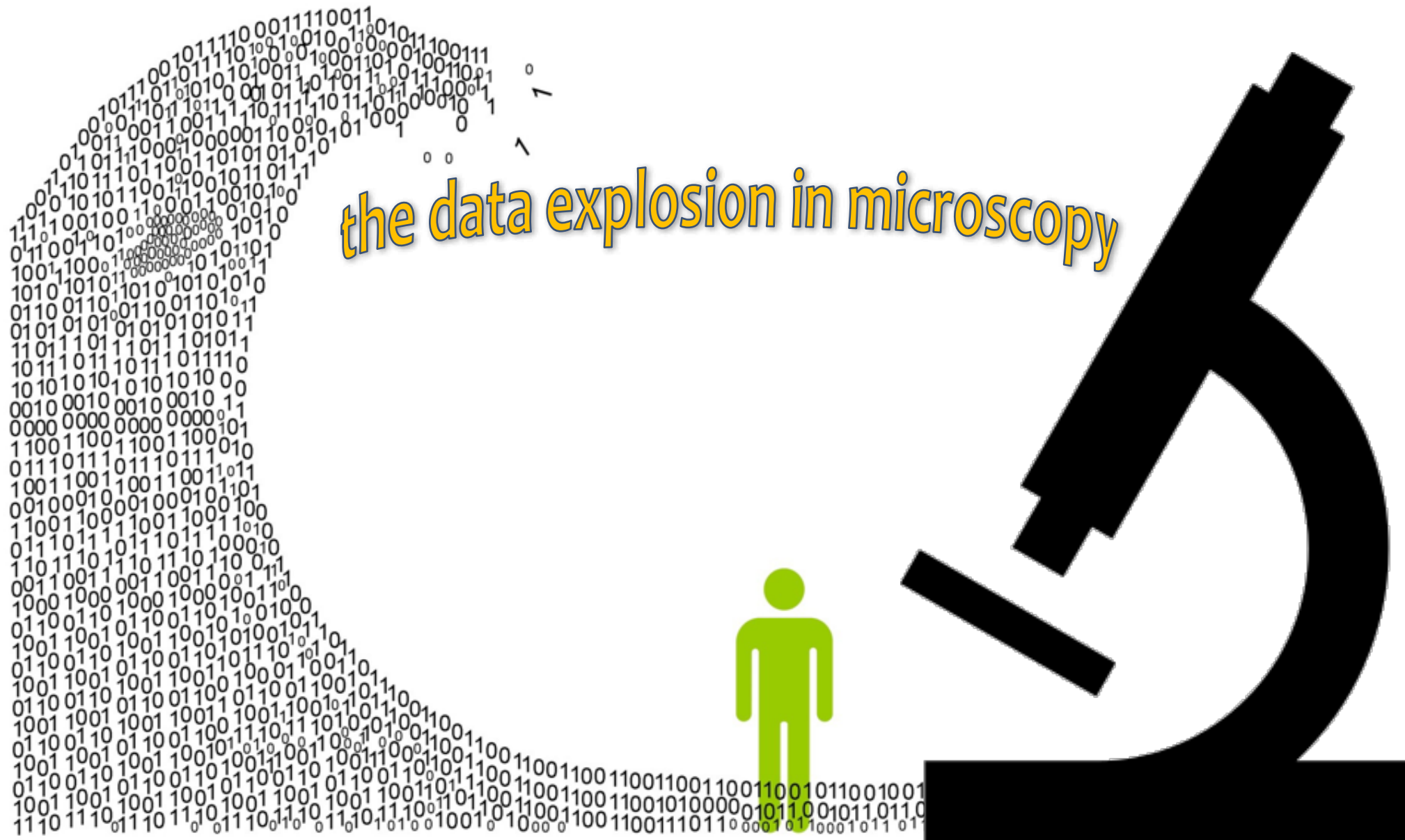
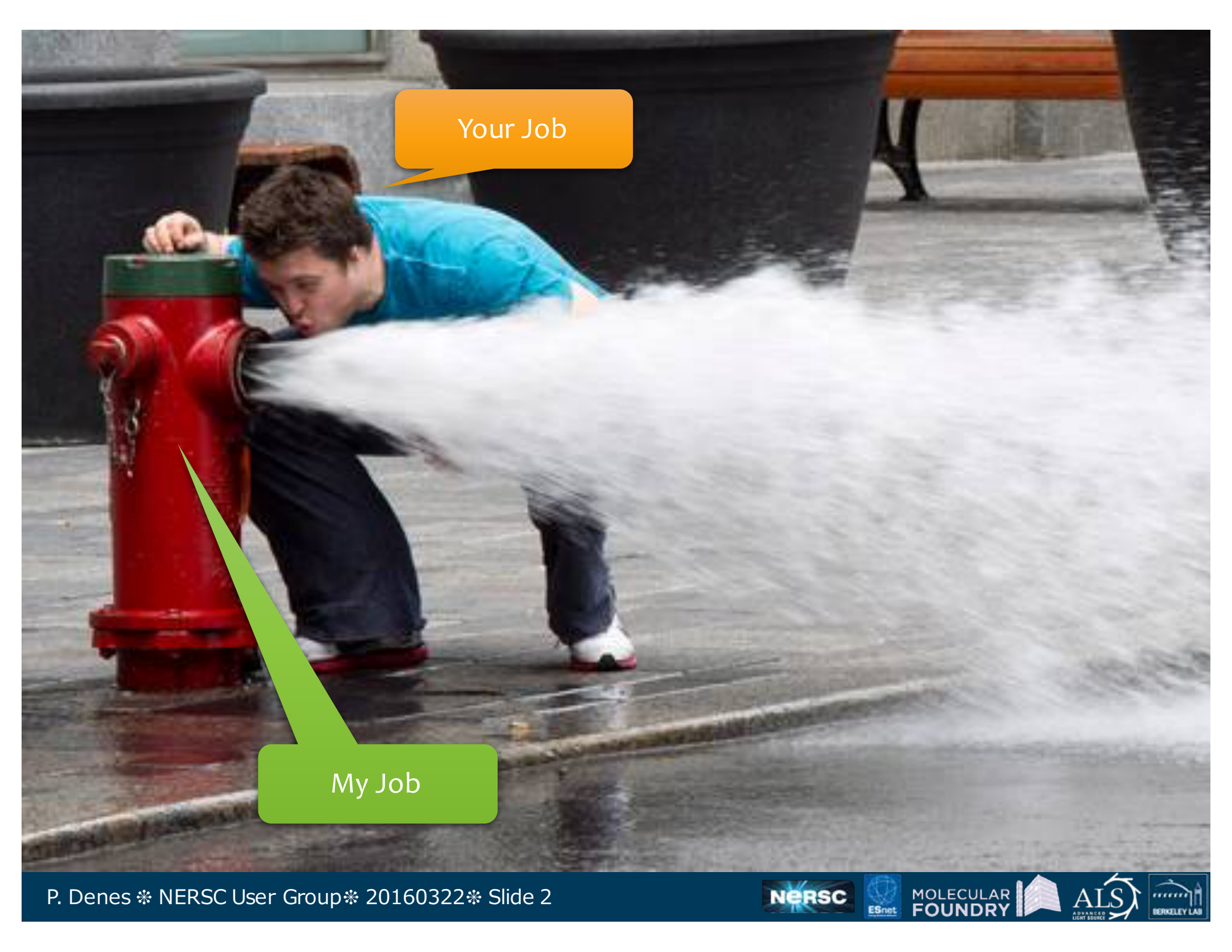


Picture Perfect

the data explosion in microscopy





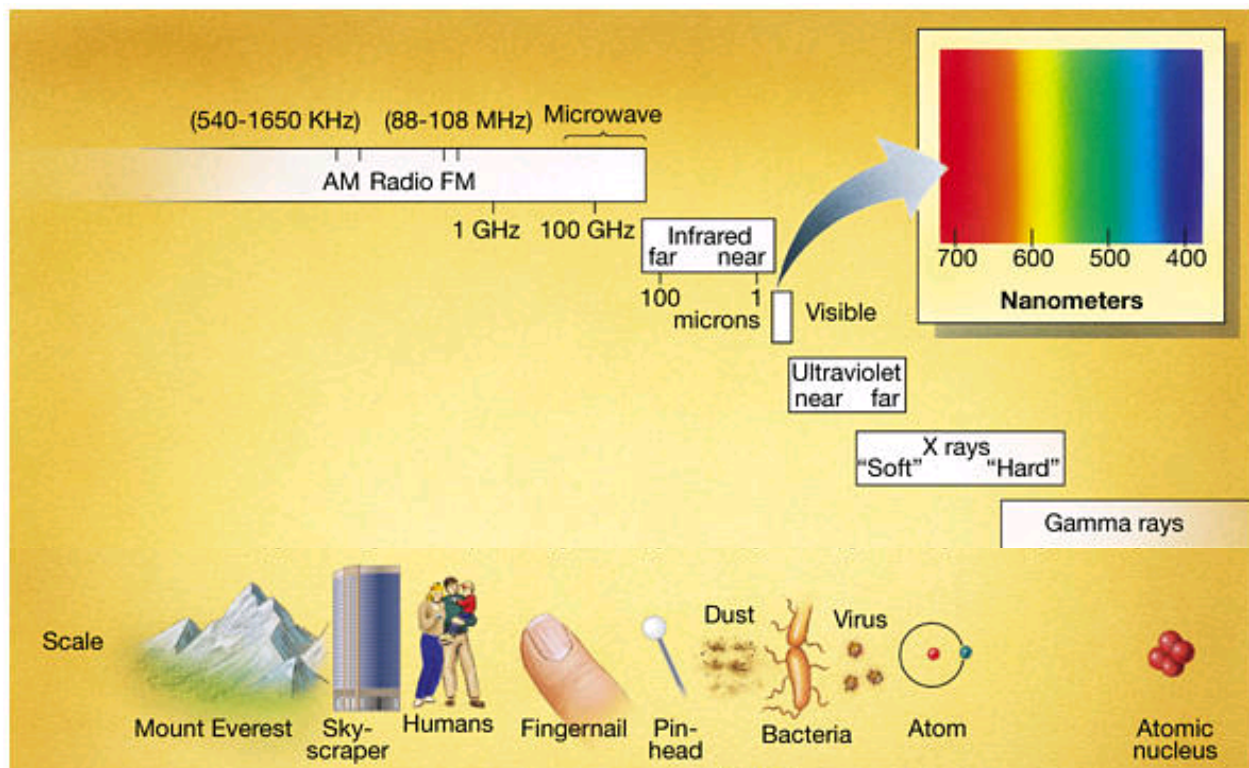
Your Job

My Job

Outline

- Microscopies and detection – challenges
- Vignettes – better detectors enable:
 - Biology (cryo-Electron Microscopy)
 - Materials (in-situ Electron Microscopy)
 - Scanning microscopies (electrons and X-rays)
 - Neural “microscopies” (a mind is a terrible thing to waste)
- “Grand Challenge” 100,000 frame/sec. “4D-STEM”
 - 400 Gb/s direct-to-NERSC
 - Science driver for a “data superfacility” – in 2017
- Synthesis – what is common to all of this?

Seeing Small Things

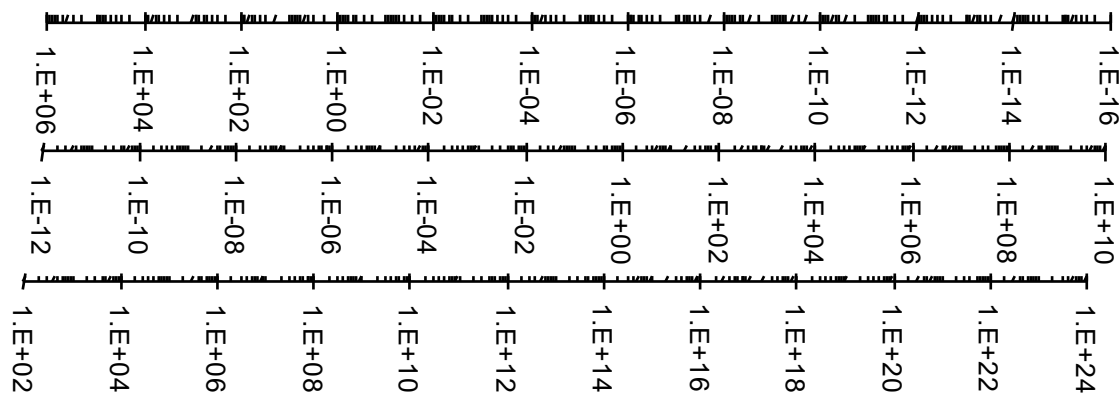


Many objects imaged with visible light

x-rays (ALS)

electrons (EM)

neutrons, ...



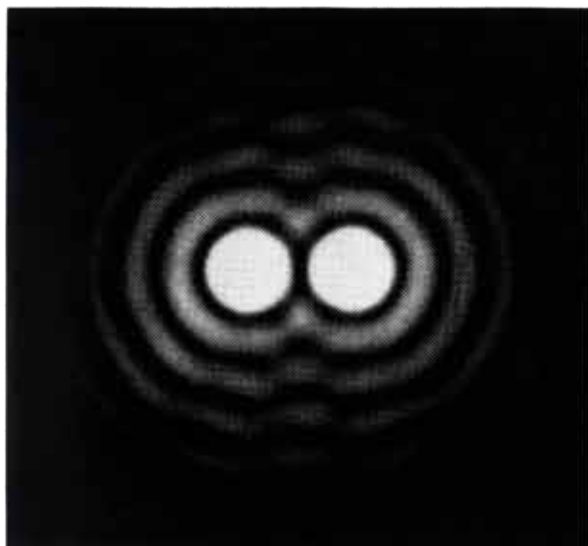
λ [m]

E [eV]

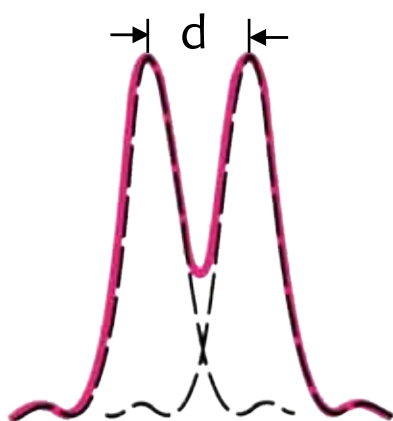
f [Hz]

Diffraction Limit

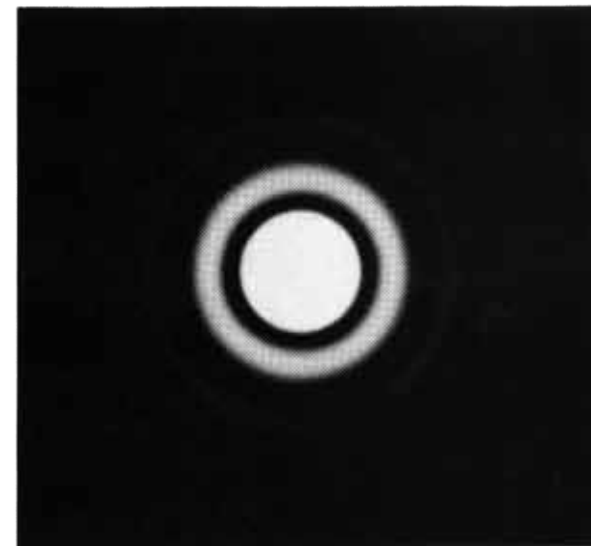
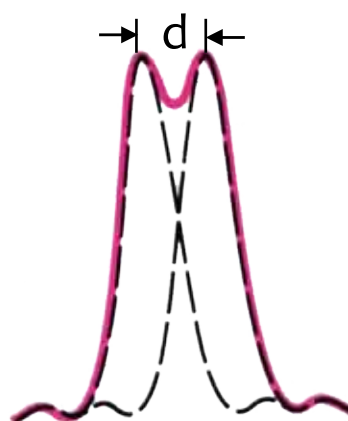
$$d \sim \lambda$$



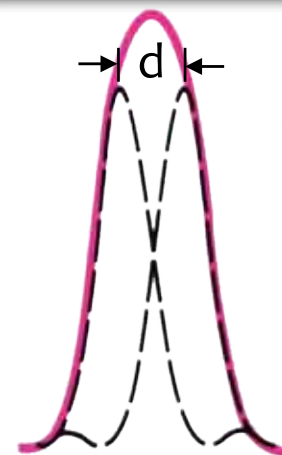
Resolved



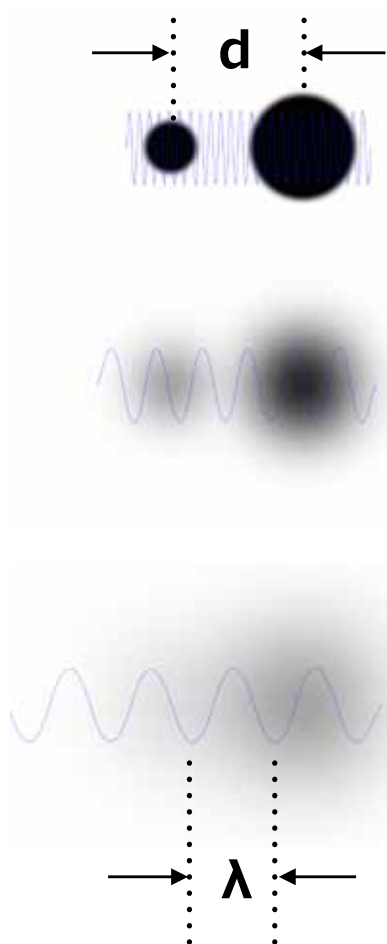
Just Resolved



Not Resolved

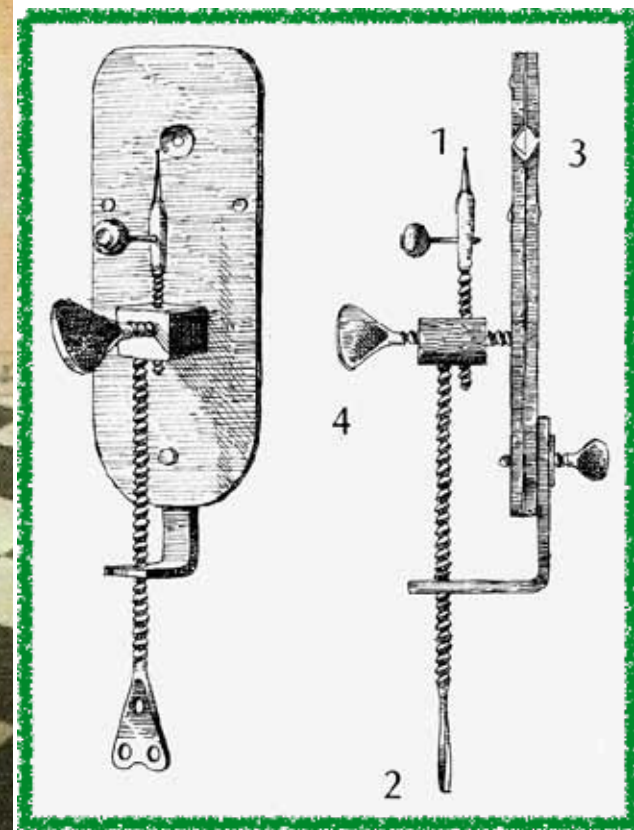
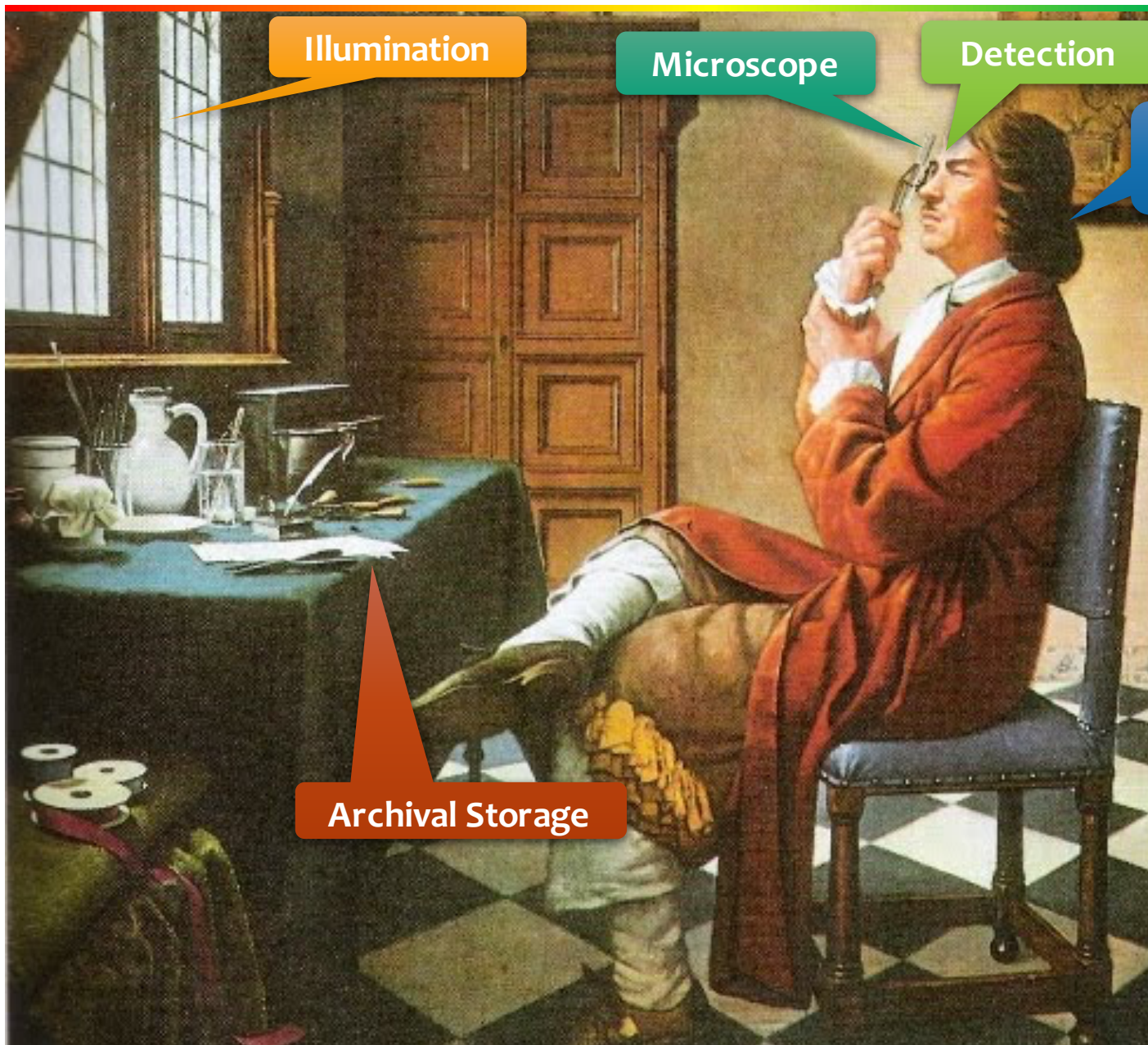


Solutions



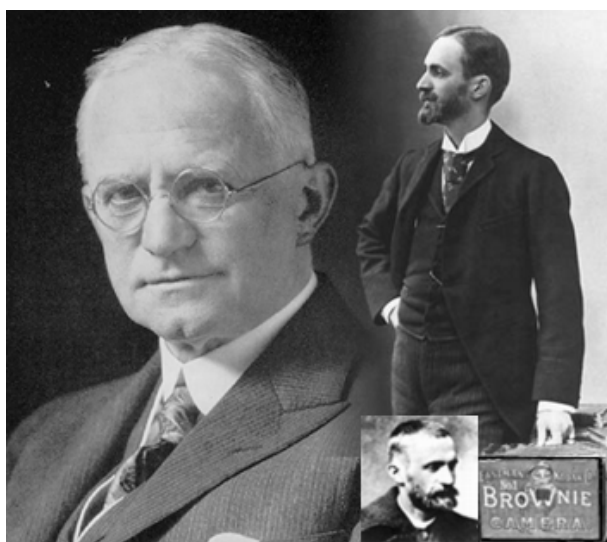
- Visible light $\lambda \sim 1000$ nm
 - Super-resolution techniques
- Shorter wavelengths
 - X-rays $\lambda \sim 10^{-1}$ Å
 - Optics
 - Electrons $\lambda \sim 0.1$ Å
 - Optics

Modern Microscopy



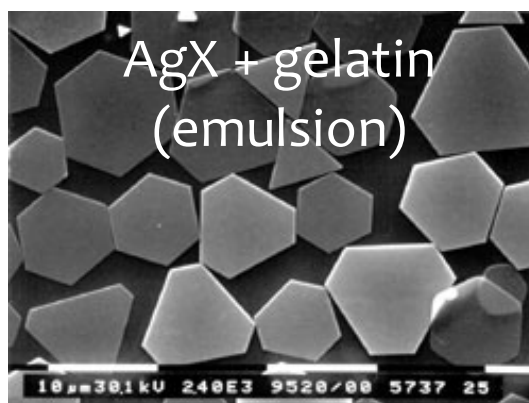
Advent of Digital Detectors

- Silver halide emulsion recording system
- Pretty good!

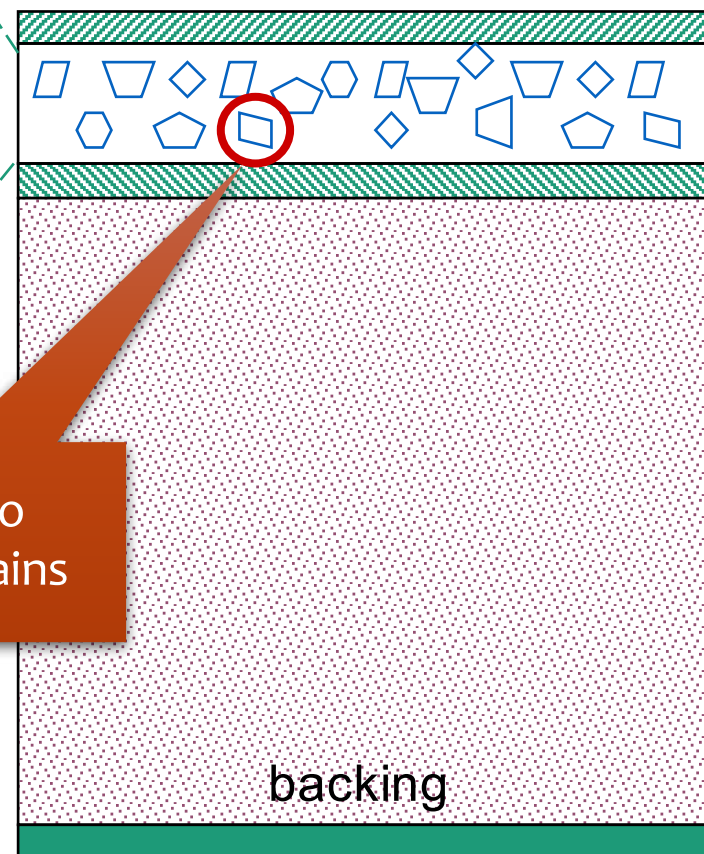


George Eastman

"You press the button, we do the rest"



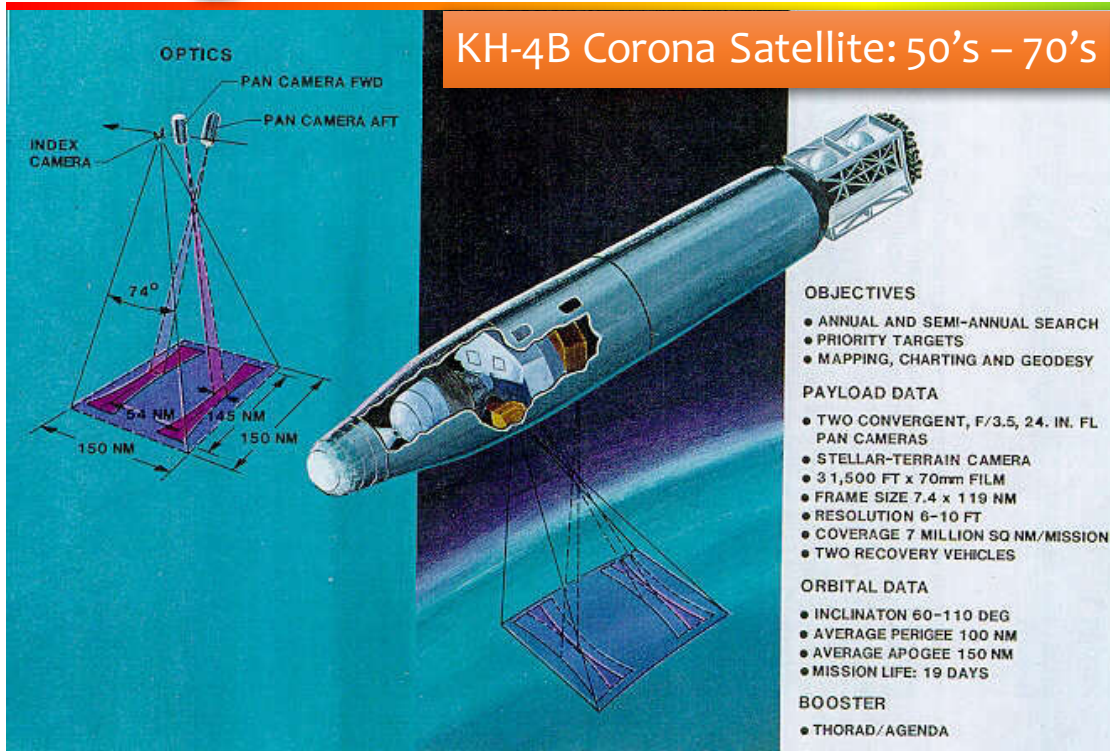
Electron micrograph of tabular grain emulsion



sub- μm to
few- μm grains

backing

Digital Detector ☺☹

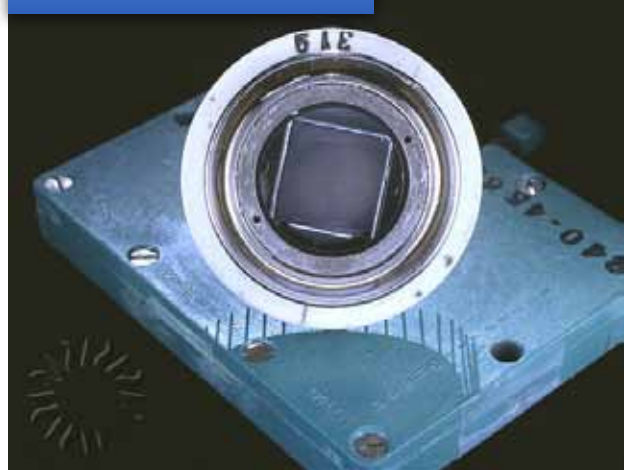


Hubble Space Telescope Mirror - 1979



- High spatial resolution
- Large area
- Easy to use
- (Cheap)
- Efficiency
- Dynamic range
- Readout time
- ...

Hubble CCD – 0.6 MPixel



Dynamics: Increase Speed and S/N

The Berkeley Lab VIEW

Search The Berkeley Lab View Archive

December 19th 2003

A Golden Opportunity: Watching the Atoms Dance

BY PAUL PRUESS

At the National Center for Electron Microscopy (NCEM) some months ago, Chris Nelson was helping a Life Sciences Division group investigate gold particles, useful for labeling protein structures. Nelson noticed unusual movement under the microscope, and when the users broke for lunch, he got their okay to record video images.

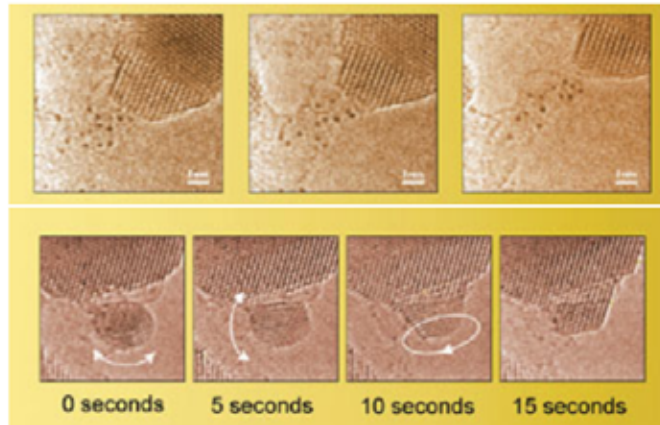
What Nelson captured on video was an extraordinary phenomenon: visible against the background of the thin carbon film that supported the samples were individual atoms of gold, moving about and attaching themselves to gold crystals.

Nelson was working with the One Angstrom Microscope (OAM), whose information limit of 0.78 angstroms allows computer reconstruction of images at better than 1-angstrom resolution, but whose 1.6-angstrom native resolution is sufficient by itself to image massive atoms. The veteran Atomic Resolution Microscope is also capable of 1.6-angstrom resolution, but, says Nelson, "We would never have seen this with the ARM."

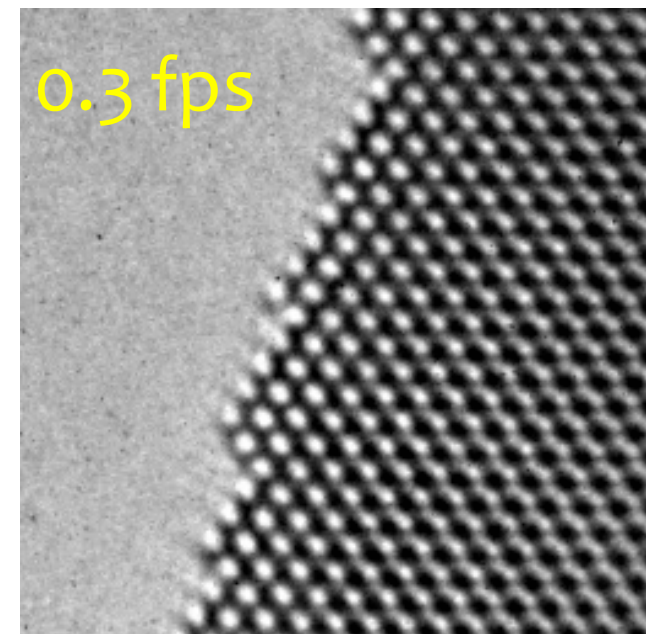
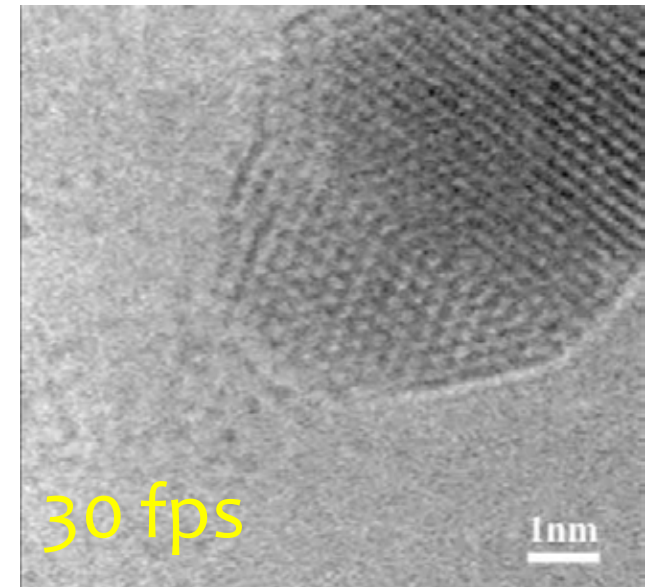
"It's the signal-to-noise ratio that determines what you see," says NCEM's Christian Kisielowski, "which is why a column of atoms looks sharper than an individual atom — the more atoms contributing to the signal, the stronger."

Kisielowski explains that the OAM incorporates recent enormous improvements in signal-to-noise ratio, made possible by computers and image-reconstruction software but also by increased stability and correction of lens aberrations. Even so, the movie's individual gold atoms were visible only because their signal was relatively strong against the "quiet" background of the lightweight carbon.

Nelson says the video has "implications for the TEAM project," the multi-institutional Transmission Electron Aberration-corrected Microscope program now underway to build a microscope with half-angstrom resolution. For one thing, the gold movie shows what can be achieved even with the short exposures (about 1/60th of a second) of a video camera running at 30 frames per second.



Top: In a video made by NCEM's Chris Nelson, individual atoms of gold (lower left) are seen migrating to join a growing gold crystal (upper right). Bottom: In this video sequence, a gold nanocrystal continually reorients itself in three dimensions before finding just the right way to fit onto the larger crystal. White arrows indicate examples of the directions of this rapid movement.



2008 – TEAM Detector (800 MB/s)

Direct Detection

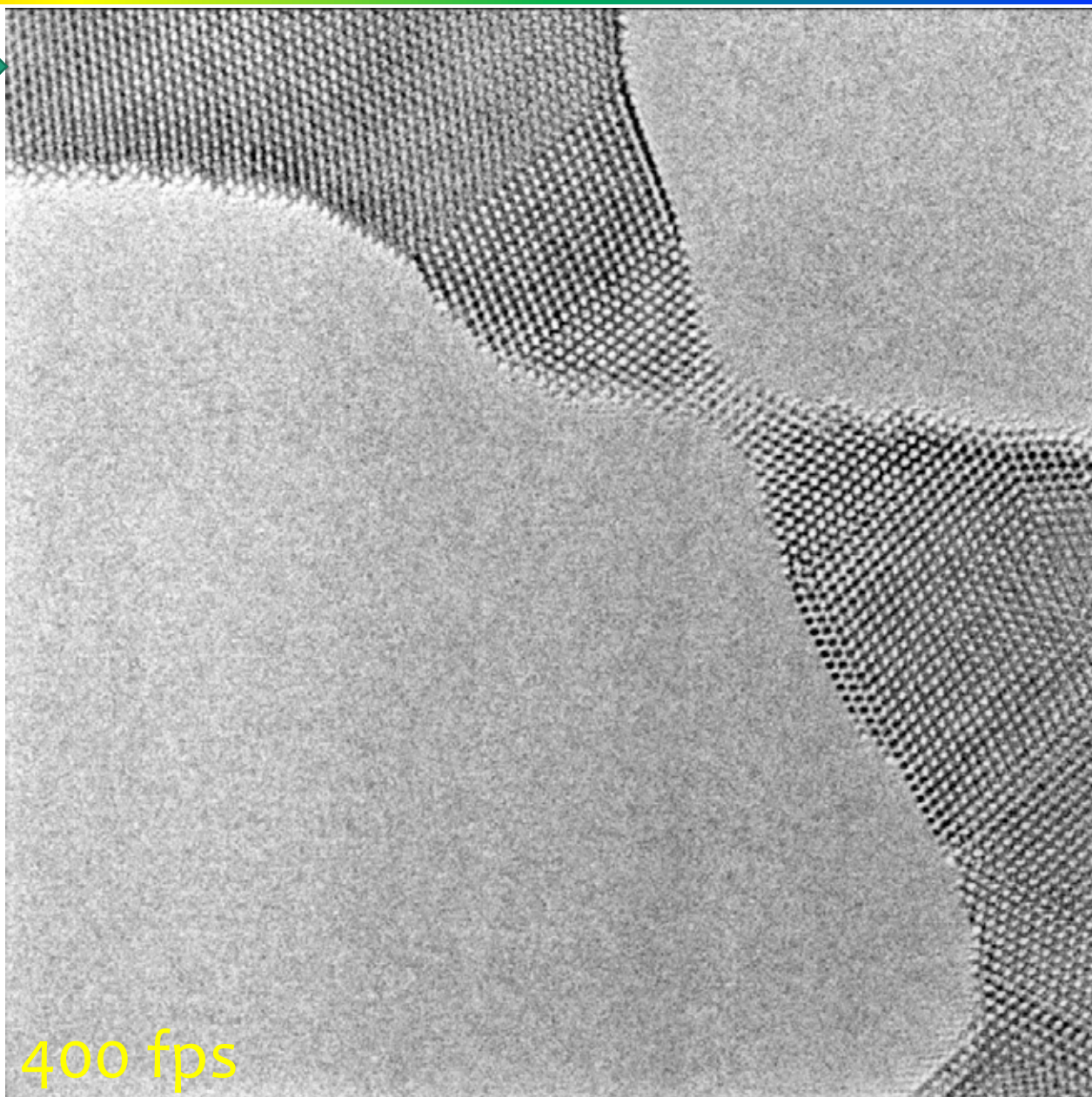
Indirect Detection

Fluorescent screen

30 fps

Phosphor CCD

0.3 fps

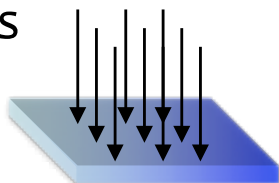


“Characterization of a CMOS active pixel sensor for use in the TEAM microscope” Battaglia et al., 2010

Factor of 2 in Spatial Resolution → 16 x Dose

- Illuminate sample

- Detect “pixel” areas
- Pixel size $\propto d$



N particles / “pixel”

Poisson statistics: $\Delta N = \sqrt{N}$

Contrast $C \propto \Delta N/N$

Rose Criterion: $C \geq k$ ($k = 5$)

- Source is Poissonian

- 2 dimensions →
- $(N^2)^2$

$$N = f \times D \times d^2$$

$$D = \text{Dose} = \text{ptcl/s/Area} \times \text{Time}$$

f = factor for scattering and detection efficiency

$$\rightarrow D \sim k^2/f \times C^2 \times d^2$$

$$\text{But } C \sim d \rightarrow \mathbf{D \sim 1/f \times d^4}$$

- Hard to beat Poisson

- “Scattering efficiency”
 e^- , X-rays, energy...

- “Detection efficiency”

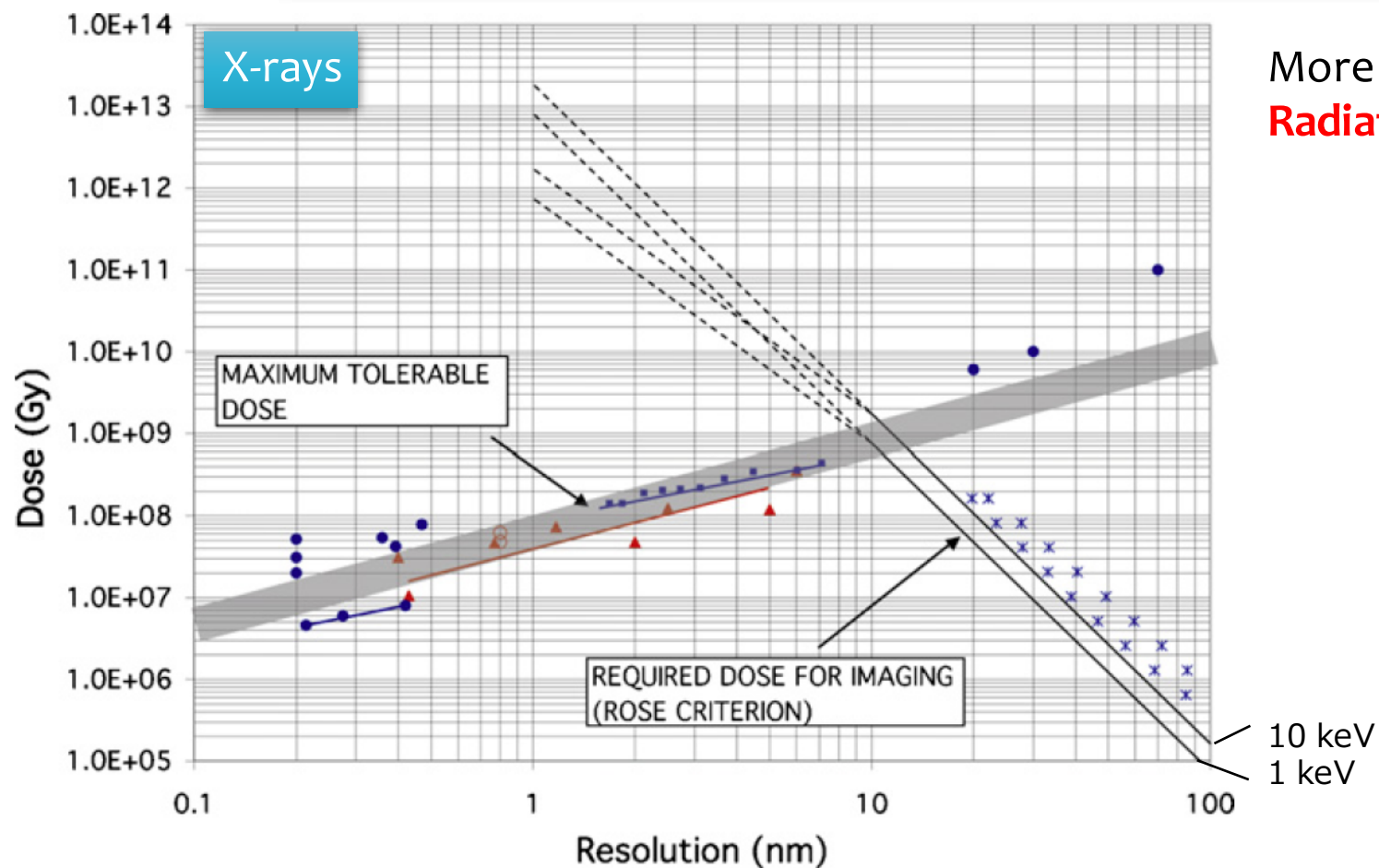
Radiation Damage $\propto$ Dose (D)

Improve f !

Imaging Biological Samples

Biological samples: Lose contrast after too much dose

Brighter source needed to see smaller objects



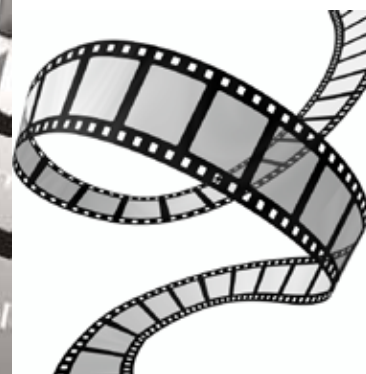
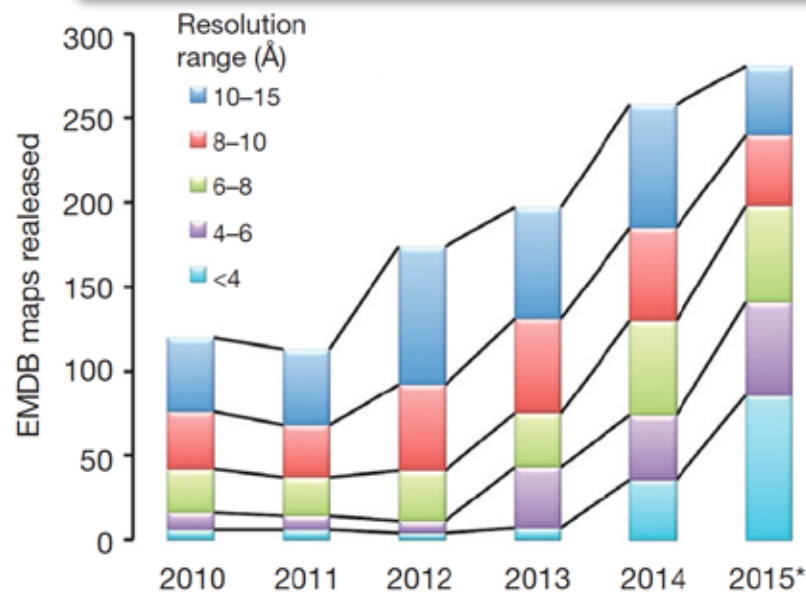
More probe particles/ $\text{\AA}^2$
Radiation Damage

“An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy”, Howells et al. 2009

Direct Detection – cryoEM



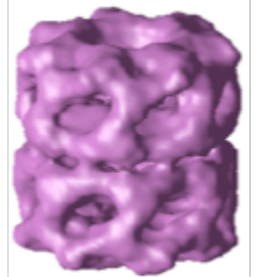
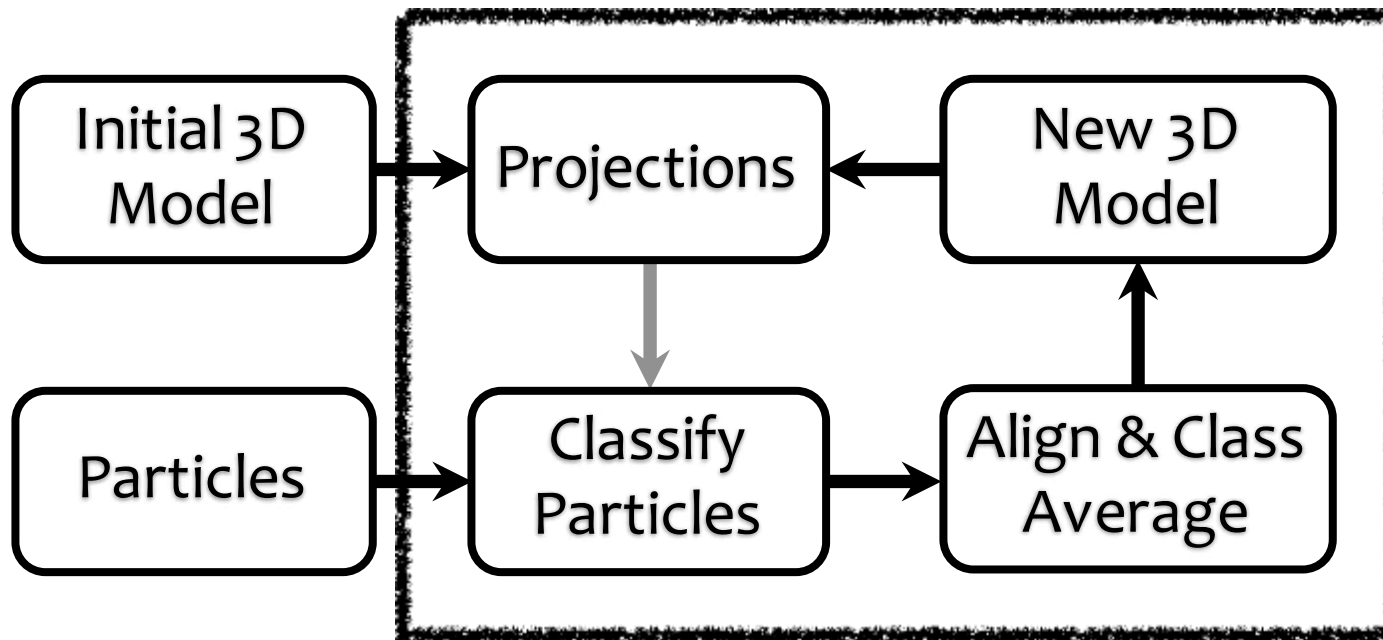
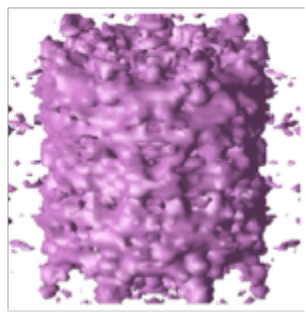
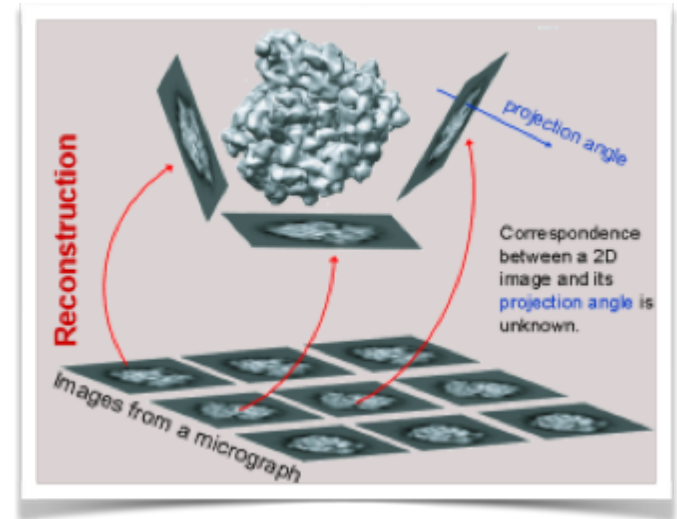
EM map depositions in the Electron Microscopy Data Bank (<http://emdatbank.org/>) over the last six years



“The development of cryo-EM into a mainstream structural biology technique” Nogales 2016

Image Reconstruction

- Image reconstruction requires assembly of the multiple projection images into a 3D model
- Particles are aligned to projections calculated from a starting model
- A new model is created and the process iterated



P. Adams, LBNL

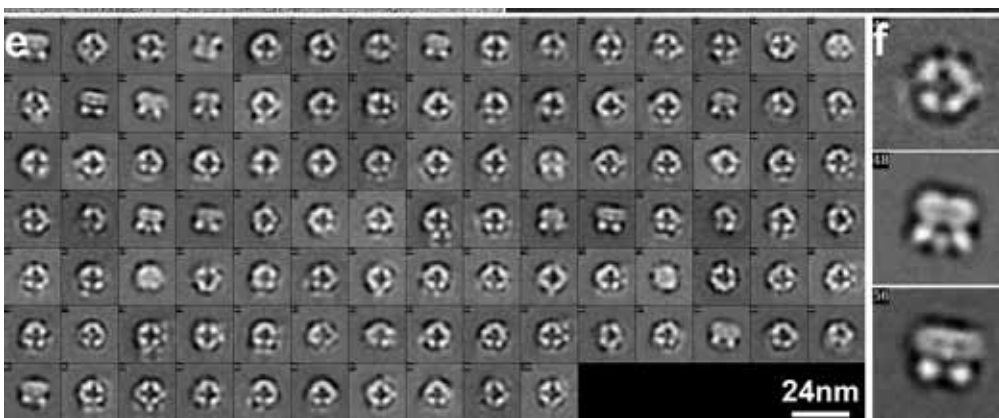
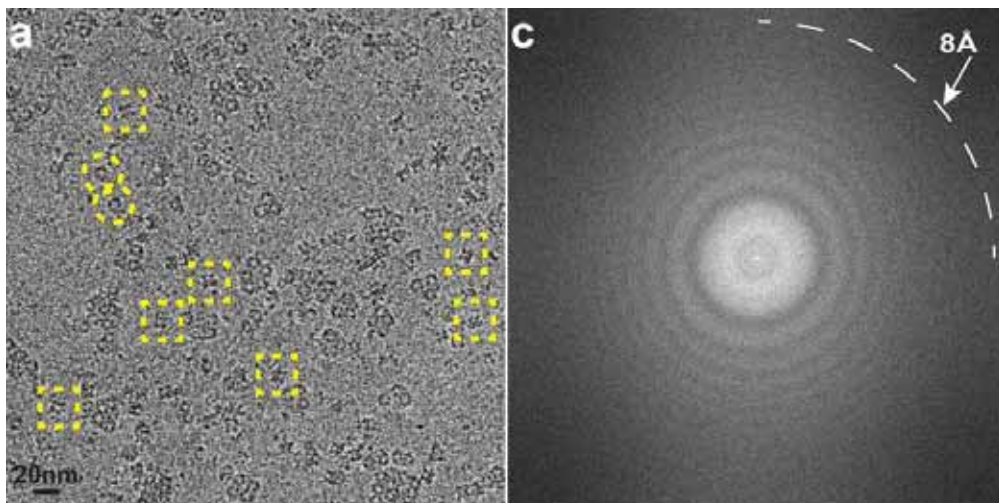
Efficiency



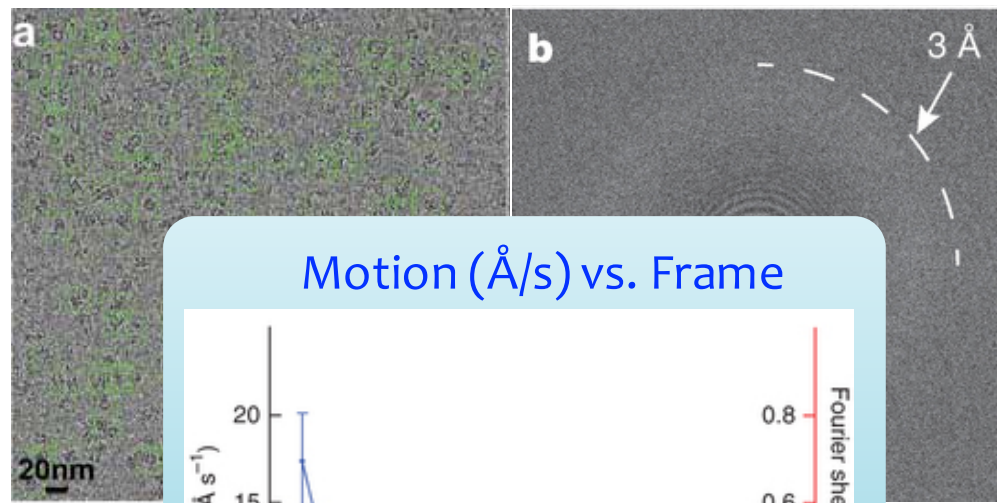
and Speed



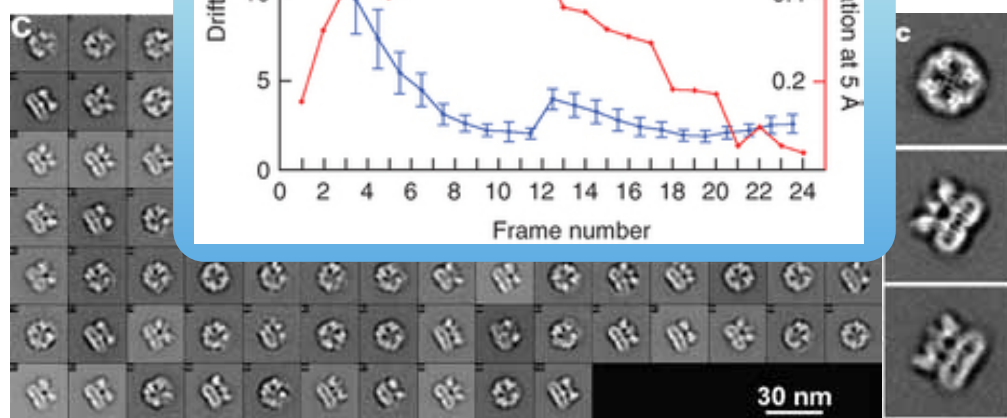
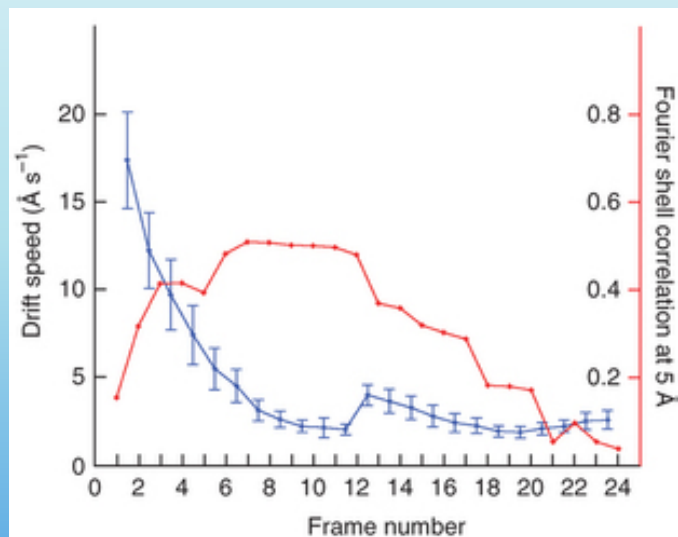
Indirect detection



Direct detection + Motion correction



Motion ($\text{\AA}/\text{s}$) vs. Frame



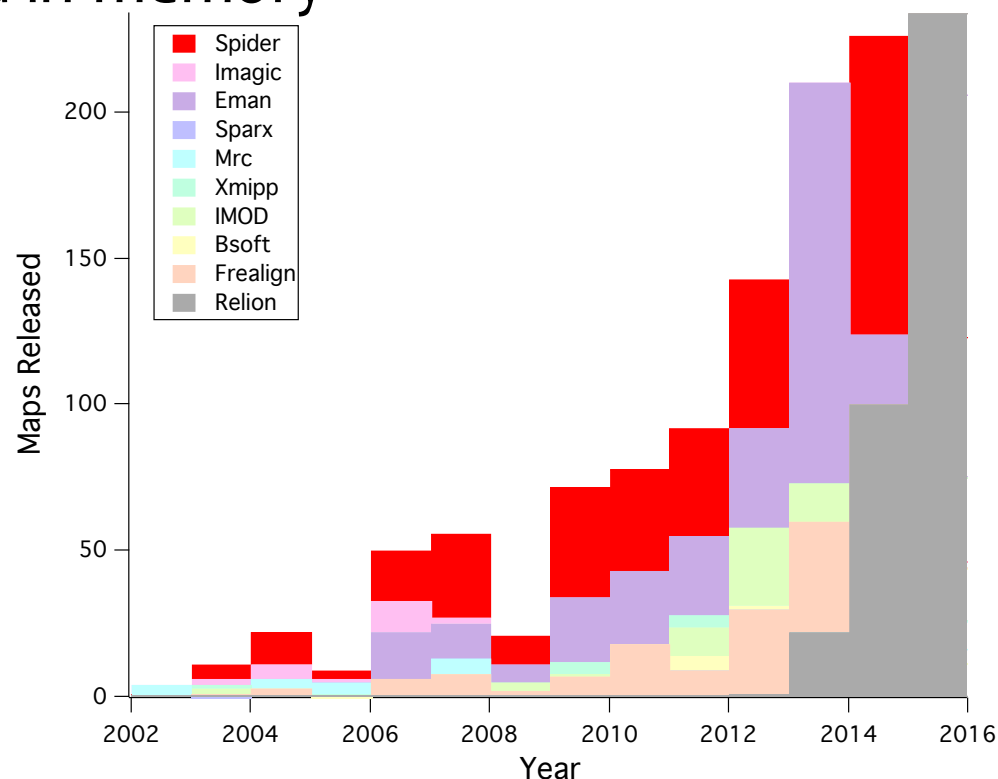
“Structure of the TRPV1 ion channel determined by electron cryo-microscopy” Liao et al., 2013

“Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM” Li et al., 2013

Software ...

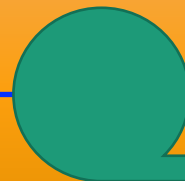
- Many packages used
 - Large numbers of images held in memory
 - Algorithm scaling (?)

- Front-end
 - “Counting/Clusterizing”

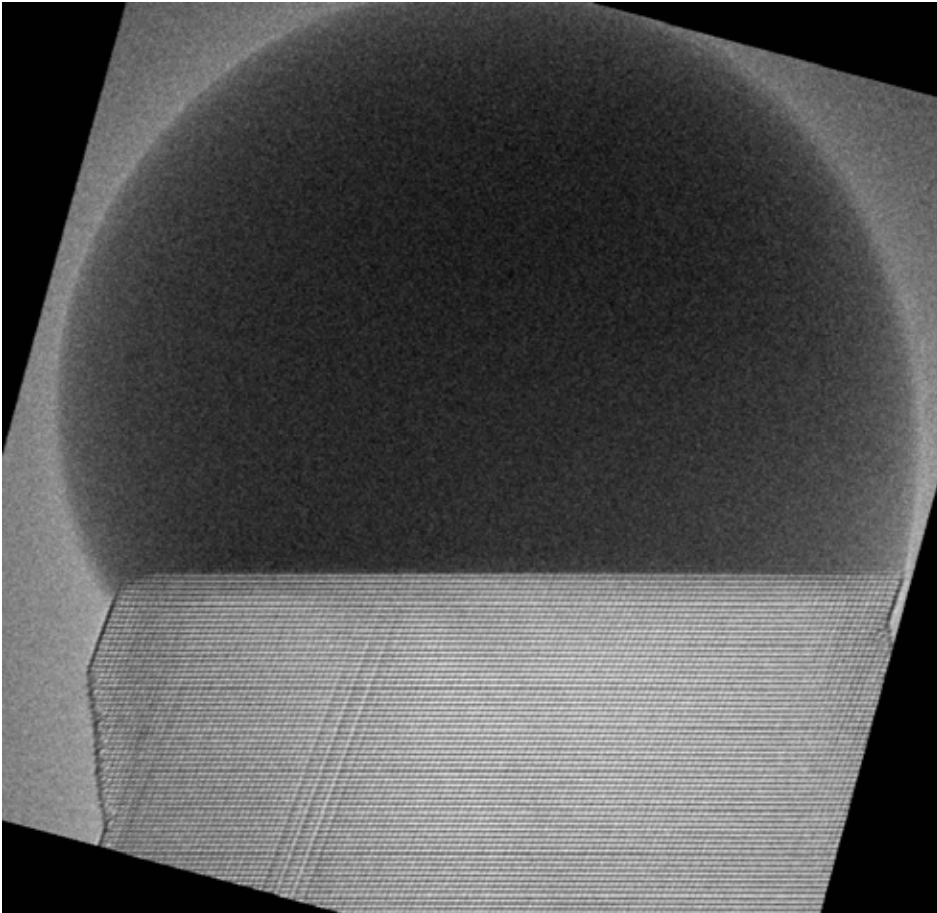


6.4 Gpix/s

f/w



Dynamics: Growth at the Nanoscale



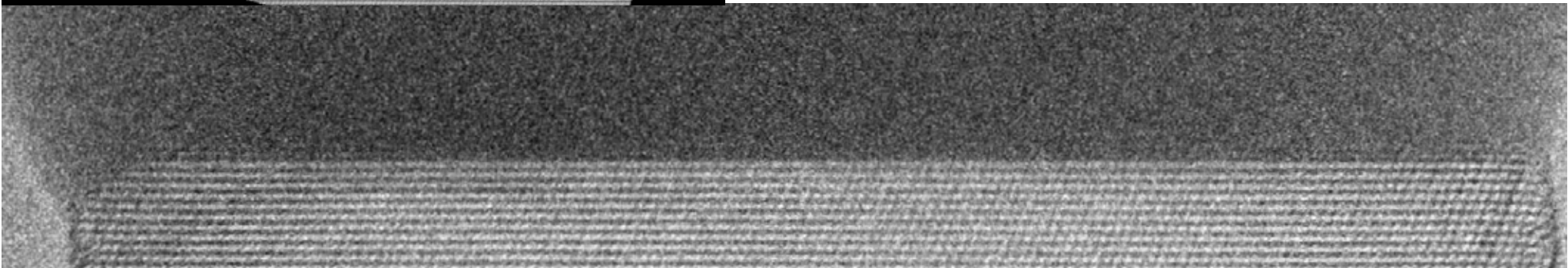
- Successful development of a direct electron detector dedicated for in-situ microscopy
- Provides unprecedented ability to image the processes of crystal growth at the atomic scale
- New high frame rate imaging capabilities allow capture of massive amounts of rich image data

Opportunities & Challenges

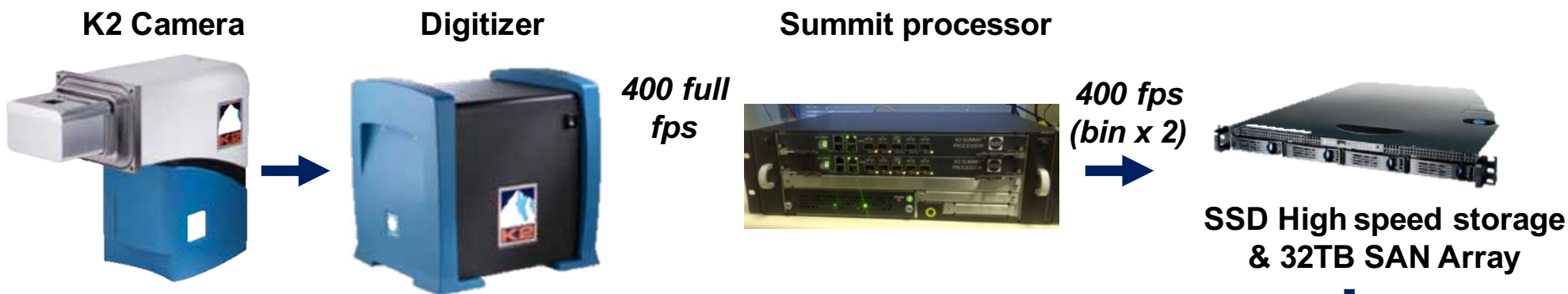
- There is a massive amount of data to be analyzed, protocols and methods to be developed

One user of E-TEM/K2 generates on average 30TB of data per successful week. This is more than 4.5 million image frames

E. Stach, et al. – CFN-BNL



A torrent of data ...



Storage

- Have 0.5 Petabytes allocated to us for 'temporary' storage

Analysis

- Super-computing facilities (e.g. "New York Blue")

Distribution

- Ultrafast data pipelines from the lab to NYC
- Connection via "Globus Online" at ANL



Install of 10 Gb/s fiber optic from CFN to RHIC/ATLAS



Working with Image Data Sets

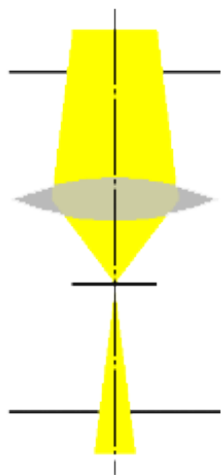
- Upload and download files
- Work with Big Data sets from remote locations
- Monitor a job, check work flow
- Get user account information
- Authenticate using user credential
- User dependent rights
 - System resources
 - Files
 - Batch jobs
 - Shell command
 - Status

What we want to do:

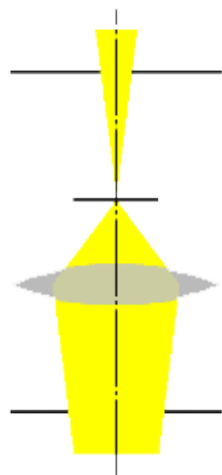
- Eliminate challenges with data transfer
 - Share data with users
 - Convert Data sets
 - Analyse data sets
 - Data sharing
 - Data management
 - Reduce latency
-
- Challenges to efficient data processing
 - High temporal and spatial resolution
 - Large Data Sets that require sophisticated storage schemes

Scanning / Transmission Microscopy

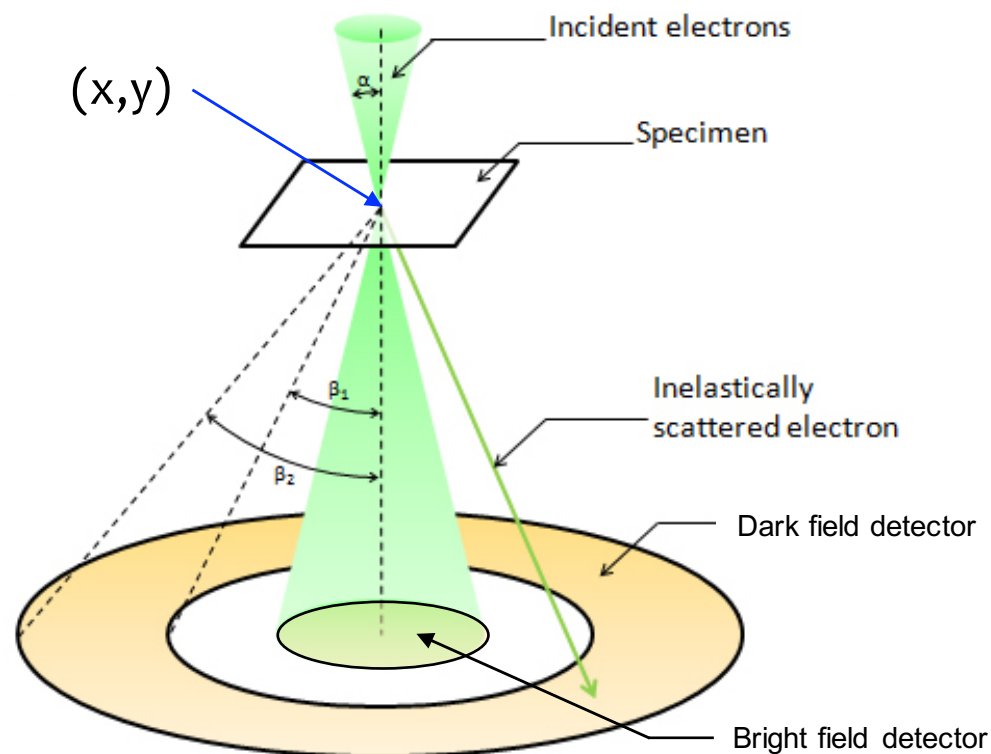
STEM



TEM



In principle, equivalent

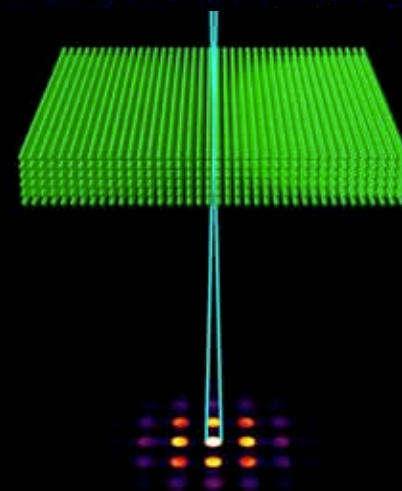
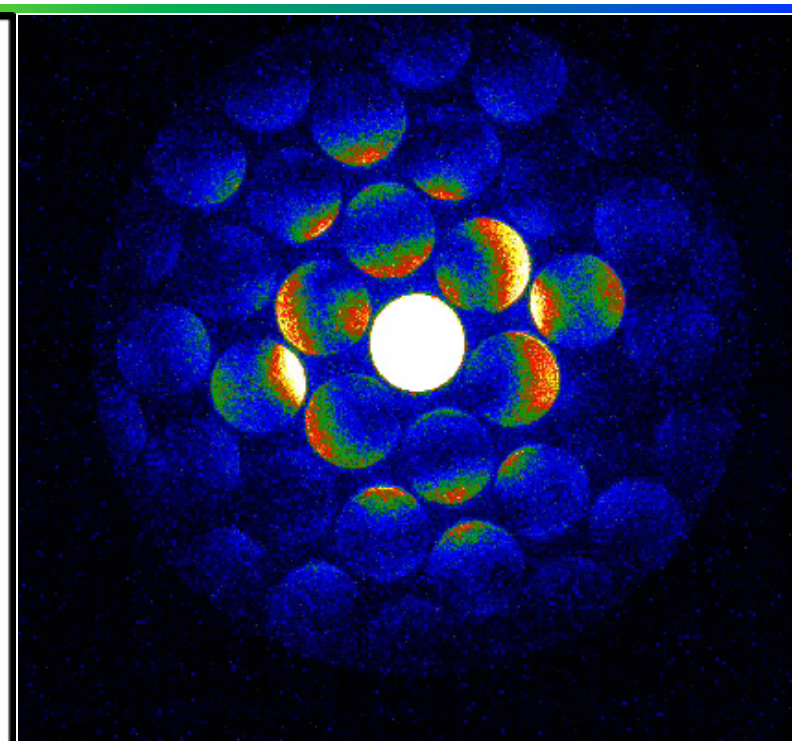
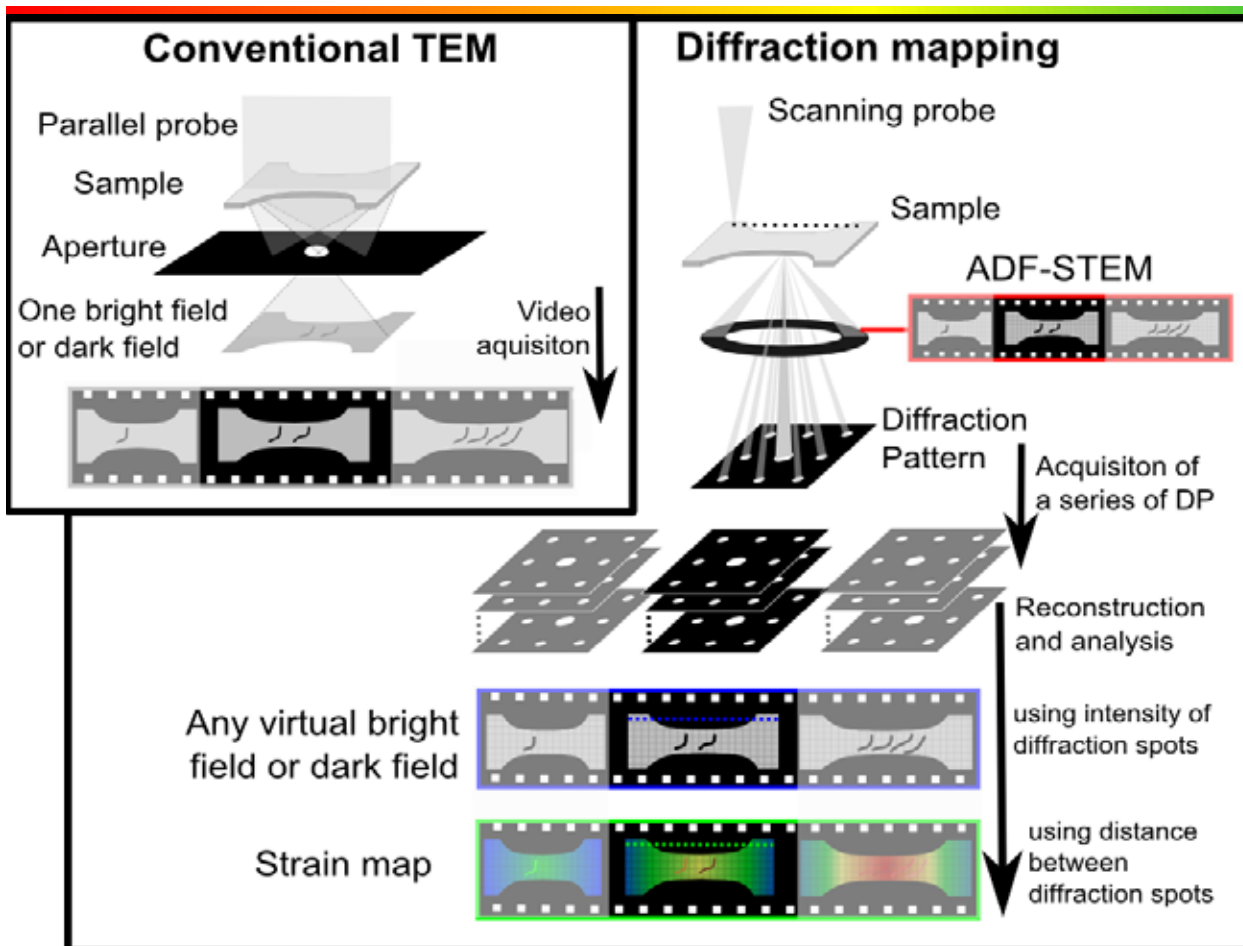


Simply recording an amplitude discards much useful information

$N(x,y)$



Nano-Beam Electron Diffraction Strain Mapping

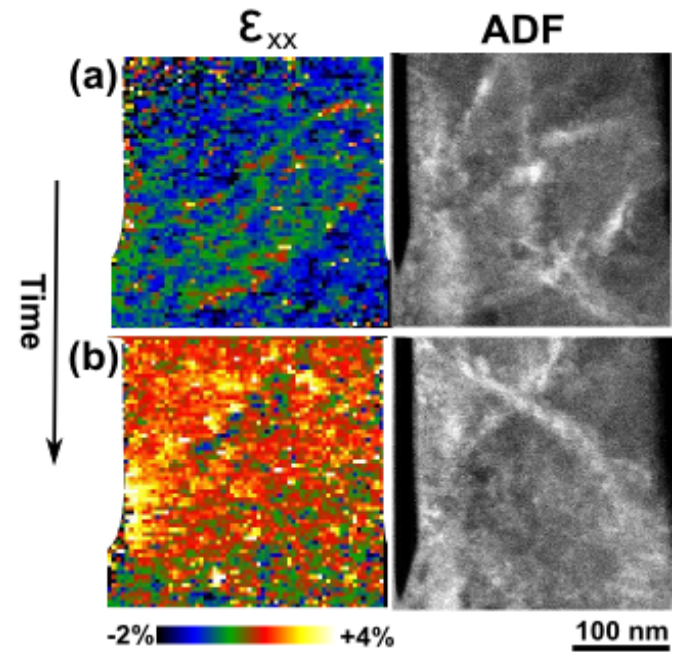


TEAM I – 300 keV
Gatan K2-IS direct detection camera
frame rate: 400 f/s

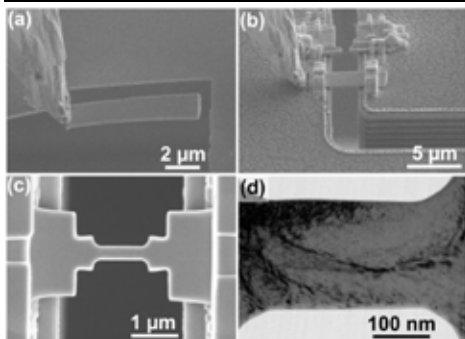
256p x 256p
1.2 nm step size
65,500 diffraction patterns
each 2k x 2k

B. Ozdol, C. Gammer, C. Ophus, P. Ercius, J. Ciston, and A. Minor, *APL* (2015)

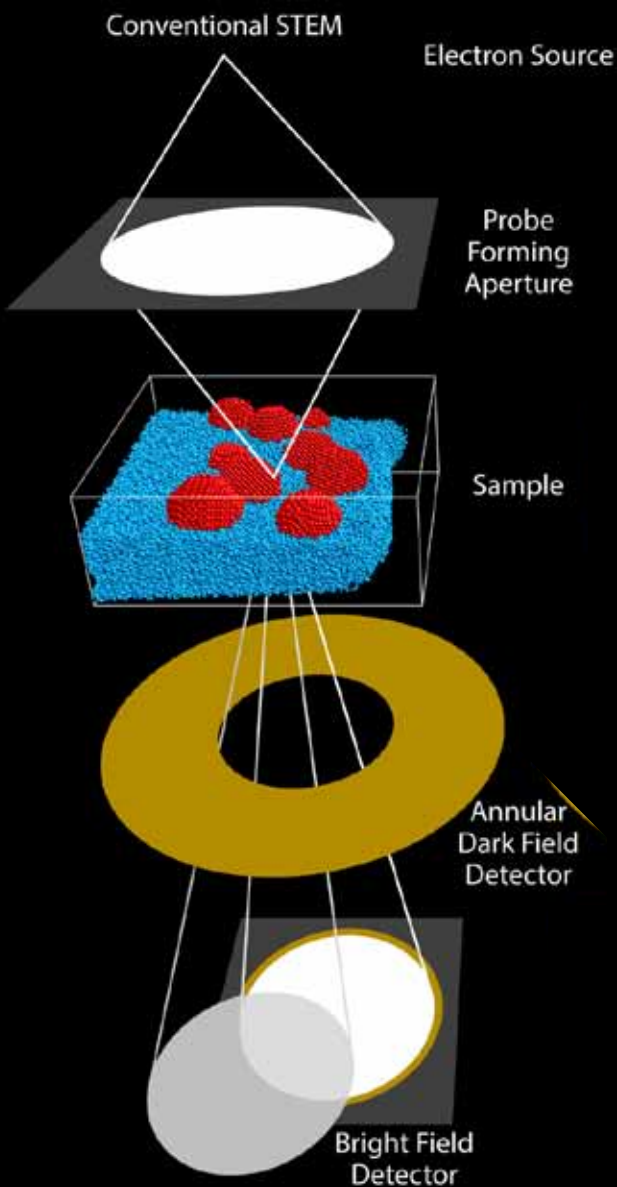
In situ Nano-Beam Strain Mapping of Al-Mg



Mapping of local strain during deformation

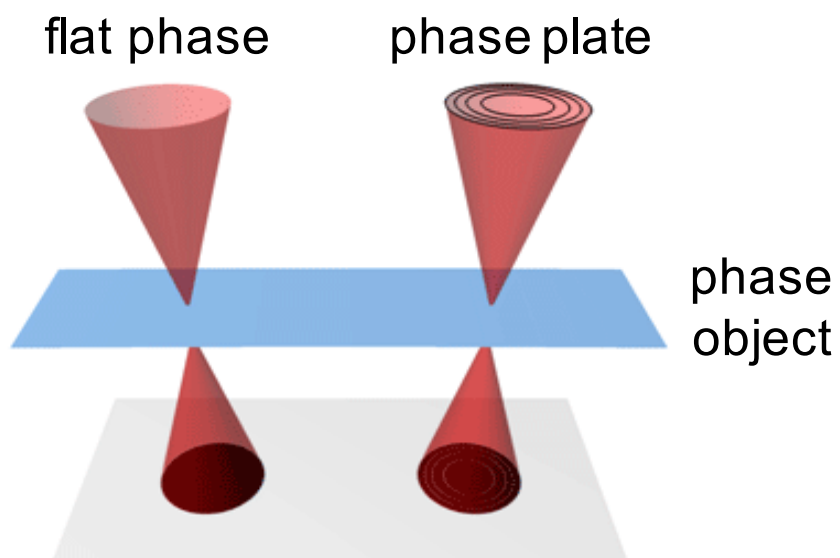


C. Gammer, J. Kacher, J. Ciston, C Czarnik, O.L. Warren, A.M. Minor, submitted



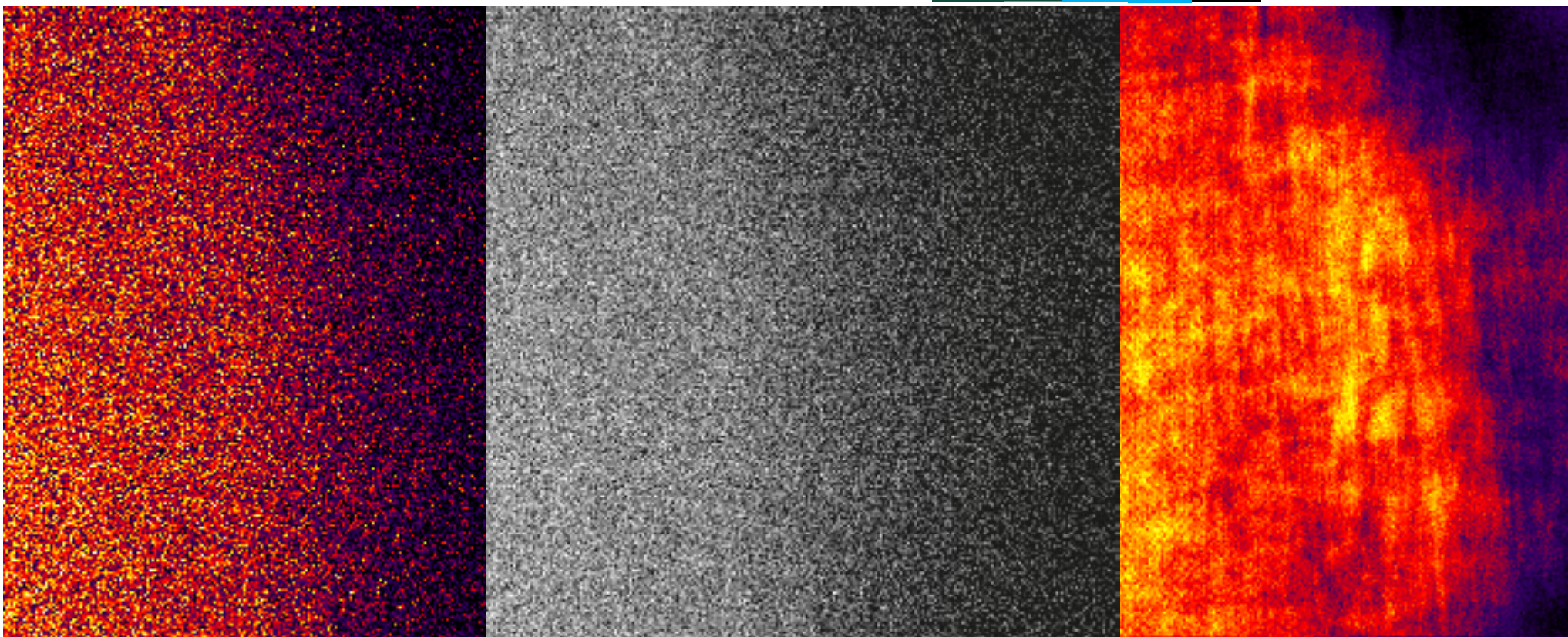
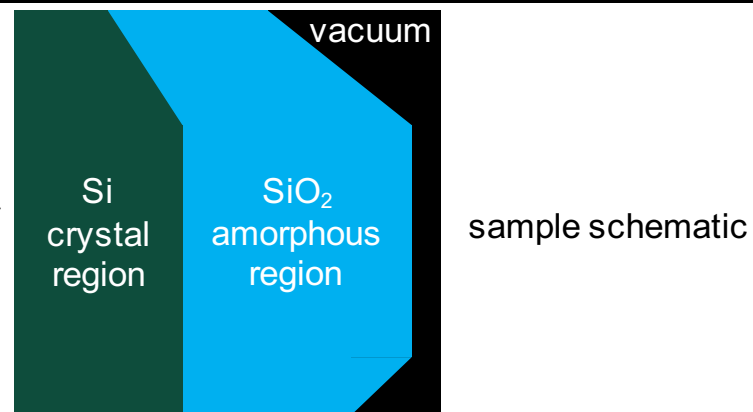
Matched Illumination and Detector Interferometry (MIDI)-STEM.

This name refers to how a virtual detector is “matched” to the phase structure imparted to the illumination, allows interferometric phase contrast.



Addition of a phase plate to the probe generates linear interference.

Low dose application of MIDI-STEM to wedge sample, containing stronger scattering and thicker region, as well as thin, weakly-scattering region:

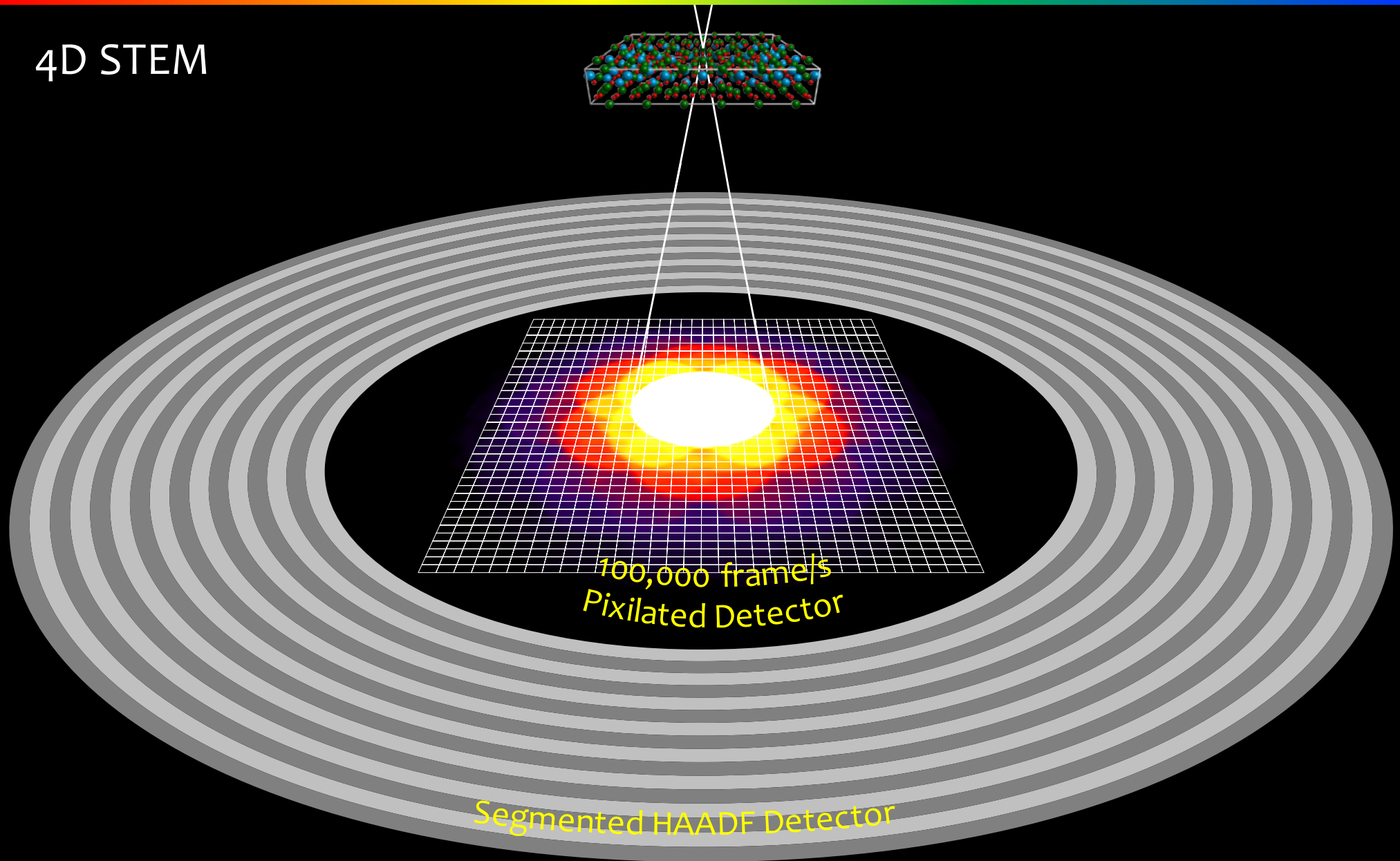


Grand Challenge

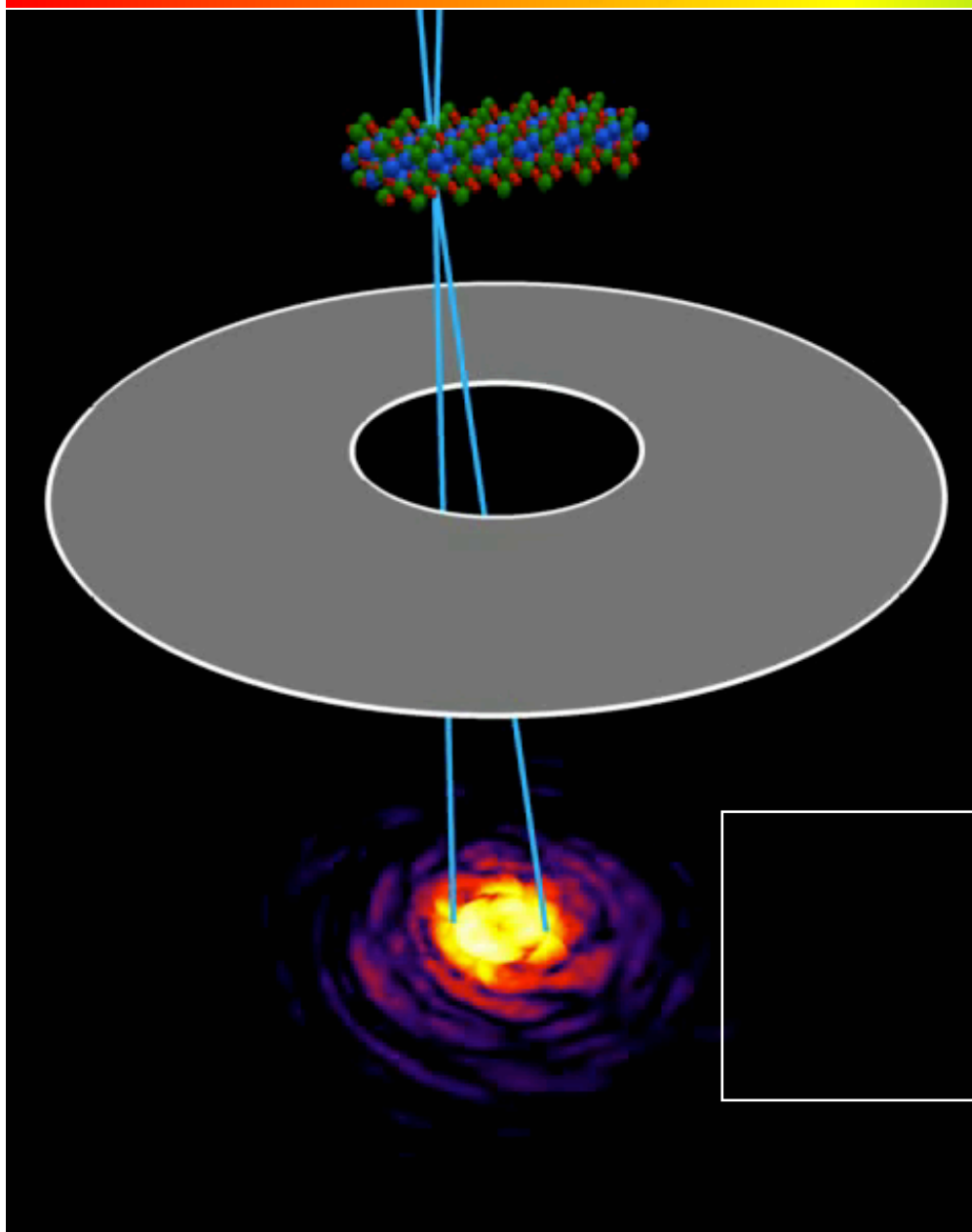
- “4D” STEM
 - 2 spatial dimensions + 2 reciprocal dimensions
 - *A diffraction pattern at every probe point*
 - **Software-driven (post hoc) experiments**
- **Data rates approach 1 Tb/s**
- **Ready in 2017**
 - A concrete science driver for a “data superfacility”
 - A prototype for LCLS-II
 - We are developing LCLS-II Soft X-ray detectors
 - 4D STEM forces us to solve problems now
 - Learn how to co-design for the future

100,000 frame/second Detector

4D STEM

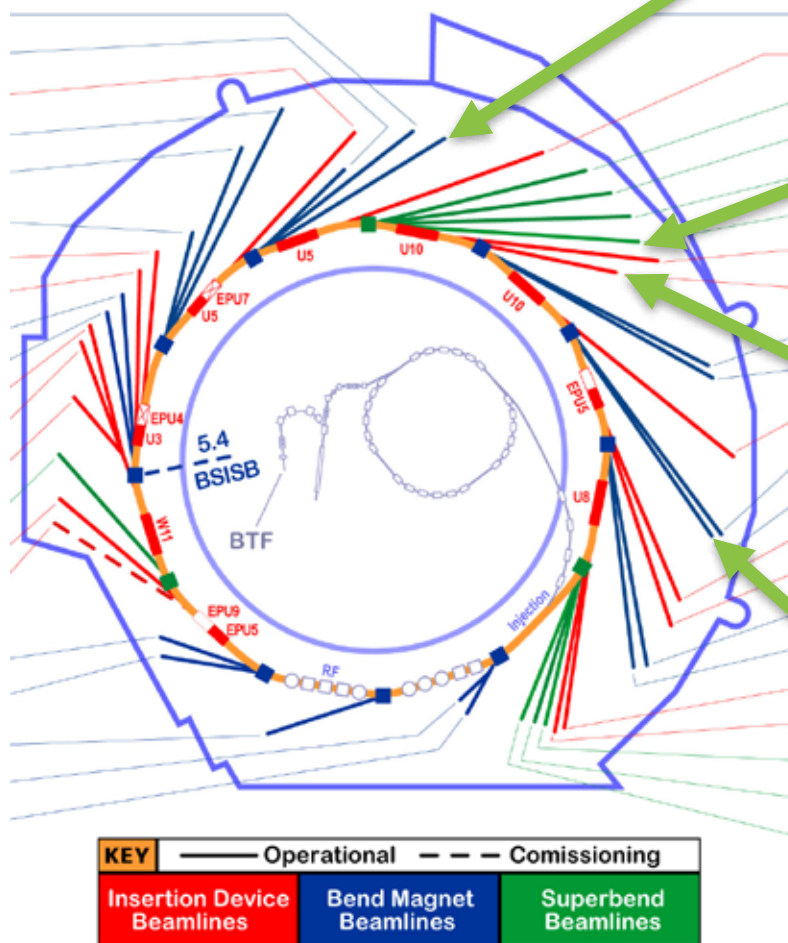


Phased Approach



- 400 Gb/s direct to Cori
 - Data in memory, not disk
 - Phase 0 – low speed
 - Learn how to deal with data
 - Phase 1 – full speed
 - Data reduction algorithm development
 - Phase 2 – FPGA processing

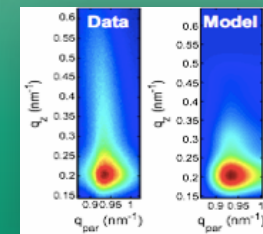
ALS



Reciprocal Space (Scattering)

(Scattering)

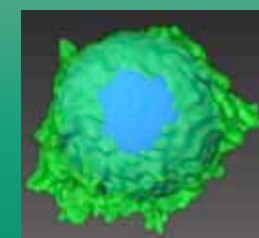
HipGISAXS/HipMC
parallel Scattering



Real Space (Tomography)

(Tomography)

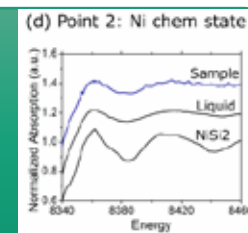
Arec3d, QuantCT,
CrunchFlow



Spectroscopy (MicroXas)

(MicroXas)

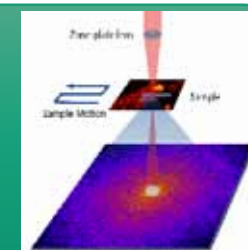
ShirleyXAS (MSD)
BerkeleyGW (NERSC)



Hybrid (COSMIC)

(COSMIC)

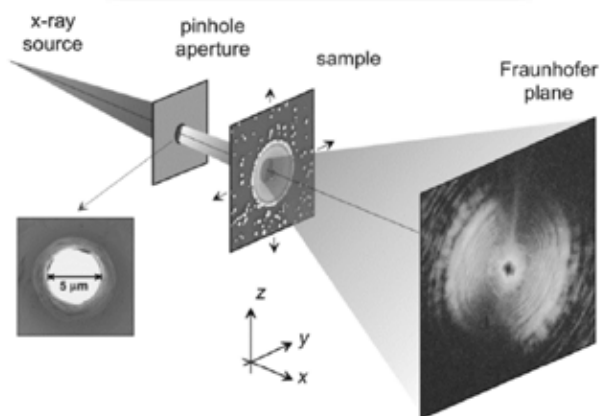
Ptychographic
reconstruction



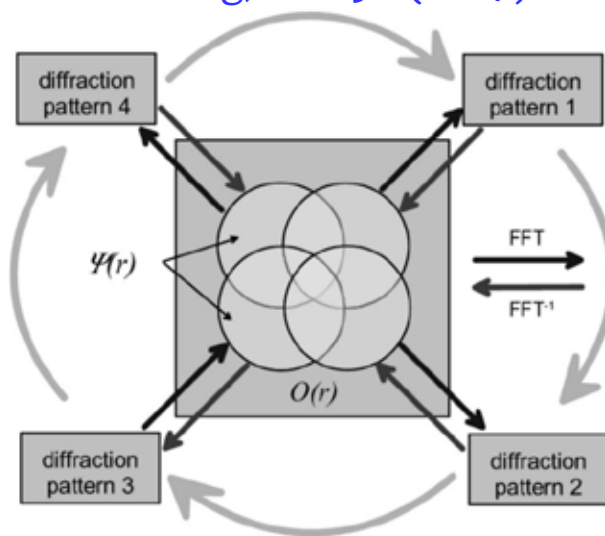
D. Parkinson, LBNL

Ptychography at ALS

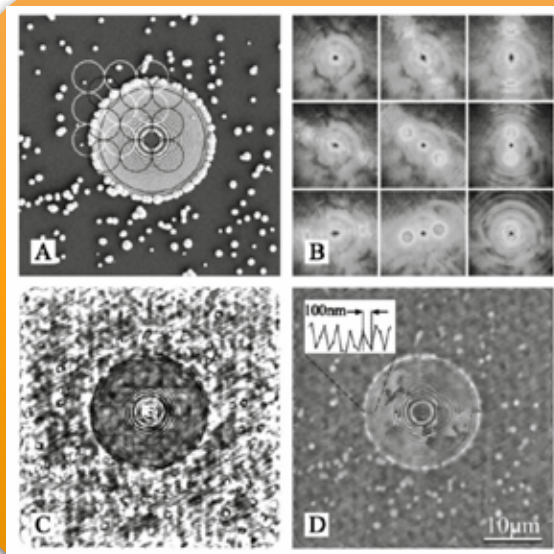
STXM (like STEM)



Rodenburg, PRL 98 (2007)

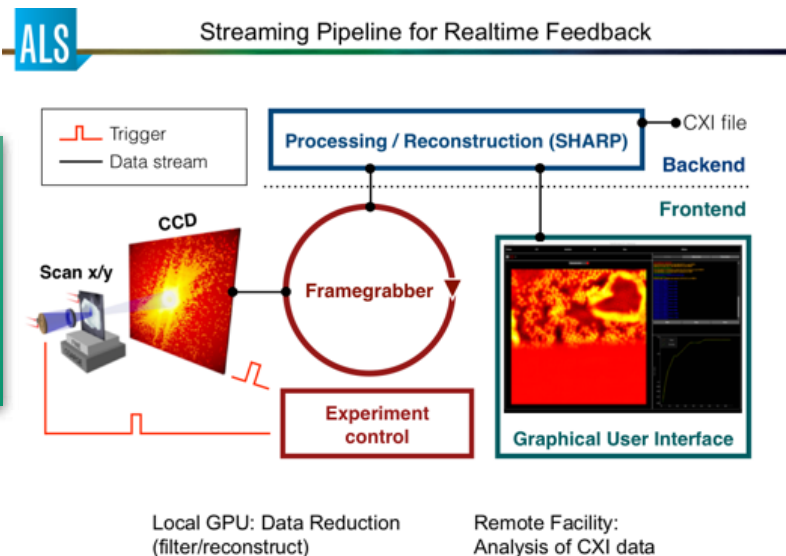


D. Shapiro, LBNL



- Ptychographical Iterative Engine algorithm (PIE)– serial reconstruction
- Non-crystalline, extended sample
- No optics

Like 4D STEM
(a) Data → NERSC
(b) Algorithm
(data reduction)

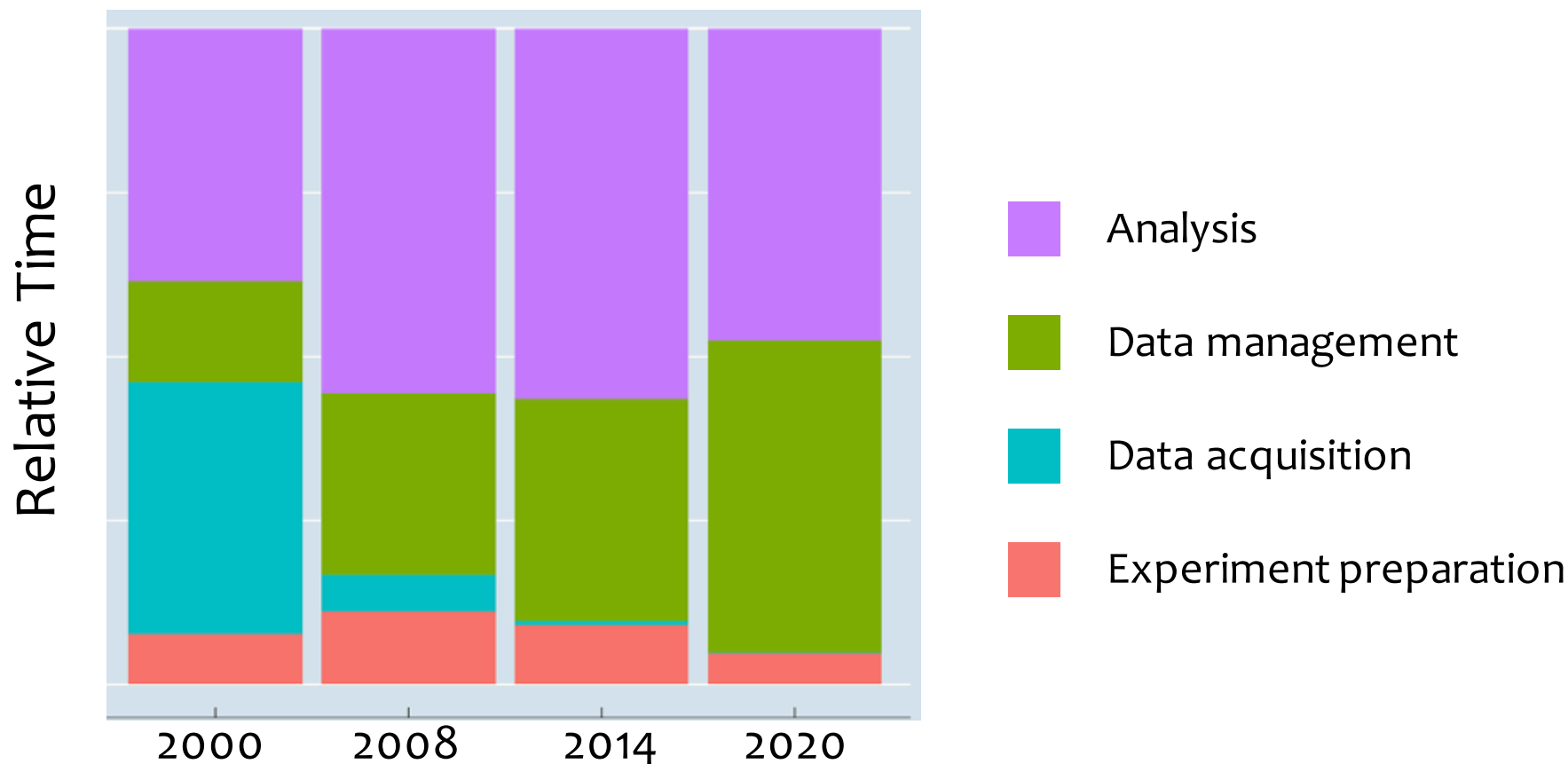


Local GPU: Data Reduction
(filter/reconstruct)

Remote Facility:
Analysis of CXI data

More Data → More Computing

Experimental Time Breakdown

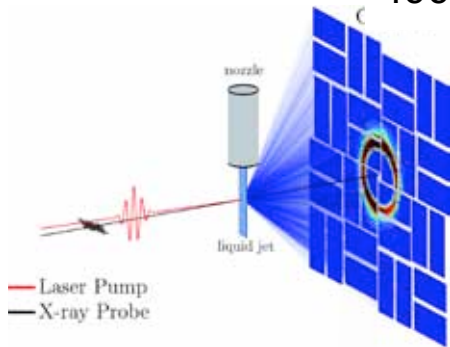


Thanks to D. Parkinson - Figure from Kevin Mader <https://rawgit.com/4Quant/SRI2015/master/SRIPres.html#/>

A “Super facility” for Photons



Beamline:
100TB bursts



ESnet @ 100Gb/s

HPC Data
Analysis

Cori Supercomputer:
500K cores, 28 PB disk

HPSS

NERSC
off-line copy

Local
Data
Analysis

HPSS

LCLS
online copy

- Today: combine ESnet's bandwidth, Cori Phase I data-centric HPC, LCLS sponsored storage, and HPSS archive

- Tomorrow: LCLS-II 10N x LCLS-I (need faster and more)



David Skinner, LBNL



BRAIN - Link structure to function across scales

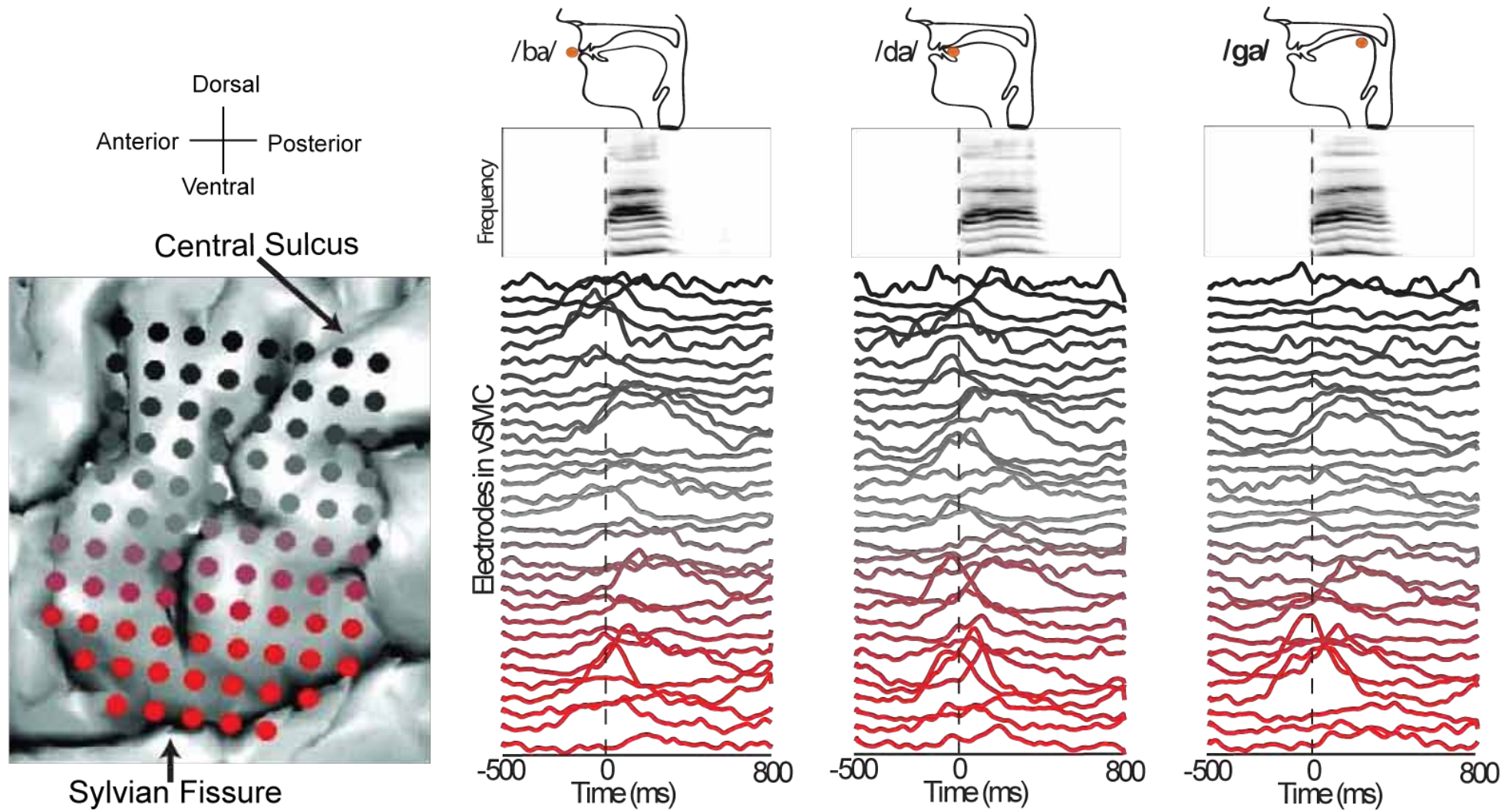
- Disparate spatio-temporal scales
 - nm to m
 - ps to years
- Diverse data types
 - genomics to (functional/structural) connectomics
 - electrical, optical and other measurements
 - behavior, sensory stimuli, perturbations
- Complex analysis issues
 - fusion of multi-modal data
 - inference robust to noise
 - provide insight into computations
 - statistical prediction of future events
 - extraction of important features



K. Bouchard, LBNL

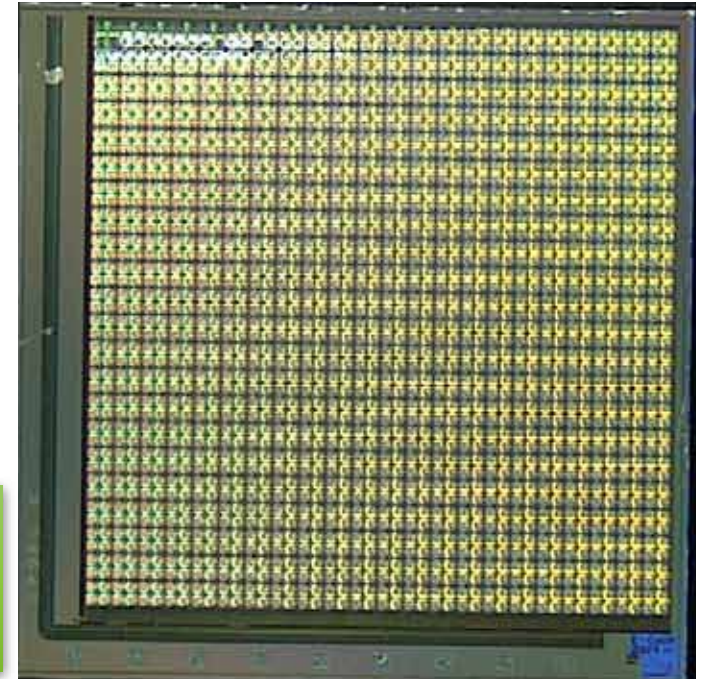
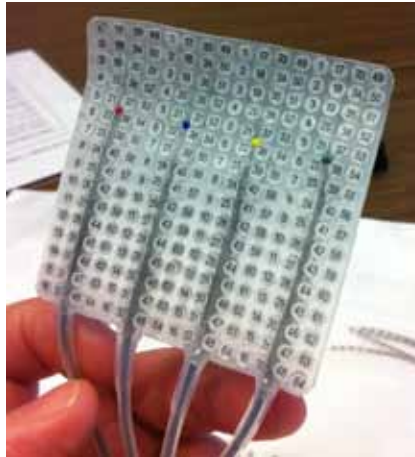
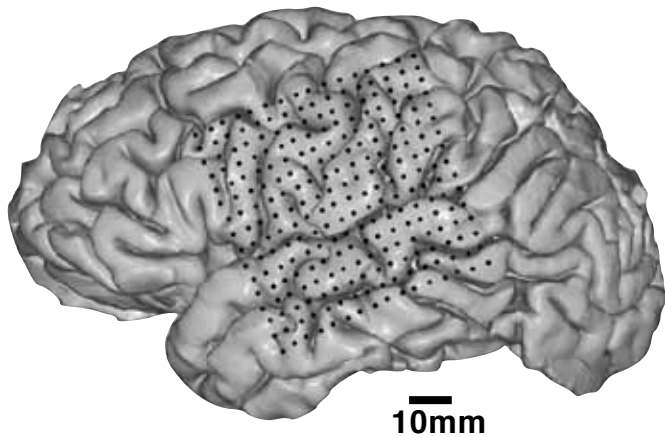
Human Brain Project, 2012

Brain Functions are Generated by Distributed Spatiotemporal Patterns of Cortical Activity



K. Bouchard, LBNL

Very-High Density Electrocorticography (ECoG)



State-of-the-art today: 4 electrodes / cm²
1 electrode / 1 wire



Detector Group



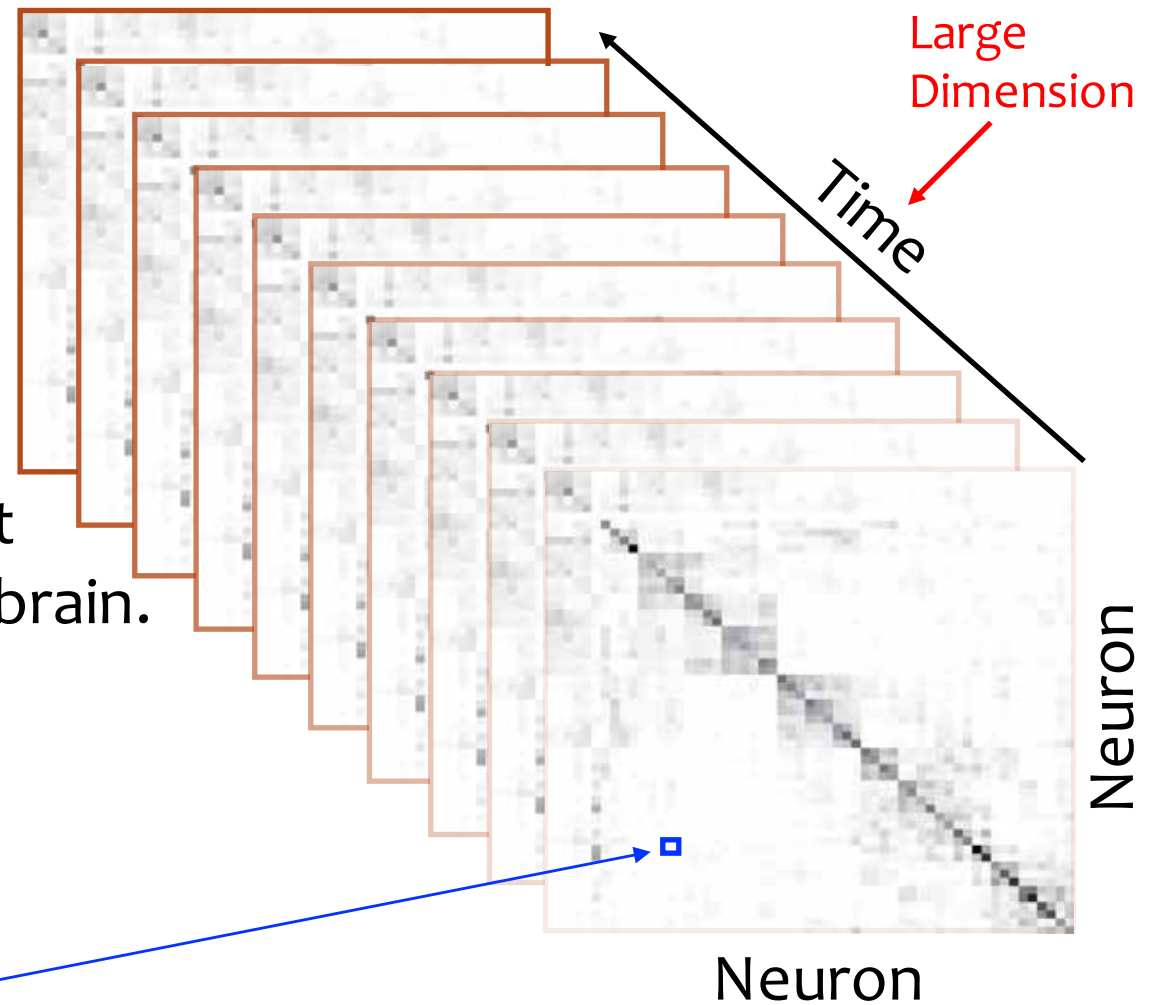
Processing

LBNL e-chip: 10,000 electrodes / cm²
1,000 electrodes / 1 wire

10^2 electrodes at 10^3 Hz $\rightarrow$ 10^6 electrodes at 10^4 Hz
Increase 10^5 in data rate

The Functional Connectome: the Structure of Brain Functions

A weighted, direct graph describing the dynamic, casual interactions amongst neurons in the functioning brain.



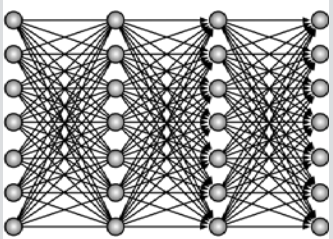
(e.g.) Each edge is **estimated** from data as partial correlation coefficient using **linear regression**.

Electrons, Photons, Neutrons

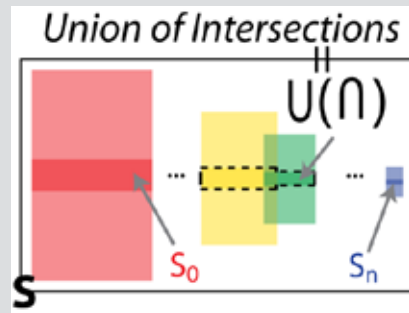
A template for data-intensive science



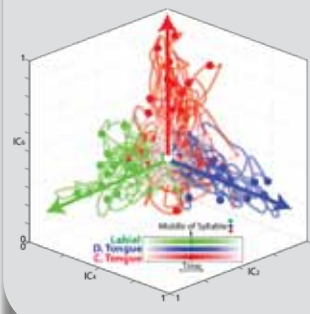
Deep Networks
for Decoding



Improved Machine
Learning Methods



Sparse
Signals



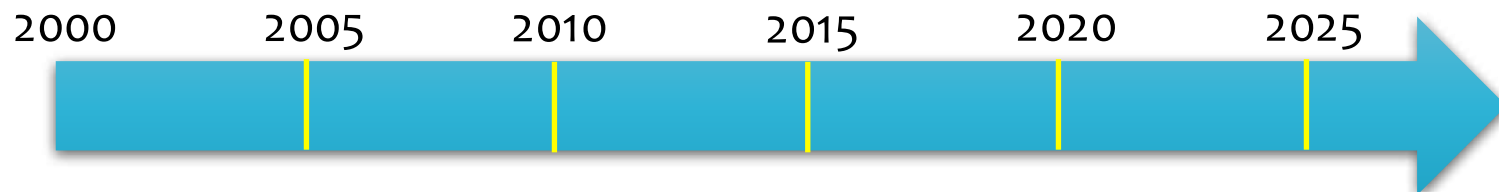
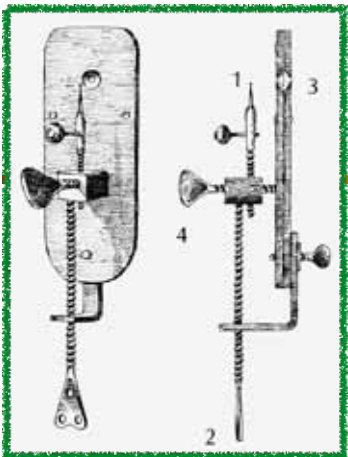
Standardized
Data Models



Hosting Data
Repositories



Data Rates / Volumes:



Data Rate [B/s]

e

10^6

10^9

10^{10}

10^{11}

TEAM Detector

K2

4D STEM

γ

10^6

10^7

10^8

10^{11}

Hybrid
Pixels

FastCCD

VeryFastCCD

n

10^6

10^9

10^{11}

e-chip

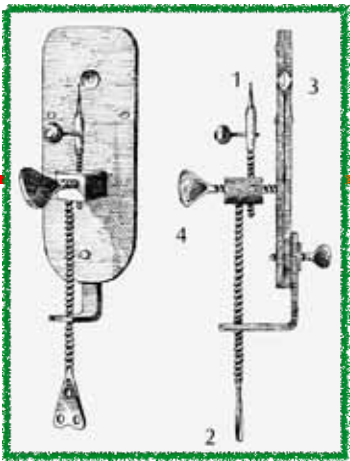
Data sets are seconds to months

Data rates are: Today: 0.1 Tb/s – 2017: Tb/s – 2020: 10 Tb/s

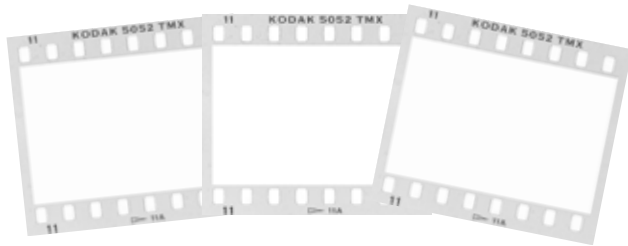
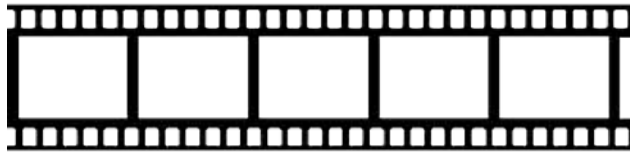
An admittedly parochial viewpoint

Analytics

	Astronomy	Cosmology	Climate	Systems Biology	Neuroscience	BioImaging	Mass-spec Imaging	Personalized Toxicology	Materials	Particle Physics	EM/X-ray Microscopy
Classification	X		X		X	X	X			X	X
Regression								X	X		
Clustering		X	X		X		X			X	X
Dimensionality Reduction			X		X		X				
Inference	X						X			X	
Model Estimation	X				X			X			
Design of Experiments		X	X						X		
Semantic Analysis			X	X					X		
Feature Learning			X		X		X	X	X	X	X
Anomaly Detection	X		X							X	X



Similarities / Differences



● Electrons / Photons

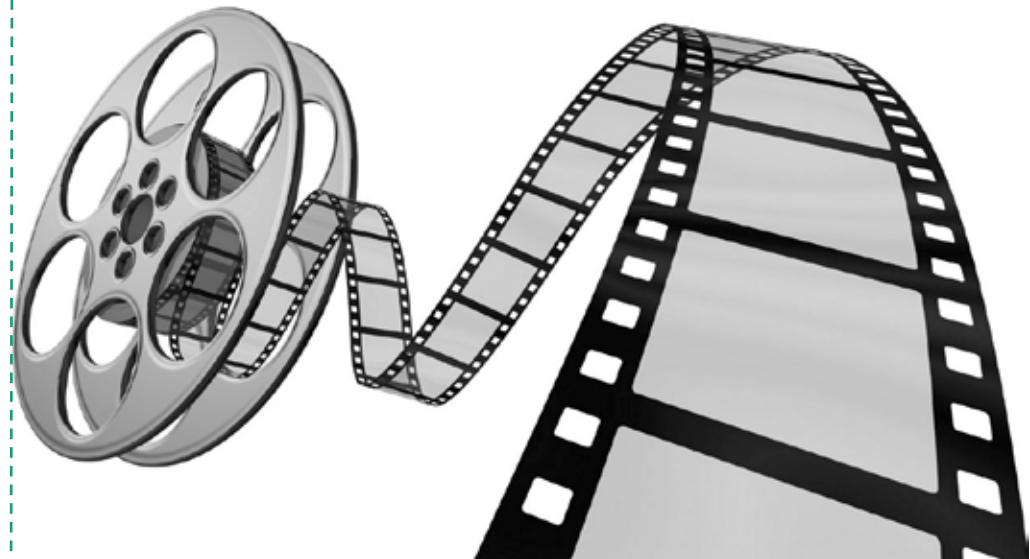
- Today: 10s Gpixel / s
- Data reduction (?)

● Frames

- Independent (XFEL)
- Δ (movie)

● Neurons

- Yesterday: 100s pixel / s
- ~Today: ~Mpixel/s
- Tomorrow: ~Gpixel/s
- Frames: complex $f(t)$
 - Data reduction (?)



Common Themes

- Data Rates ↑: (e.g. 4D STEM – saturate 400 Gb/s link)
- A priori data reduction technique often unknown
 - Initially need to “look at” all the data
 - Move, analyze, store
- Then move only reduced data
- User data (of one form or another)
- All need some variant of “data superfacility”

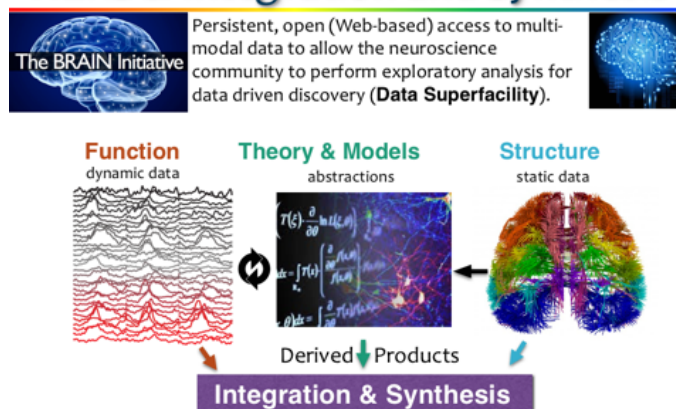


Connect the building blocks to create a “Super-Facility”

- Facilities (e.g. synchrotrons)
 - Advanced Light Source
- Network
 - LBLnet, Etnet, etc. give 10Gb+
- Computers
 - DTN's, GPU clusters, HPC centers
- Applied Math, Algorithms, Software
 - CAMERA (led by Jamie Sethian, Mathematics Group)
- Data management, workflow, visualization
 - SPOT Suite (led by Craig Tull, Computational Research Division)



NERSC: Integration and Synthesis



Working with Image Data Sets

- Upload and download files
- Work with Big Data sets from remote locations
- Monitor a job, check work flow
- Get user account information
- Authenticate using user credential
- User dependent rights
 - System resources
 - Files
 - Batch jobs
 - Shell command
 - Status

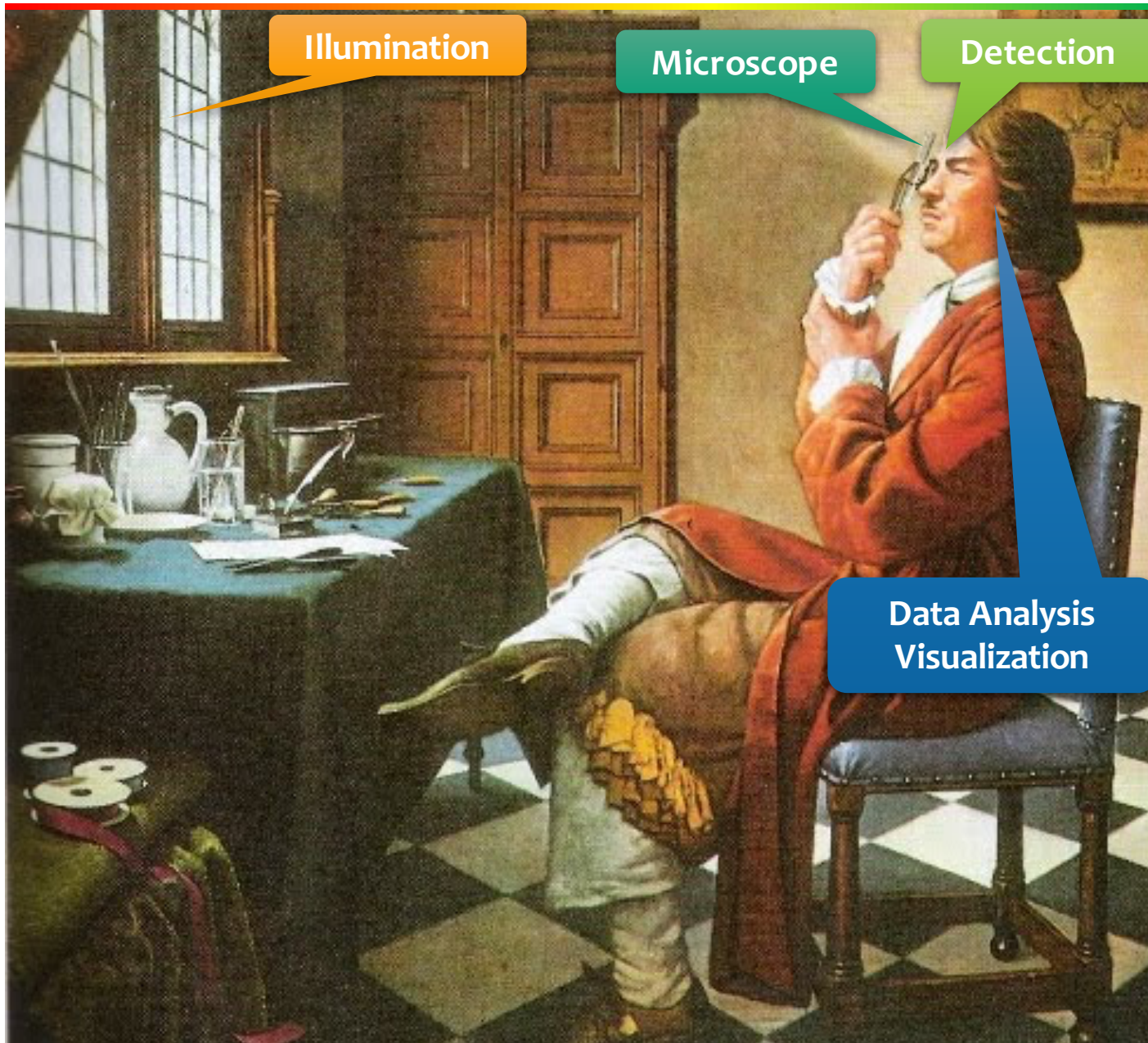
- What we want to do:
- Eliminate challenges with data transfer
 - Share data with users
 - Convert Data sets
 - Analyse data sets
 - Data sharing
 - Data management
 - Reduce latency
- Challenges to efficient data processing
- High temporal and spatial resolution
 - Large Data Sets that require sophisticated storage schemes

E. Stach, CFN-BNL

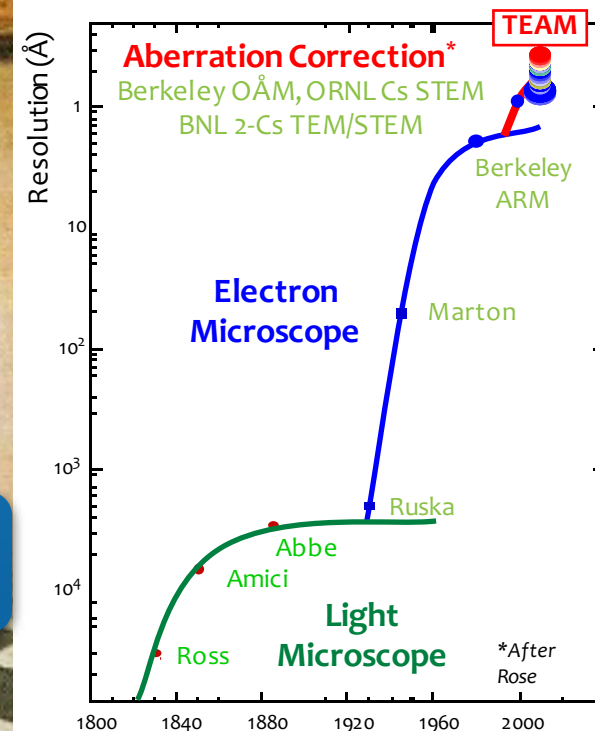
P. Denes @ NERSC User Group @ 20160322 @ Slide 36



Modern Microscopy



- 200 years of microscopy



- 20 years of detectors

- Computing?