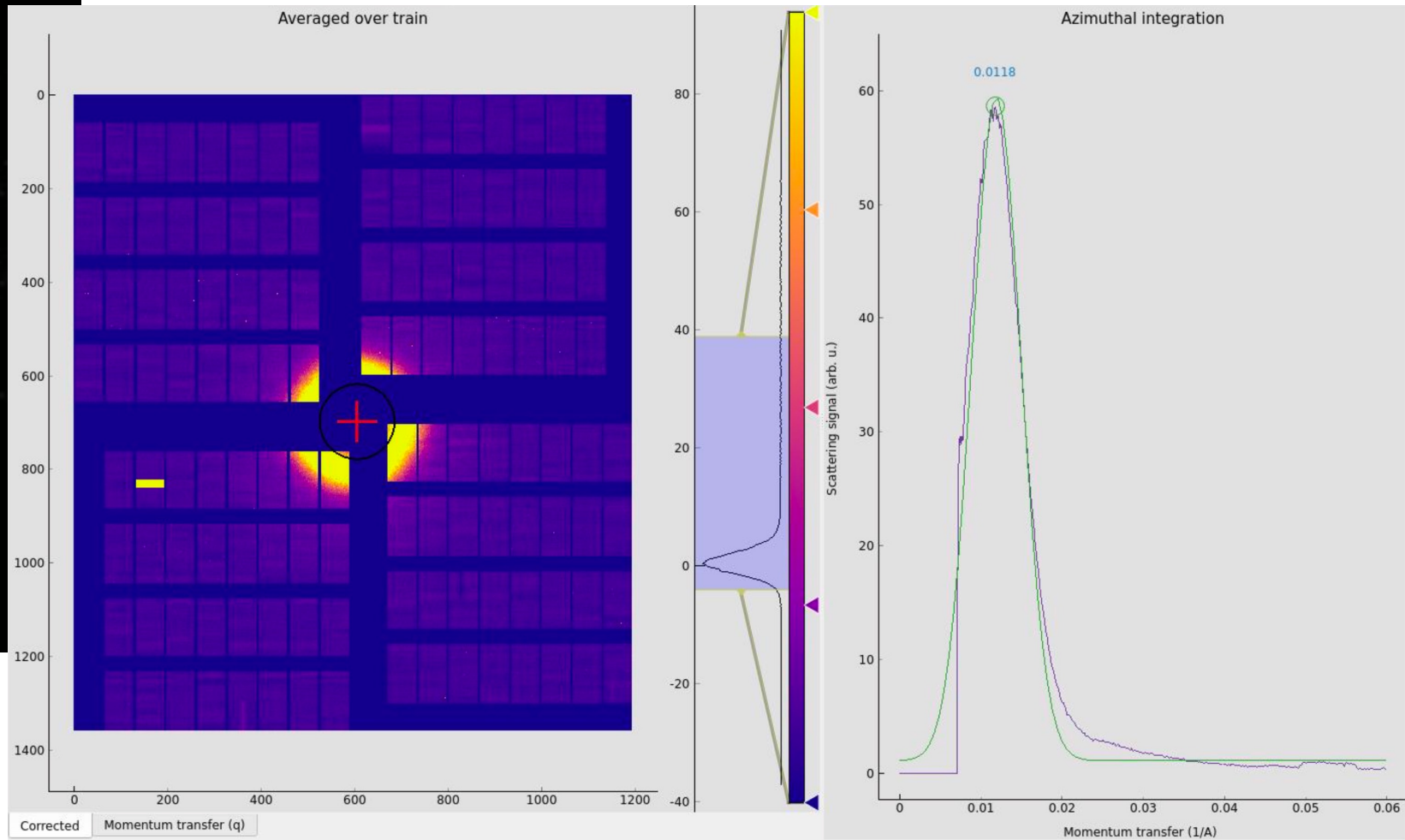
# Making pictures without a camera lens



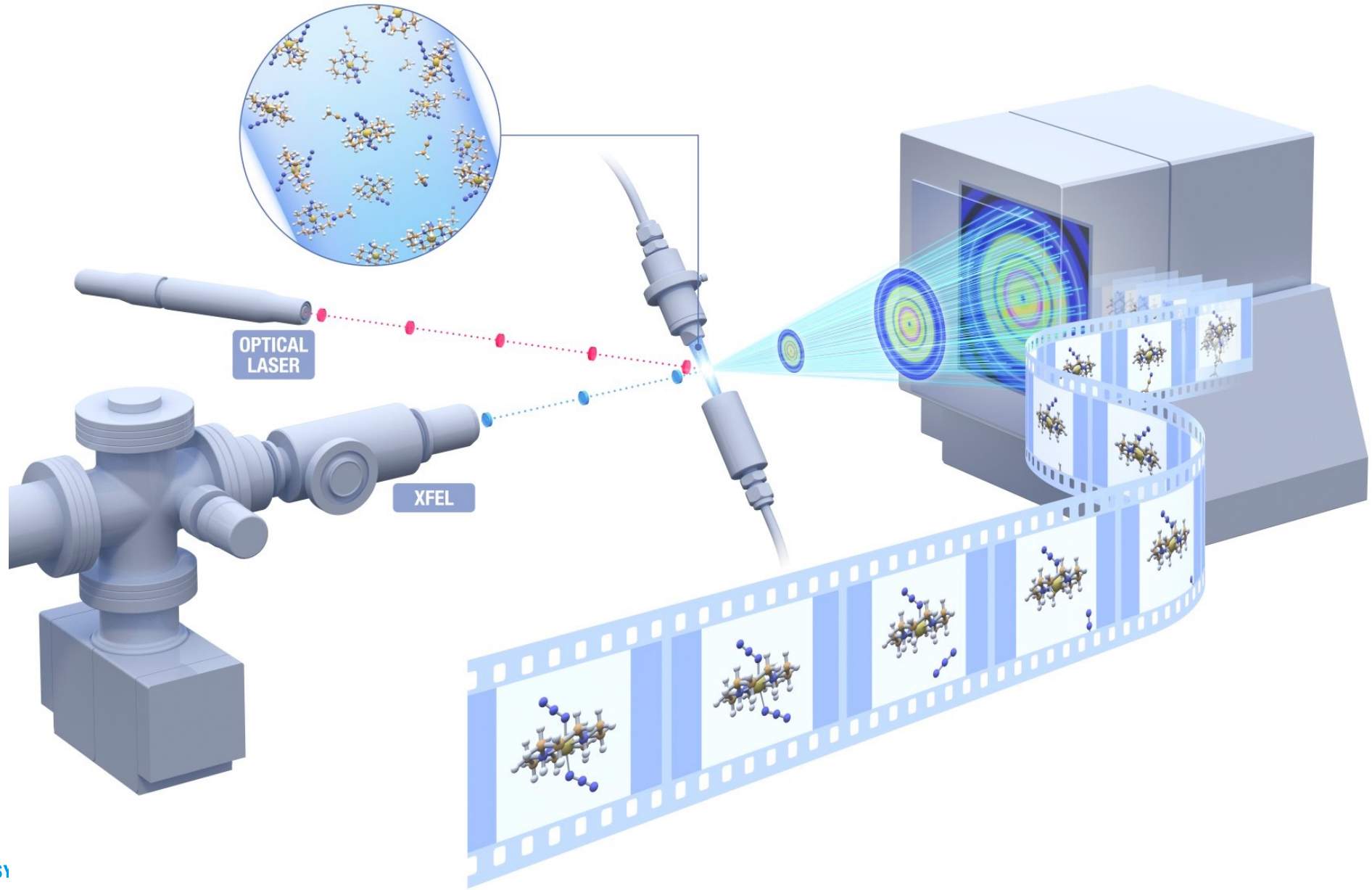

- Samples injected as liquids into vacuum chamber
- X-ray pulse hits sample and diffracts onto detector
- Sample is destroyed, but diffraction pattern is recorded beforehand

# how it looks like





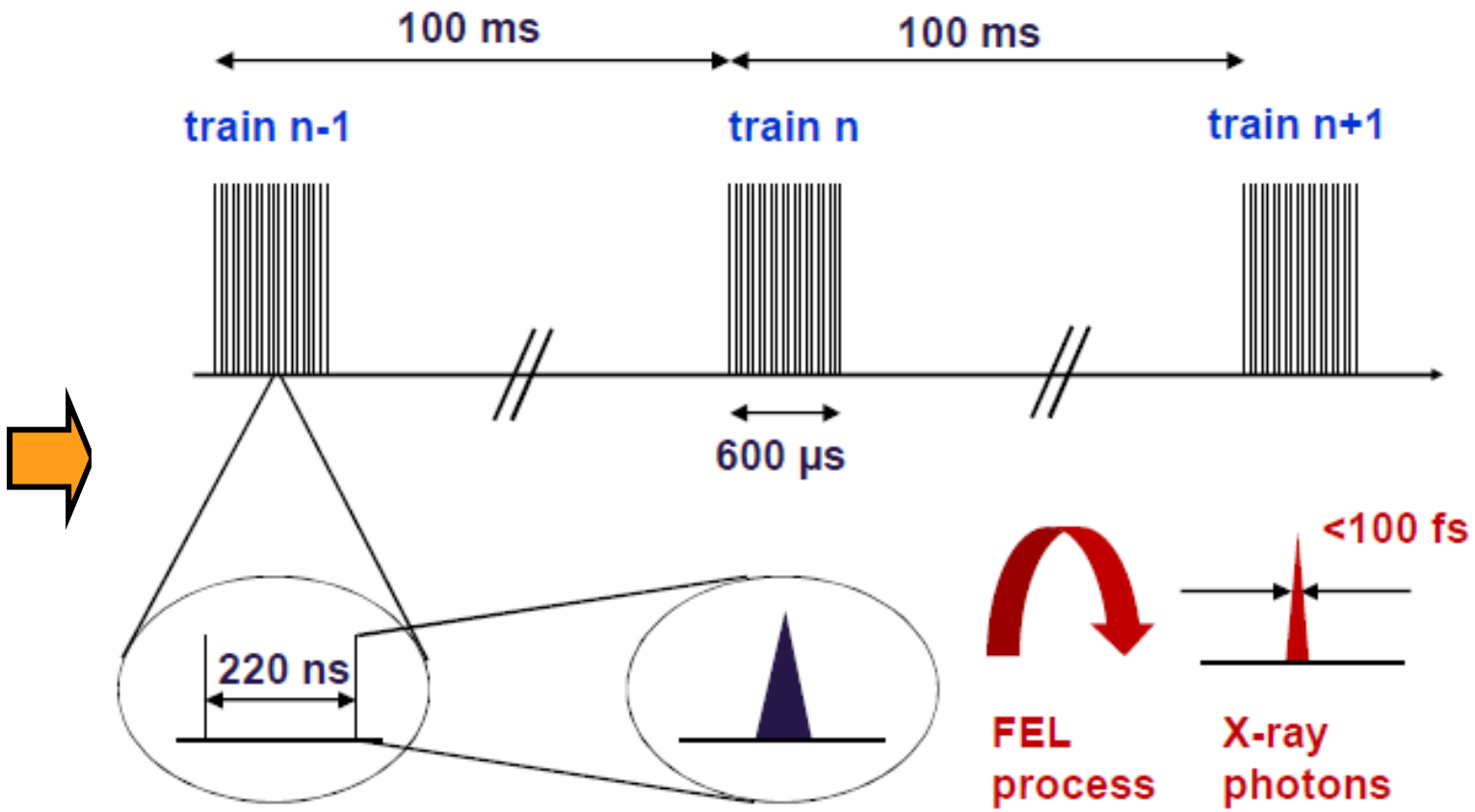
# at the end…



# short movie from the FXE instrument @EuXFEL

https://zitwowza4.desy.de/CumulusDB/_definst_/mp4:ConFilm/FXE-animation_EN.mp4/playlist.m3u8

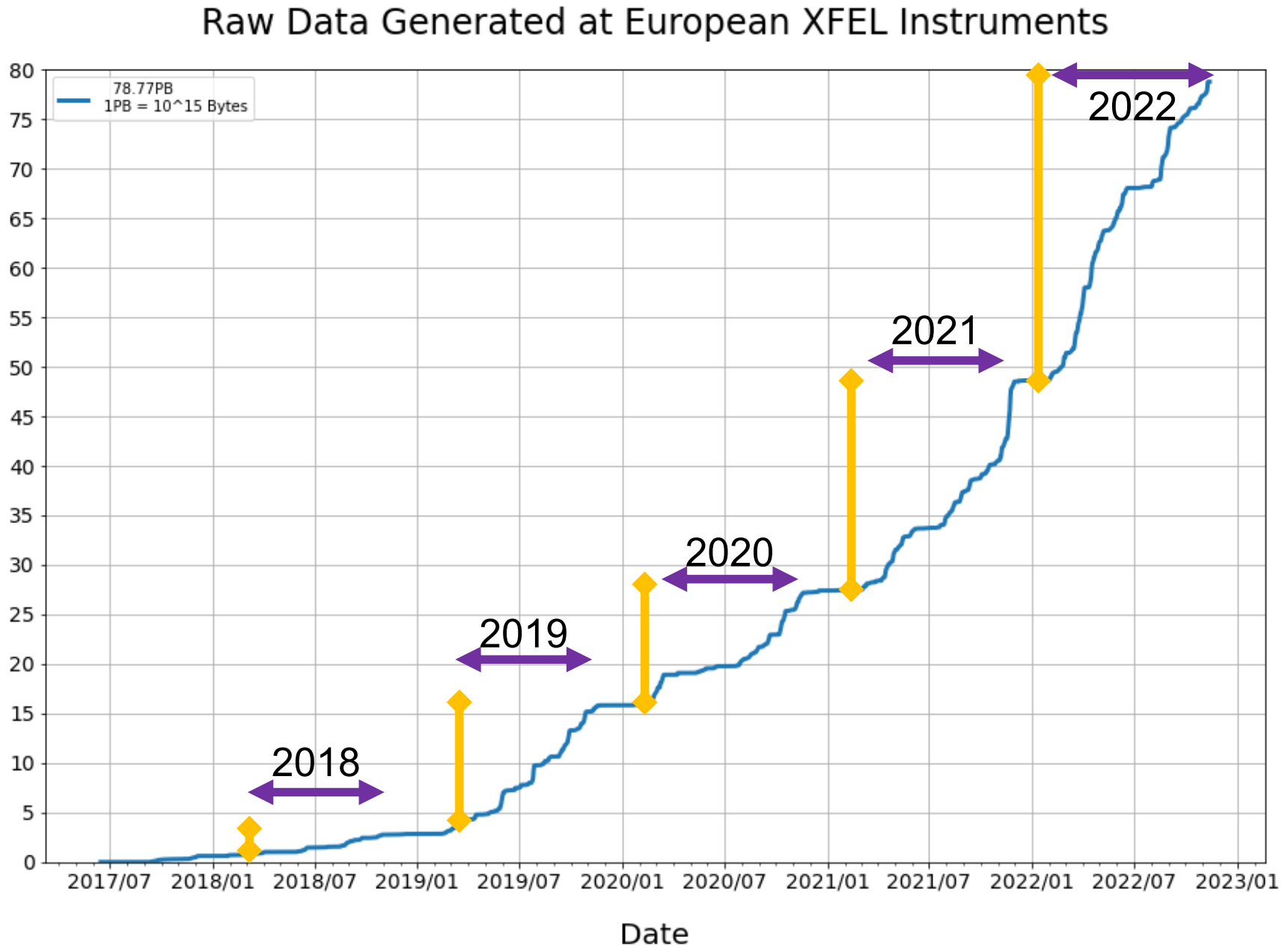
- Readout rate driven by bunch structure
  - 10 Hz train of pulses
  - 4.5 MHz pulses in train
- Data volume driven by detector type



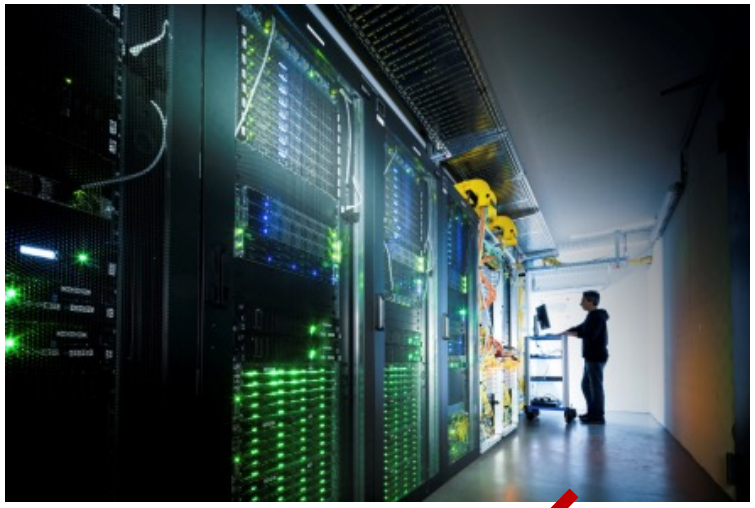

| Detector type | Sampling | Data/pulse | Data/train | Data/sec |
|---|---|---|---|---|
| 1 channel digitizer | 5 GS/s | ~2 kB | ~6 MB | ~60 MB |
| 1 Mpxl 2D camera | 4.5 MHz | ~2 MB | ~1 GB | ~10 GB |
| 4 Mpxl 2D camera | 4.5 MHz | ~8 MB | ~3 GB | ~30 GB* |

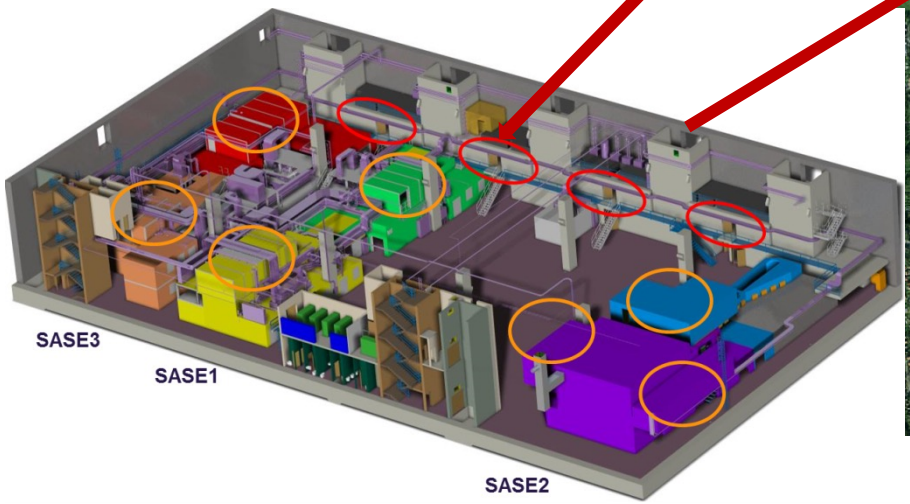
the obvious ***unsustainable*** mode

- becoming regular ‘1 PB per day’
  - ~50 TB/h
- further processing adds >60%
  - i.e. calibration



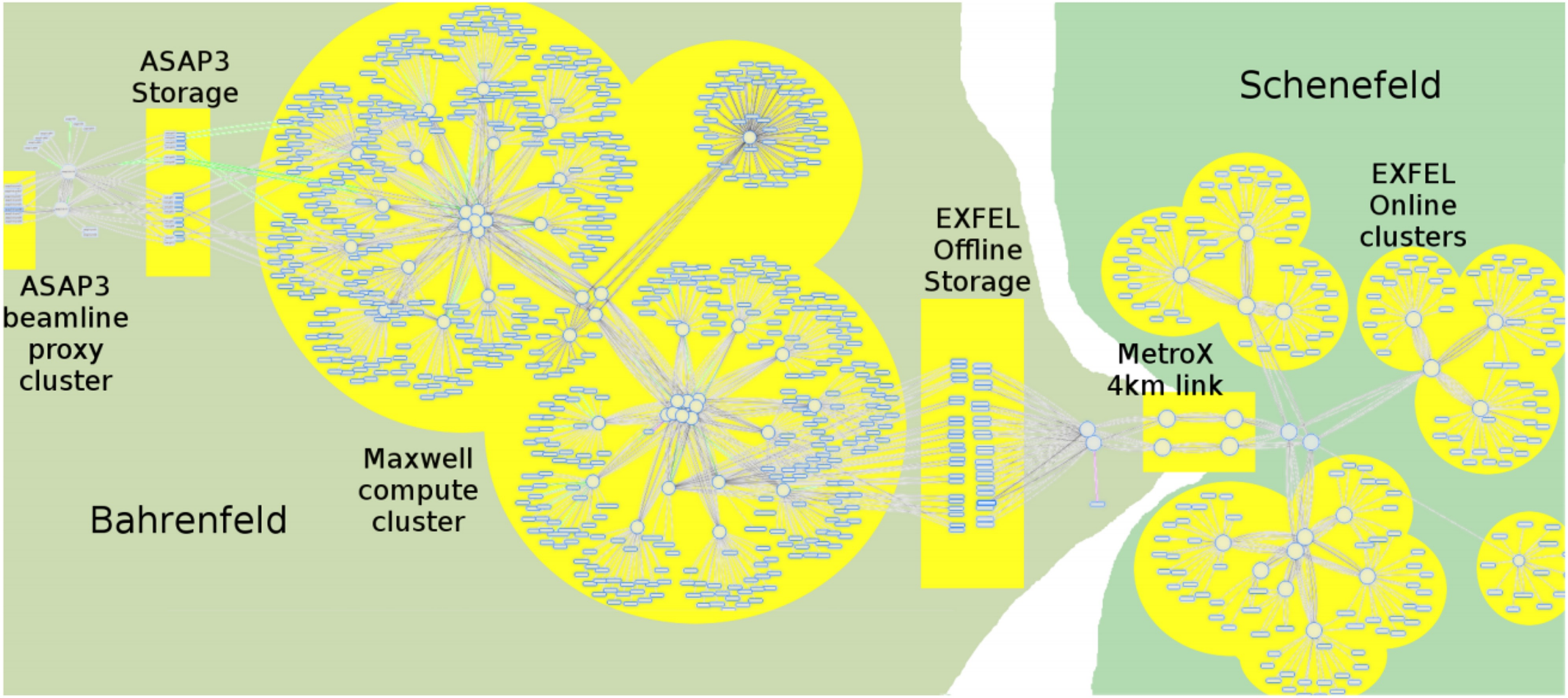
# computing stuff locations

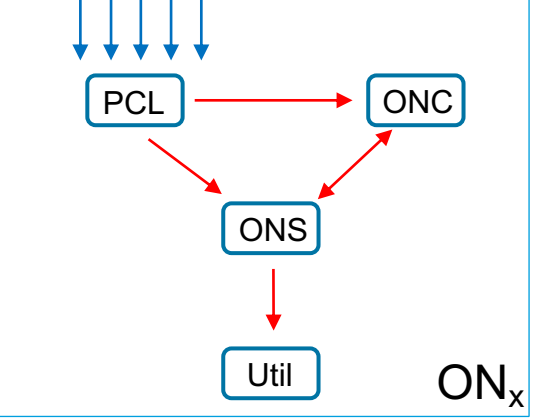


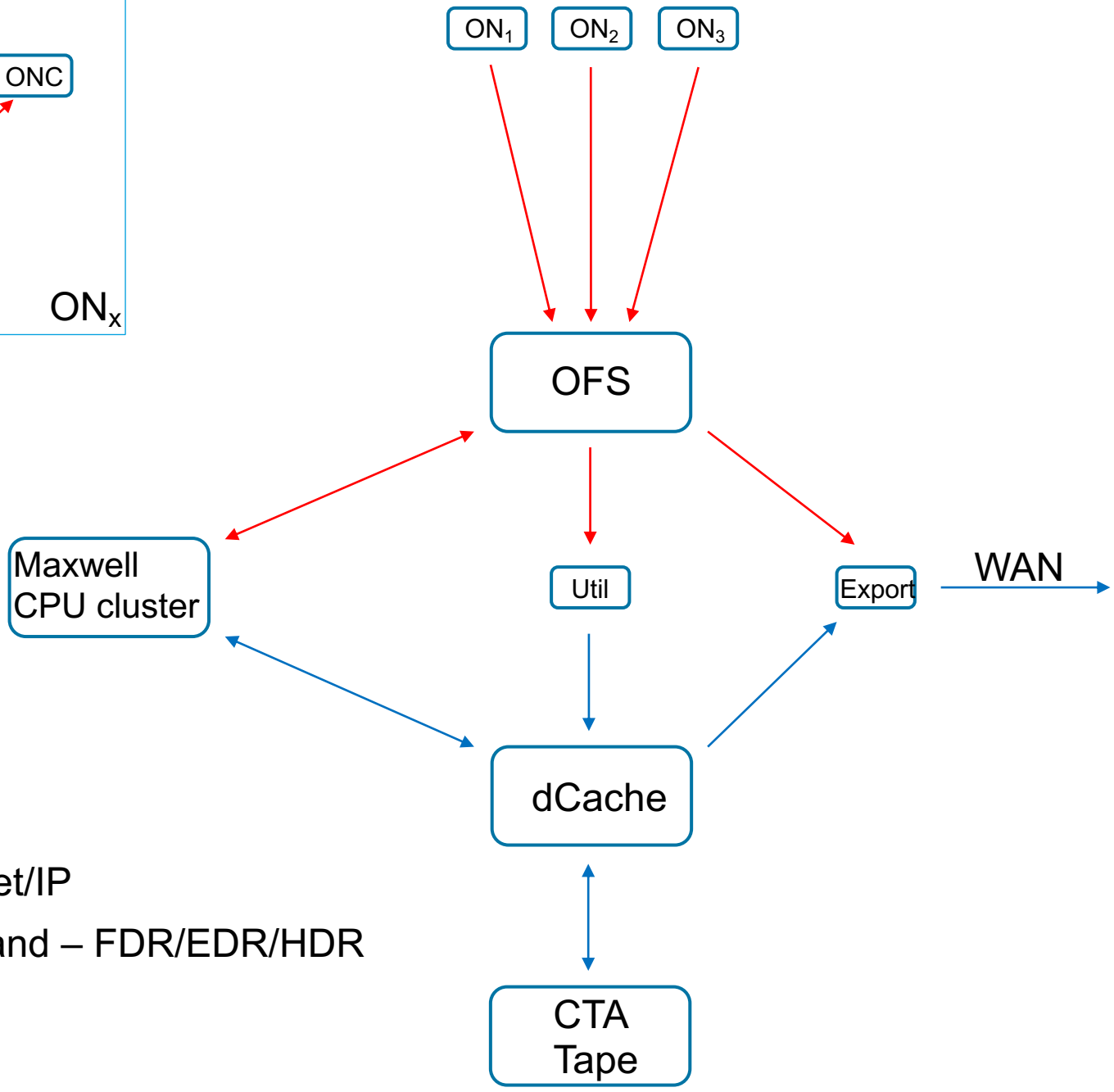



# strategic basics we start with

- InfiniBand as the data transport network for all RDMA traffic
  - only slow and non-demanding access via NFS & SMB
- GNR only
  - for disks and SSD based systems
- simple fault tolerant ‘building blocks’ with in-built redundancy

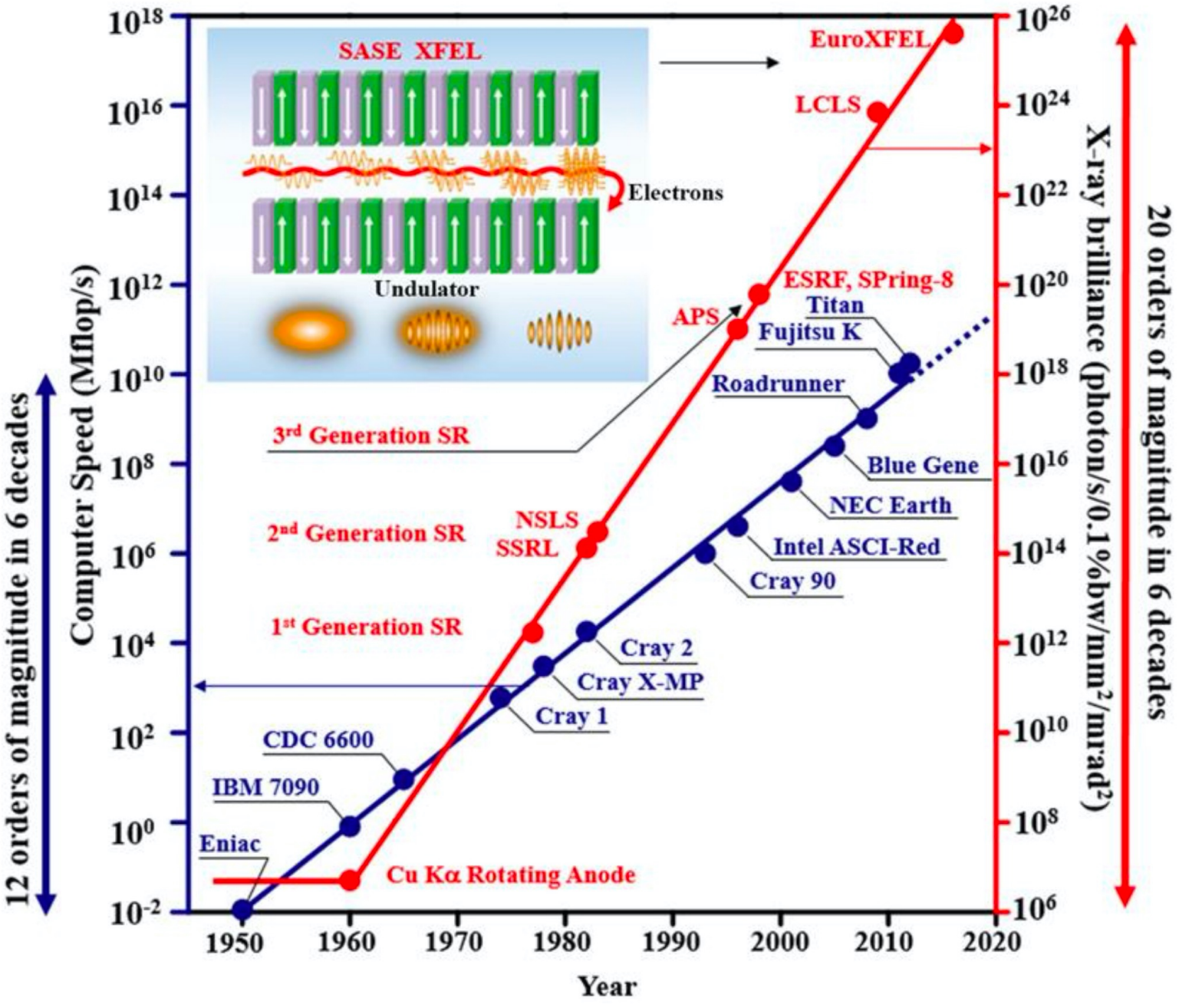

ASAP3
Storage
ASAP3
beamline
proxy
cluster
Maxwell
compute
cluster
Bahrenfeld
EXFEL
Offline
Storage
Schenefeld
EXFEL
Online
clusters
MetroX
4km link




ONS: OnliNe Storage / GPFS / SSD+Disk
~600TiB / soon 2.6PB
OFS: OFfline Storage / GPFS / Disk
>52 PiB, ~4000 disks
regularly: >180GB/s read, >90GB/s write
dCache: Disk
>100 PiB
CTA/Tape:
>300 PiB

Maxwell:
nearly 1000 nodes
FDR and HDR100 connected




| | Data rate[GB/s] | fps | fSize[MB] |
|---|---|---|---|
| Lambda9M | 28.6 | 2000 | 1.3 |
| Lambda(12b)/HDF5 | 2.6 | 2000 | 1.3 |
| Lambda2M/HDF5 | 7.8 | 2000 | 1.3 |
| Jungfrau, HDF5, BIN | 0.5 | 500 | 1 |
| PCO Edge | 0.8 | 100 | 8 |
| Eiger2X16M | 4.7 | 130 | 18 |
| Eiger2X9M | 4.6 | 230 | 10 |
| Eiger2X4M | 4.5 | 500 | 4.5 |
| Eiger4M,HDF5,LZ4 | 1 | 750 | 9/18 |
| Pilatus2 6M, CBF/TIF | 0.15/0.6 | 25 | 6 |
| Pilatus3 2M, CBF/TIF | 0.6/2 | 250 | 2.5/11 |
| Pilatus2 1M, CBF/TIF | 0.05/0.2 | 50 | 1 |
| Pilatus2 300k, CBF/TIF | 0.05/0.2 | 200 | 1 |
| Perkin Elmer XS 1621 | 0.25 | 15 | 16 |

# summary

- initial key ingredients were correct and good – and are still good
  - alternative ingredients might also do the job ;-) – there are plenty ways to …

- near future challenges
  - more experimental lines
  - again – newer (faster) detectors
  - higher detector rates

- Disks do not scale – technically
- Flash/SSD do not scale – economically