Online Data Analysis

– Users' Story from Göttingen –

Markus Osterhoff
Institut für Röntgenphysik,
Uni Göttingen

AG Salditt + AG Köster
GINIX @ P10

Who we are

- AG Tim Salditt
- AG Sarah Köster
- Markus Osterhoff
- PhD students
- Master and Bachelor students

Acknowledgements

- DESY Photon Science
  - FS-EC, FS-PE
  - especially Michael and P10 team

- Funding:
  - BMBF
  - SFB 755

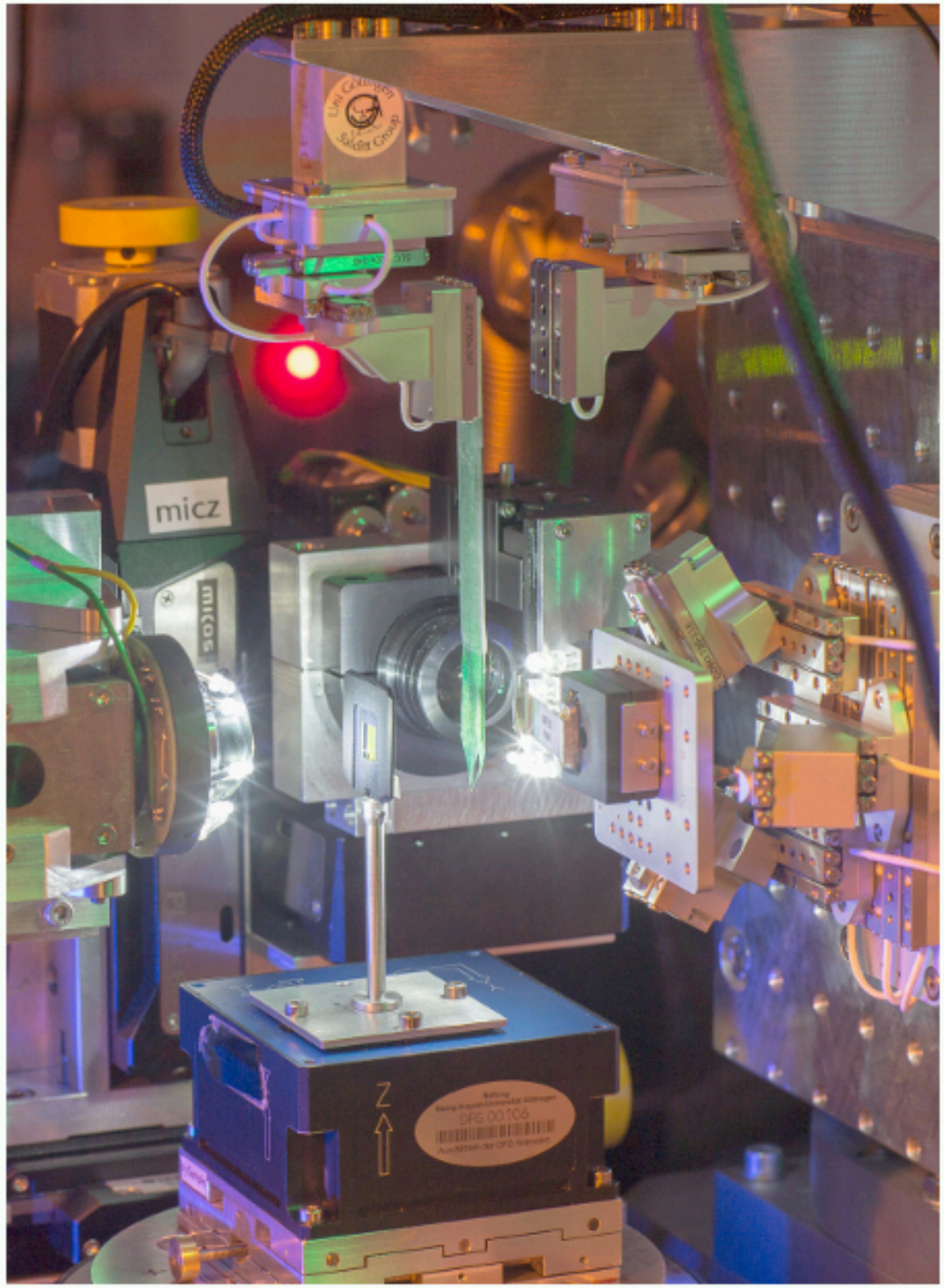
- Sarah and Tim for great atmosphere
  - all our students

- workshops @ Göttingen
  mechanical and electrical

Where we are

everywhere :)

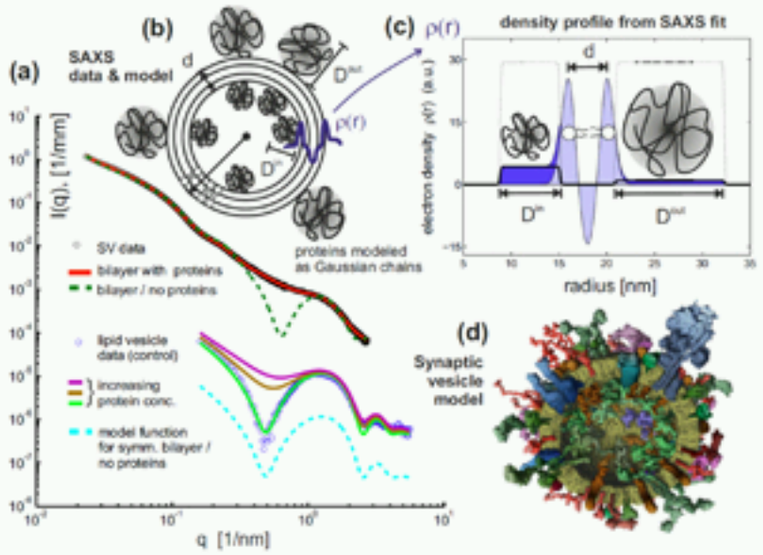
- Lab sources
- PETRA III
  mainly GINIX @ P10
- ESRF, SLS, Elettra

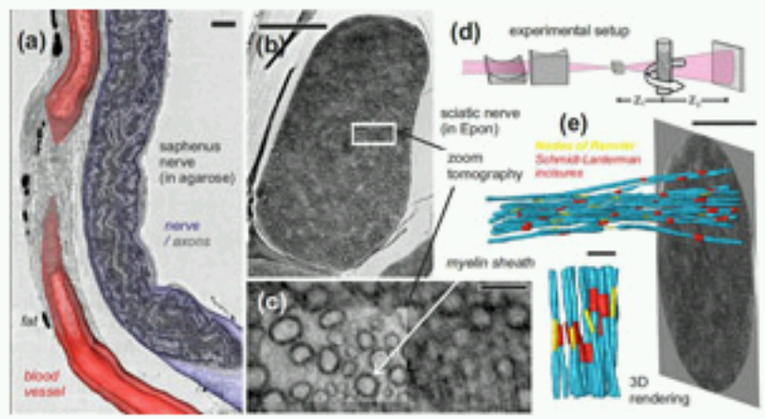


Scientific Questions and Methods

Questions

- soft matter, biomolecular assemblies from molecular to cells to tissue
- quantitative physical description
- nanoscale structures + dynamics -> function





nanoscale imaging of cellular dynamics

Scientific Questions and Methods

Imaging

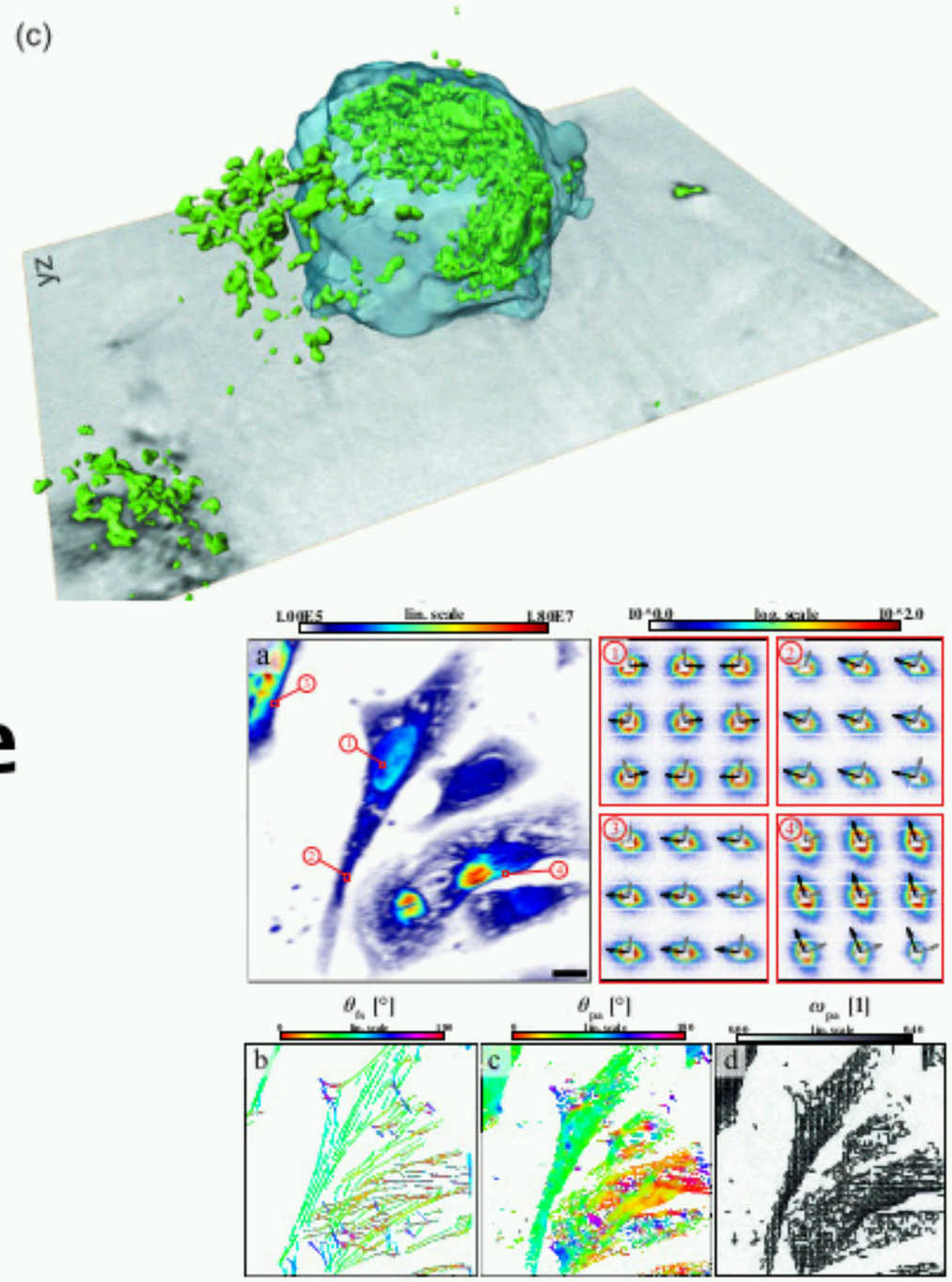
- holography, tomography

Diffraction

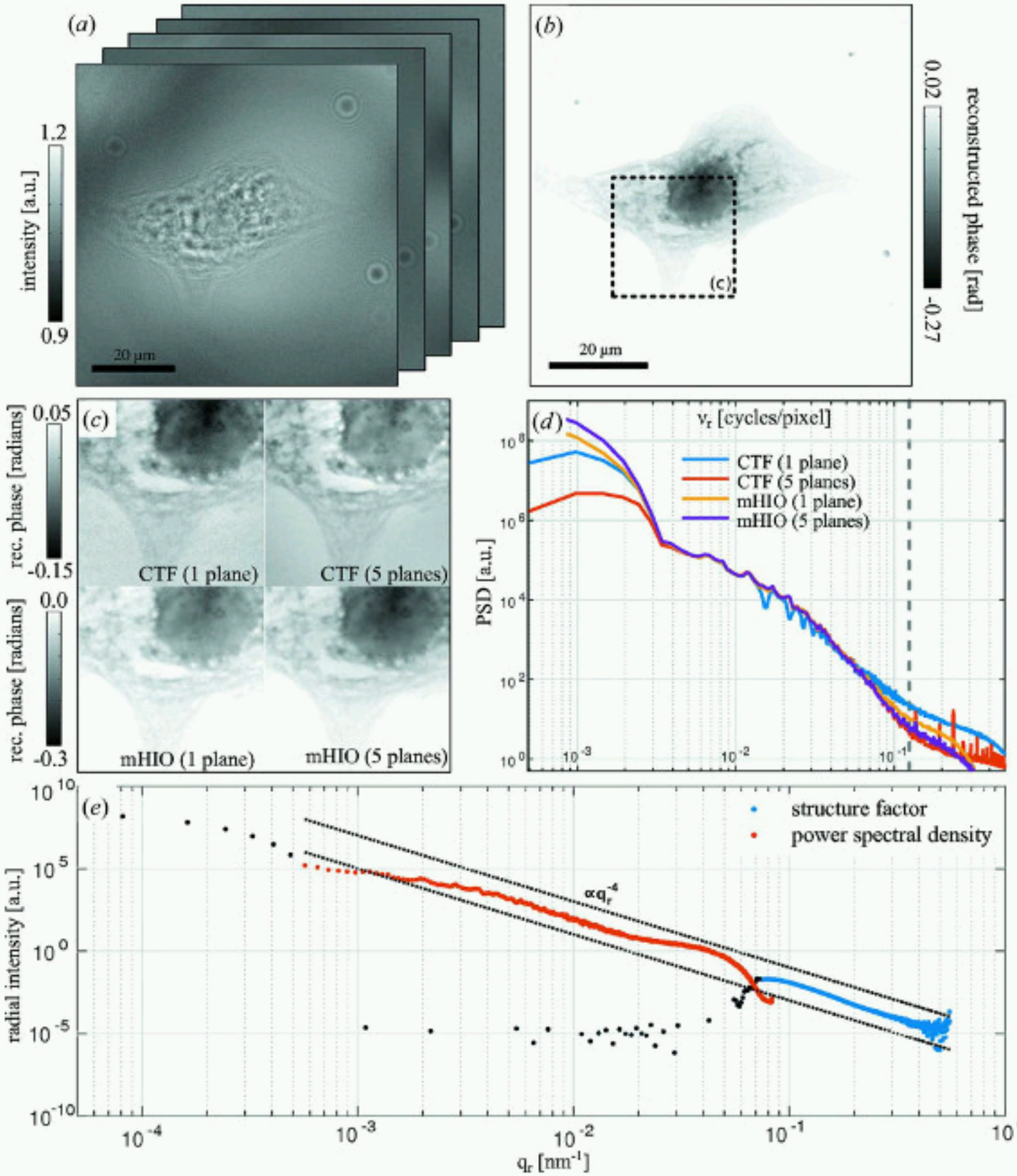
- scanning nano-SAXS

Visible-light fluorescence

Mechanics of cells



Scientific Questions and Methods



Data Rates

Tomography:

- 4 distances
  × 1500 angles
  × 4 MP (sCMOS)
  × 16 Bit

- 45 GiB in < 1 hour
- during analysis:
  - × 4 (complex float) during 3D phase-retrieval

- has to fit into memory
- after analysis: volume rendering

Data Rates

Scanning:

- 255 × 255
  × 4 MP (Eiger)
  × 0.1 Byte (LZ4)

- 24 GiB
  in 3 minutes … 30 minutes … 120 minutes
  (750 Hz … 50 Hz … 10 Hz)
- during analysis: 480 GiB raw data

- all images can be processed individually
- massive information compaction
  4MP -> few numbers

STXM analysis

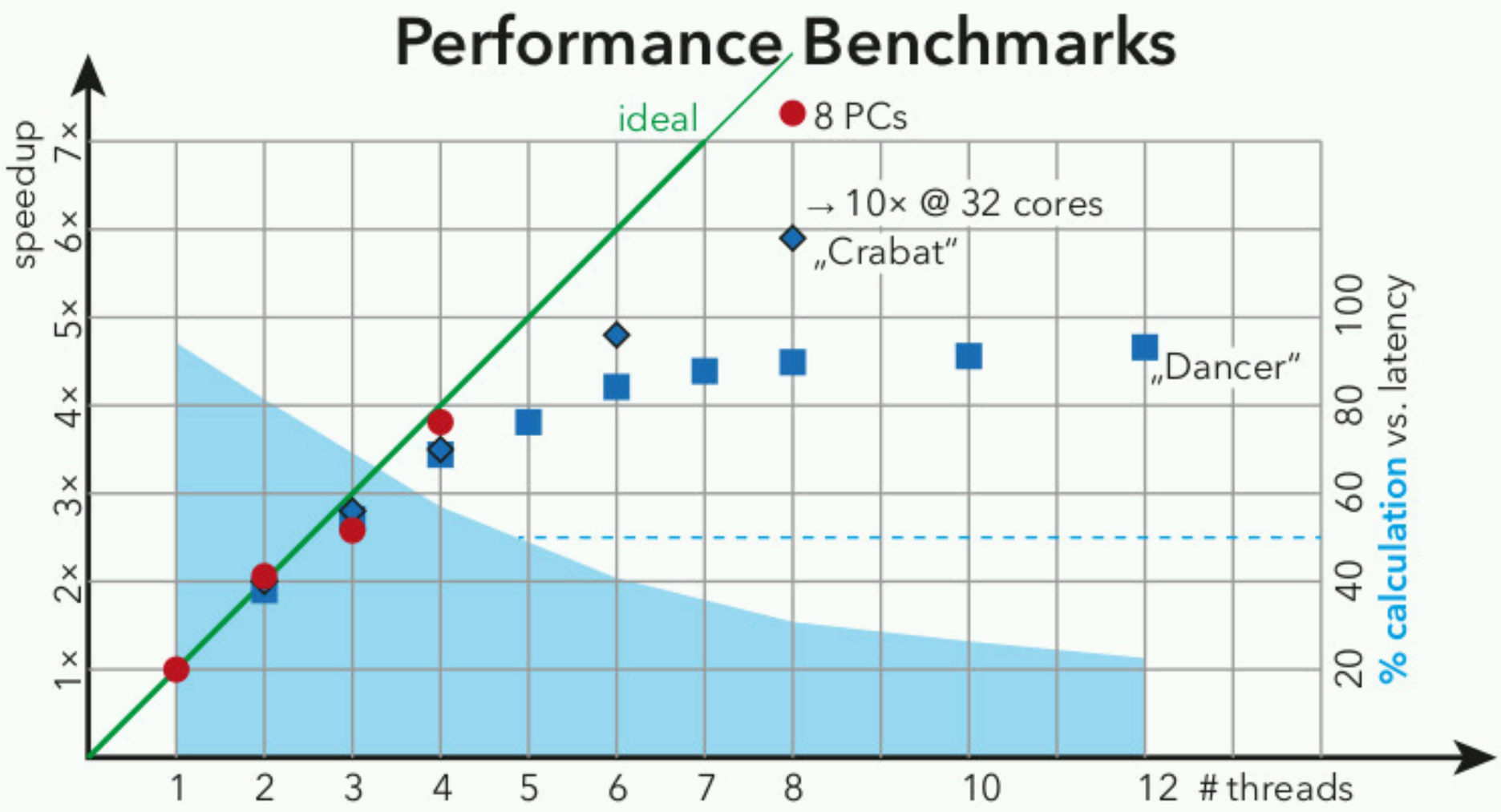
many independent images

- "simple" analysis
  - Itot, comh, comv (darkfield)
  - simple radial analysis (q-ROI)
  - principal component analysis
- structure factors
  - azimuthal average, I(|q|) fitting
- trivial parallelisation!

scaling on parallel CPUs

bad scaling

- CPU core has finished analysis of cached data before new bunch from RAM has arrived (latency ~100 ns)



Solution (within our budget)

“Heinzelmännchen” (24×450€):

- not one large workstation, but many small cheap boards,

- 1 “Heinzelfrau”:

Solution (within our budget)

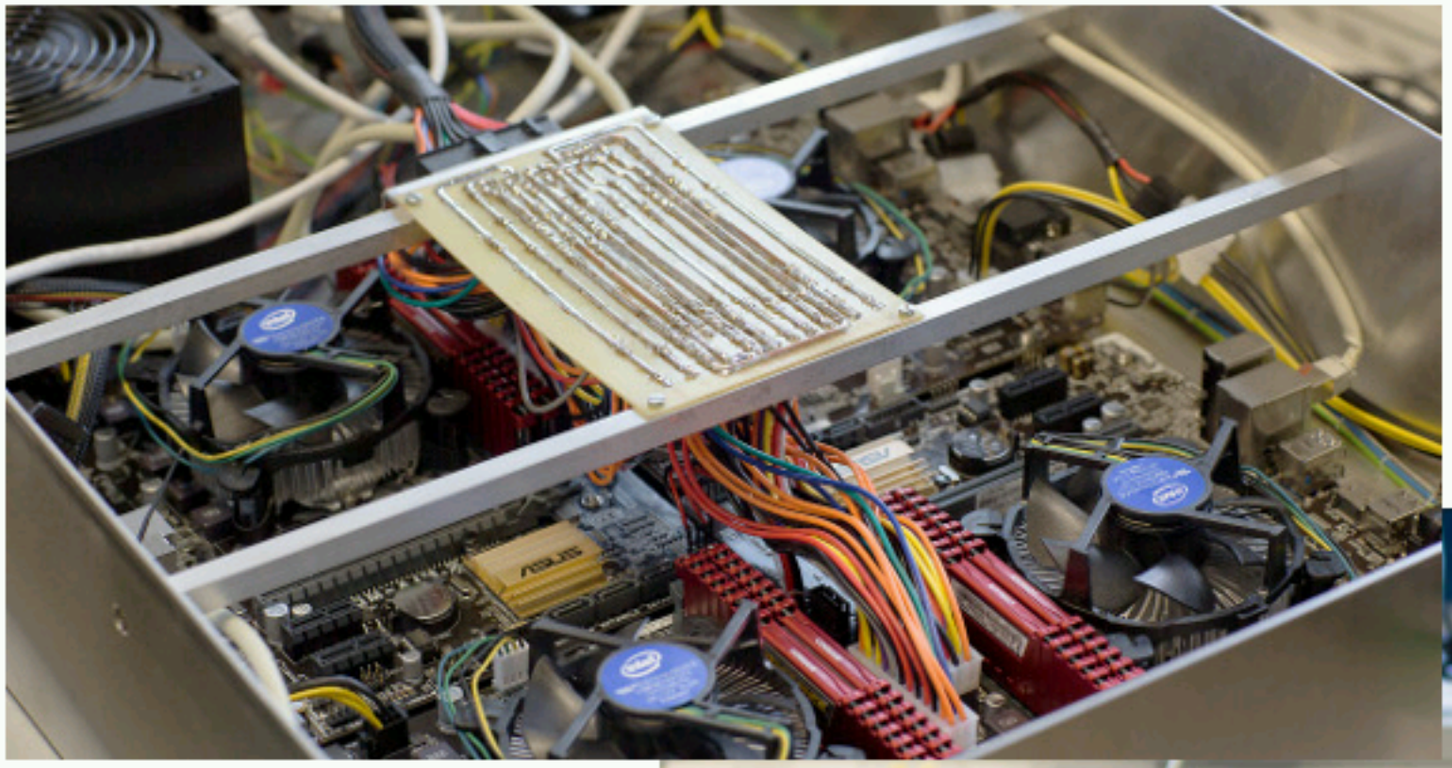
“Heinzelmännchen” (24×450€):

- not one large workstation, but many small cheap boards,
  - 24 boards
  - 24 Intel i7-7700
  - 24 × 8 GiB RAM
  - 6 power supplies; self-made 19" drawers (now)

- 1 “Heinzelfrau”:
  - SSD cache between NFS storage and nodes
  - load balancer + webGUI
  - 20 GBit uplink, 24 × 1 GBit downlinks

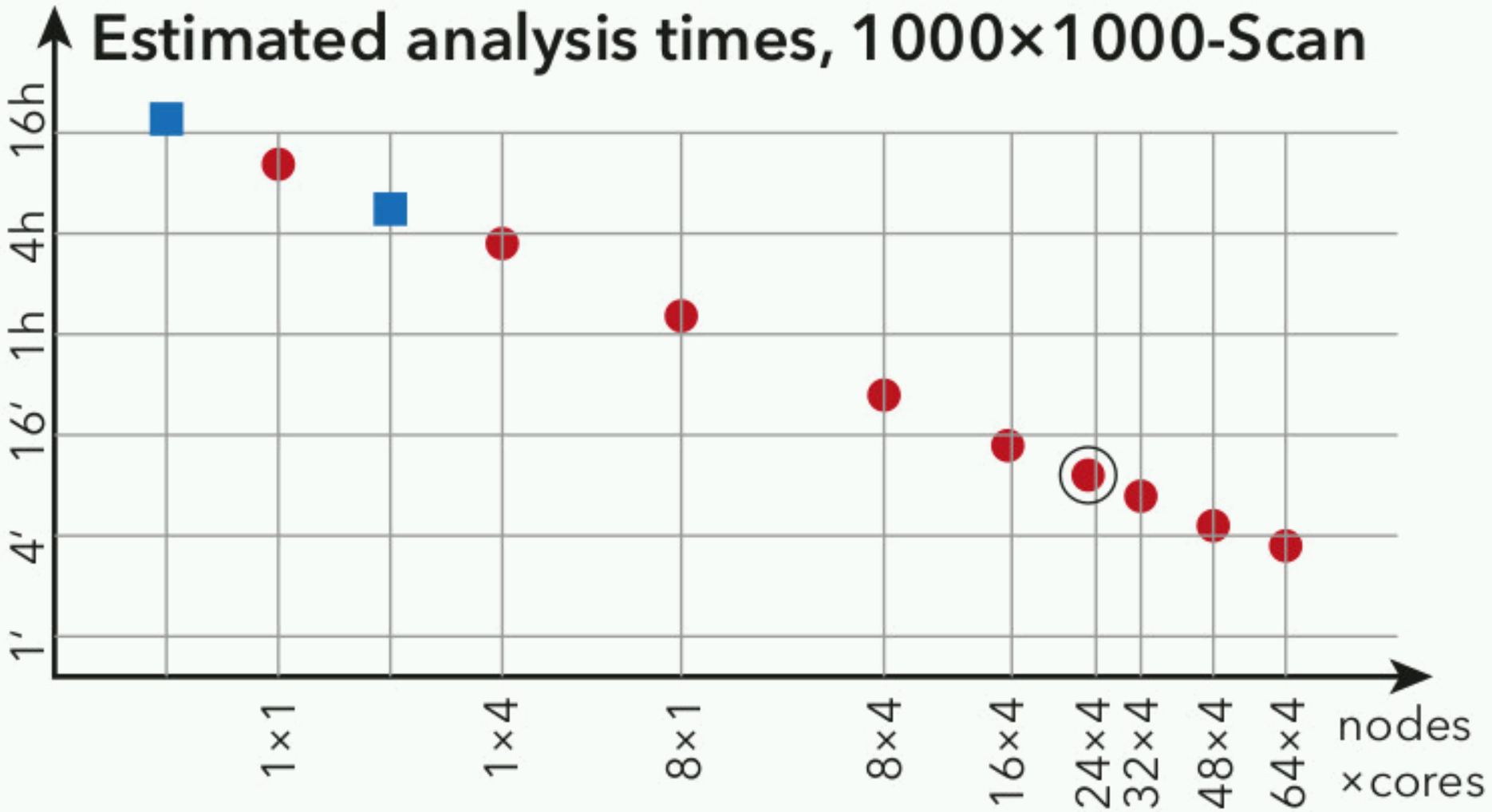
Solution (within our budget)

“Heinzelmännchen” (24×450€):



Solution (within our budget)

“Heinzelmännchen” (24×450€):



Solution (within our budget)

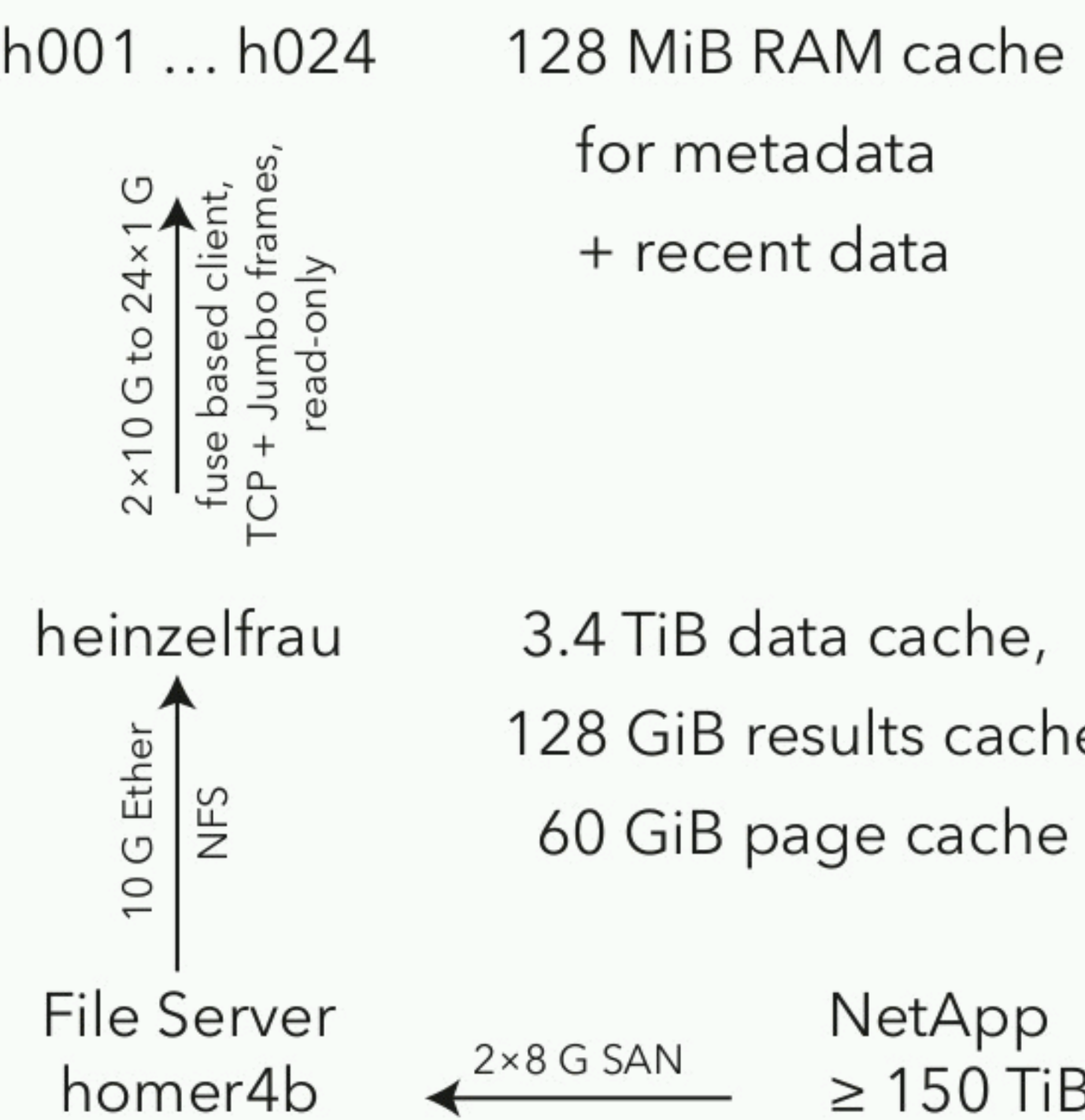
dadafs (network file system)

- read only
- optimal caching

- data storage on NetApp

- SSD Cache @ Heinzelfrau
  - data fragments
  - FS meta data

- up to 20 GBit of LZ4 HDF5



At the beamline

GINIX @ P10

- Control: SPEC
- 2D detectors:
  - Pilatus 300k, EigerX 4M (DESY);
  - sCMOS (different types, IRP)

- live viewer + counter for SPEC
  - image on top screen
    including empty beam correction
  - ct; dscan + pic/cen
  - live_print for (paper) lab book
    also: mic_print et al

direct feedback, visual and quantitative

At the beamline

direct feedback, visual and quantitative

- detector images
  - check alignment of optics
    + beamstop, apertures, general status
  - check alignment of sample

- counting on 2D detectors (the “modern diode”)
  - alignment of optics
  - alignment of sample
  - estimation of flux

- online-analysis: check for …
  - alignment, beam damage
  - resolution goal

decide, how to proceed

Online analysis

dada:

- image browser,
- composites,
- STXM analysis

Matlab:

- advanced analysis,
- workflow under development,
- especially for tomography;
- once routine work -> implement in dada

Online analysis

Important: direct visual feedback

using Maxwell,
it was crucial to mount
/beamline/p10 via NFS,

waiting for sync -> core
interrupted experiments

Thanks for making that possible!

Thanks for your attention.