COMPUTING cHALLENGES IN LIGHT SOURCES (CHILS 2016???)

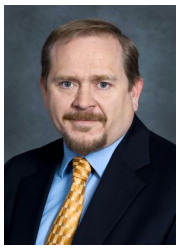
Dula Parkinson
Advanced Light Source, Lawrence Berkeley National Laboratory

Thanks to:





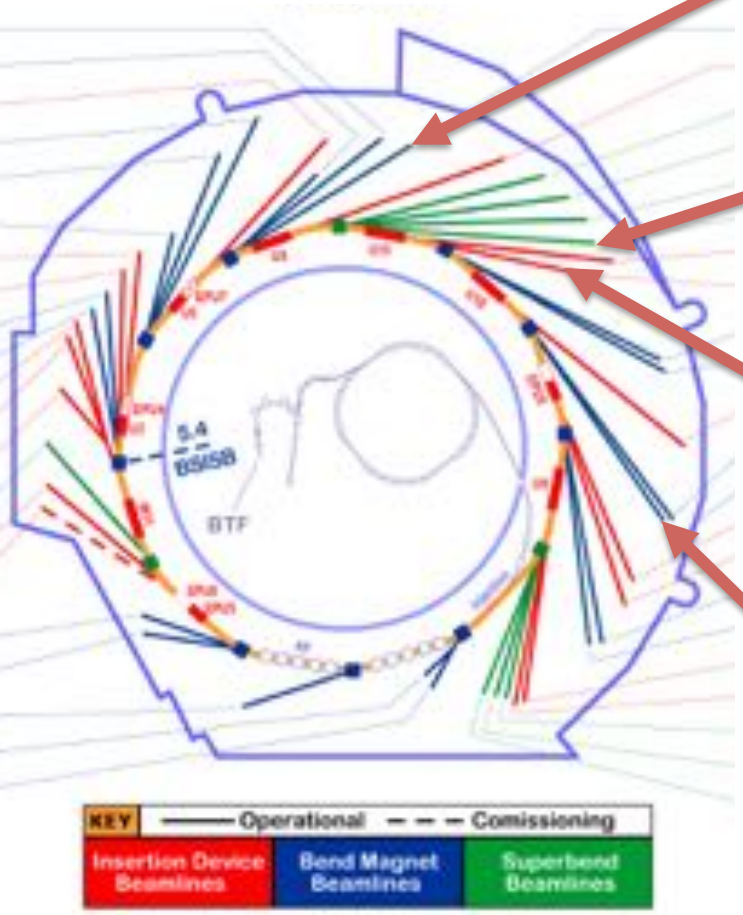




A few of the light source user facilities

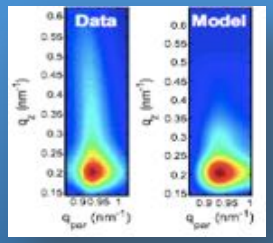
Canadian Light Source: https://www.flickr.com/photos/clsresoff/8498996062

ALS



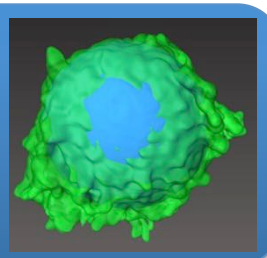
Reciprocal Space
(Scattering)

HipGISAXS/HipMC
parallel Scattering



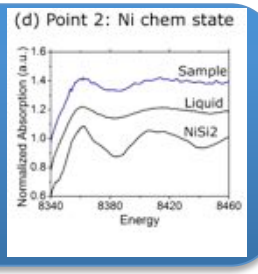
Real Space
(Tomography)

Arec3d, QuantCT,
CrunchFlow



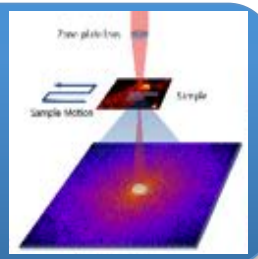
Spectroscopy
(MicroXas)

ShirleyXAS (MSD)
BerkeleyGW (NERSC)



Hybrid
(COSMIC)

Ptychographic
reconstruction



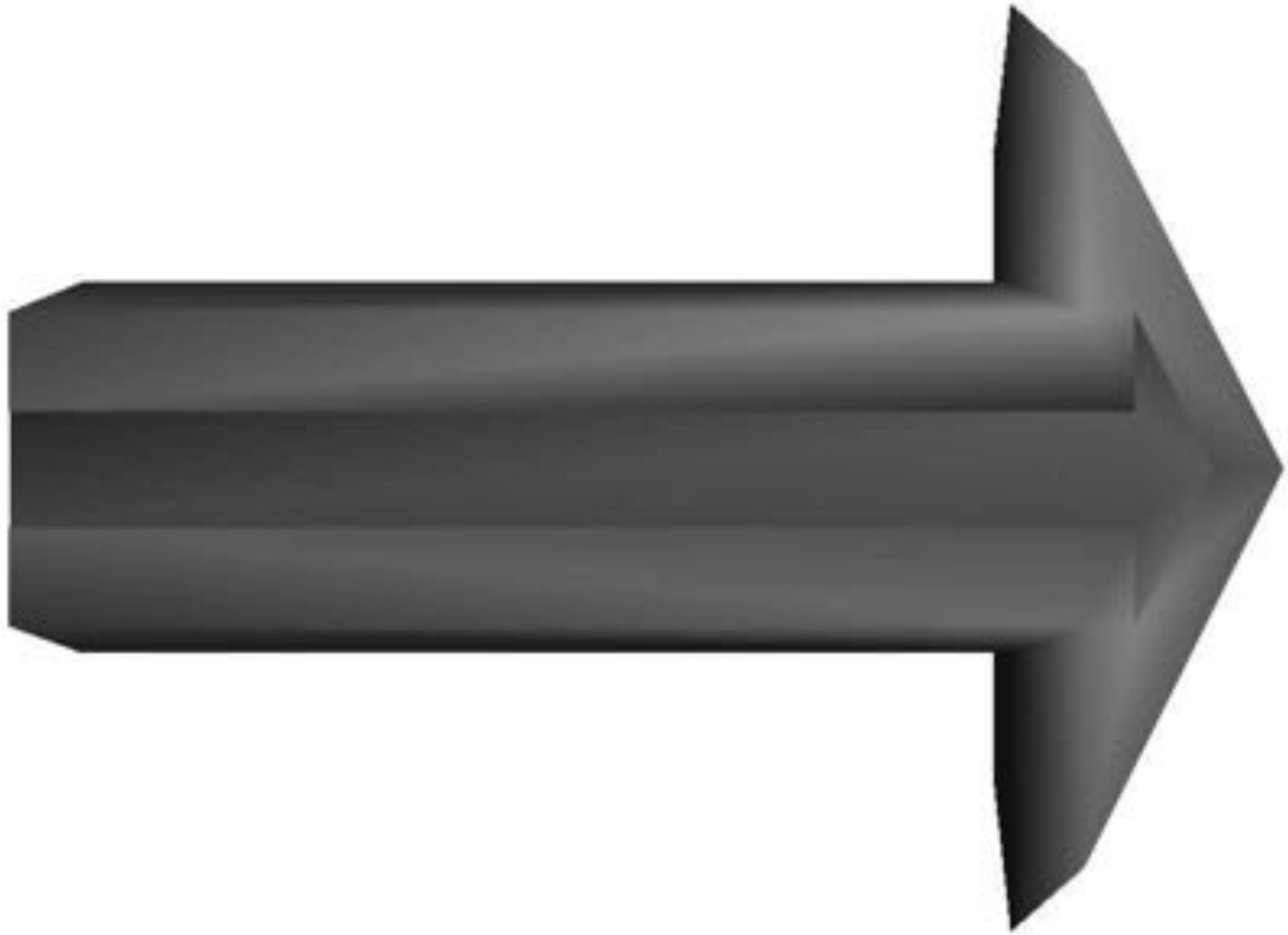
ALS, NCXT, CXRO, BCSB, etc.

BERKELEY LAB
Lawrence Berkeley National Laboratory
MICROCT
.LBL.GOV

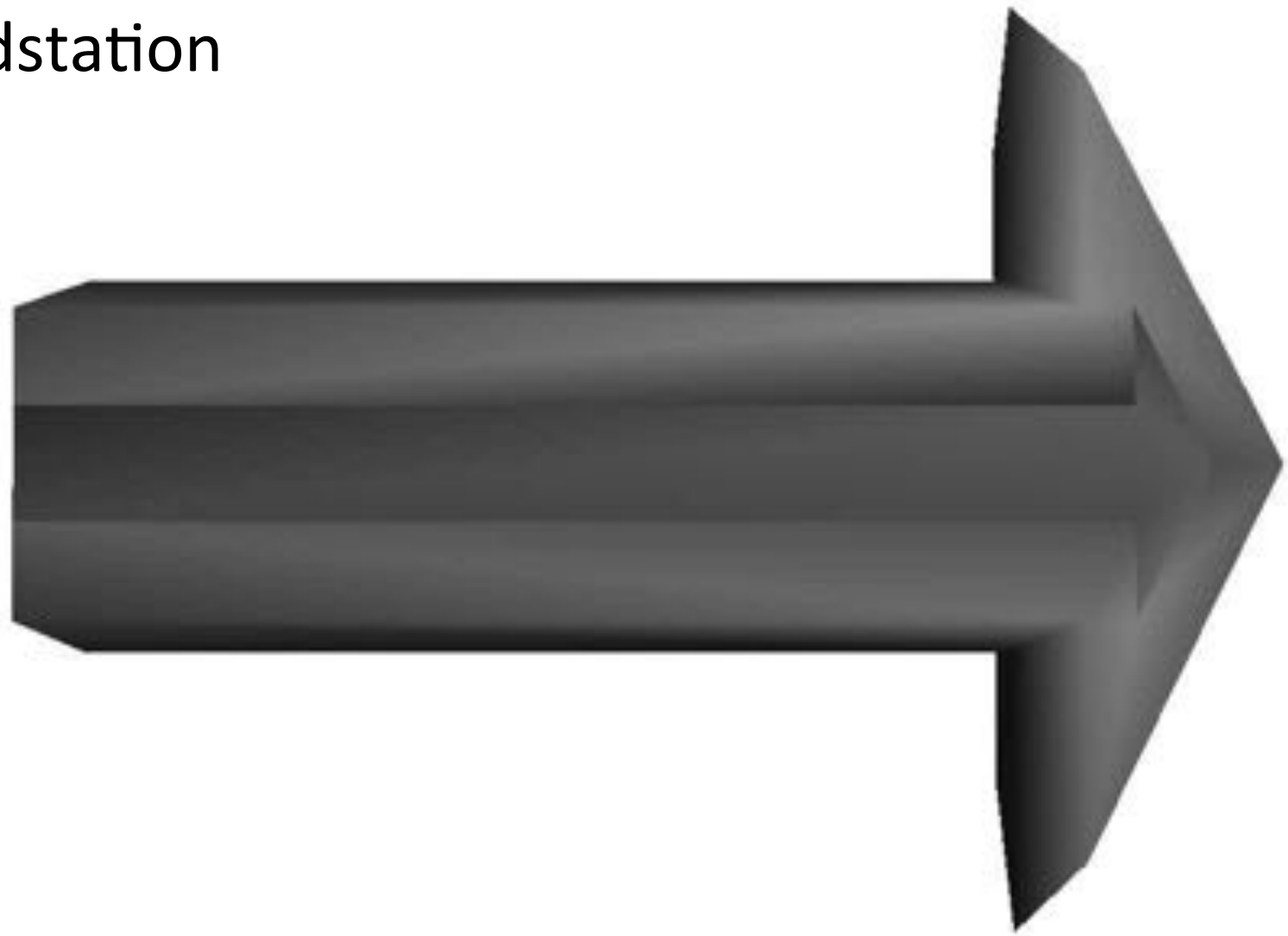
(Oral history, as told to me by Alastair MacDowell)

A BRIEF HISTORY OF SYNCHROTRON USER FACILITIES

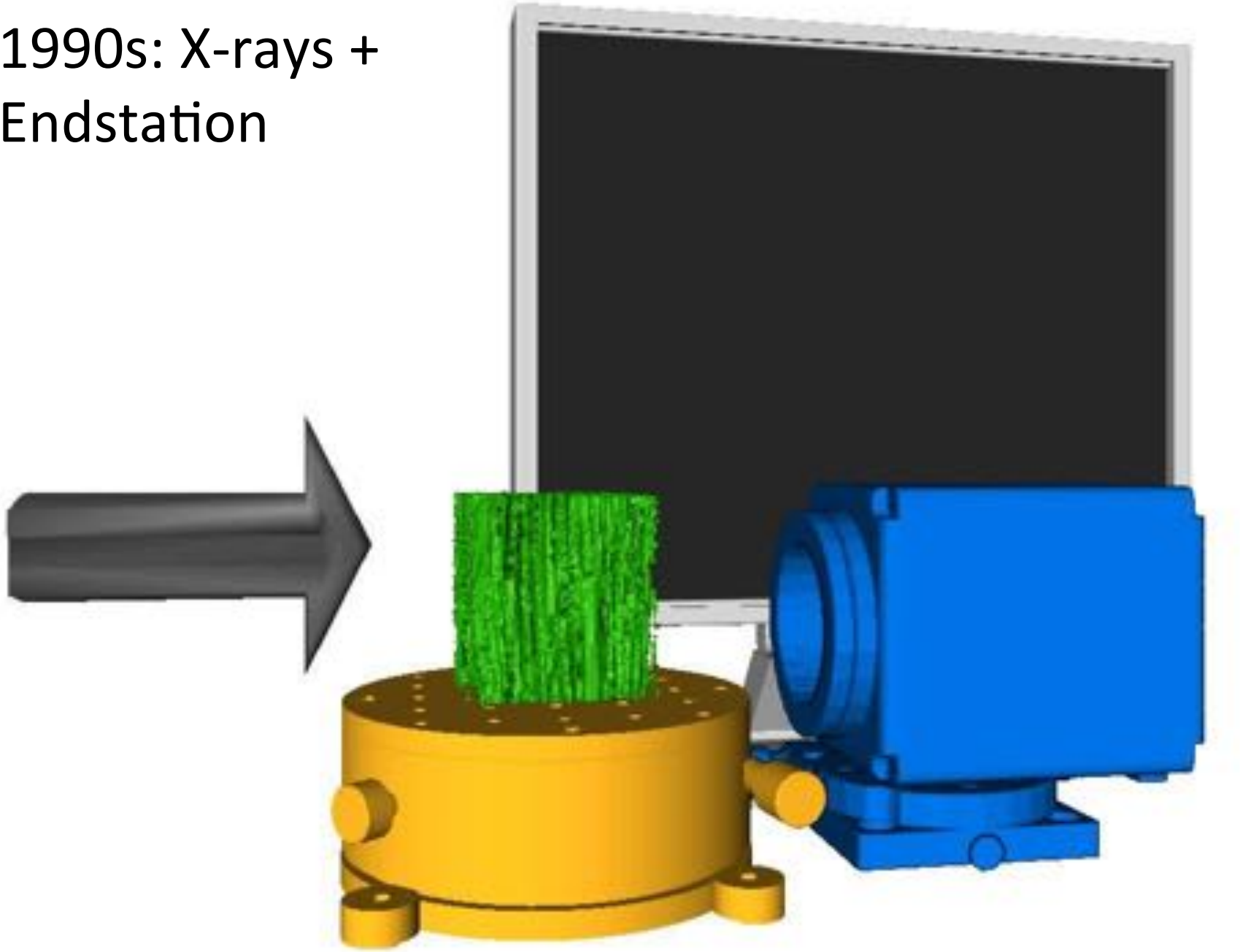
1980s: X-rays



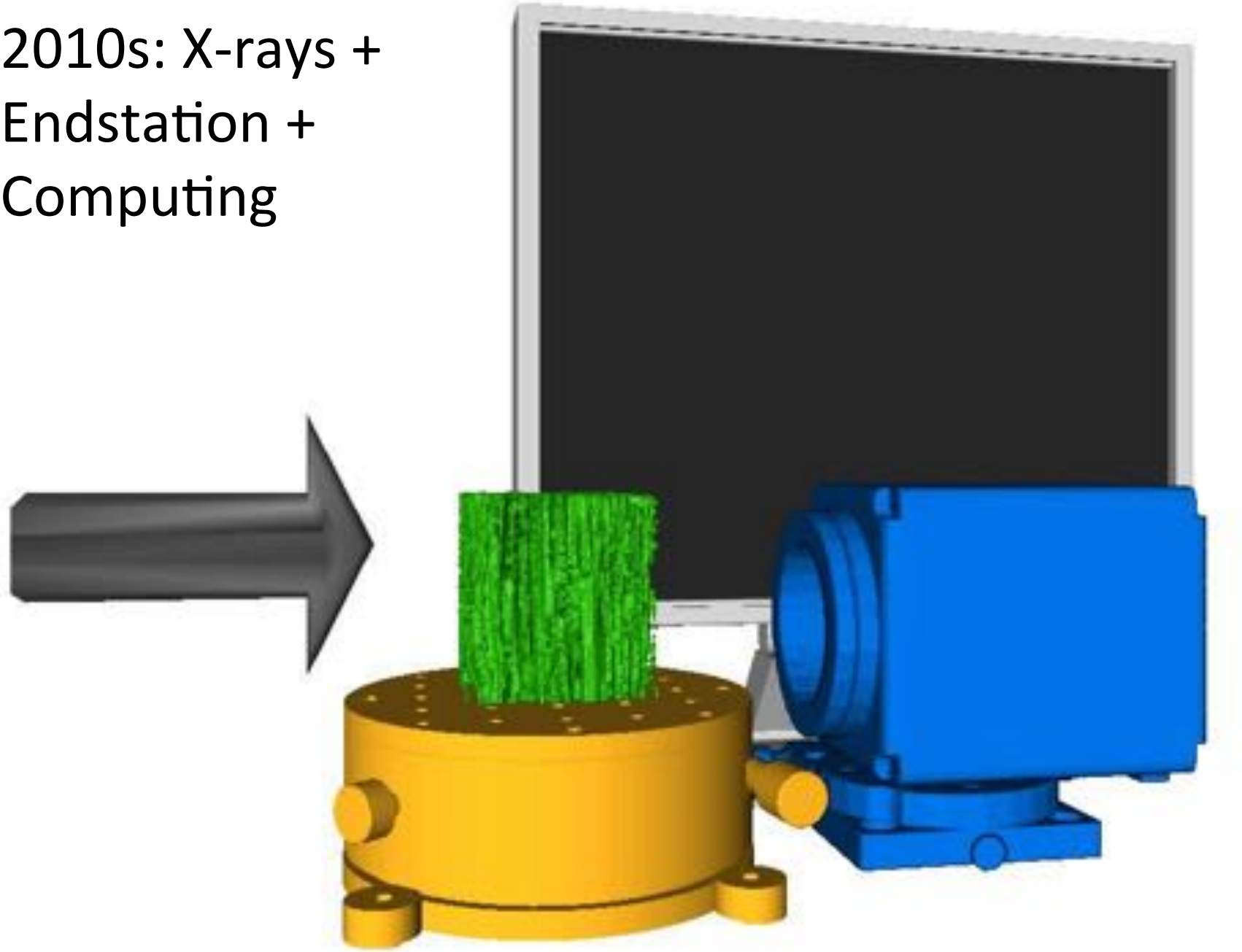
1990s: X-rays + Endstation



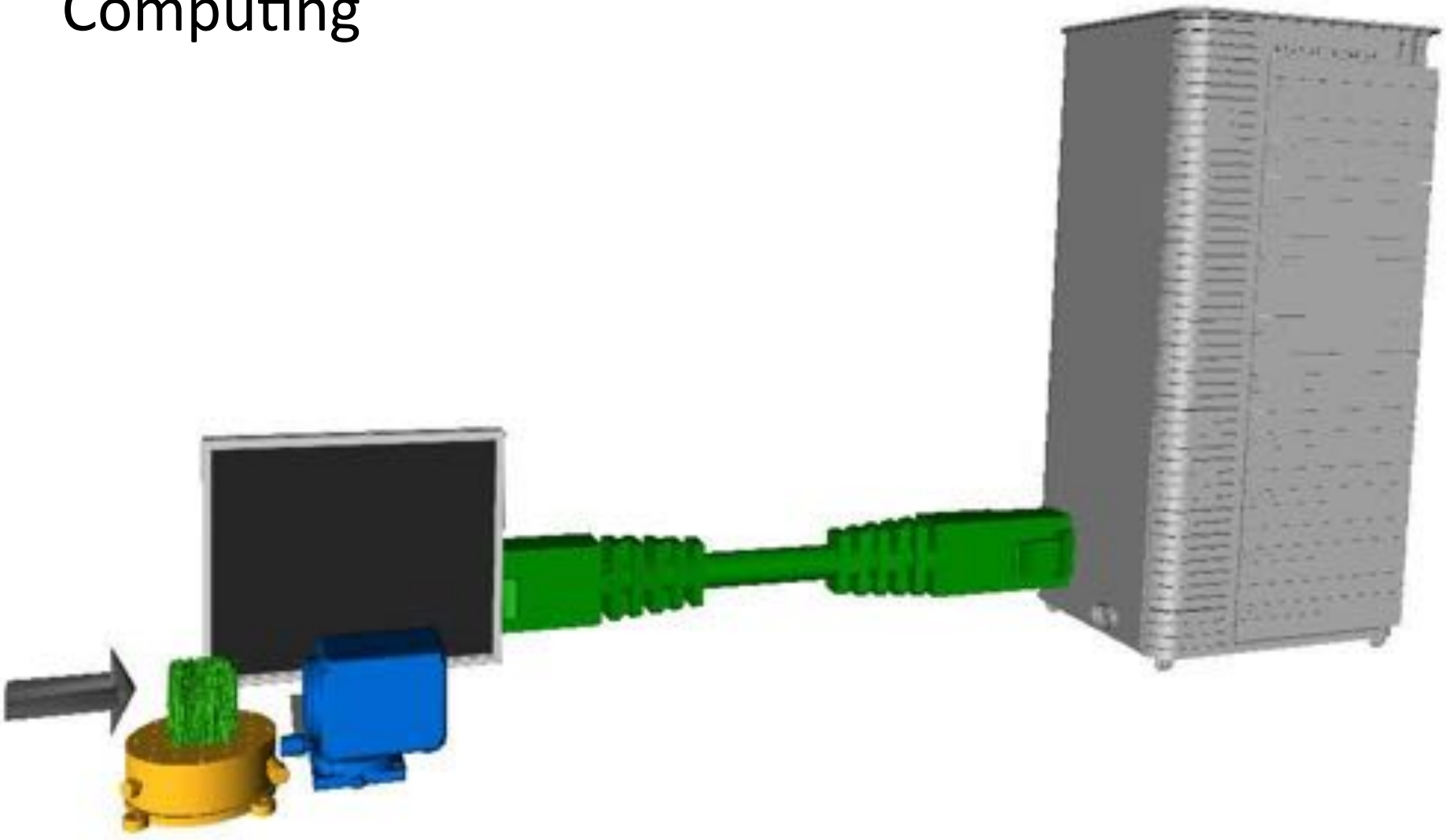
1990s: X-rays + Endstation



2010s: X-rays + Endstation + Computing

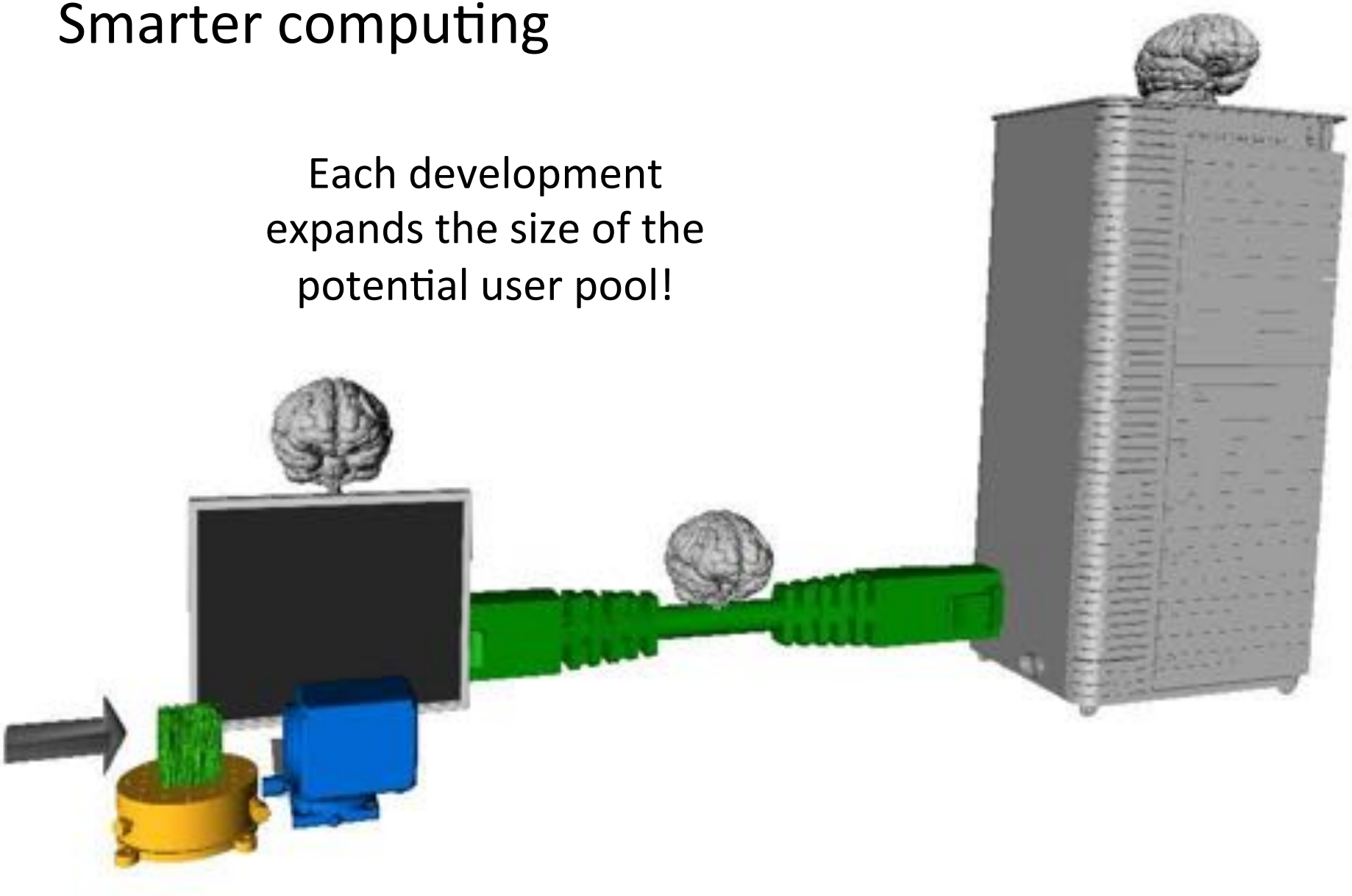


2010s: X-rays +
Endstation +
Computing

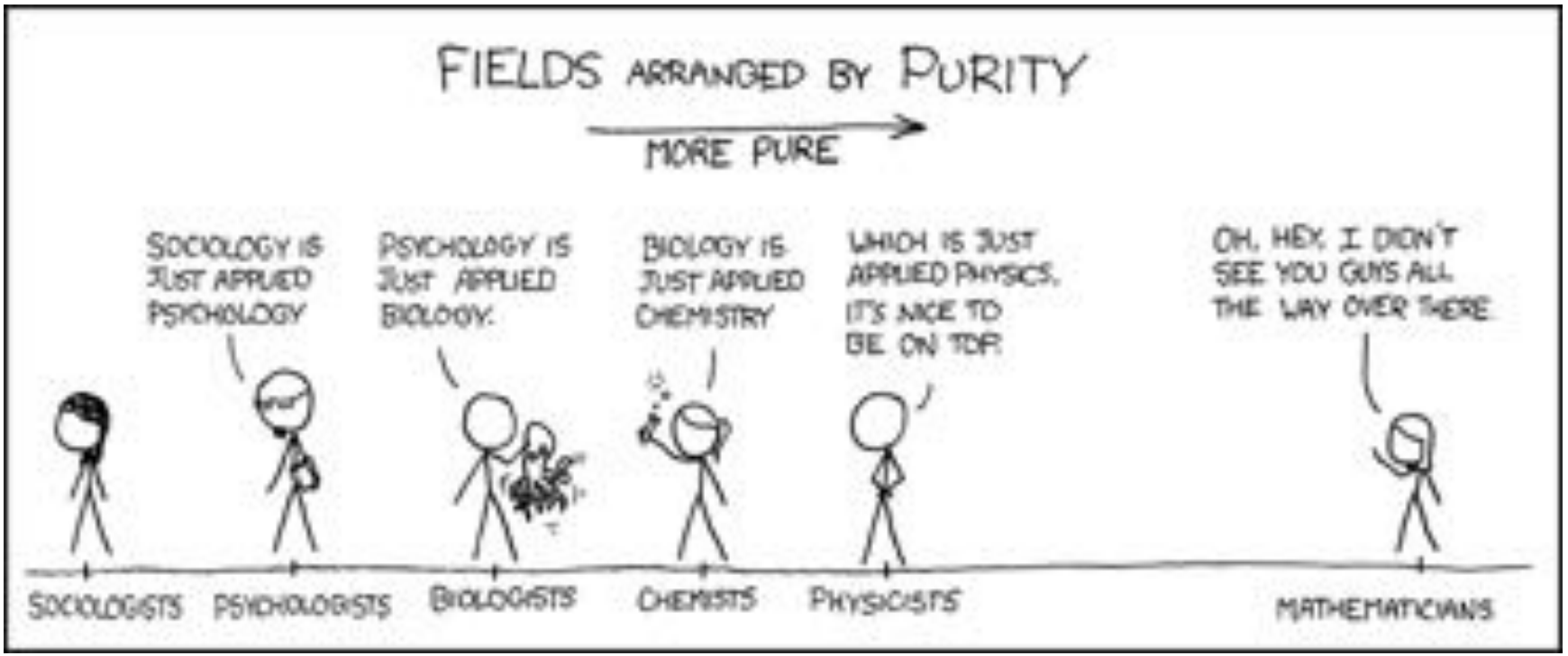


The Future:
Smarter computing

Each development expands the size of the potential user pool!



Bringing in increasingly impure users…



“Physicists seem to phrase things in a solvable way.” -Alastair MacDowell

ALS User Community

Oct	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	S	Su	M	T	W	Th	F	S	Su	M	T	W	Th	F	S	Su	M	T	W	Th	F	S	Su	M	T	W	Th	F	S	Su	M
0100-0900	DP	IZ	IT	M/I	M/I	M/I	S/T	IT	IT		NB	MM	TW	TW	PM	AM	AP	AM	AM	AM	LW	JA	JA	AP	M/I		AW	MC	MC	YG	YG
0900-1700	IZ	MH	M/I	M/I	M/I	M/I	IT	IT	IT	NB	MM	TW	TW	JS	AM	AP	AM	AM	AM	LW	JA	JA	AP	M/I	M/I	AW	MC	MC	YG	YG	AP
1700-0100	IZ	MH	M/I	M/I	M/I	S/T	IT	IT	IT	NB	MM	TW	TW	TWc	AM	AP	AM	AM	AM	LW	JA	JA	AP	M/I	S/T	AW	MC	MC	YG	YG	AP

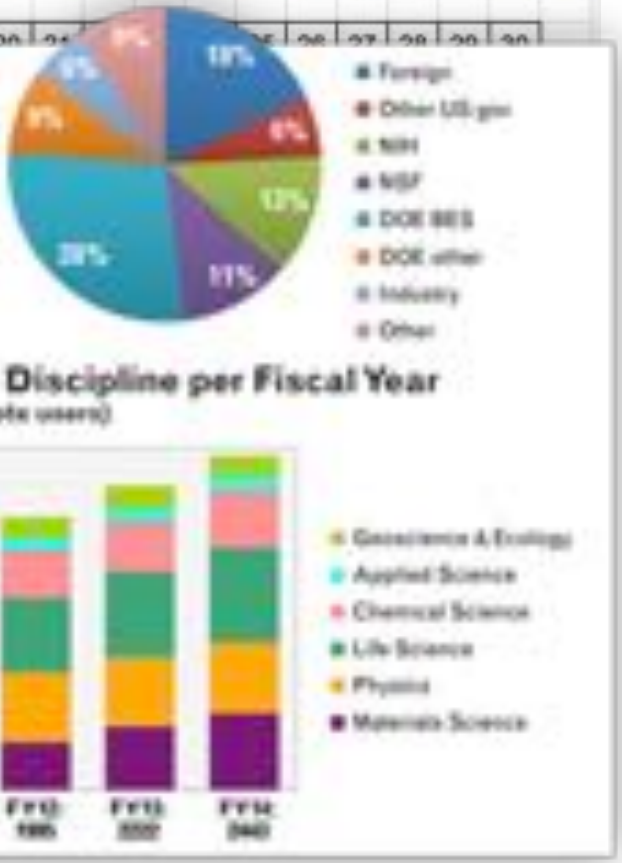
Nov	1	2	3	4	5	6	7	8
	T	W	Th	F	S	Su	M	T
0100-0900	AP	TL	TL	DM	DM	FZ	IT	M/I
0900-1700	TL	TL	DM	DM	FZ	MH	M/I	M/I
1700-0100	TL	TL	DM	DM	FZ	MH	M/I	S/T

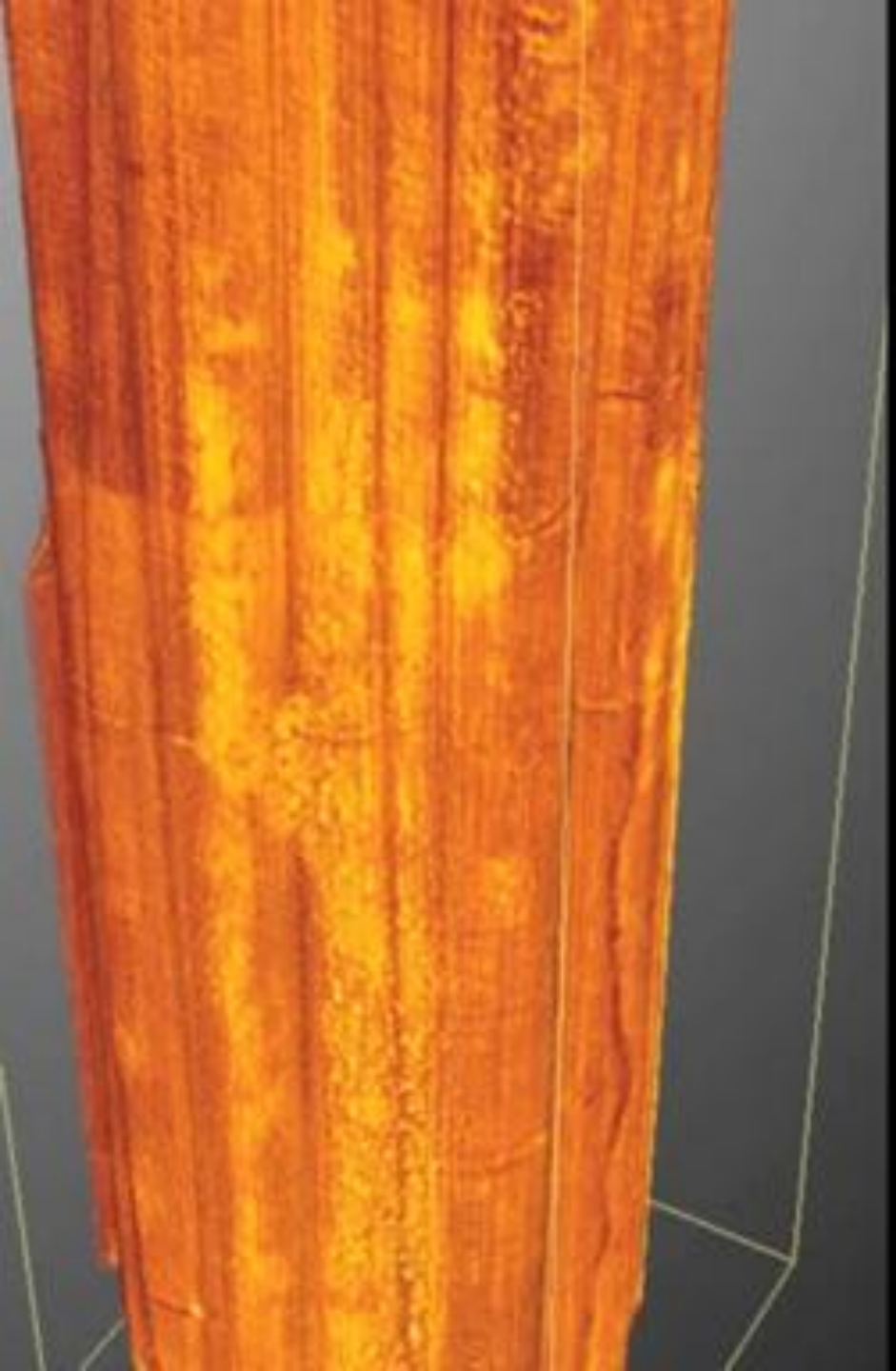
Dec	1	2	3	4	5	6	7	8	9
	Th	F	S	Su	M	T	W	Th	F
0100-0900	NB	AW	JA	JA	JA	AP	RR	RR	RR
0900-1700	AW	JA	JA	JA	AP	RR	RR	RR	SCh
1700-0100	AW	JA	JA	JA	AP	RR	RR	RR	SCh

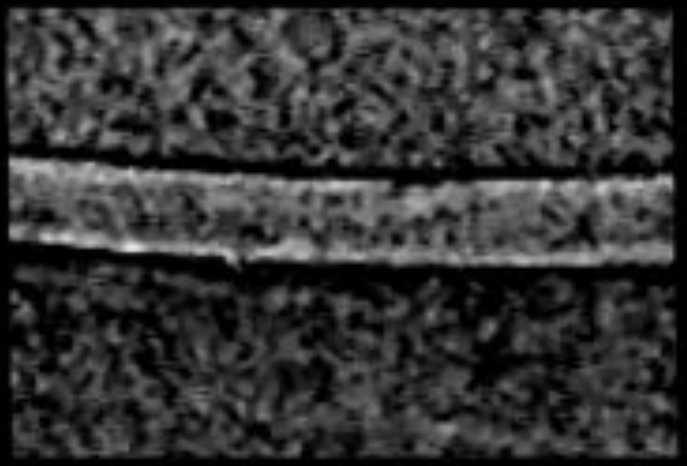
USER STATS

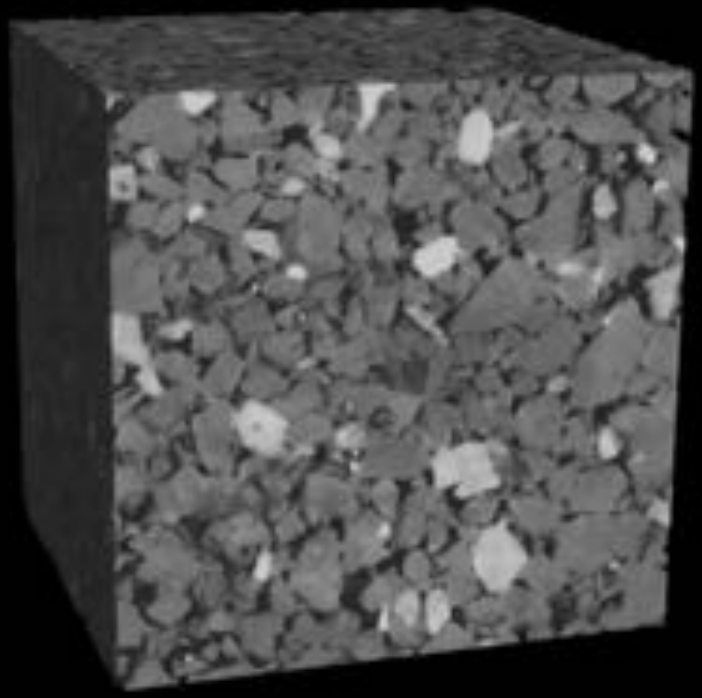
50-100
Users on site at any one time

1 hour to 10 days
Average stay of users









User time is increasingly spent on computing (with decreasing expertise)

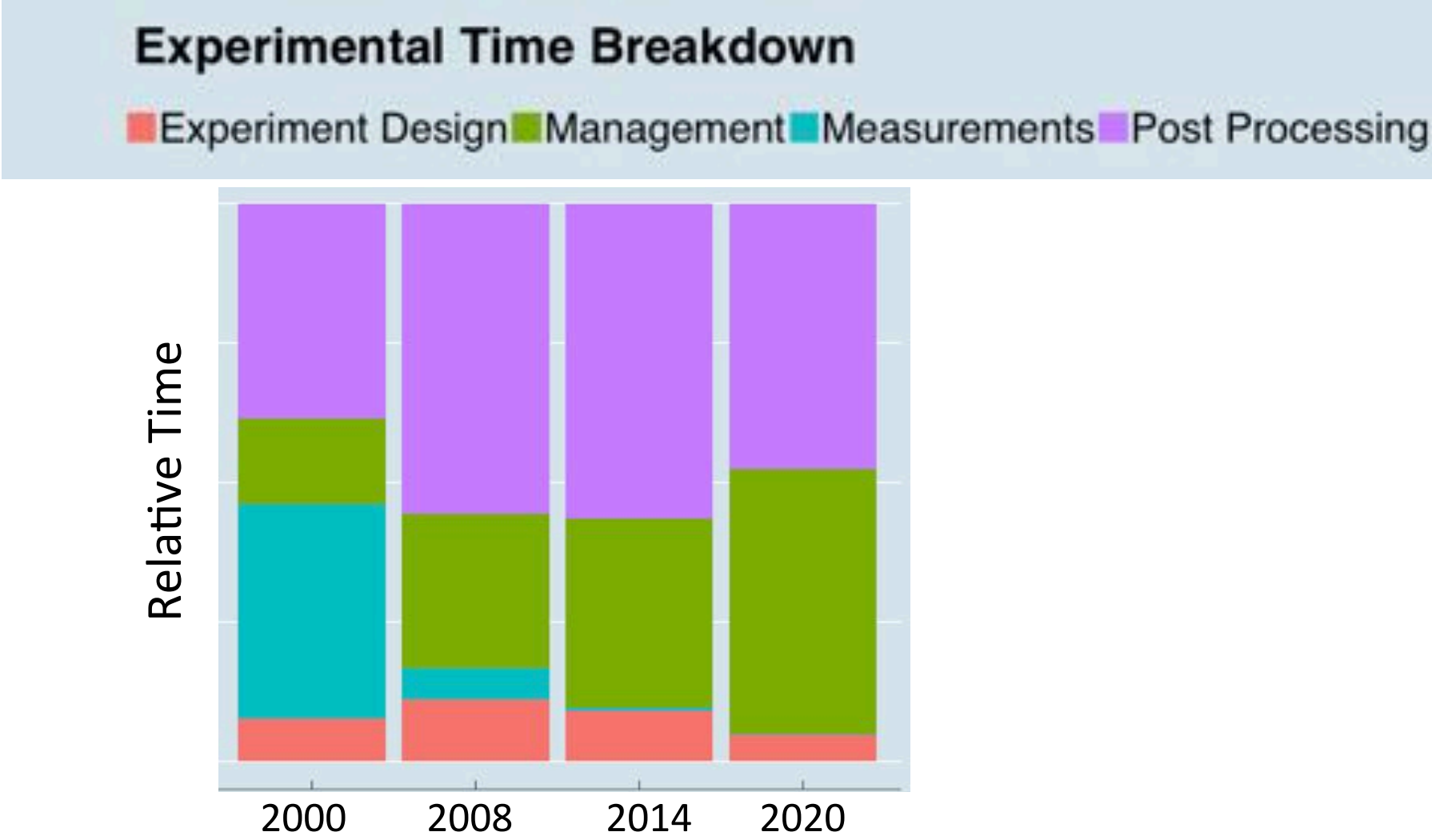


Figure from Kevin Mader
https://rawgit.com/4Quant/SRI2015/master/SRIPres.html#/

User time is increasingly spent on computing (with decreasing expertise)

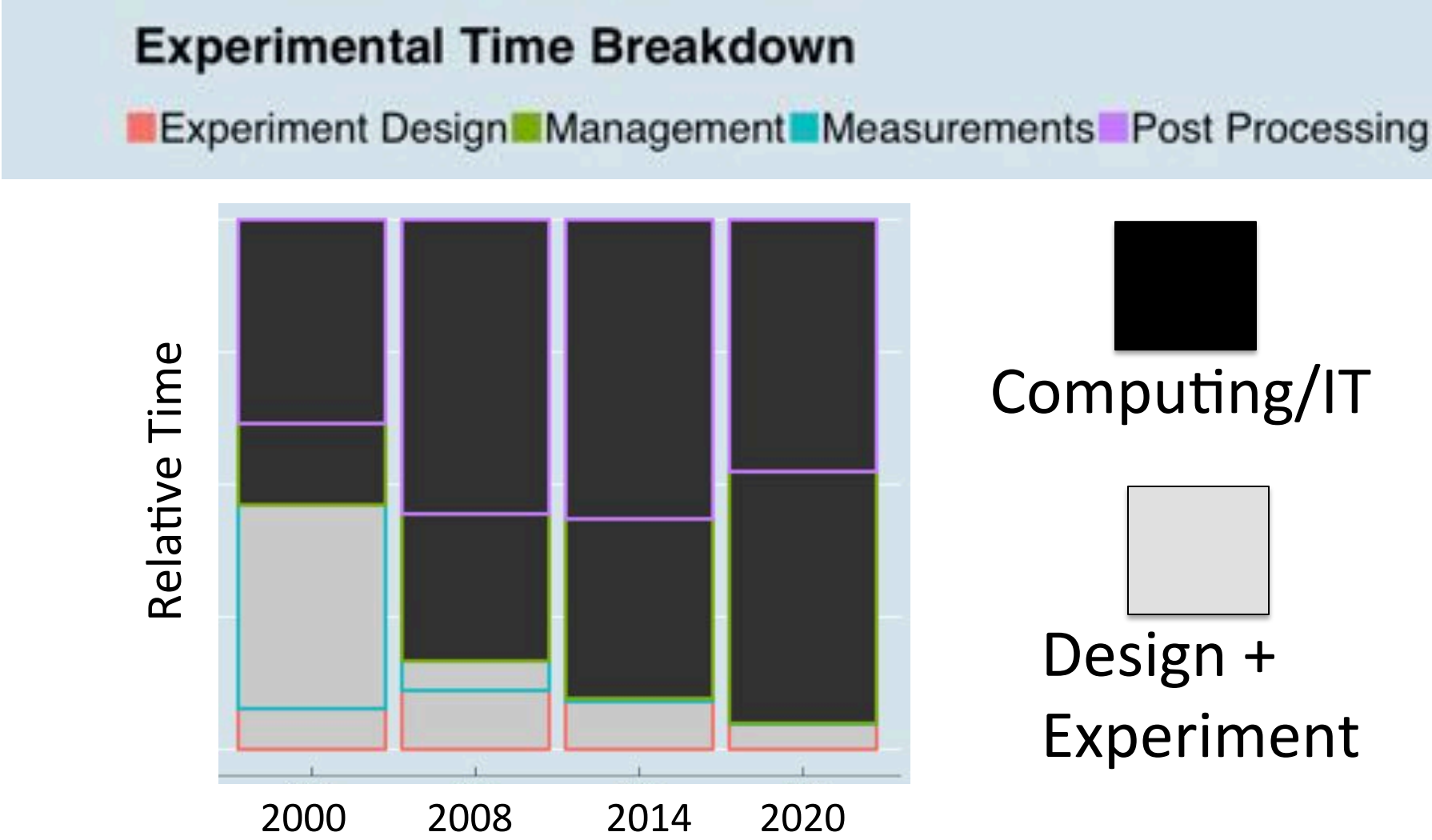


Figure from Kevin Mader
https://rawgit.com/4Quant/SRI2015/master/SRIPres.html#/

BERKELEY LAB
Lawrence Berkeley National Laboratory
MICROCT
.LBL.GOV

COMMON COMPUTING SCENARIOS AT LIGHT SOURCES

Case 0: Users who collect data

"Not processing your data in real time is the first step towards not processing your data at all."

– Peter Zwart

- Many users don't have access to the computers/software on their own
- >20% new users

Babylonian tablet; Image never published

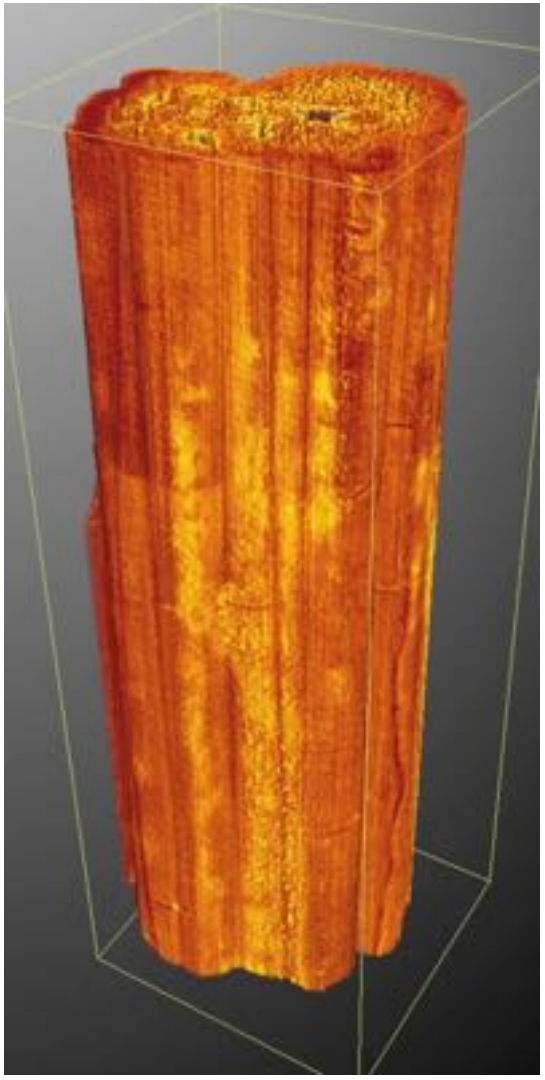
Case 1: Users who collect lots of data

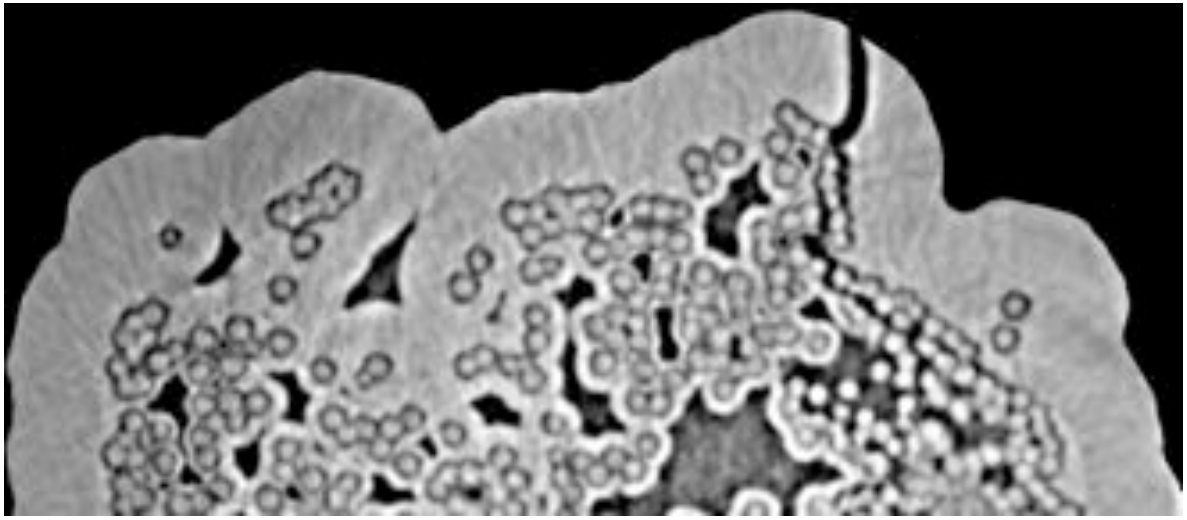
- >900, 3D time points (~3TB) from 1 day of data collection

Lava in situ, Wim de Gruyter, Benoit Cordonnier, Michael Manga (UC Berkeley)

Case 2: Experiment feedback

- Does the data look ok?
- When should I change parameters (speed up, slow down, zoom in)?
- Acquisition automation

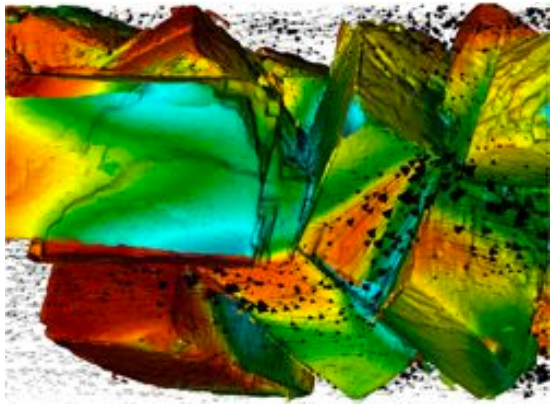




HA Bale, RO Ritchie, et al, Nature Materials, 2013, 12(1) 40-46

Case 3: The "Digital Twin"

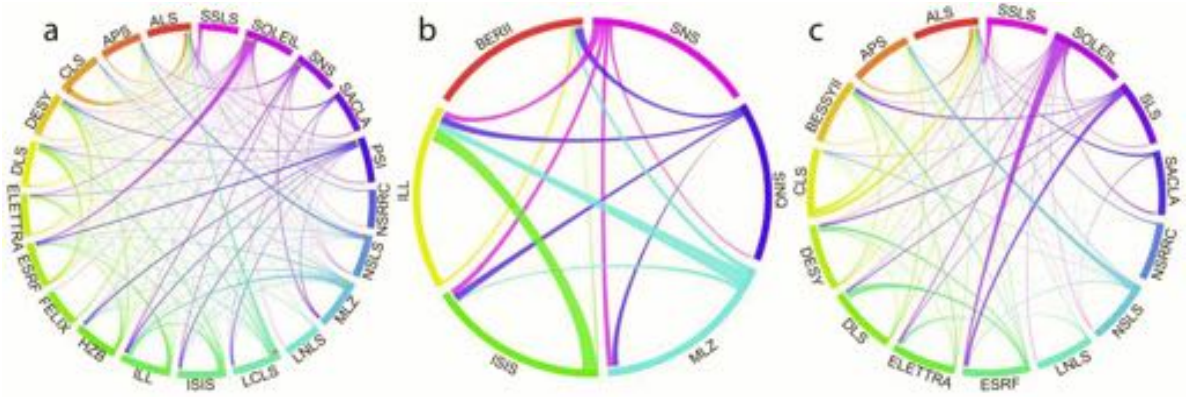
- After collecting initial data, create model and launch simulation
- The simulation gives the direction for experiment control and automation

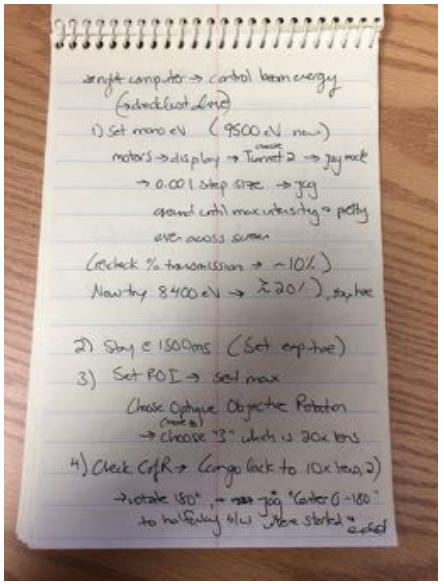


EJ Tuegel, et al, Intl. J Aerospace Engineering, Vol. 2011, Article 154789
HA Bale, RO Ritchie, et al, Nature Materials, 2013, 12(1) 40-46
S Molins, D Trebotich, et al. Environ. Sci. Technol., 2014, 48 (13), pp 7453–7460

More Scenarios

- Collecting fast or low quality scans, rely on algorithms to rescue
- Using machine learning to sort/search/mine features in data
- Sharing/comparing data across techniques and facilities
- Adding and linking data and metadata from many sources





F Schluenzen et al, Resesarch Data Alliance

BERKELEY LAB
Lawrence Berkeley National Laboratory
MICROCT
.LBL.GOV

Challenges

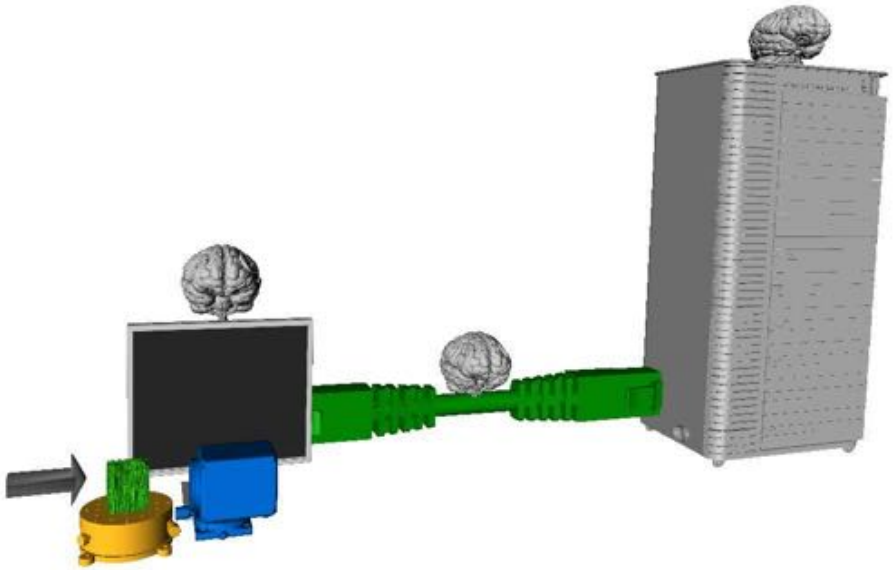
- Lots of small groups
  - hard for individual users to justify development expense
  - Average computing hardware and skill of light source users not high
    - laptops, Windows
    - 20% new users
  - Large variety of computing needs, best matched to different hardware (e.g. GPU, etc.)
- In U.S., not enough computing investment and planning by BES
  - many light sources are retrofitting computing/network for working instruments
- Lots of existing code that is difficult to re-use (or not available)
  - lots of reinventing the wheel
- Not enough communication
  - amongst light sources
  - light sources and users
  - light sources with computing, HEP, etc.

BERKELEY LAB
Lawrence Berkeley National Laboratory
MICROCT
.LBL.GOV

How to proceed?

- Some standards (HDF5, Message Queues, Globus)
- Small reusable tools (rather than monolithic)
- Develop software with an eye to sharing/collaborating (open frameworks and architectures)
- Continued resources towards developing computing, working together and sharing code
  - start Computing cHallenges in Light Sources (CHILS) conferences?
- Collaborate with CHEP in future areas?
  - machine learning/neuromorphic computing
  - streaming data
- Interchange of staff between fields

BERKELEY LAB
Lawrence Berkeley National Laboratory
MICROCT
.LBL.GOV



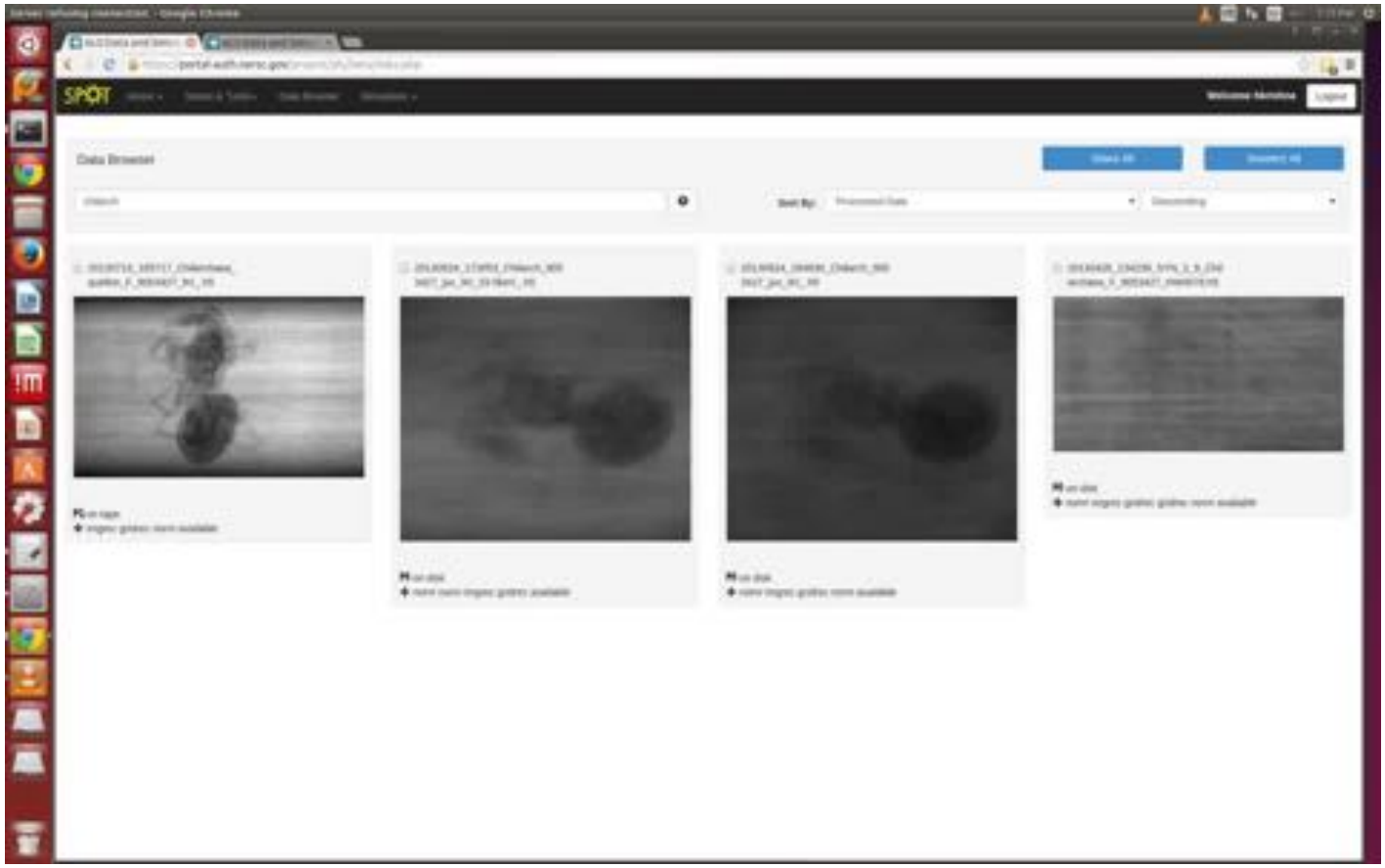
EFFORTS AT ALS/BERKELEY LAB

Collaboration

- Light Source Facilities
  - Advanced Light Source
- Network
  - LBLnet, ESnet, 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)
- Users

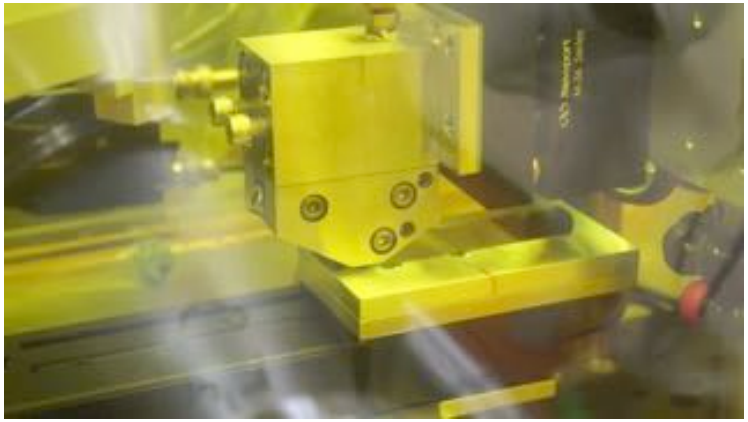
SPOT Suite: data management, automated workflow, sharing, data access, remote 3D viz

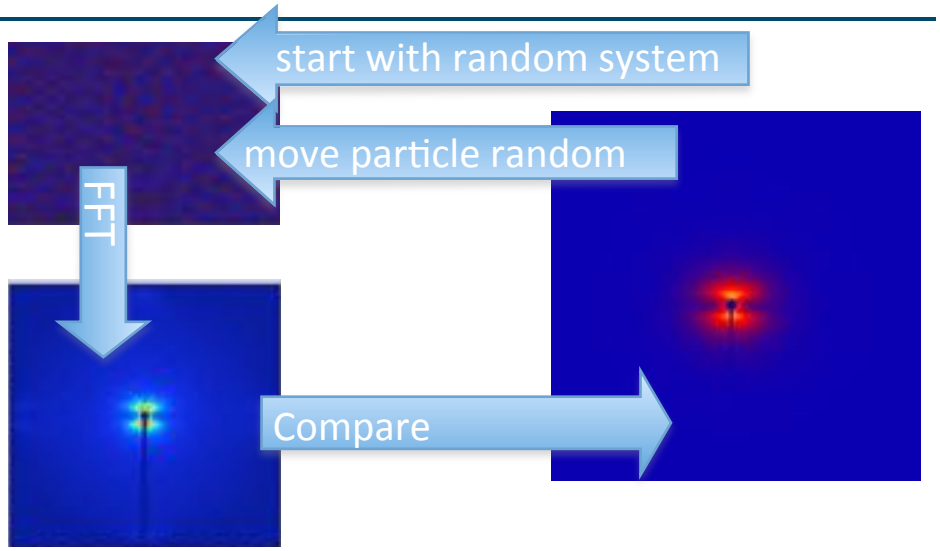
- Uses HDF5, RabbitMQ, SPADE
- Built by CHEP'ers, Craig Tull, Simon Patton

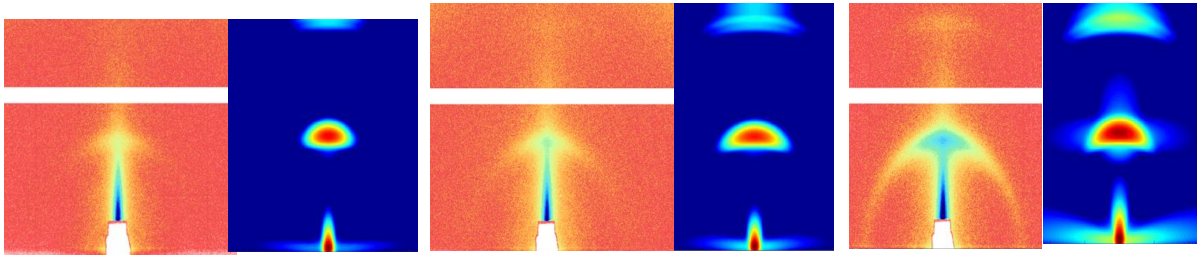


Screen capture courtesy Hari Krishnan

High Performance Scattering Simulations During Beamtime



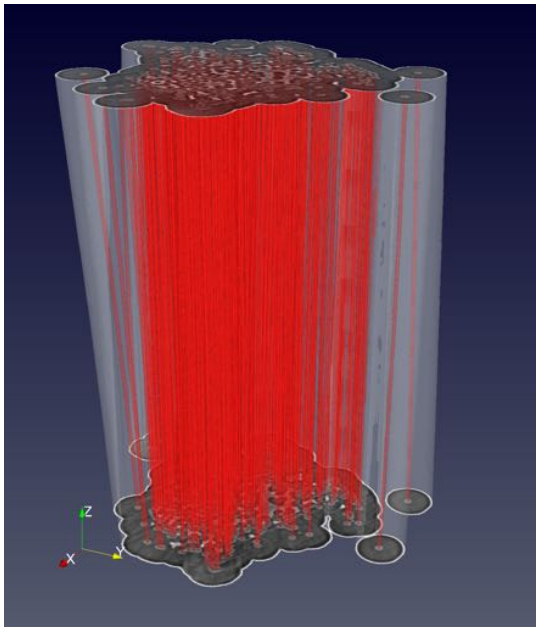




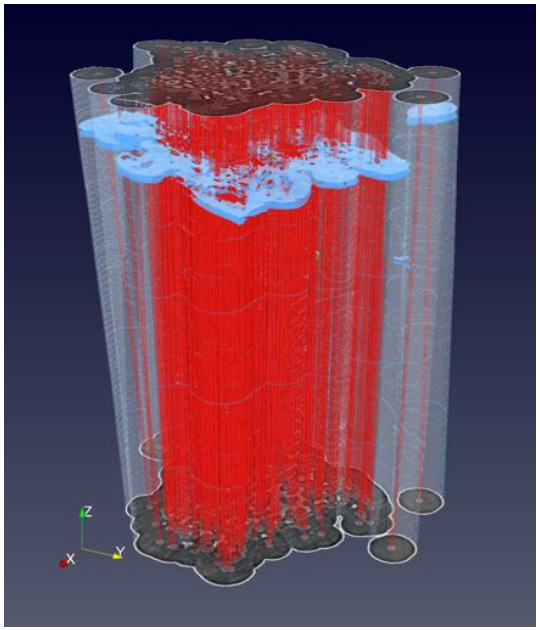
A. Hexemer (LBNL/CAMERA), C.E.Tull (LBNL), J. Deslippe (NERSC), R.S. Canon (NERSC), E. Dart (ESnet), I.Foster (ANL), J.A. Sethian (LBNL/CAMERA), G. Shipman (ORNL), J. Wells (ORNL), K. Kleese van Dam (PNNL), T.P. Russell (UMass), E. Gomez (PennState)

Facilities: ALS (BES), NERSC (ASCR), ANL(ASCR), OLCF (ASCR), ESnet (ASCR), CAMERA (ASCR

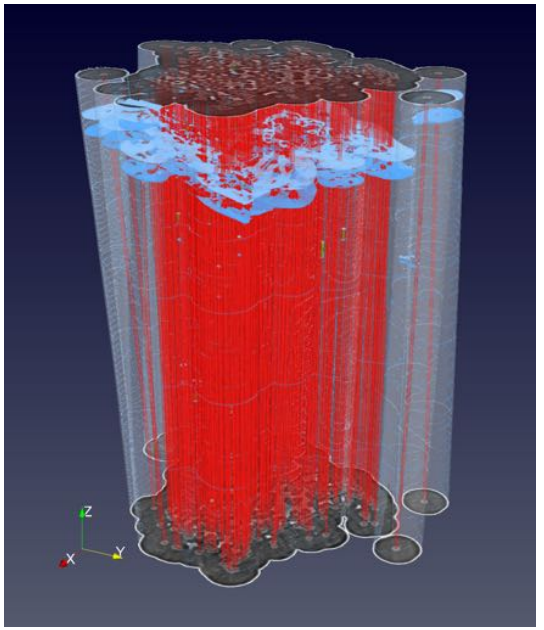
Fiber and crack analysis through pattern matching



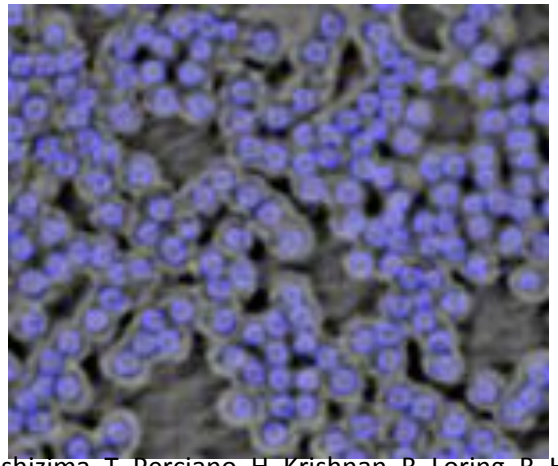
93N

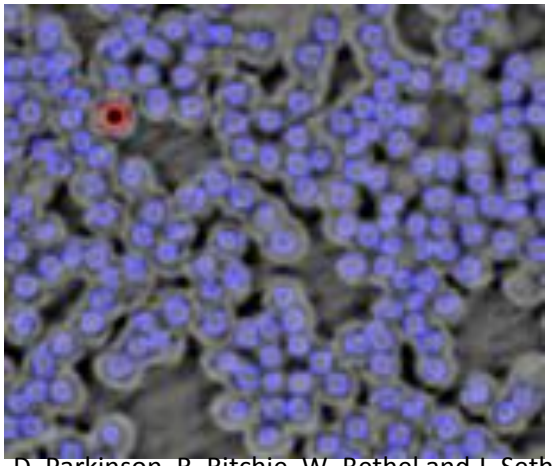


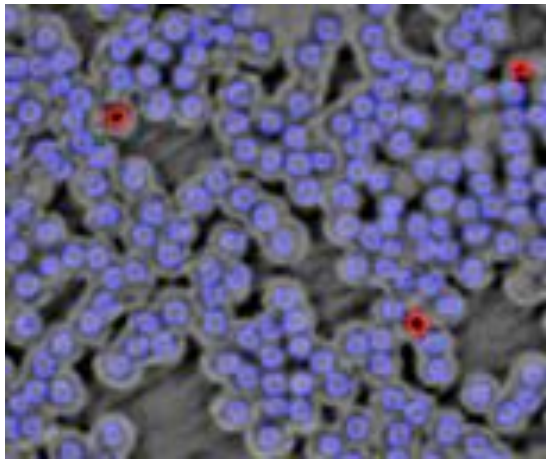
133N



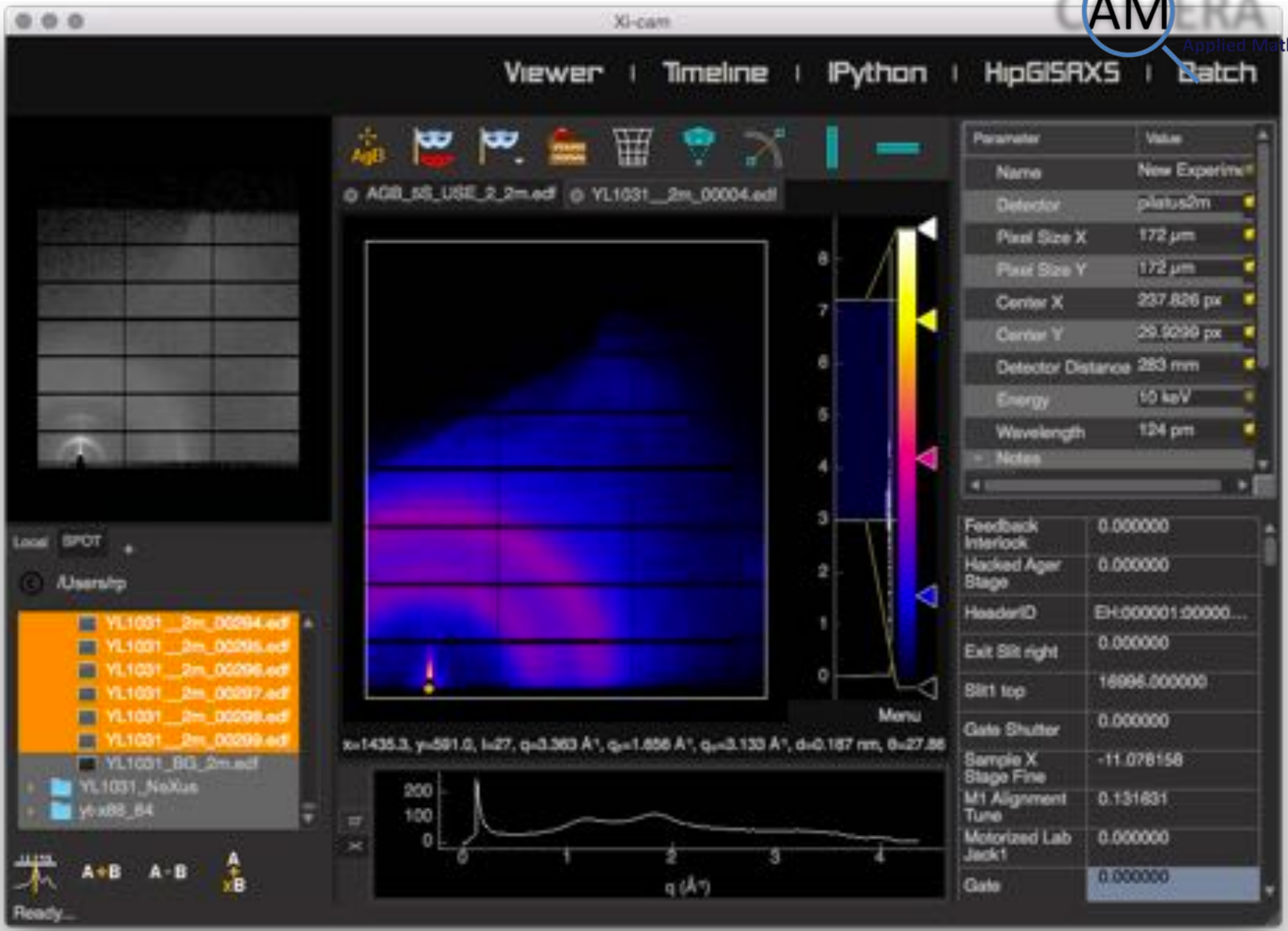
151N



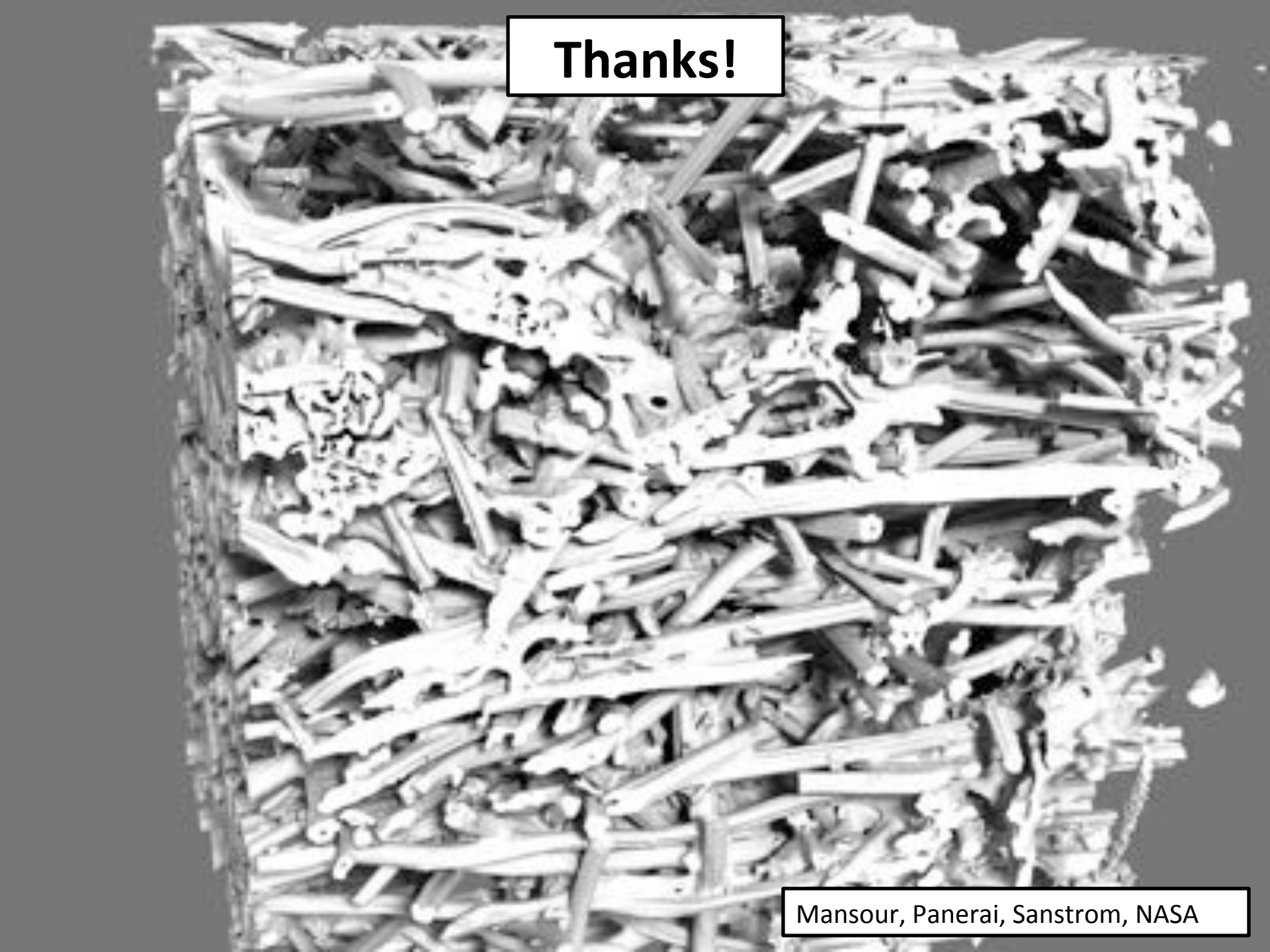




D. Ushizima, T. Perciano, H. Krishnan, B. Loring, B. Hrish, D. Parkinson, R. Ritchie, W. Bethel and J. Sethian,

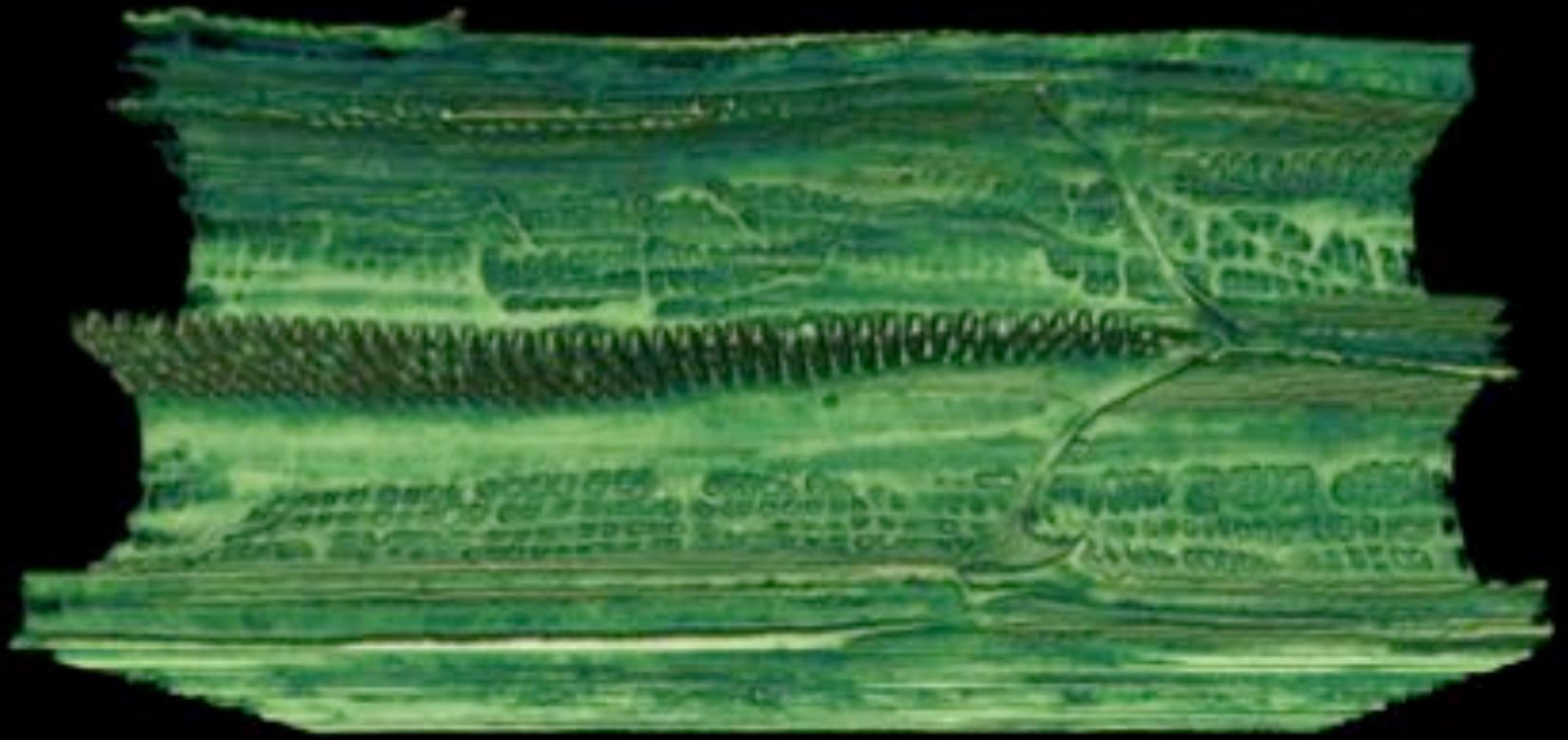

Xi-cam
Viewer
Timeline
IPython
HipGISAXS
Batch
AM
Applied Math
AgB
AGB_sS_USE_2_2m.edf
YL1031__2m_00004.edf
Parameter
Value
Name
New Experim
Detector
pilatus2m
Pixel Size X
172 µm
Pixel Size Y
172 µm
Center X
237.826 px
Center Y
29.9299 px
Detector Distance
283 mm
Energy
10 keV
Wavelength
124 pm
Notes
Local
SPOT
/Users/rp
YL1031__2m_00294.edf
YL1031__2m_00295.edf
YL1031__2m_00296.edf
YL1031__2m_00297.edf
YL1031__2m_00298.edf
YL1031__2m_00299.edf
YL1031_BG_2m.edf
YL1031_NeXus
yb-x86_64
Feedback Interlock
0.000000
Hacked Ager Stage
0.000000
HeaderID
EH:000001:00000...
Exit Slit right
0.000000
Slit1 top
16996.000000
Gate Shutter
0.000000
Sample X Stage Fine
-11.078158
M1 Alignment Tune
0.131631
Motorized Lab Jack1
0.000000
Gate
0.000000
Menu
A+B
A-B
A xB
Ready...
q (Å⁻¹)

Thanks!



Mansour, Panerai, Sanstrom, NASA

Thanks!



McElrone (Davis), Brodersen (Yale)