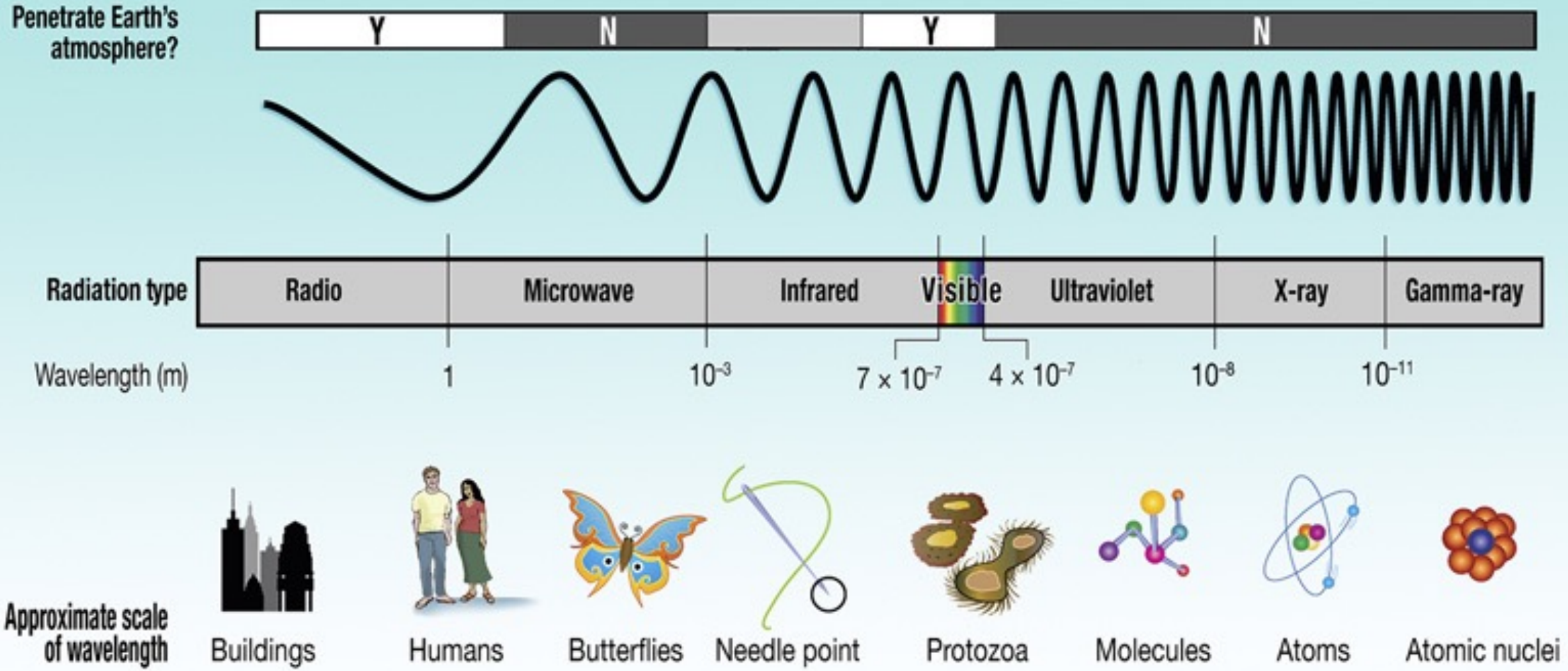
Workflows for X-ray micro-CT at the ALS + NERSC “Superfacility”

Dula Parkinson

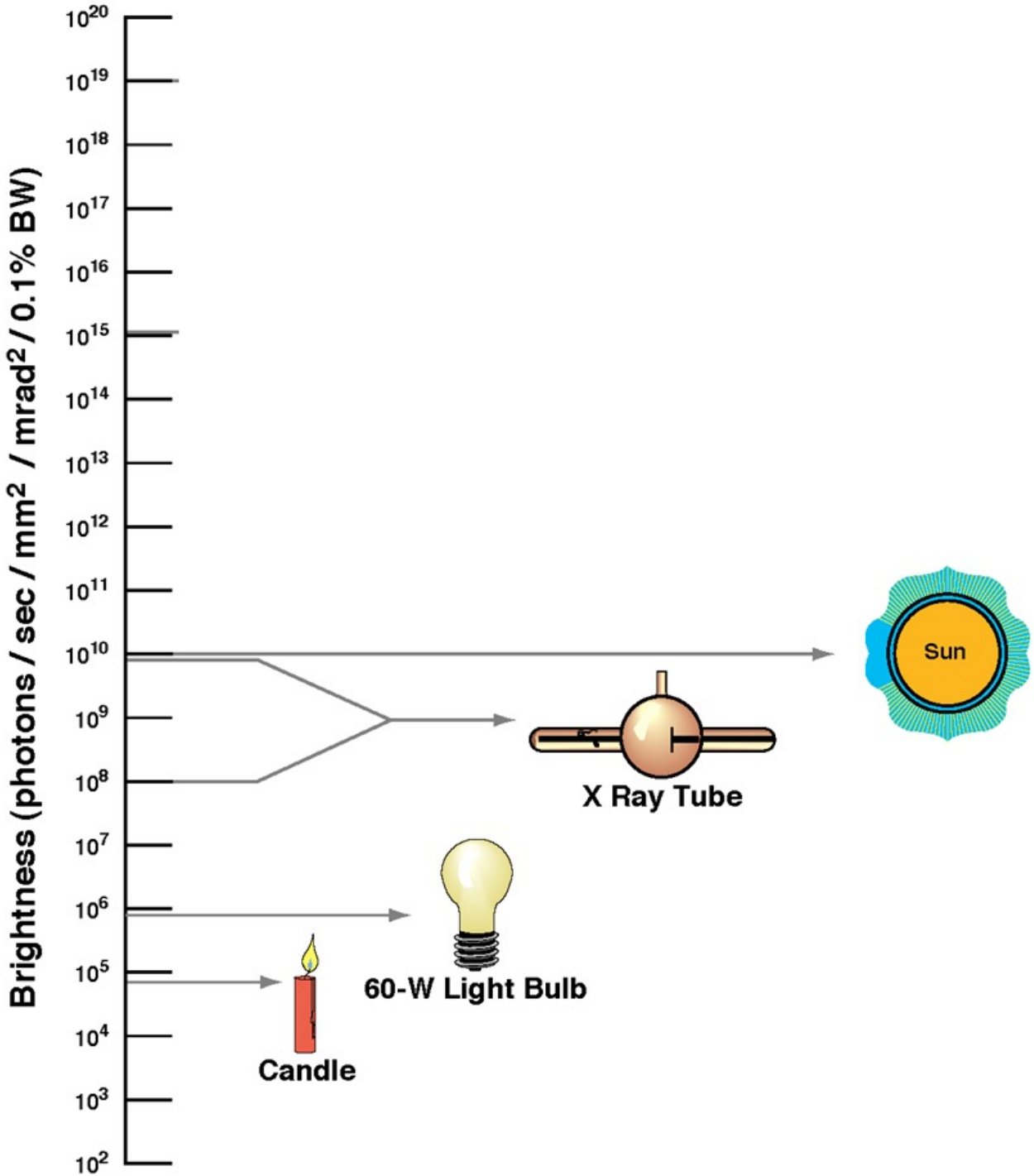
May 2023

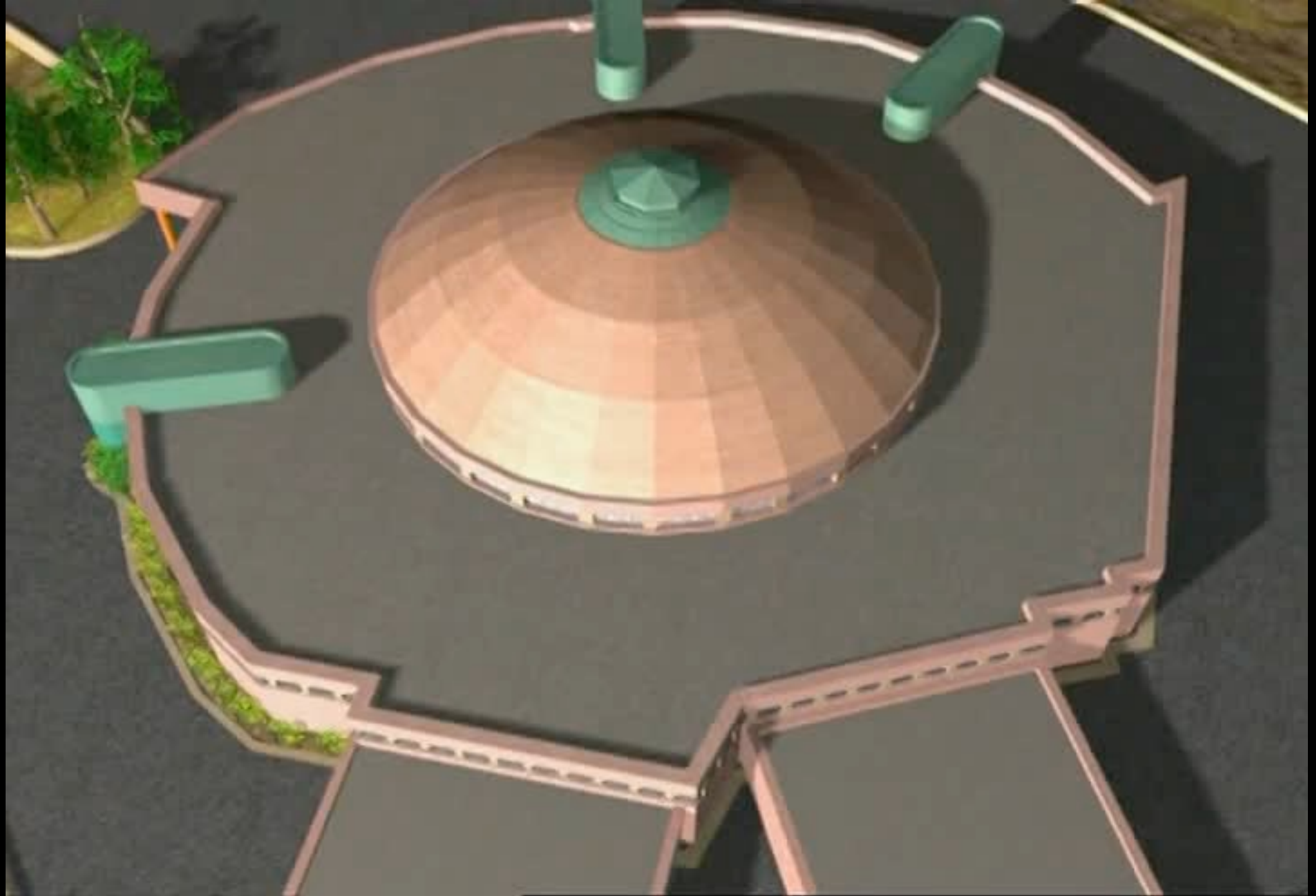
What is the Advanced Light Source (ALS)


Penetrate Earth's atmosphere?
Y
N
Y
N
Radiation type
Radio
Microwave
Infrared
Visible
Ultraviolet
X-ray
Gamma-ray
Wavelength (m)
1
10^{-3}
7×10^{-7}
4×10^{-7}
10^{-8}
10^{-11}
Approximate scale of wavelength
Buildings
Humans
Butterflies
Needle point
Protozoa
Molecules
Atoms
Atomic nuclei

X-rays are useful!

- 19 Nobel prizes have been awarded for X-ray-related work
  - 9 in chemistry (of which four used synchrotron X-rays)
  - 7 in physics
  - 3 in medicine


Brightness (photons / sec / mm² / mrad² / 0.1% BW)
10²⁰
10¹⁹
10¹⁸
10¹⁷
10¹⁶
10¹⁵
10¹⁴
10¹³
10¹²
10¹¹
10¹⁰
10⁹
10⁸
10⁷
10⁶
10⁵
10⁴
10³
10²
Sun
X Ray Tube
60-W Light Bulb
Candle



Synchrotron facilities across the world (around 50)

(Pic source: Google)

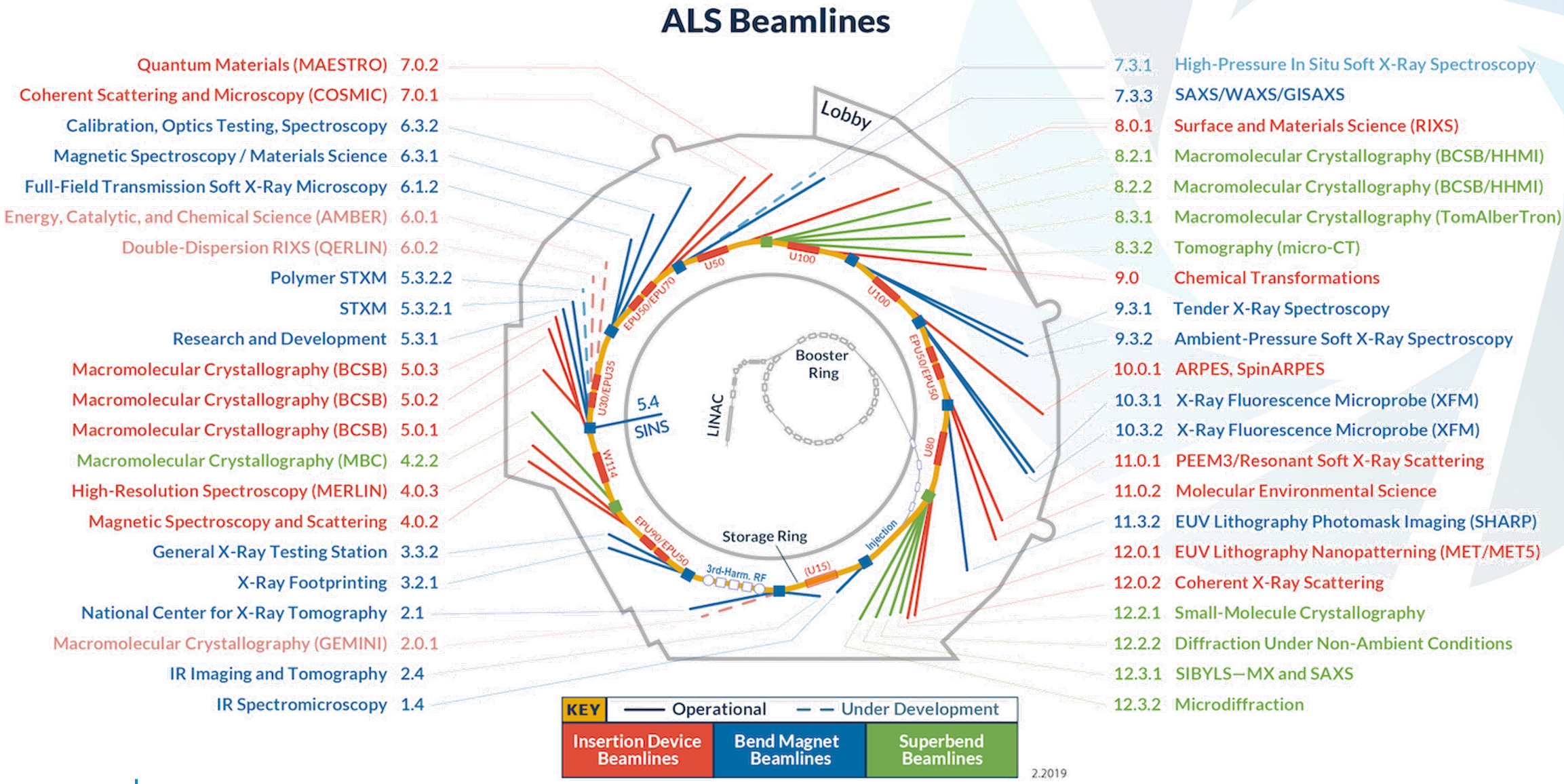
ADVANCED LIGHT SOURCE (ALS)

Department of Energy-funded synchrotron facility in Berkeley, CA

Lots of lead and concrete shielding help make it safe to work there!


CAPACITY
30
TONS
LOAD
LIFTER
AED
NOT
AN
EXIT

40+ beamlines serving about 2500 users/year



USER STATS

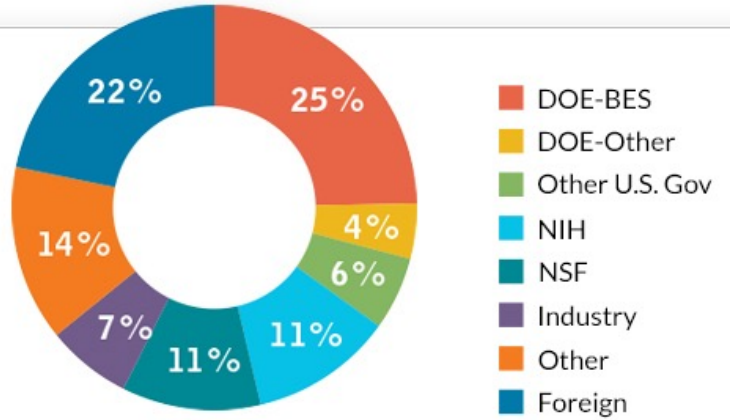
50-100

Users on site at any one time

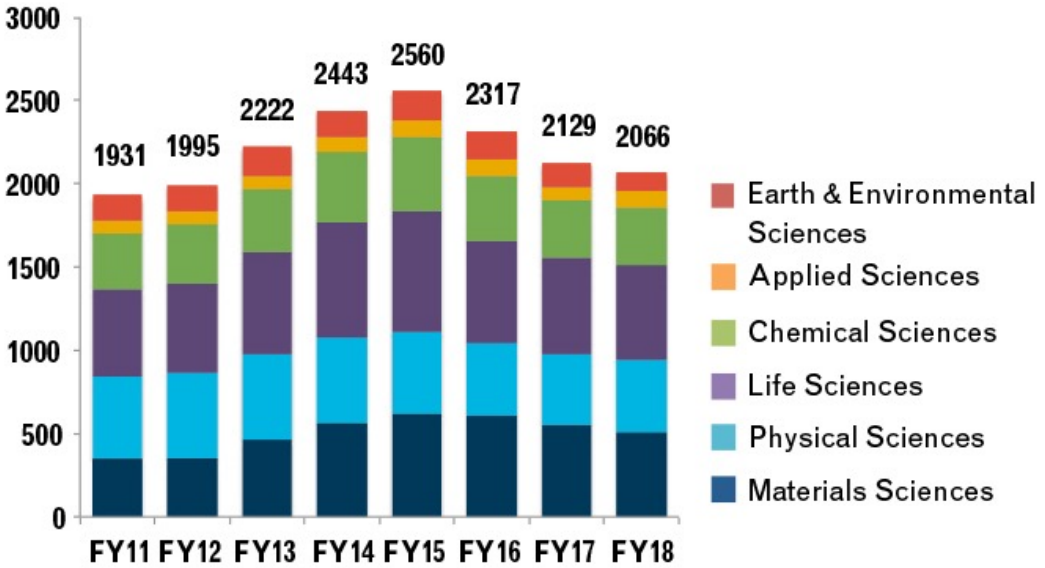
1 hour to 10 days

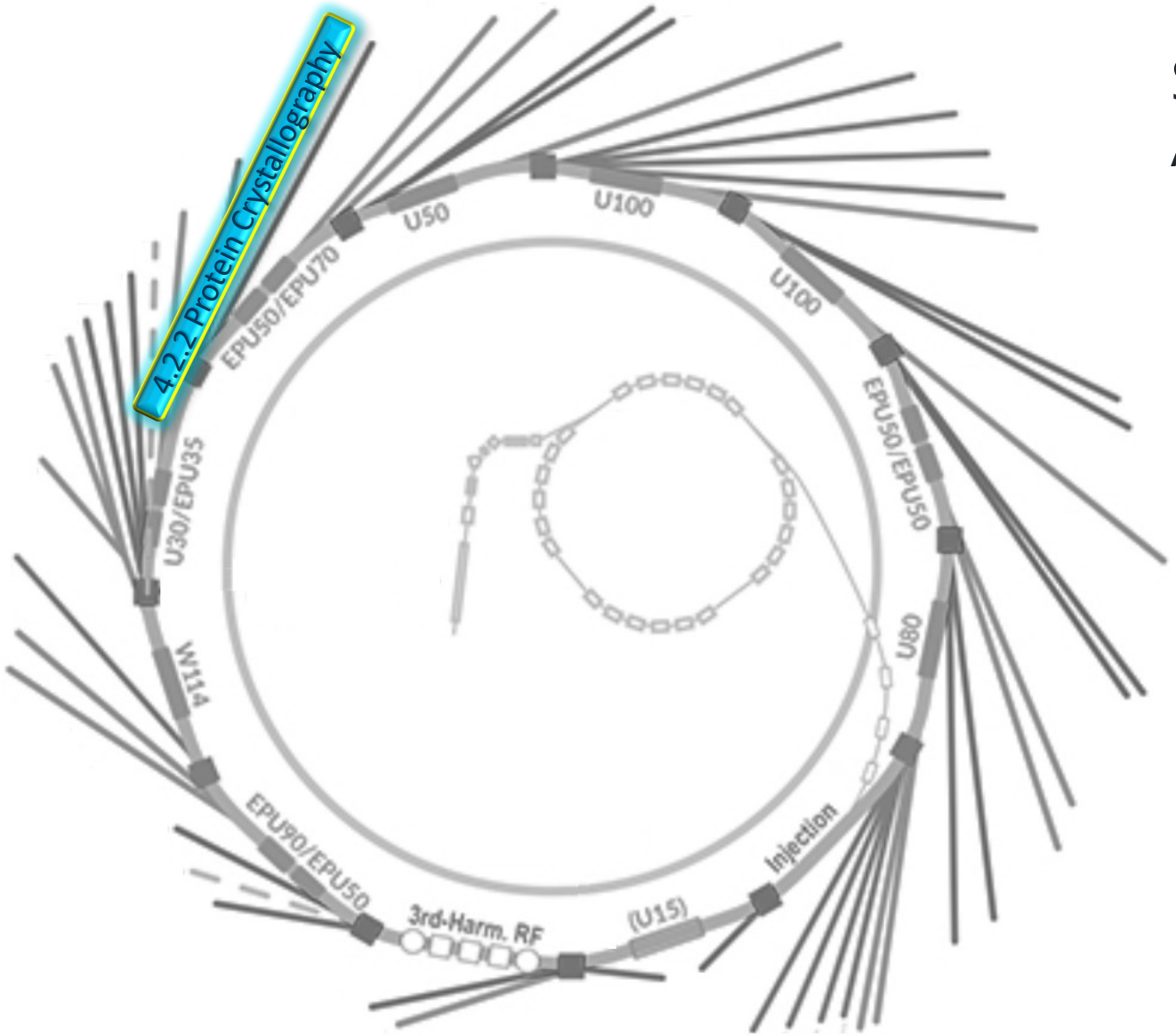
Typical stay of users

Users by Funding



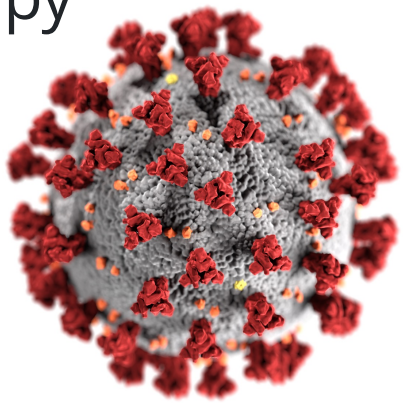
Users per Discipline per Fiscal Year

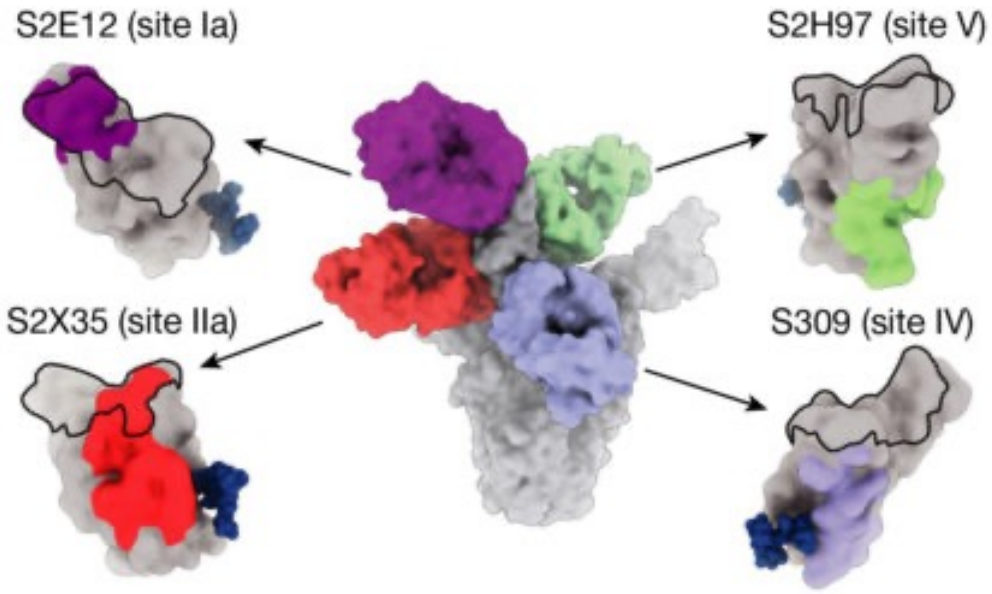




ADVANCED LIGHT SOURCE BEAM LINES

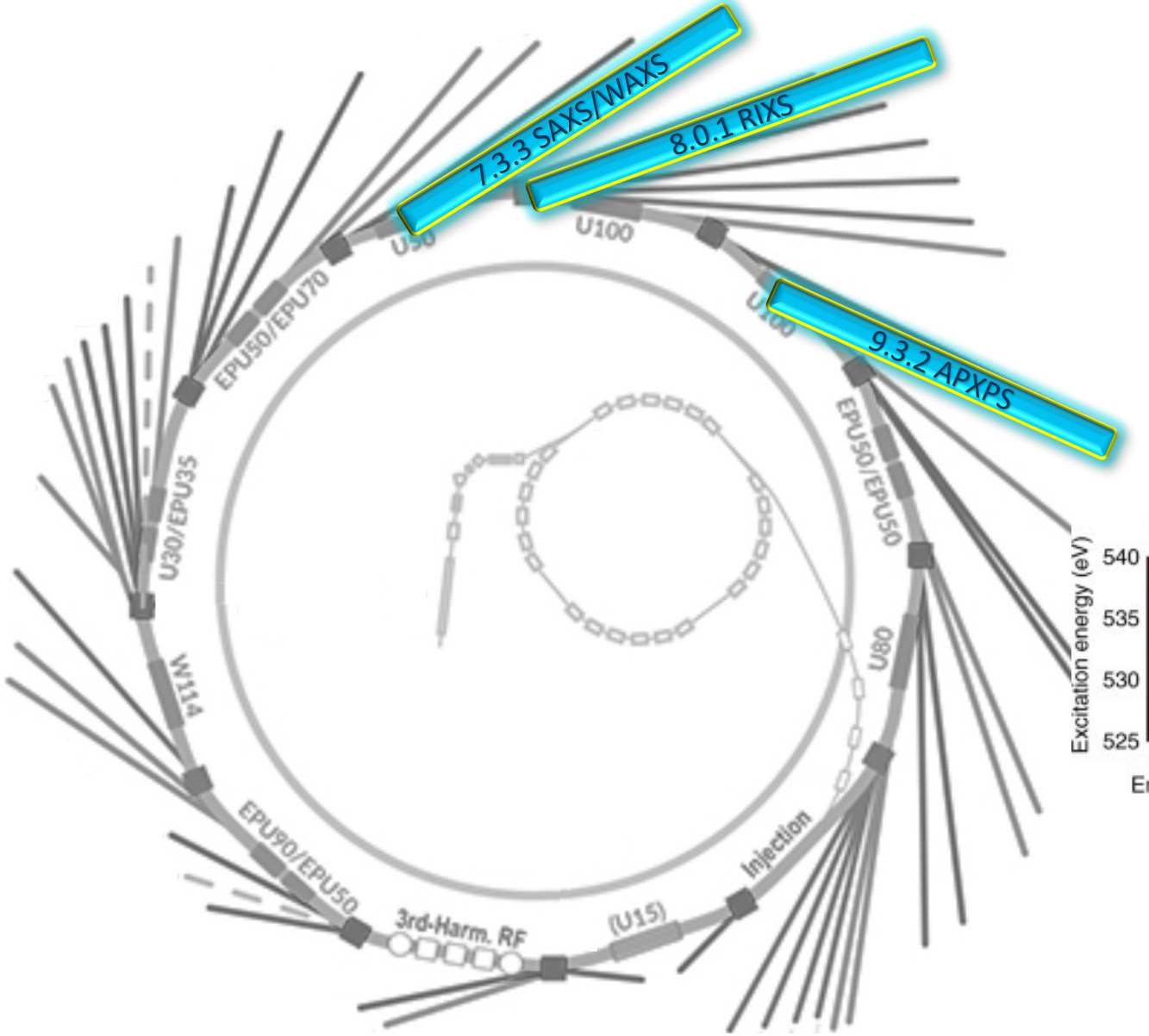
Structures for development of COVID Antibody Therapy





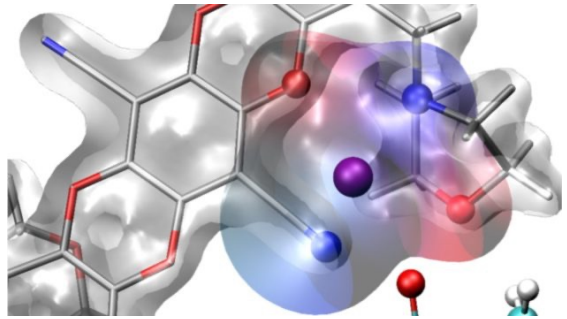
CDC/ Alissa Eckert, MSMI; Dan Higgins, MAMS

T.N. Starr, et al., "SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape," *Nature* **597**, 97 (2021

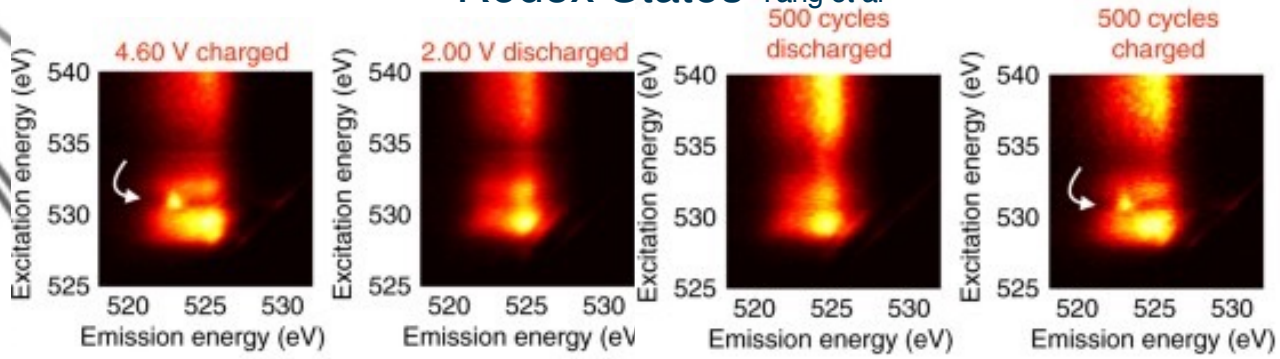


ADVANCED LIGHT SOURCE BEAM LINES

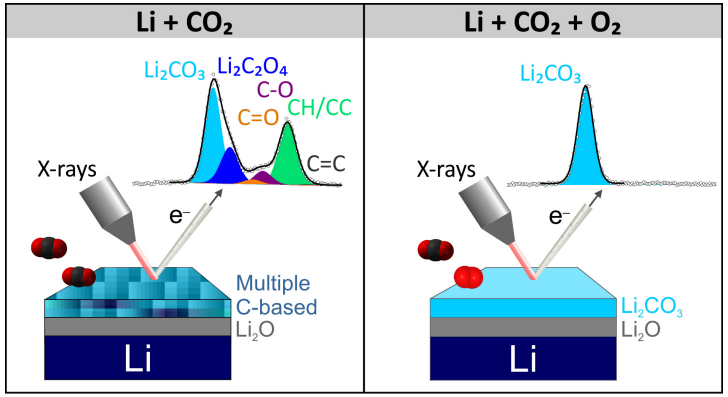
Selective membranes Helms et al

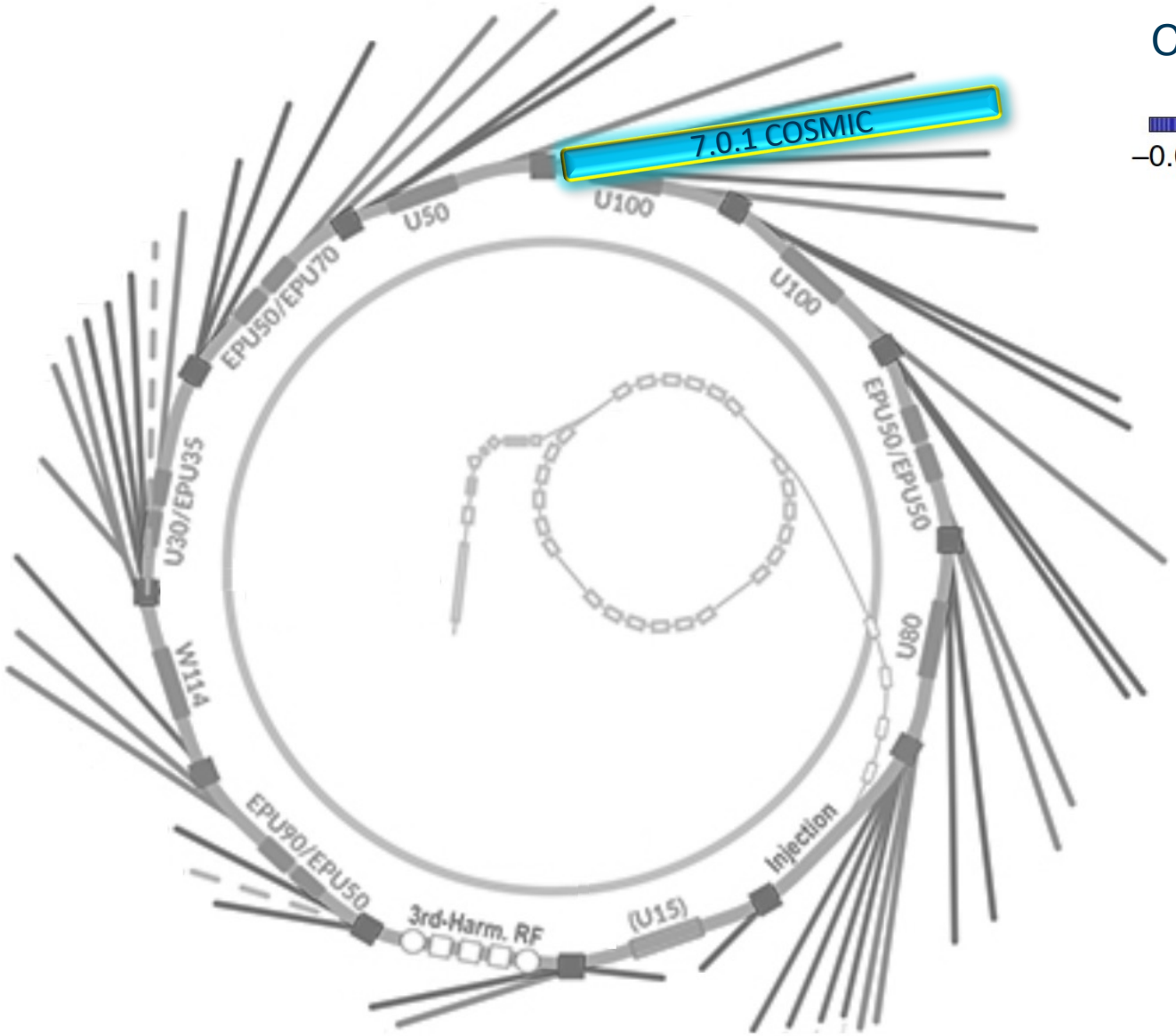


Redox States Yang et al



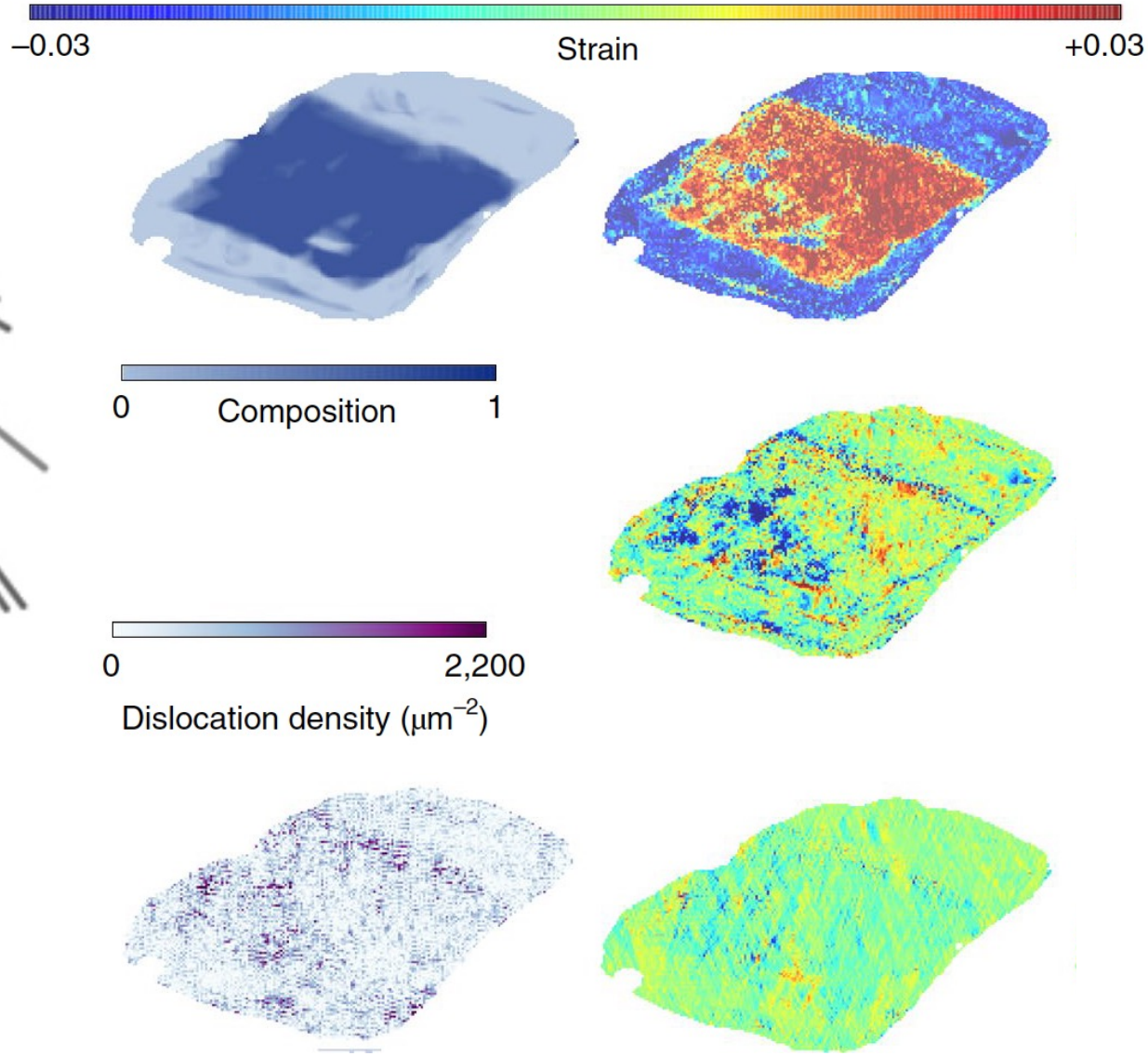
Lithium surface reactions Crumlin et al



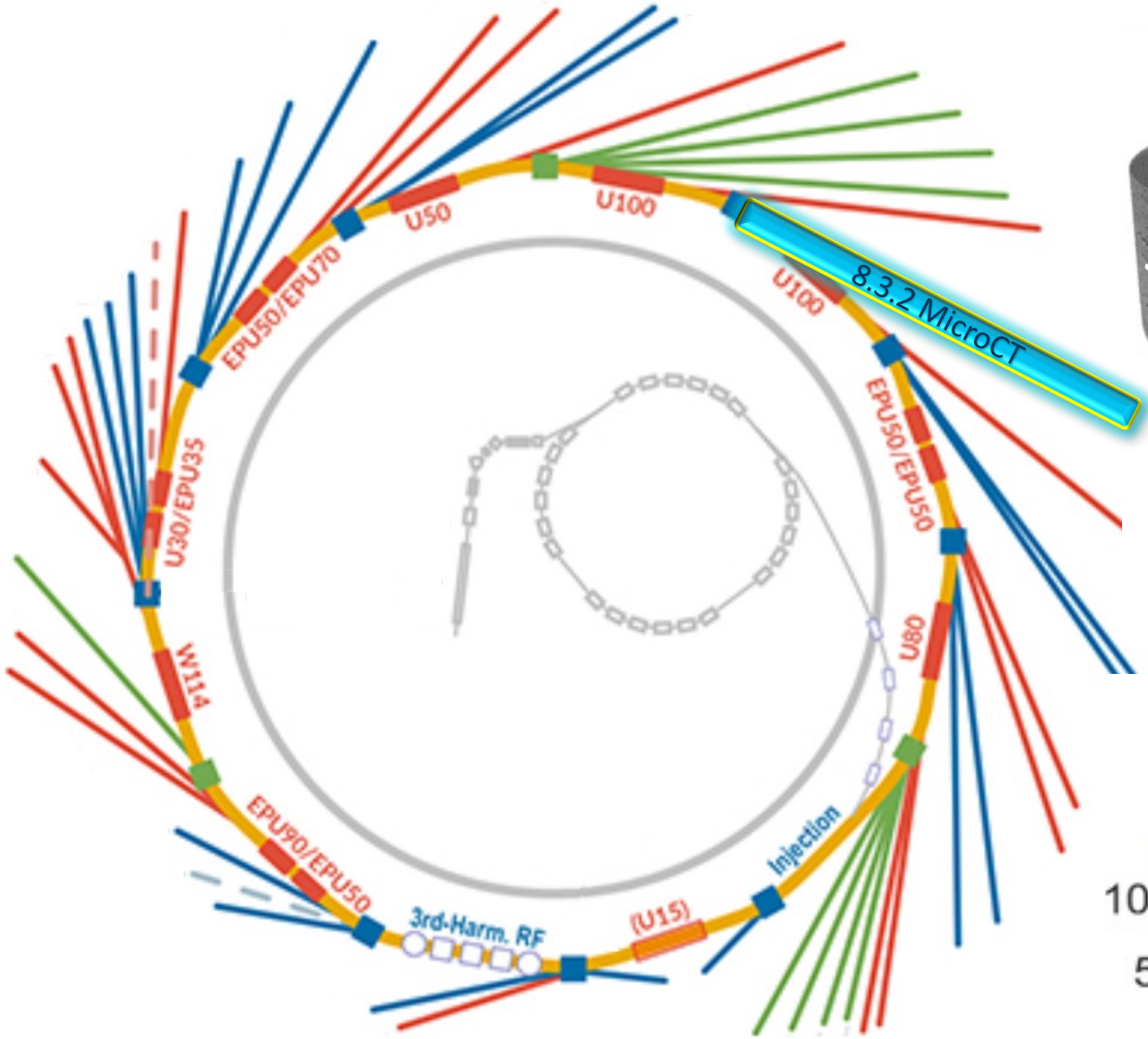


ADVANCED LIGHT SOURCE BEAM LINES

Composition, strain, and defects in cathode material

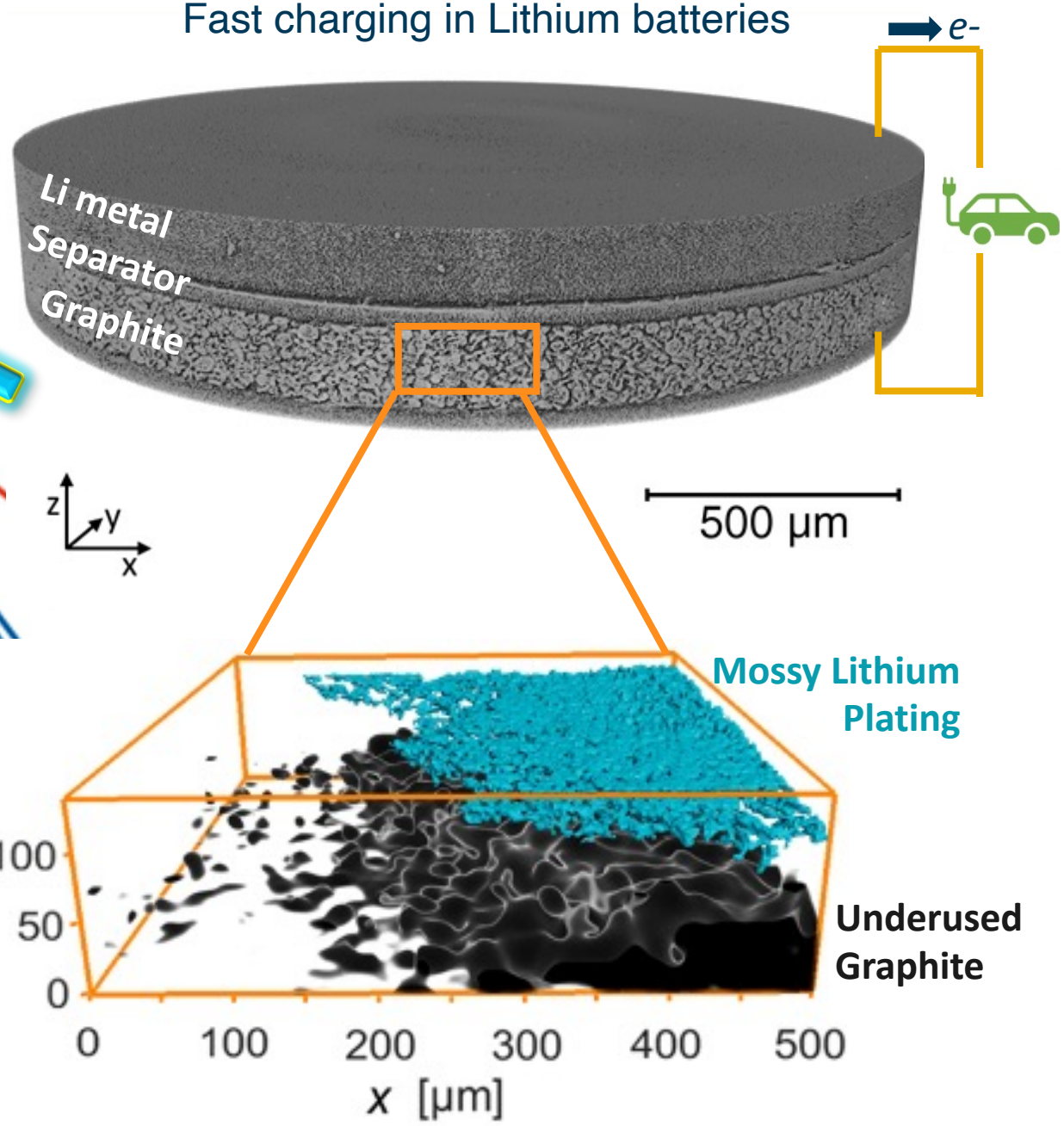


Chueh et al.

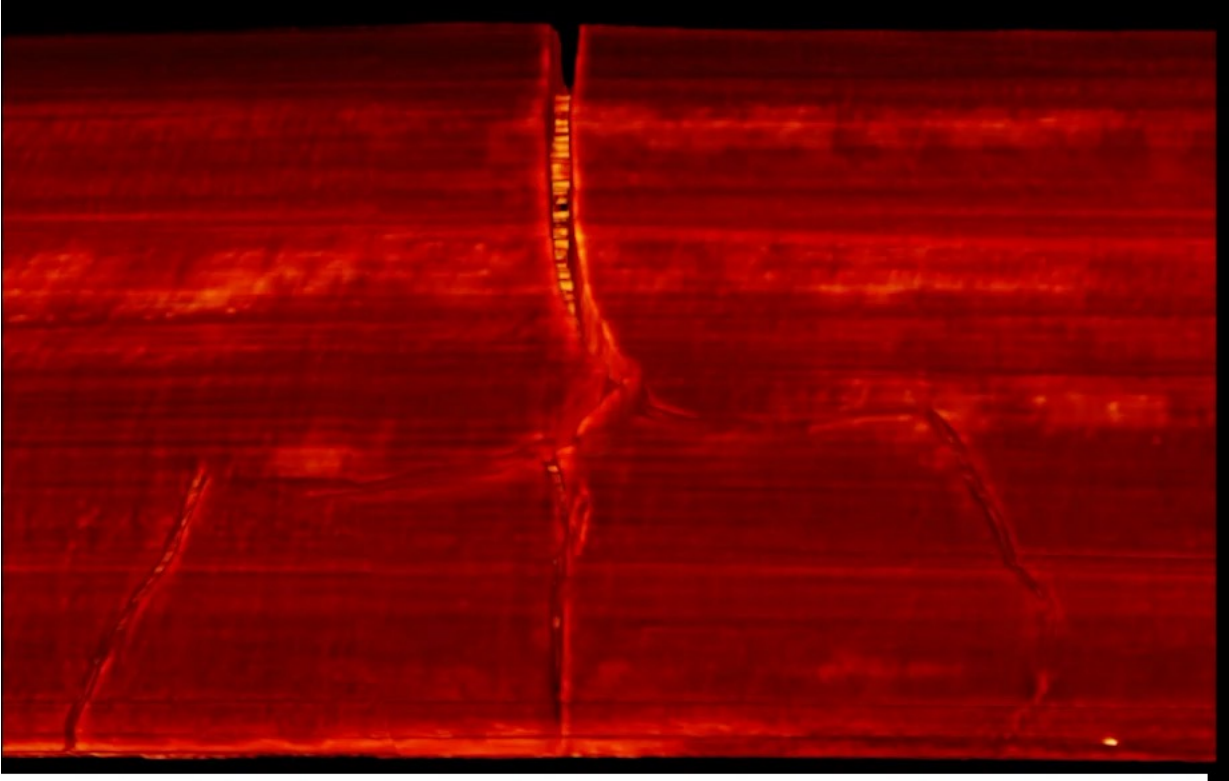


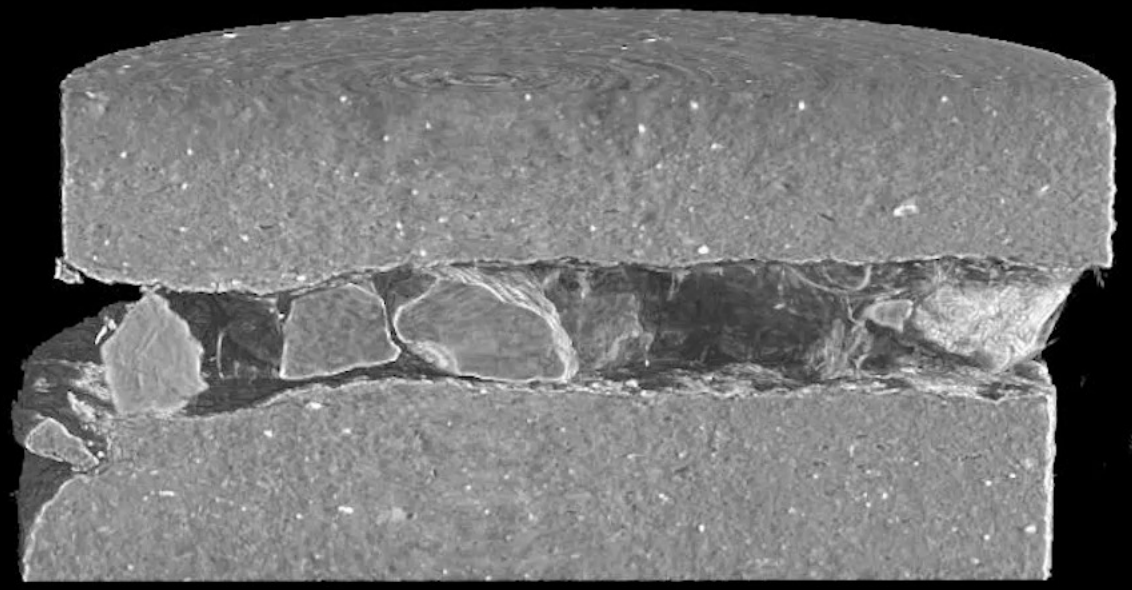
ADVANCED LIGHT SOURCE BEAM LINES

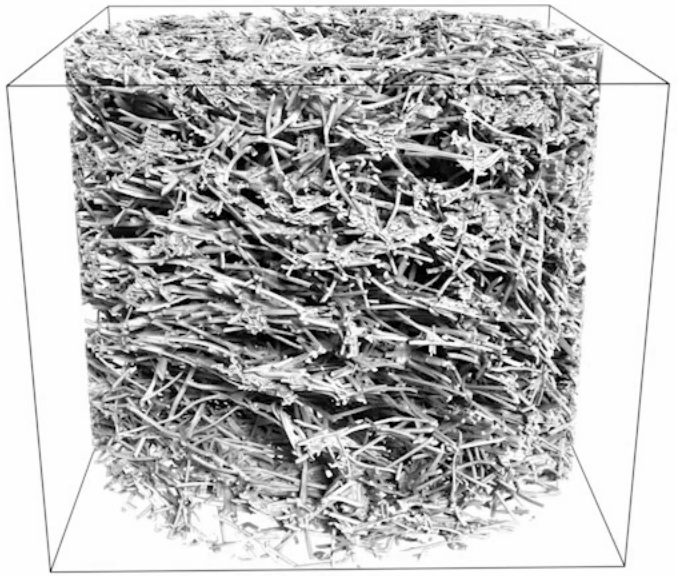
Fast charging in Lithium batteries



Balsara, Ho, et al.

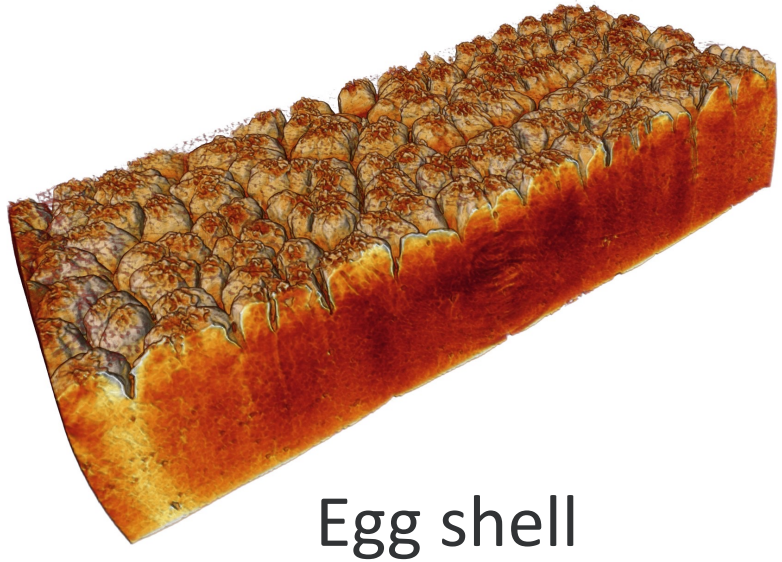






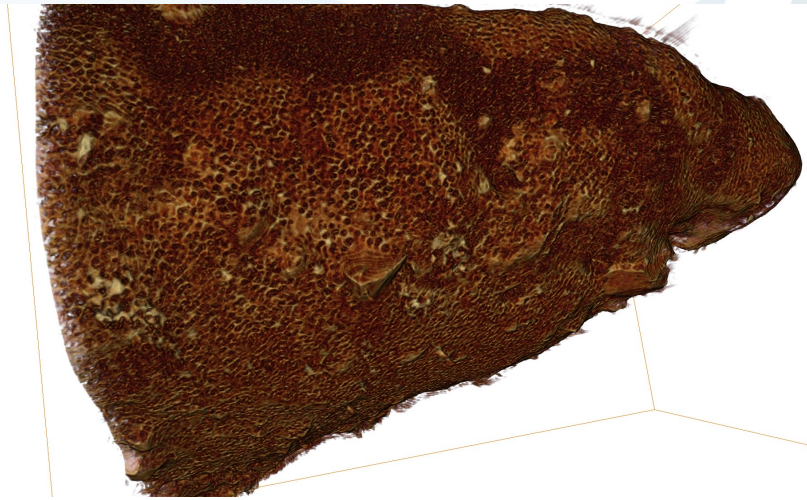
Wanna X-ray something?

- Send in a proposal!
  - Rapid access: 1 page
  - General user: 3 pages



Egg shell

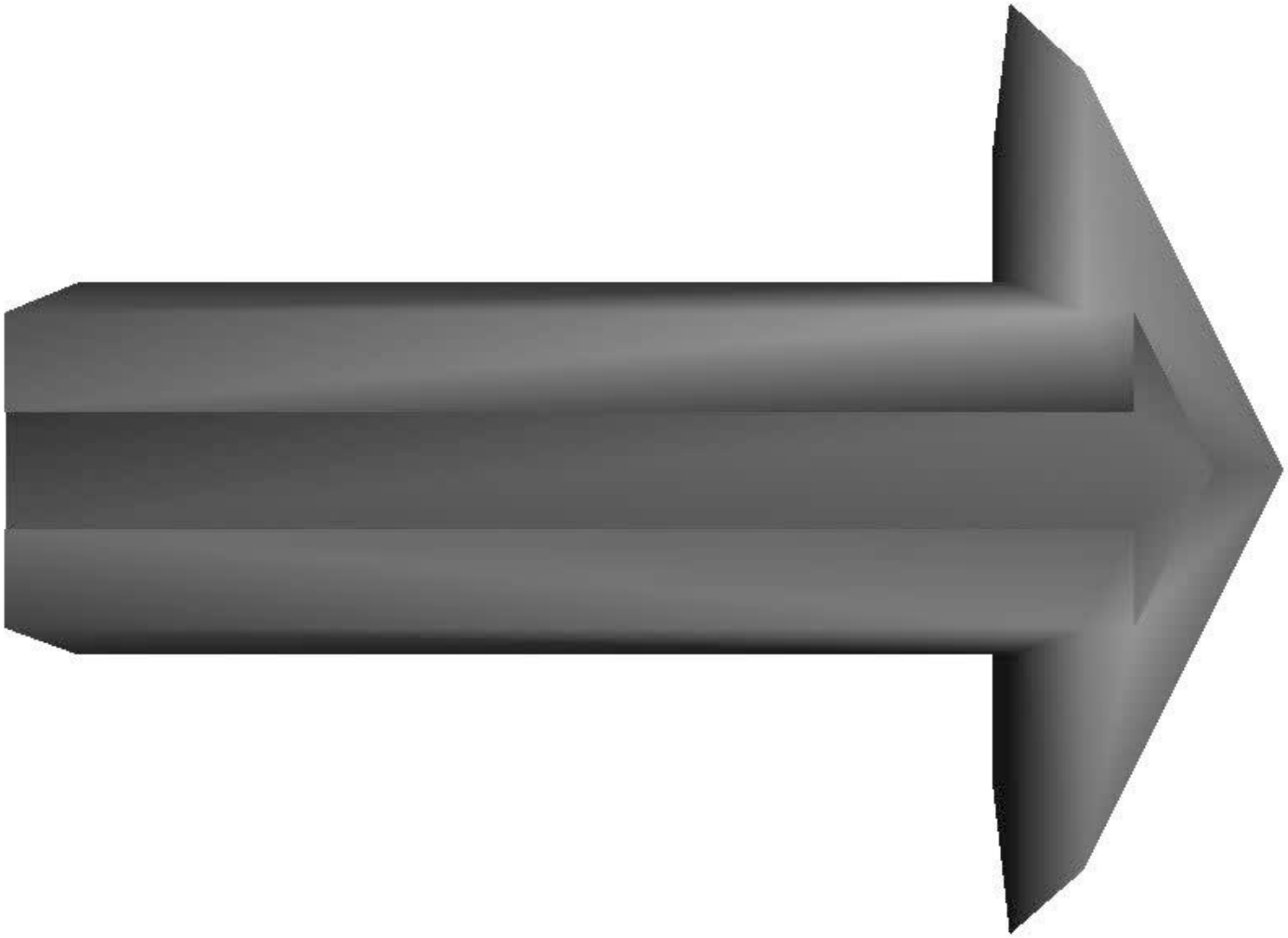
Butterfly wing



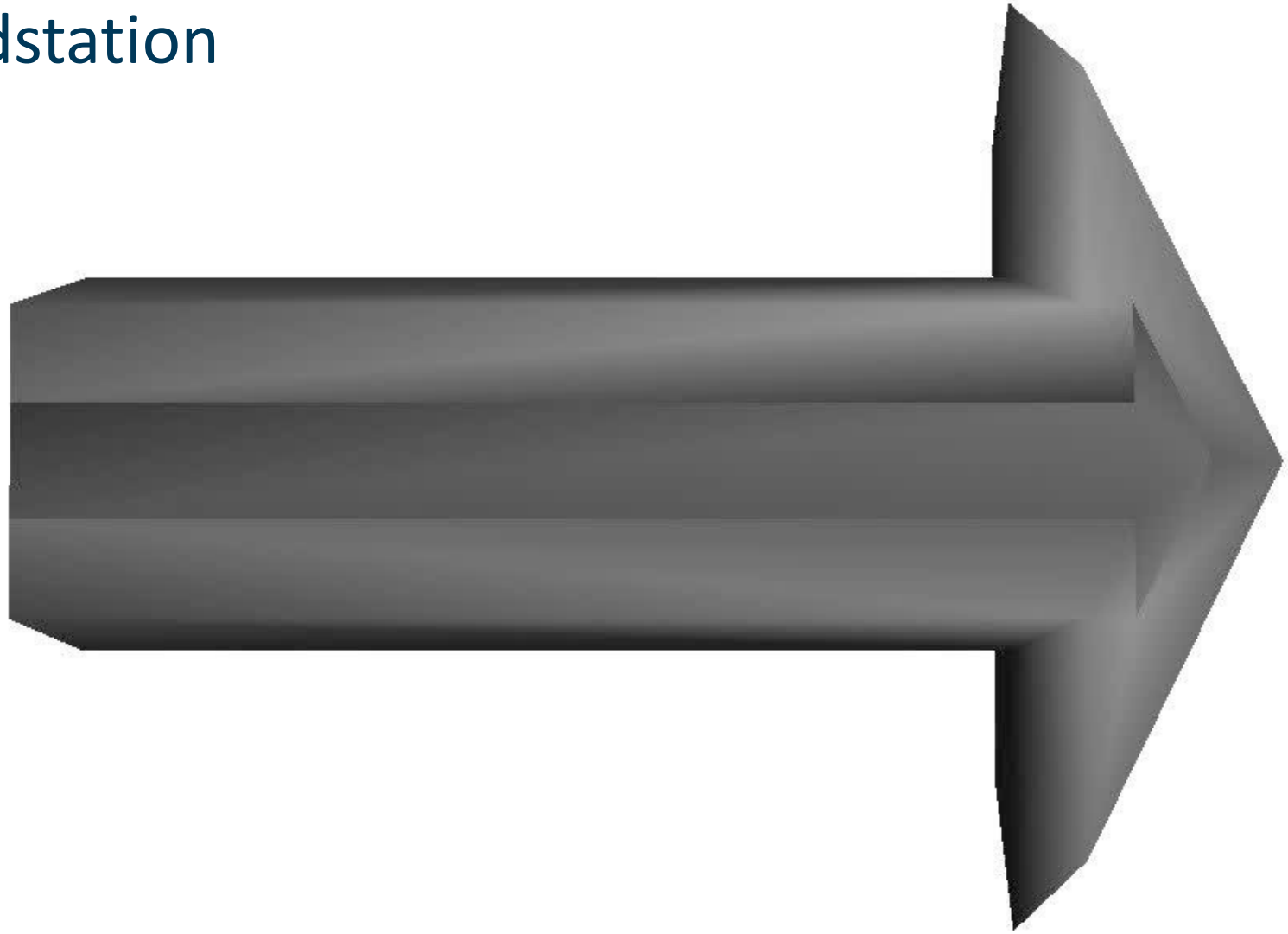
Mentos (pre-Diet Coke)

Changes that are leading us to use workflows

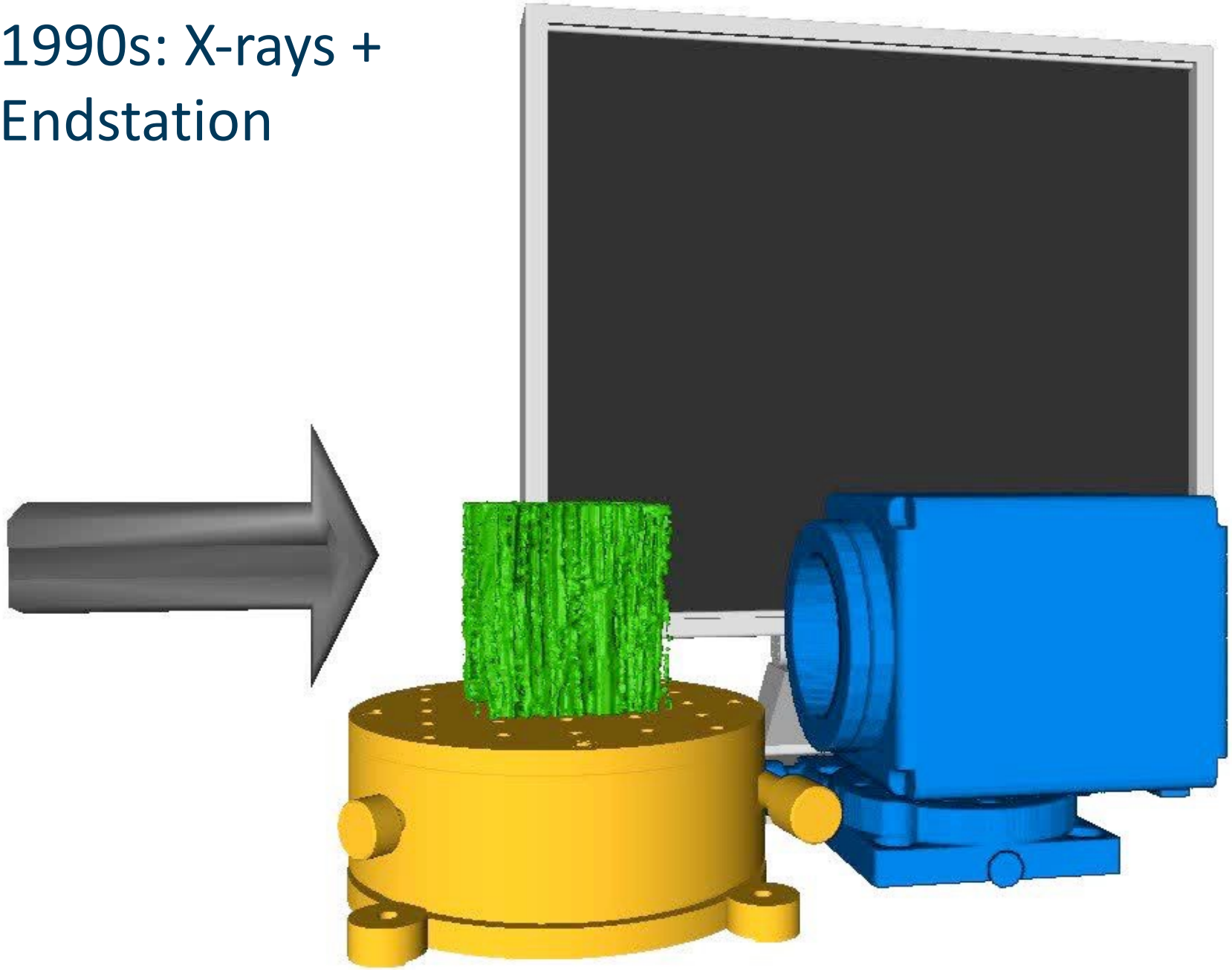
1980s: X-rays



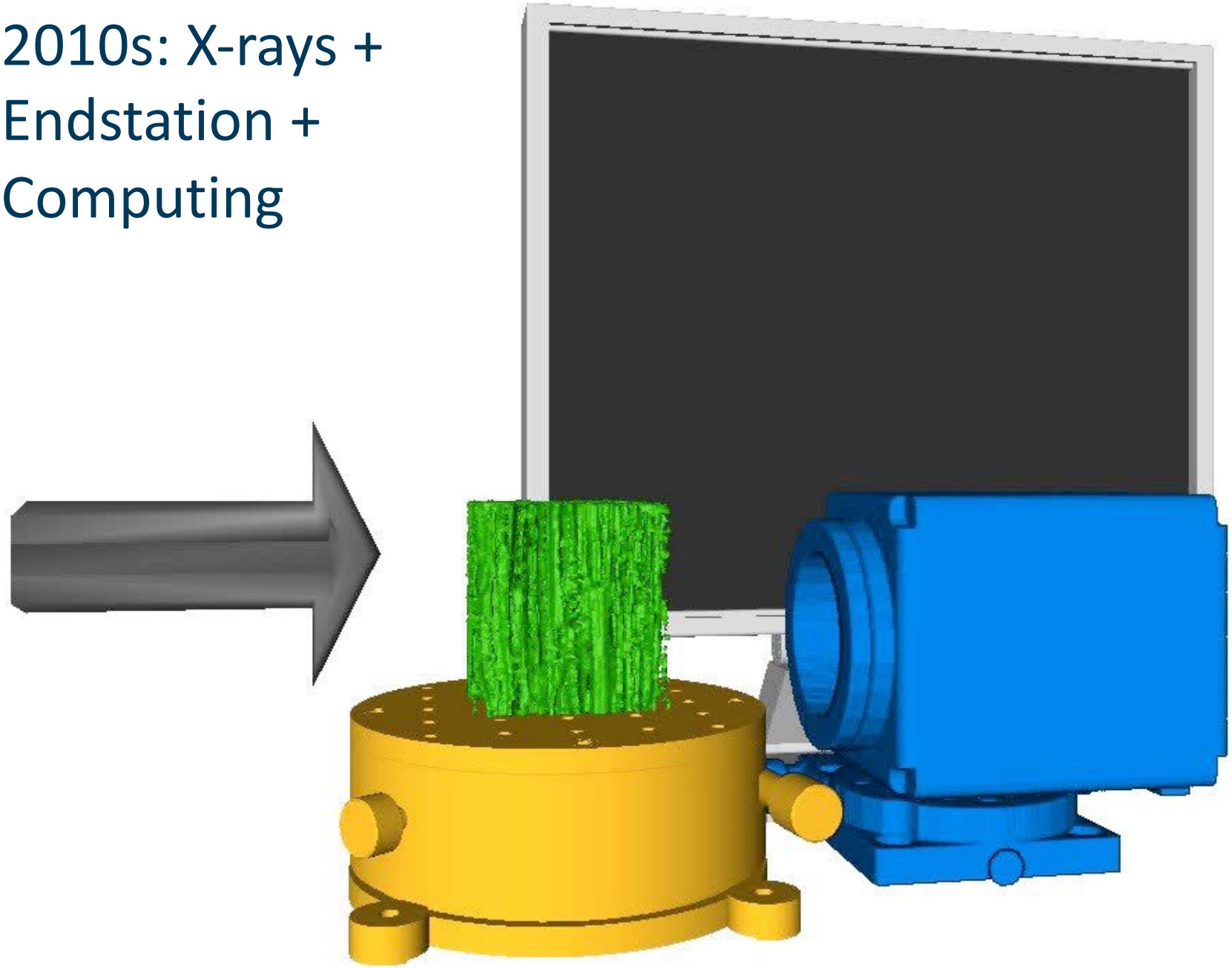
1990s: X-rays + Endstation



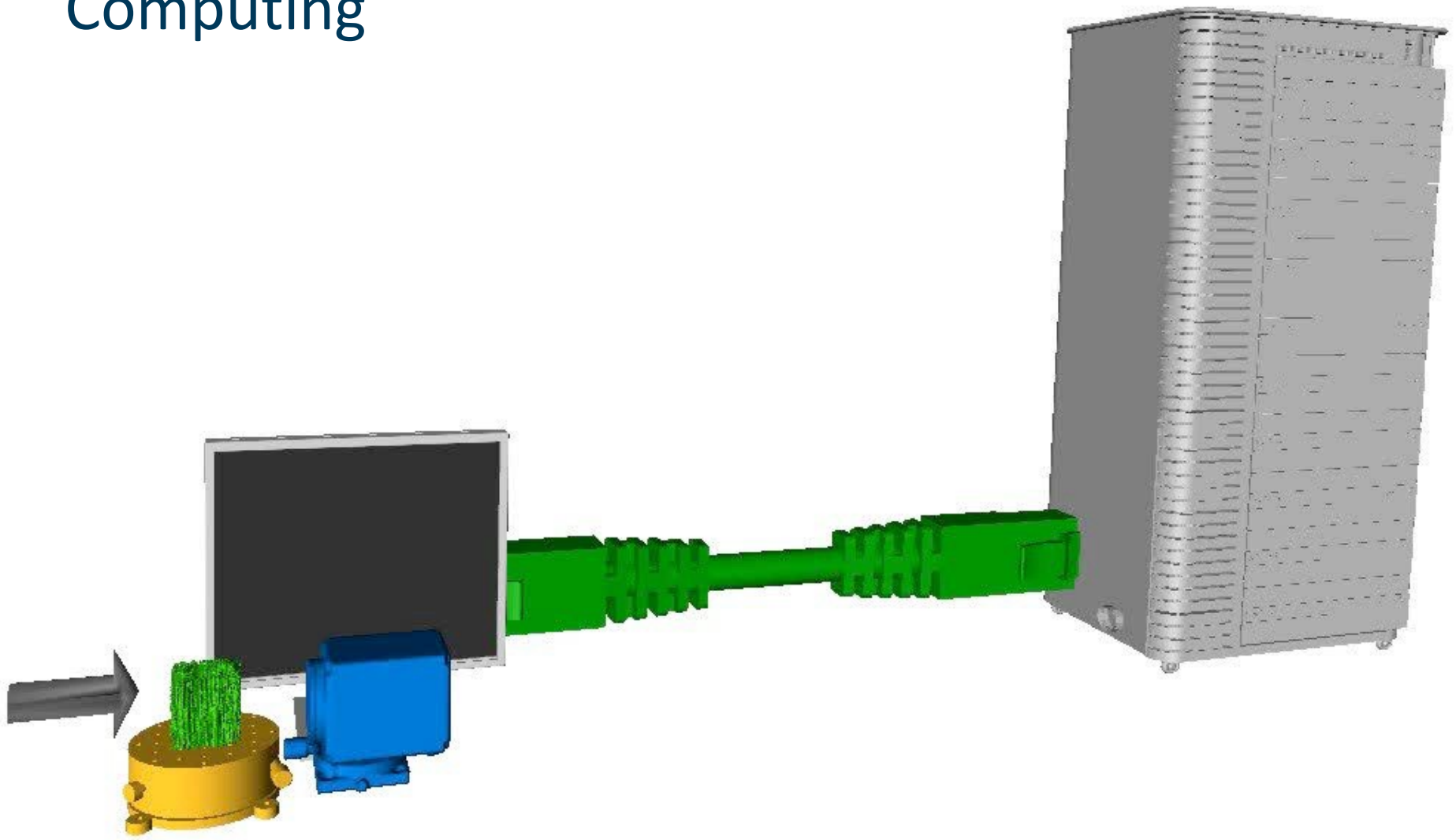
1990s: X-rays + Endstation



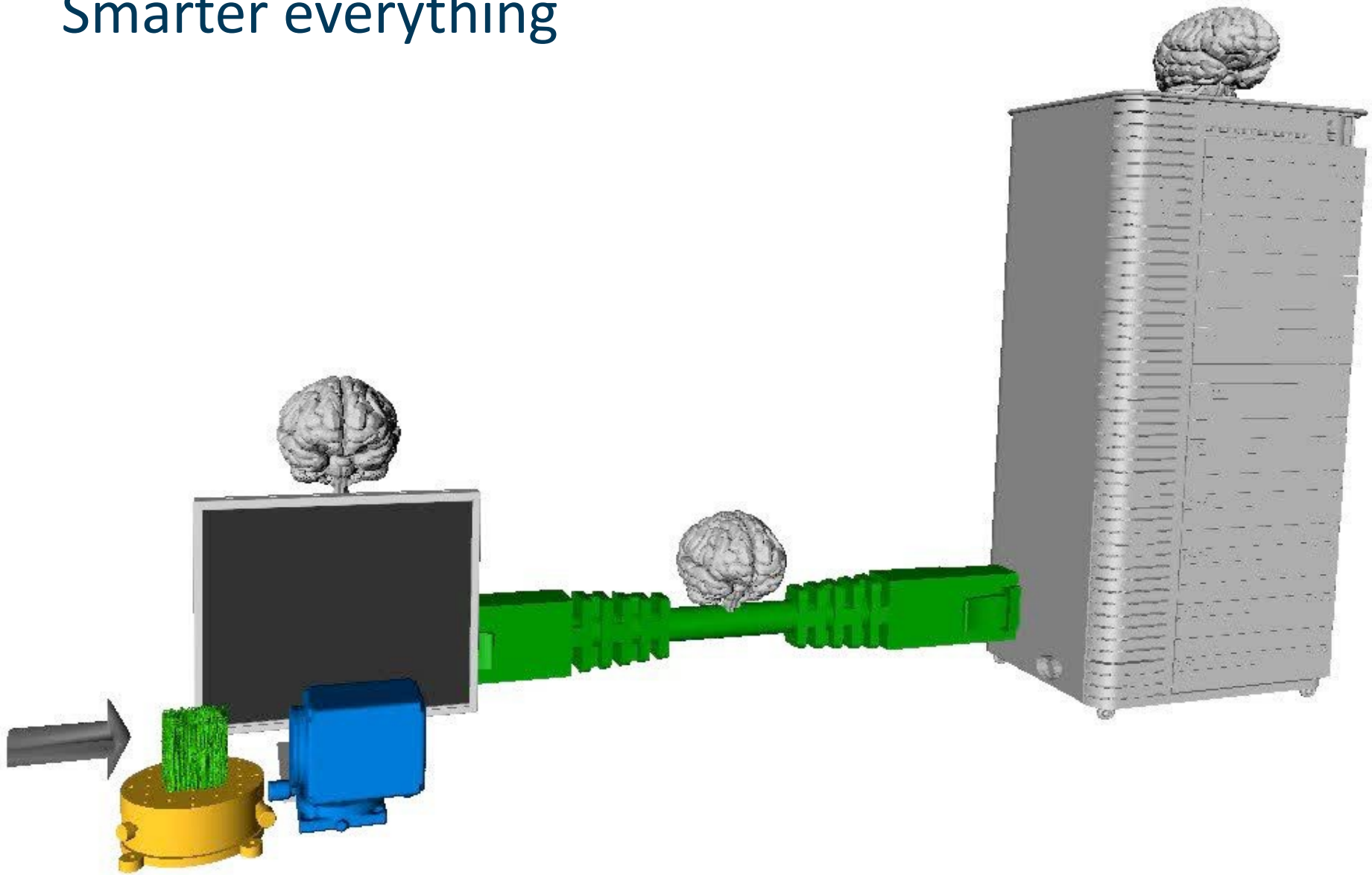
2010s: X-rays + Endstation + Computing



2010s: X-rays + Endstation + Computing

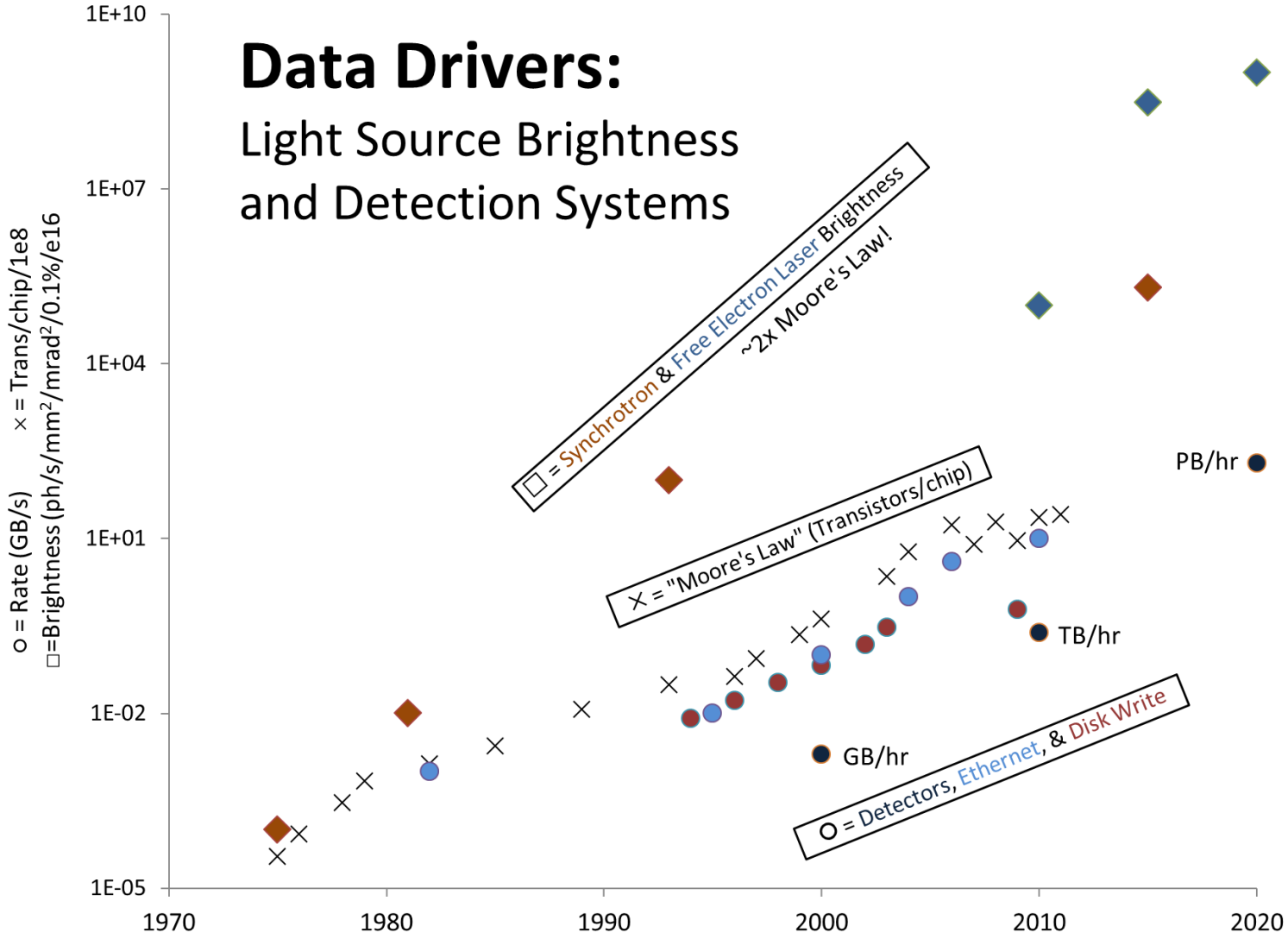


Next:
Smarter everything



Data volumes are increasing

- Brighter X-rays beams
- Faster detectors
- Robots/automation



Developments in computing, algorithms, and software

Massively Parallel 3D Image Reconstruction

Description: Computed Tomography (CT) image reconstruction is an important technique used in a wide range of applications. Among reconstruction methods, Model-Based Iterative Reconstruction (MBIR) generally produces higher quality images. However, the irregular data access pattern, the difficulty of effective parallelization and slow algorithmic convergence have made MBIR impractical for many applications. This paper presents a new algorithm for MBIR, Non-Uniform Parallel Super-Voxel (NU-PSV), that regularizes the data access pattern, enables massive parallelism and ensures fast convergence. We compare the NU-PSV algorithm with two state-of-the-art implementations on a 69632-core distributed system. Results indicate that the NU-PSV algorithm has an average speedup of 1665 compared to the fastest state-of-the-art implementations.

Authors: Xiao Wang, Amit Sabne, Putt Sakdhnagool, Sherman J. Kisner, Charles A. Bouman, Samuel P. Midkiff
https://sc17.supercomputing.org/presentation/?id=gb103&sess=sess147

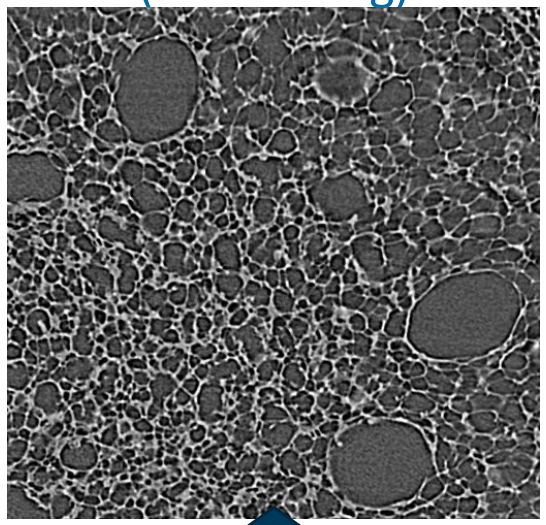
Developments in ML

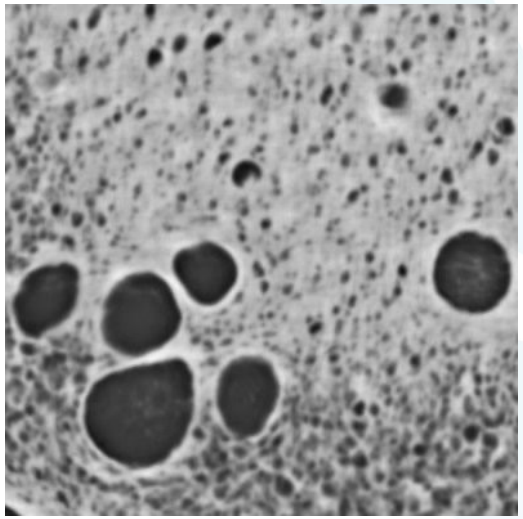
Pelt and Sethian, A Mixed-Scale dense convolution neural network for image analysis, PNAS (2018) v115n2, p254

Static scan (for training)

Time-resolved scan

1024 angles (original)

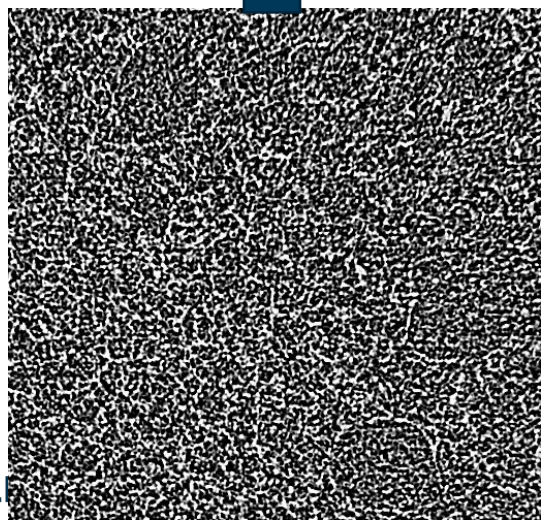


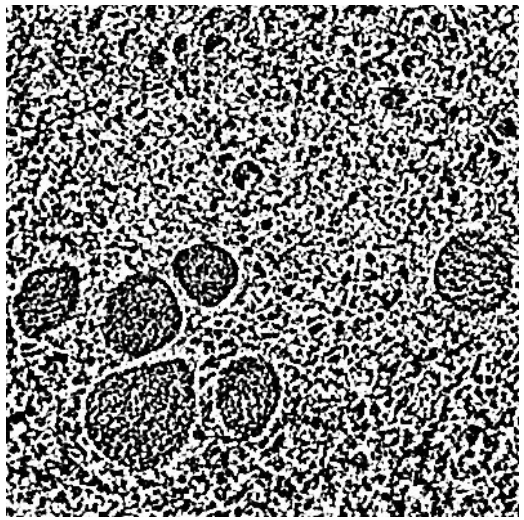


Output of Mixed-Scale Dense Network

Train

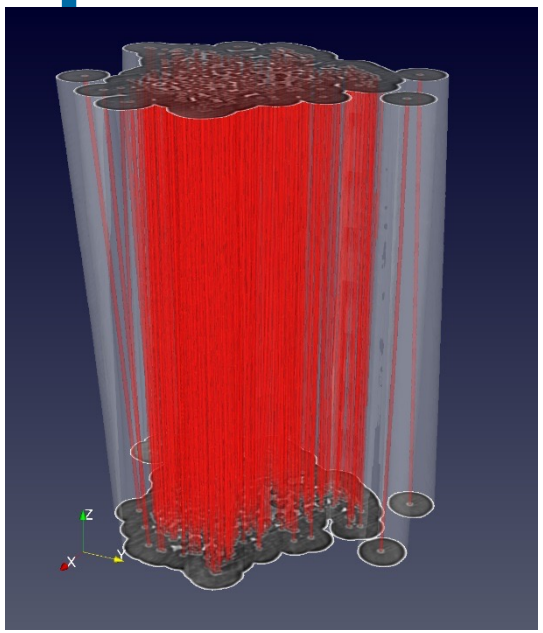
128 angles (downsampled)



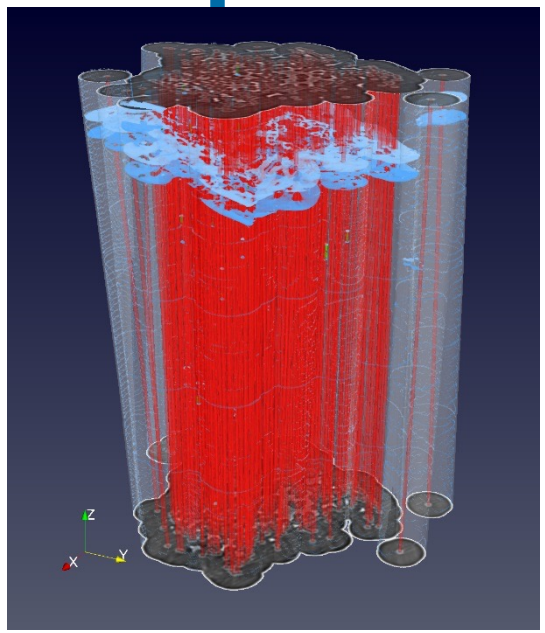


128 angles

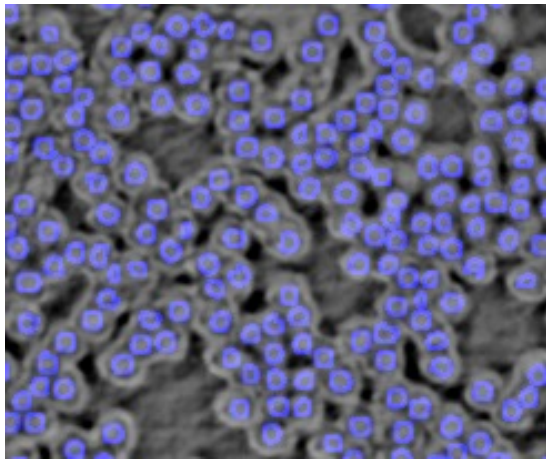
Developments in computer vision

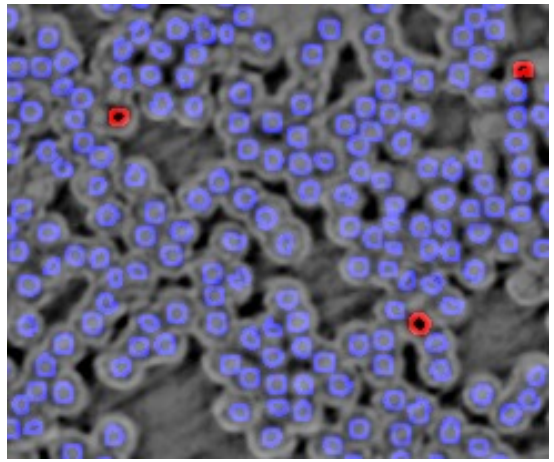


93N



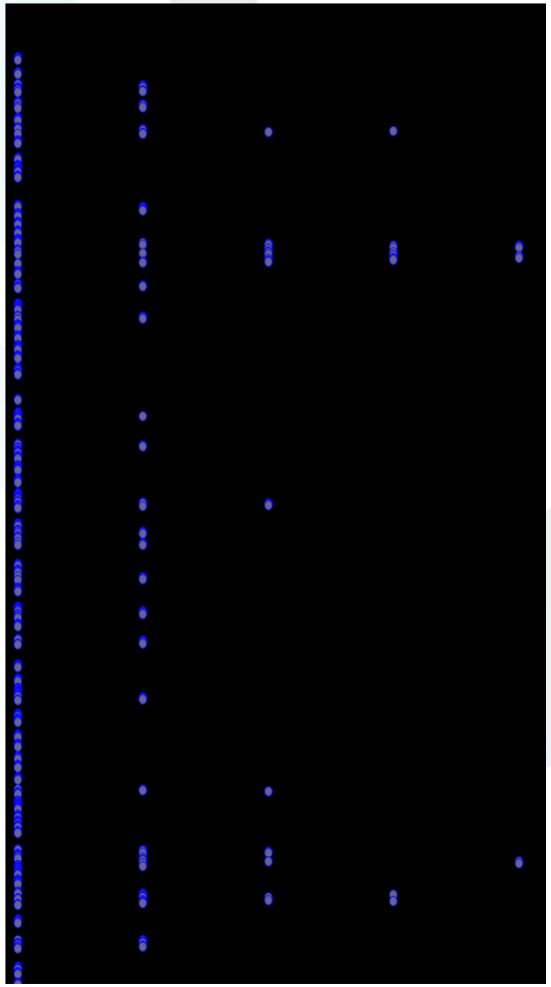
151N



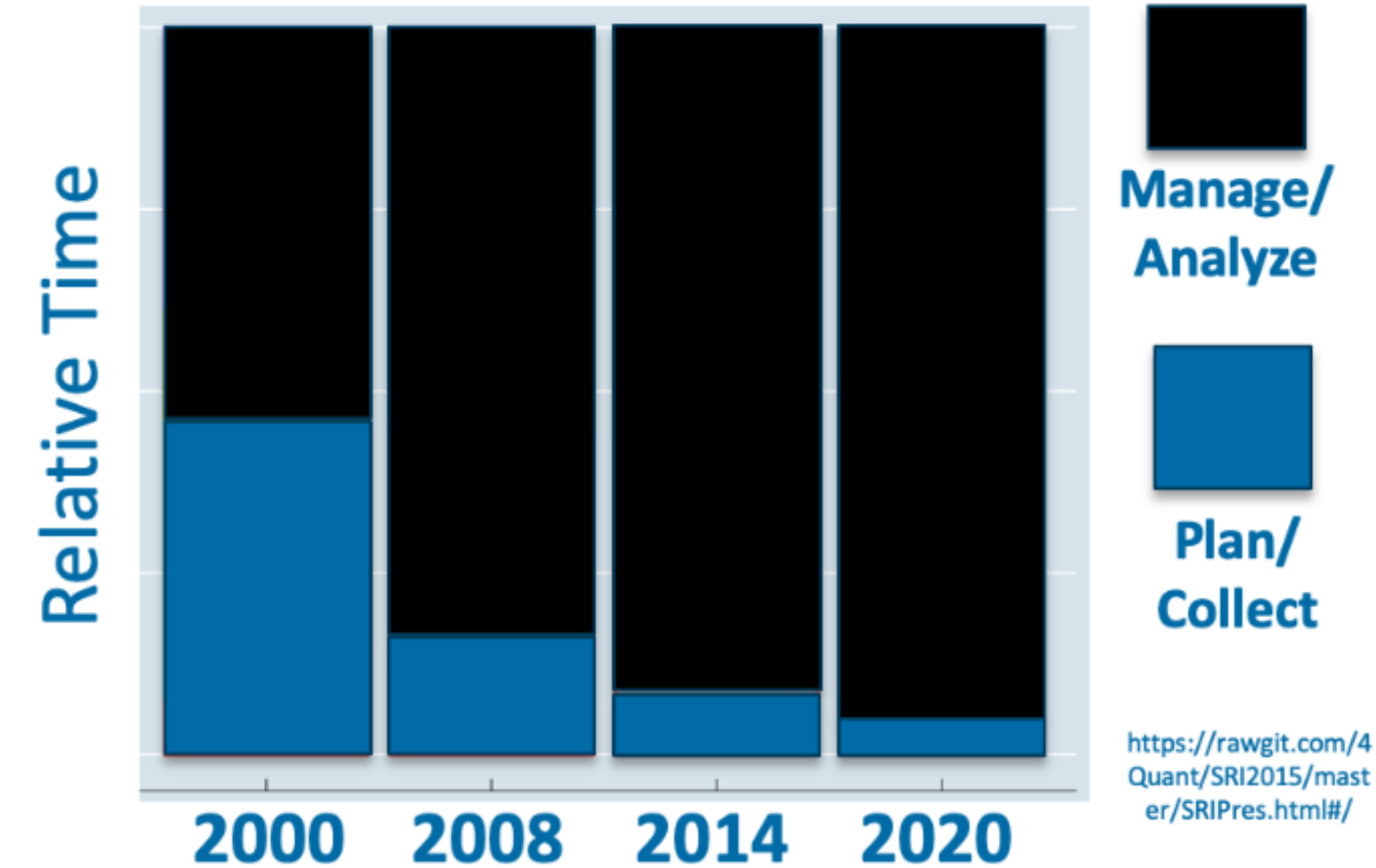


Fiber Breaks

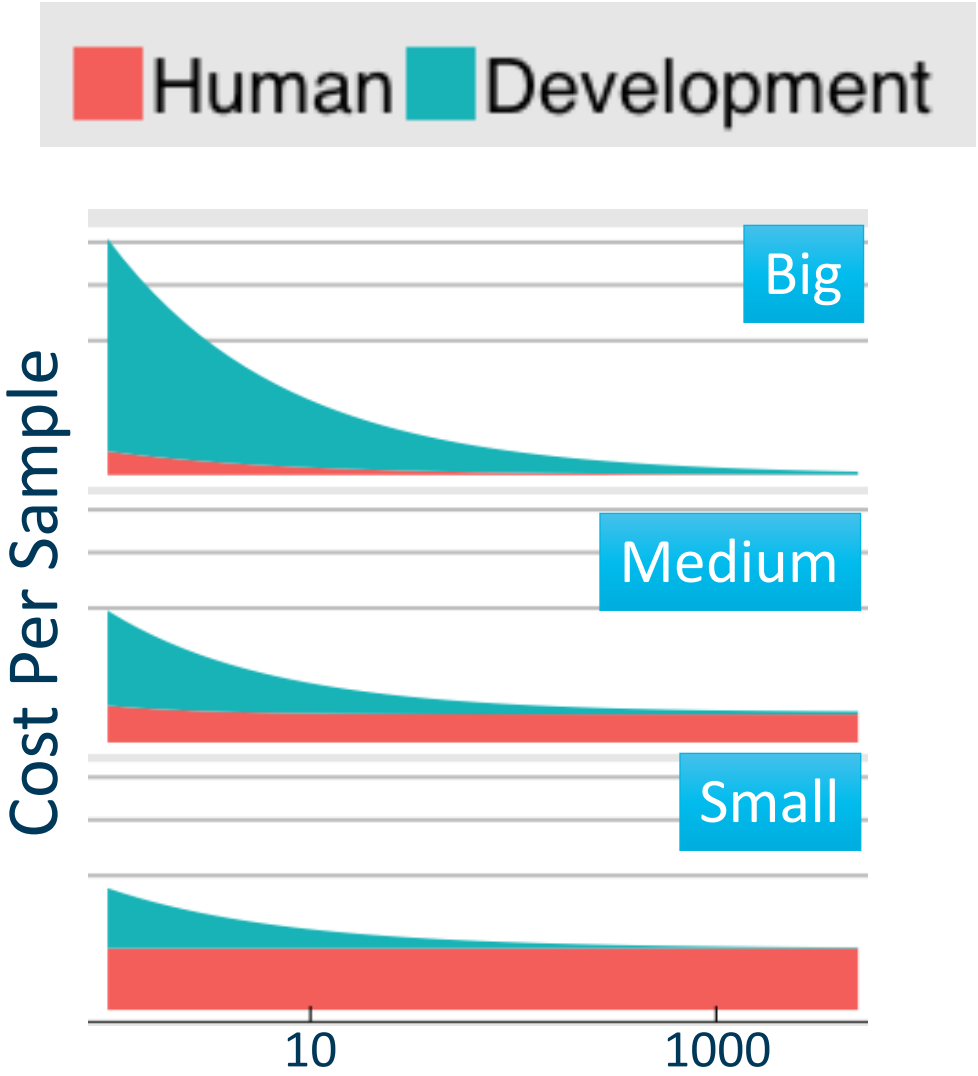
Slice #



D. Ushizima, T. Perciano, H. Krishnan, B. Loring, B. Hrish, D. Parkinson, R. Ritchie, W. Bethel and J. Sethian,

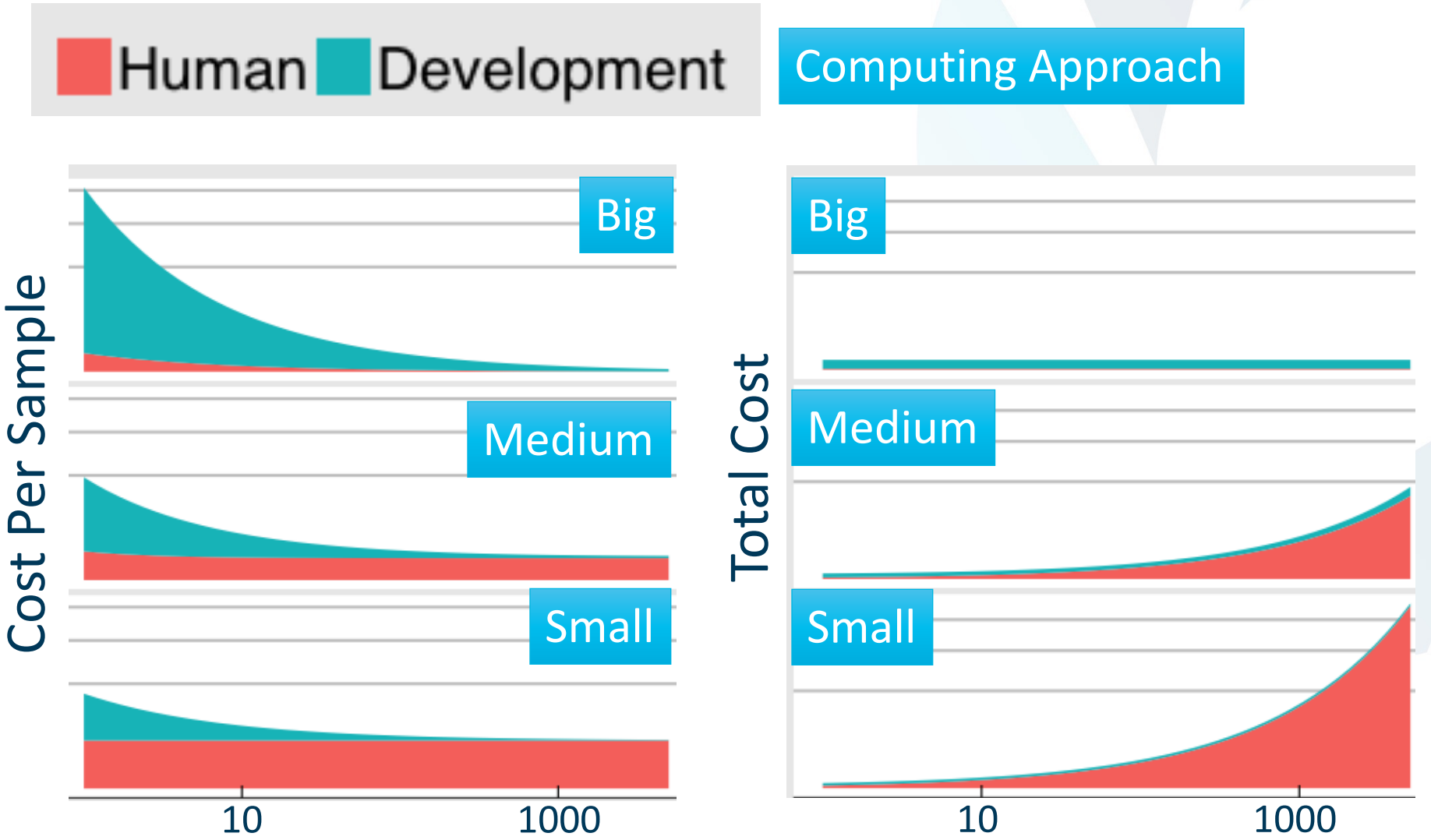

Relative Time
Manage/
Analyze
Plan/
Collect
2000
2008
2014
2020
https://rawgit.com/4Quant/SRI2015/master/SRIPres.html#/
Data Tasks Are Increasingly Important!

Should we build a workflow system?



Figures from Kevin Mader
https://rawgit.com/4Quant/SRI2015/master/SRIPres.html#/

Should we build a workflow system?



Figures from Kevin Mader

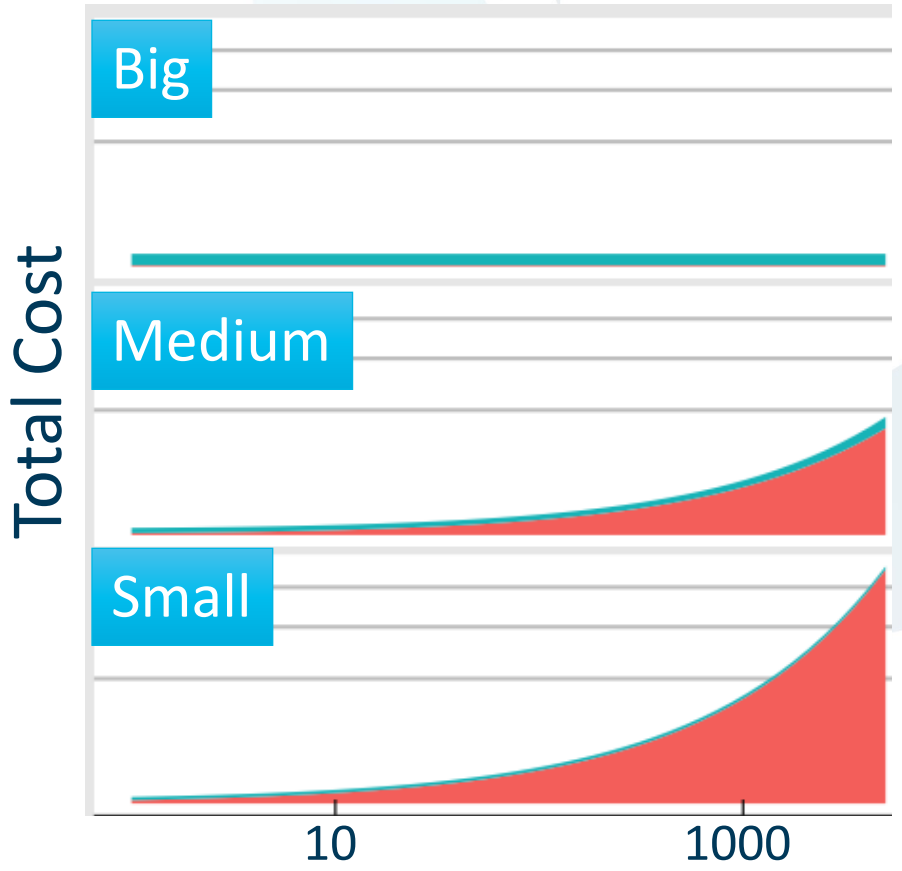
https://rawgit.com/4Quant/SRI2015/master/SRIPres.html#/

Should we build a workflow system?

Human Development

Computing Approach

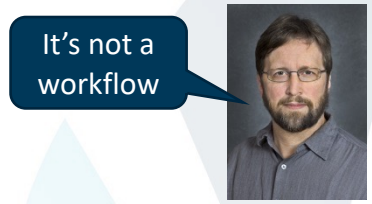
Most individual users don't have enough samples to justify investing in the required development time for Big Data approaches, but facilities can!



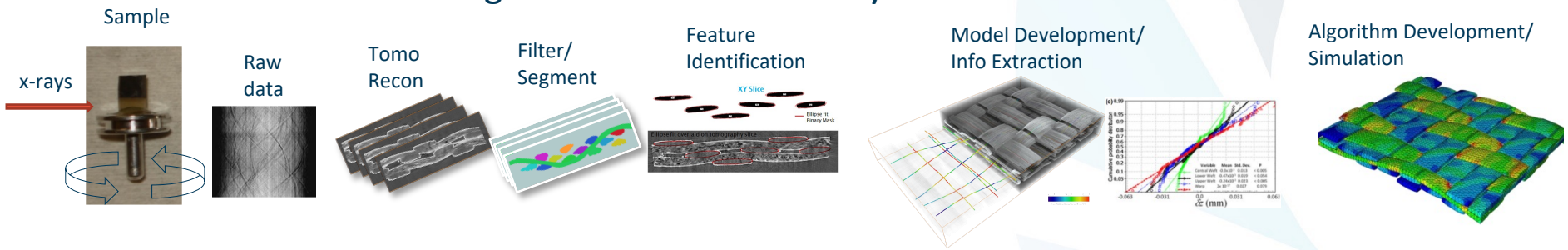
Figures from Kevin Mader

https://rawgit.com/4Quant/SRI2015/master/SRIPres.html#/

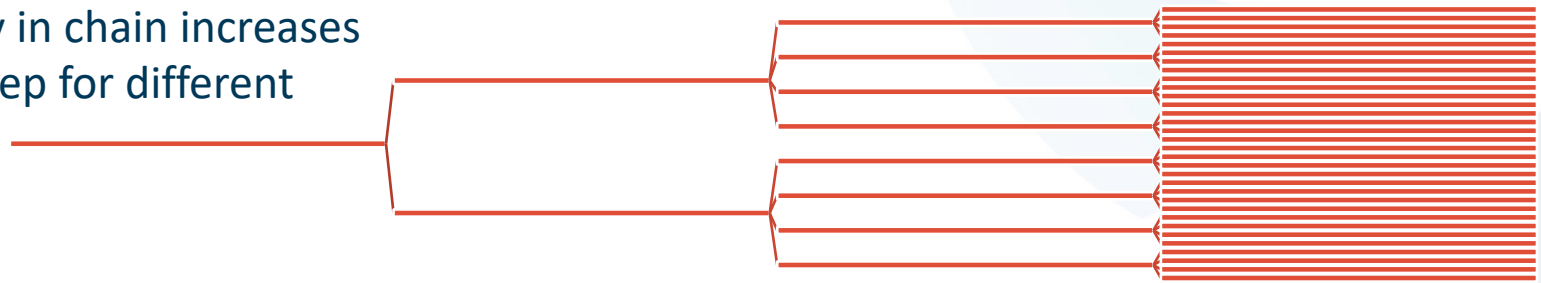
MicroCT Analysis chain

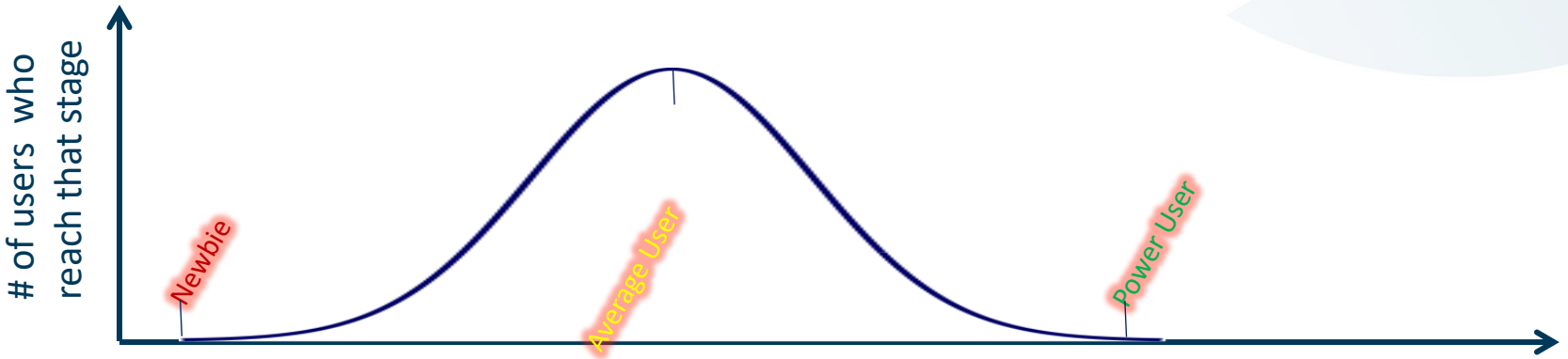


"Processing" "Analysis" "Simulation"



Variability in chain increases at each step for different useres





MicroCT Analysis chain

"Processing" "Analysis" "Simulation"

Sample

x-rays

Raw data

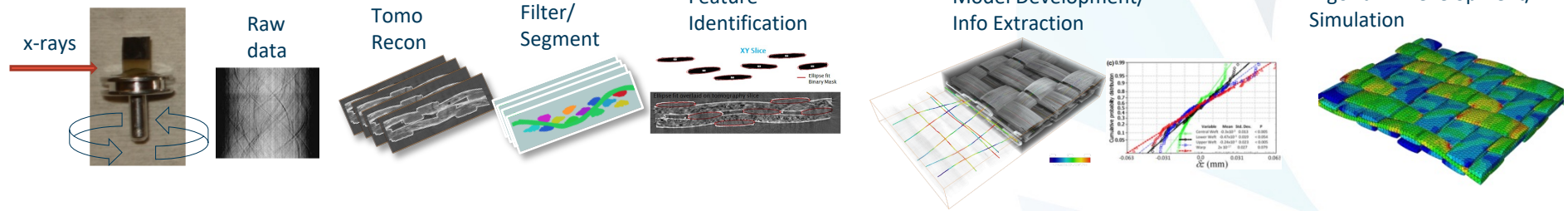
Tomo Recon

Filter/ Segment

Feature Identification

Model Development/ Info Extraction

Algorithm Development/ Simulation



Variability in chain increases at each step for different users

I could do so much more if all the software was in one place, running on a supercomputer.

I will only use my own analysis code.

HELP!

I hate the framework you chose.

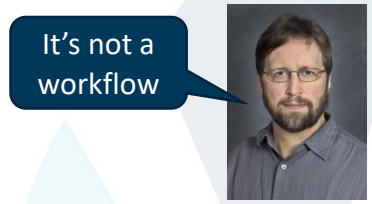
# of users

Newbie

Average User

Power User

MicroCT Analysis chain



"Processing" "Analysis" "Simulation"

Sample

x-rays

Raw data

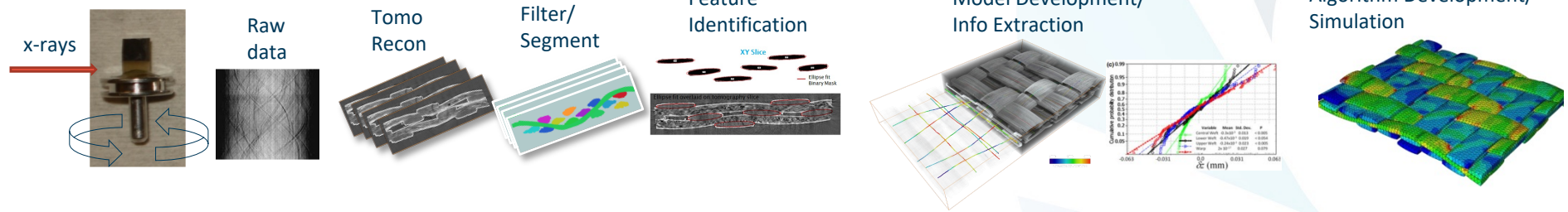
Tomo Recon

Filter/ Segment

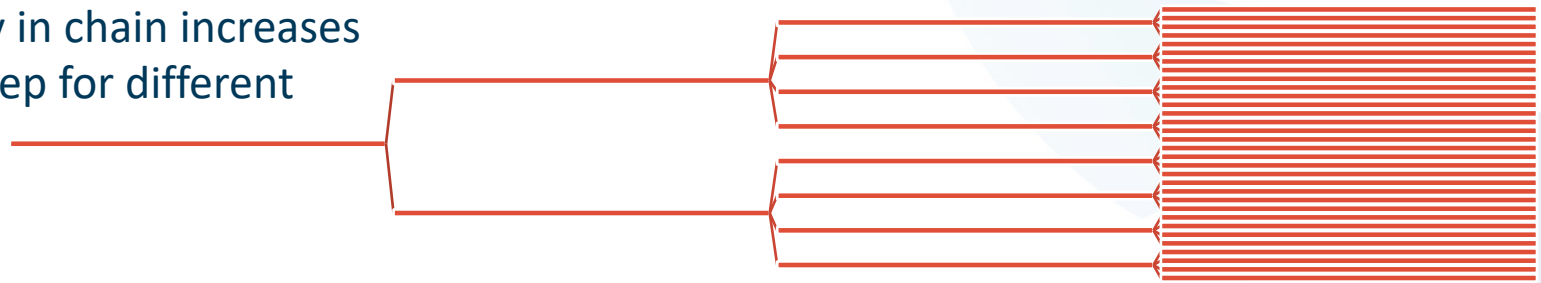
Feature Identification

Model Development/ Info Extraction

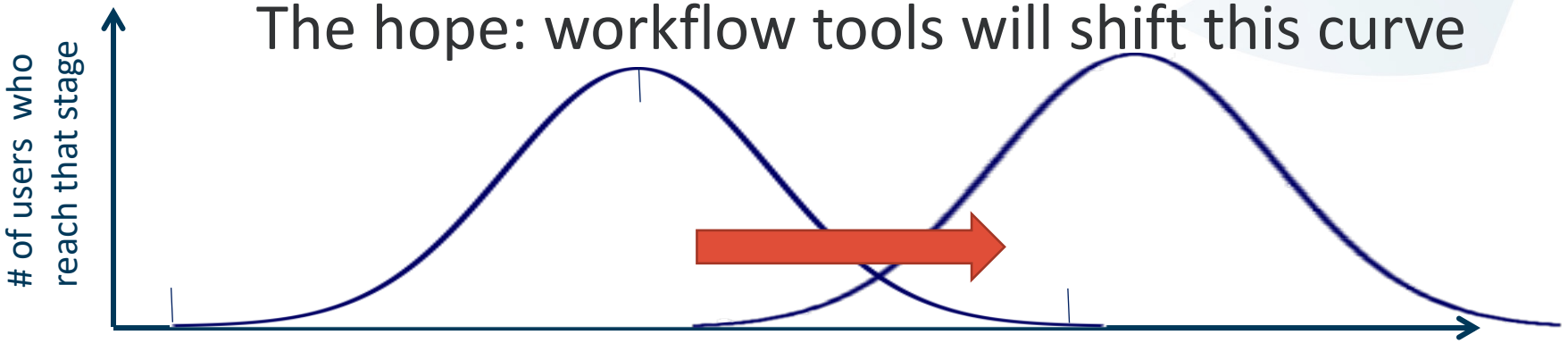
Algorithm Development/ Simulation



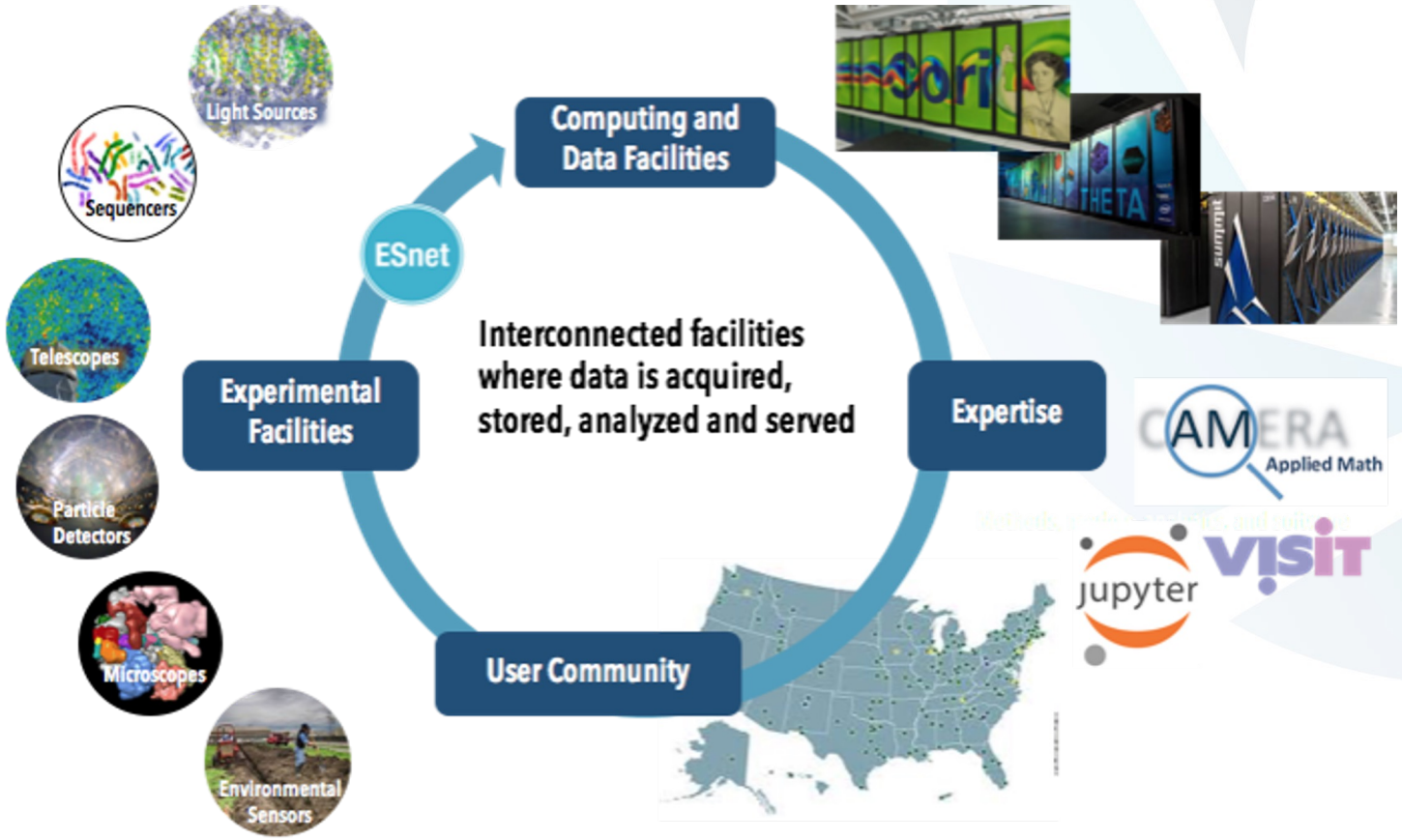
Variability in chain increases at each step for different users



The hope: workflow tools will shift this curve



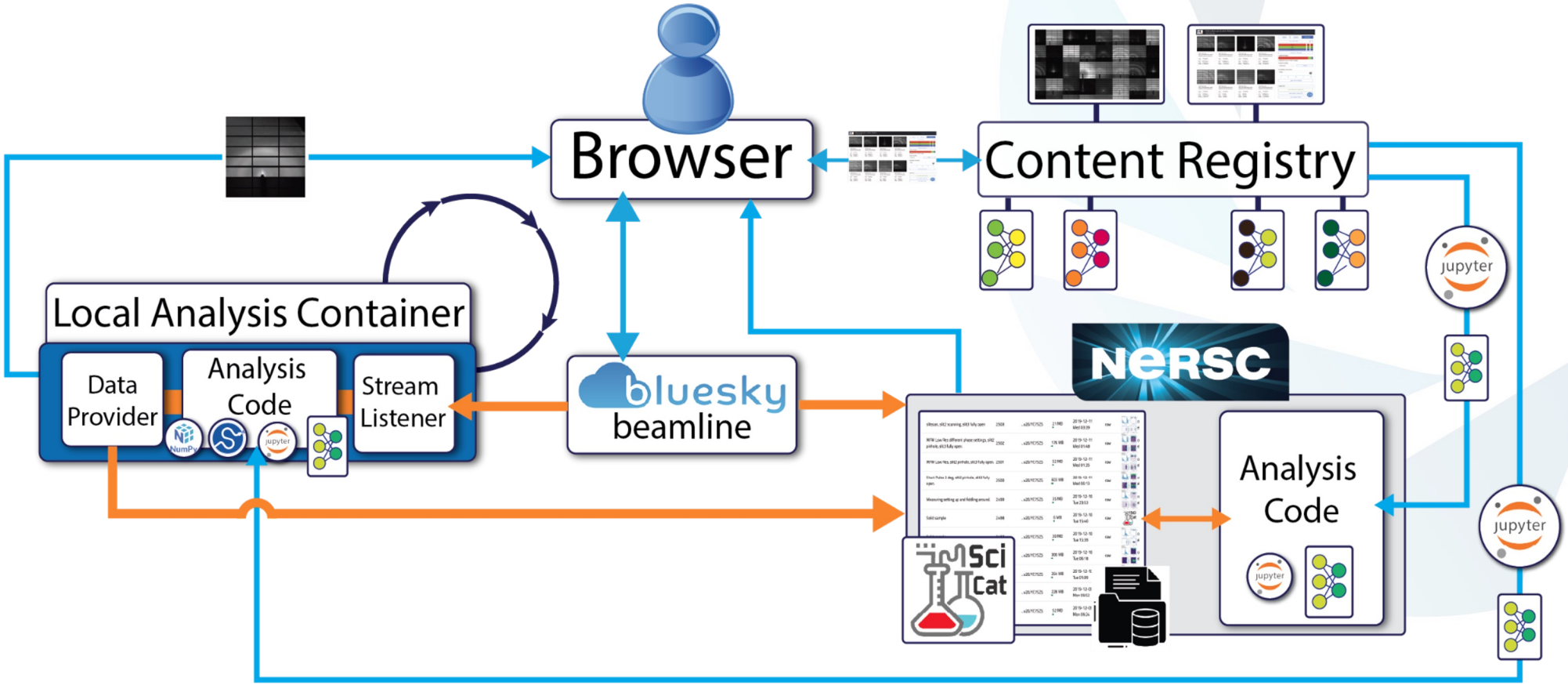
The Superfacility Model: an ecosystem of connected facilities, software and expertise to enable new modes of discovery



Workflows from different perspectives

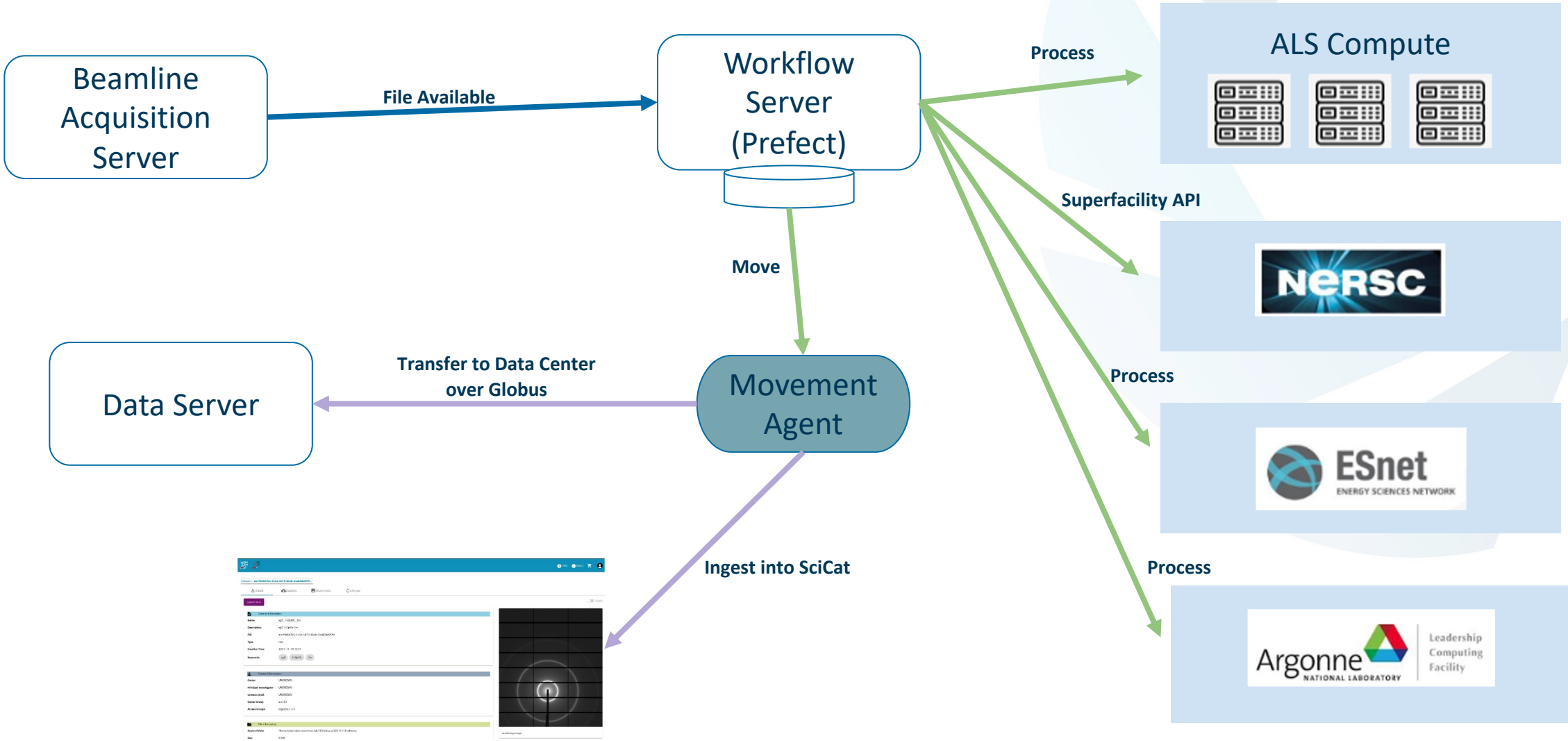
ALS Computing Group perspective

Vision for full Integration of Compute at Beamlines



Scalable, Maintainable, and Solutions have to be Shareable between beamline.

Centralized Workflow Management



User perspective

ALSHub | Remote ALS | https://rem | Flow Runs - | SciCat ALS | feb2023_sta | File Manage

alshub.als.lbl.gov/dashboard

a | LBL | home | bW | R | Other bookmarks

ALSHub

BERKELEY LAB

Home | Profile | Contact | ALS User Guide | ALS Facility Homepage | Staff Access

Welcome Dilworth Y. Parkinson

Important Log In and Links Information for Users

New users: Create an account by logging in using one of the login options (we encourage using

Your Status

Access Status

LBNL#:**005481**

Appointment Status: ✓ Active

Appointment Start/End Dates:

07/07/2010 -

LBNL Badge status: ✓ Active

Badge Expiration Date: 07/11/2024

ALS Remote access: ✓ Active

General required on-site training complete: Yes

Arrival Timeline

Your requested arrival date: Update

None

User Agreement

Non-proprietary: Expires 12/31/2099

Proprietary: None

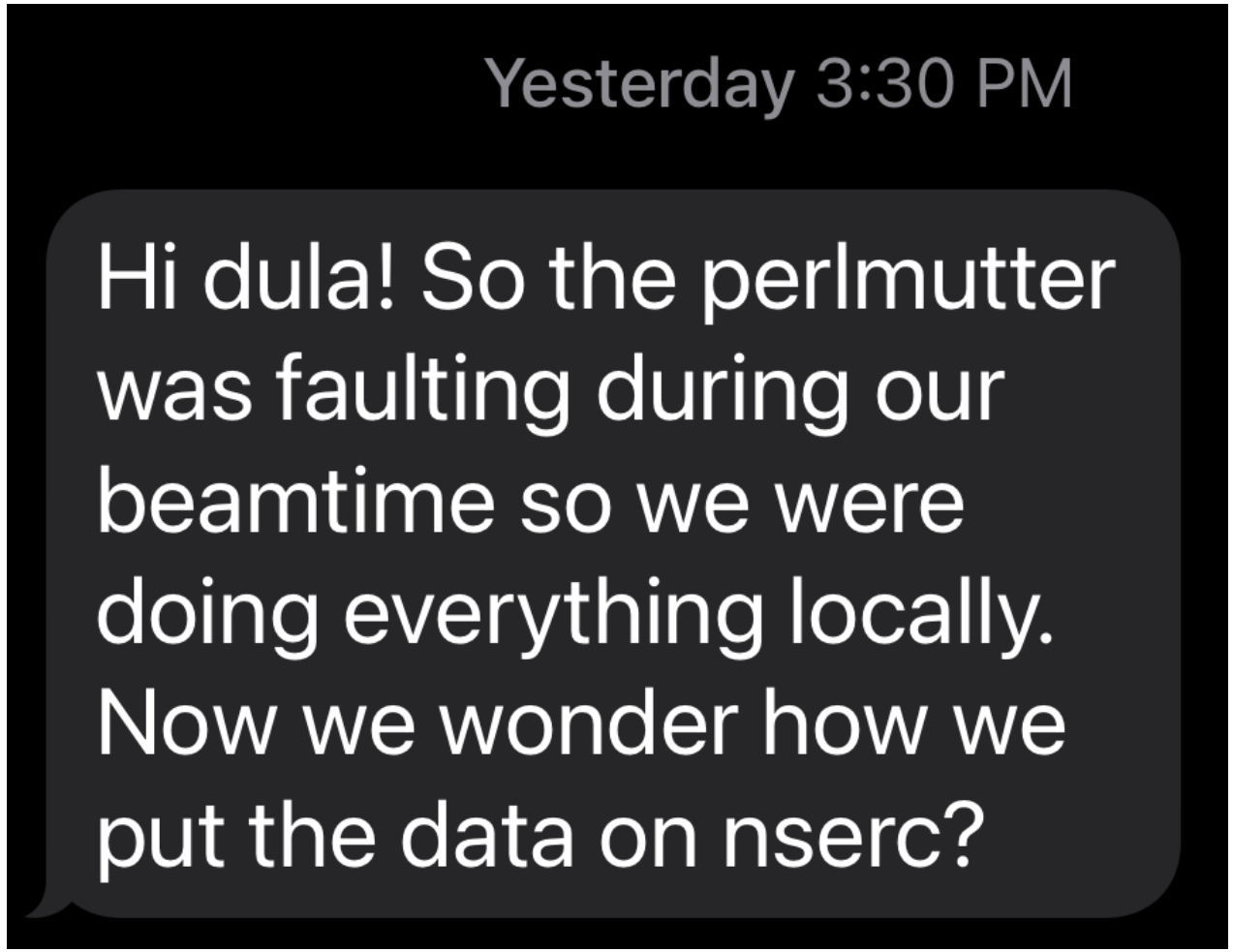
Safety Training (Do not do without an LBNL#)

LBNL # is issued in Human Resources welcome email

My perspective

Superfacililty Dream! (Dula 2015)

Superfacility reality (Dula 2023)



- acquisition system
- zmq stream
- file writing
- workflow orchestrator
- globus transfer
- network
- spin on NERSC
- jupyter on NERSC
- cfs on NERSC
- Perlmutter compute
- data permissions settings
- user error

Principles of and Notes on our Workflows

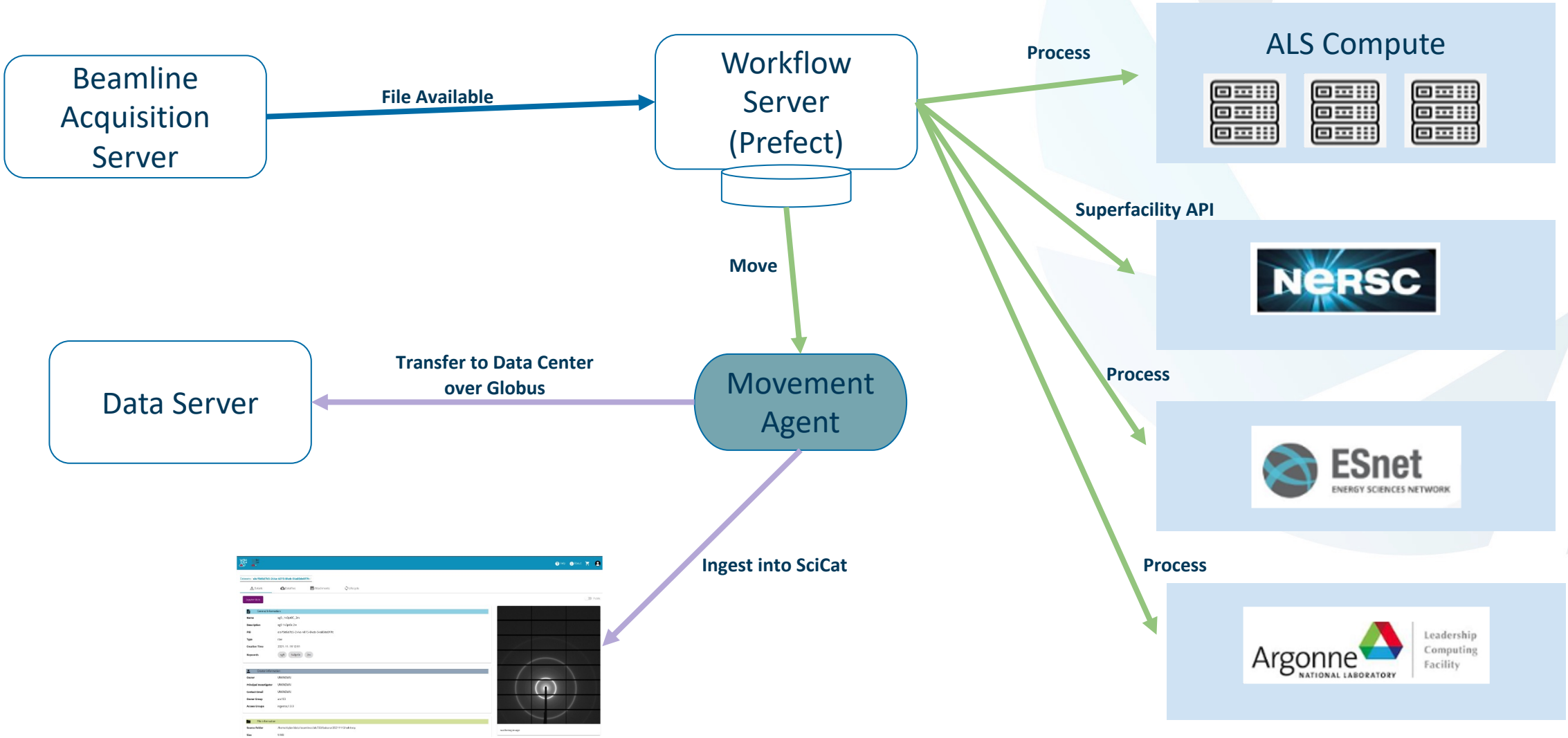
Web interfaces

- No installation for users
- No computing requirements for users beyond laptop
- Remote participants use the same interface

Docker/Shifter

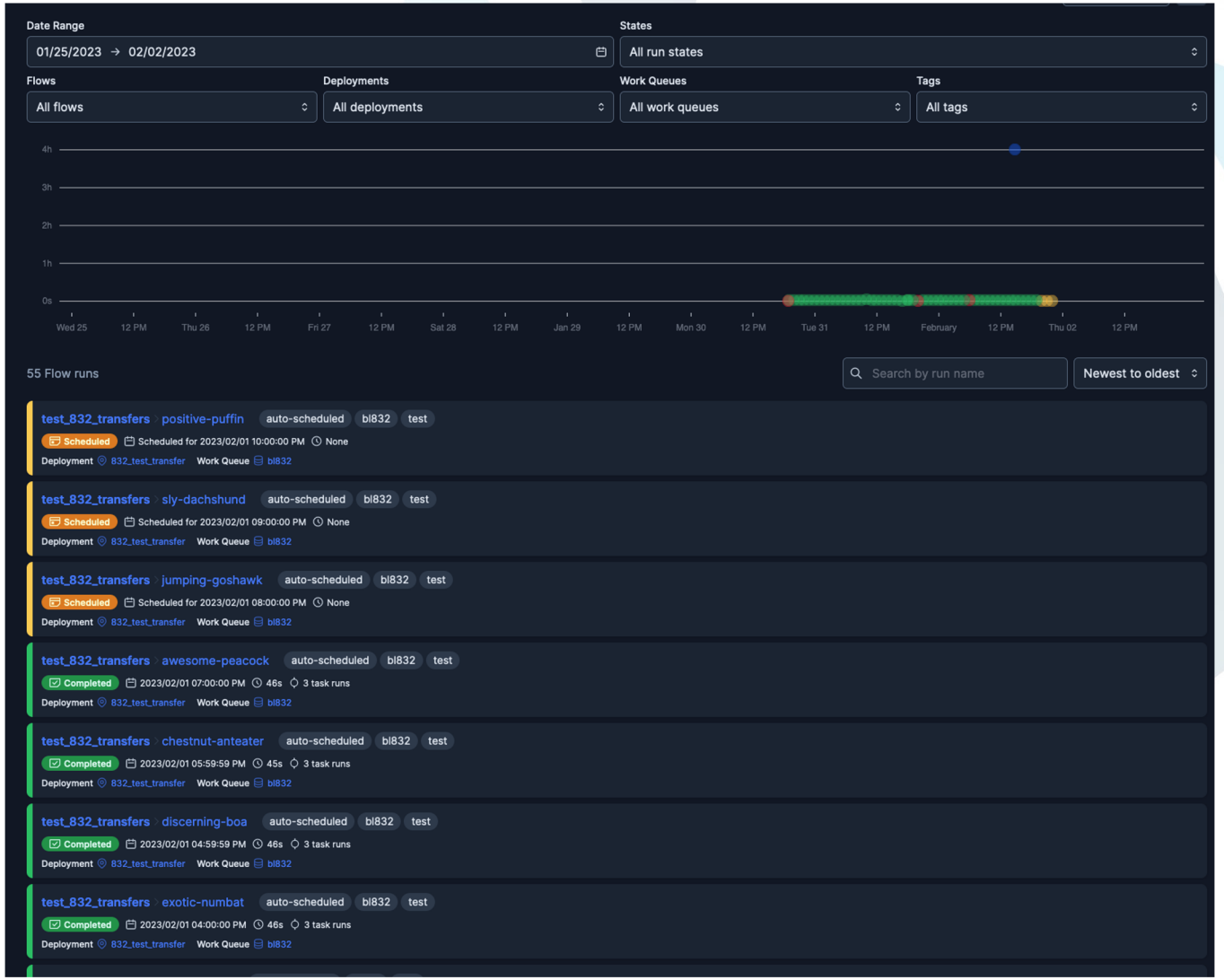
- Same environment and software at NERSC and ALS servers
- Script to launch docker on NERSC allows custom volume mounting to simplify what users see

Loosely coupled components, redundancy



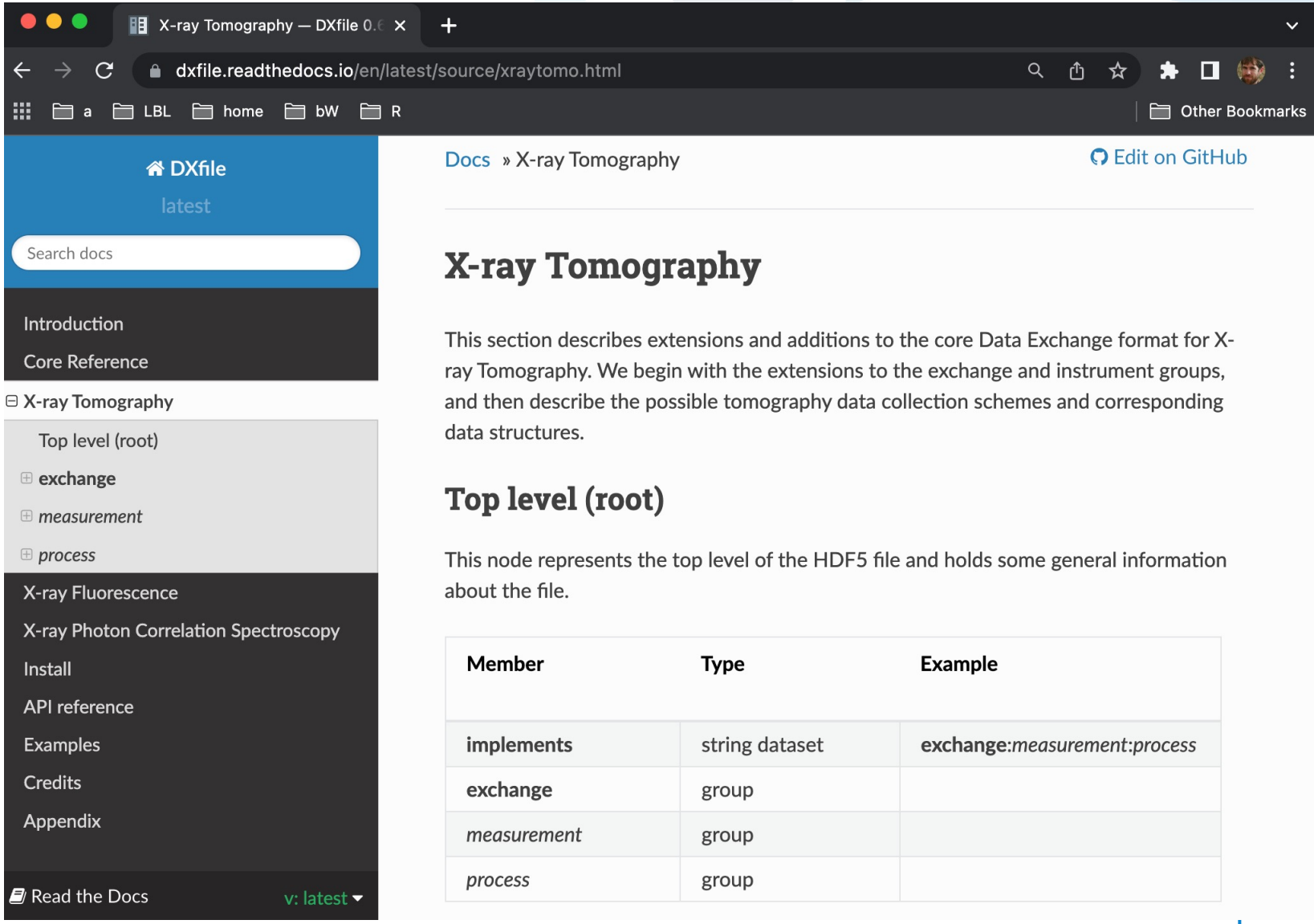
Prefect Workflow Orchestration

- Large community with industry backing (financial companies) and NSLS II.
- Easy monitoring, notifications
- Retries.
- Flexible compute.



Interoperable metadata

- Nexus and other standards for metadata naming
- Information about
  - users
  - experiment
  - data processing



Still painful when working across facilities/systems

- Security
- Accounts
- Passwords
- Certificates
- ACL’s and data access

Conclusion

Workflows at the ALS

- The need for workflows is increasing, and more people are open to doing the work to adopt them
- A few good examples exist and are promising!
- Still a lot of work to do to make them easy to deploy and robust across many beamlines
- The infrastructure we build for workflows will be useful for autonomous experiments

Thanks

- ALS and LBL computing, controls, and IT groups
  - Alex, Dylan, Lee, Tanny, Wiebke, Kevan, Jason, Cobber, Steve,Tibbers, Kuldeep, Karen
- LBL computing
  - Dani, David
- NERSC
  - Bjoern, Lipi, Debbie, Rolli, Shreyas, Matthew
- ESnet and LBLnet