Workflows for experimental facilities and the digital twin approach

IPAM Long Program: New Mathematics for the Exascale: Applications to Materials Science

Workshop III: Complex Scientific Workflows at Extreme Computational Scales

Amedeo Perazzo, Jana Thayer
May 1st, 2023

SLAC NATIONAL ACCELERATOR LABORATORY

Stanford University | U.S. DEPARTMENT OF ENERGY

Contents

Introduction to the Linac Coherent Light Source (LCLS) - the most powerful X-ray Free electron Laser in the world

- Accelerator and instruments
- Beam parameters

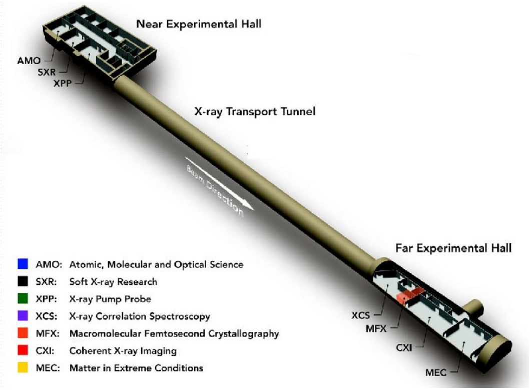
Material science experiments at an XFEL

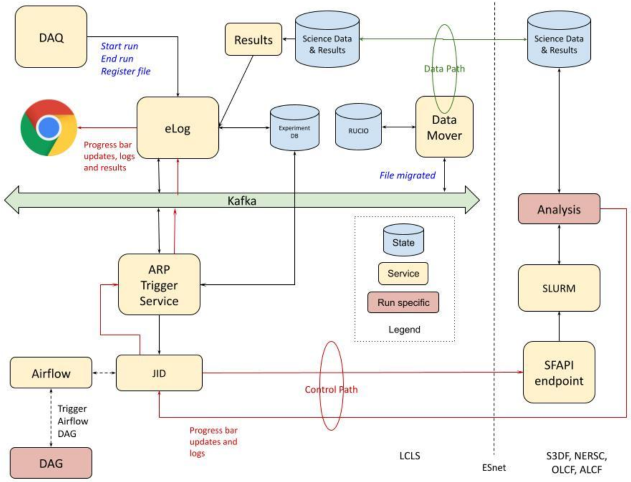
Complex workflows for XFELs

- Interfacing the LCLS beamlines to a supercomputer

Introduction to the digital twin approach

- Integrating simulations with experiments in real-time

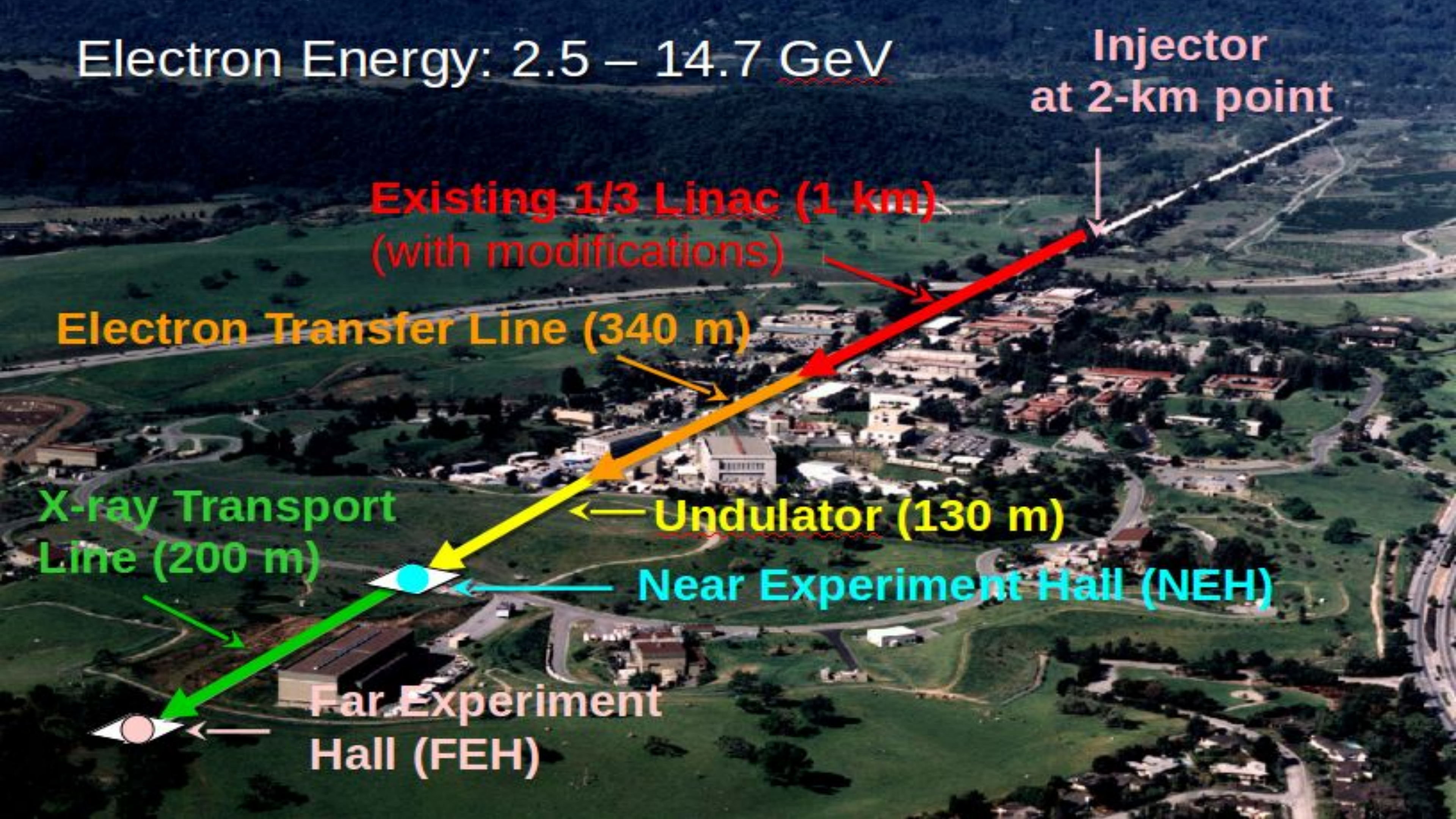




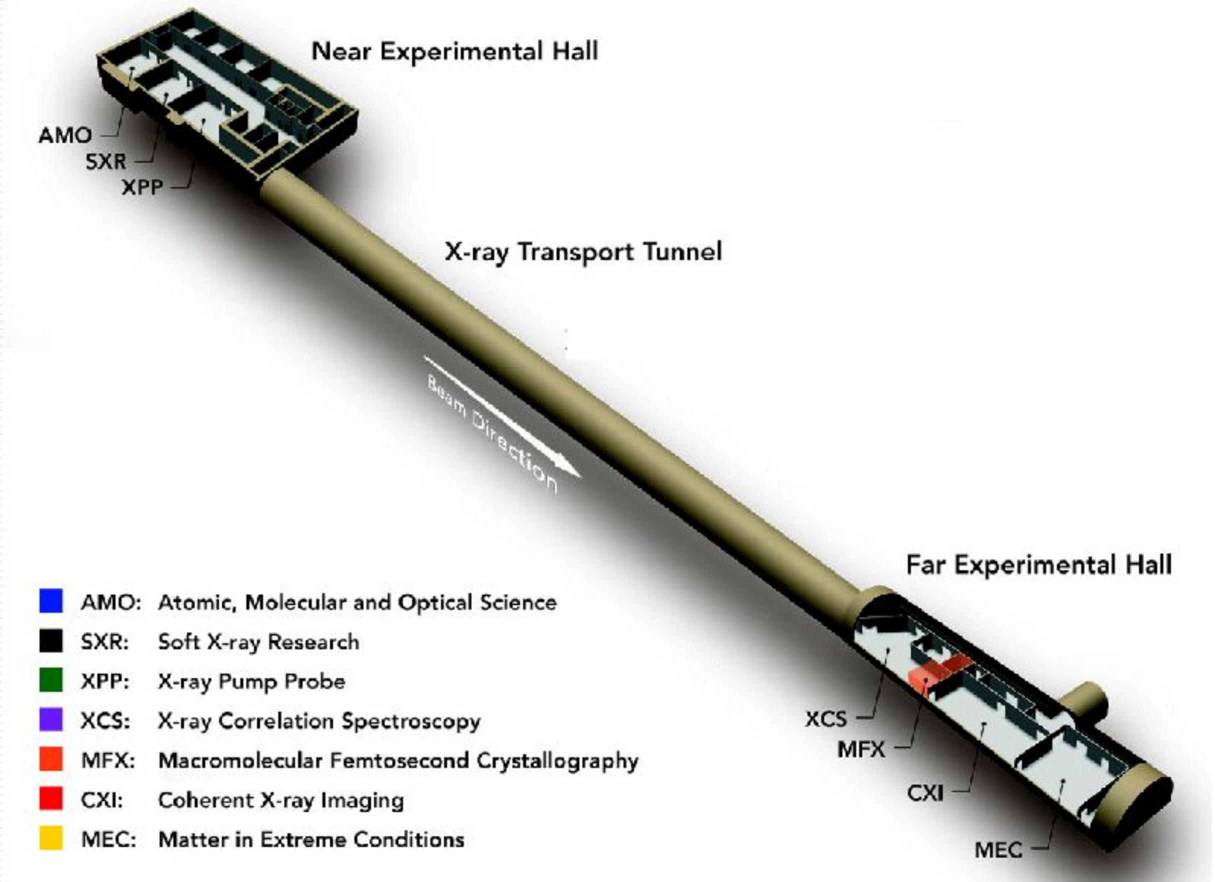
SLAC

Introduction to the Linac Coherent Light Source (LCLS)

The most powerful X-ray Free electron Laser in the world


Electron Energy: 2.5 – 14.7 GeV
Injector
at 2-km point
Existing 1/3 Linac (1 km)
(with modifications)
Electron Transfer Line (340 m)
X-ray Transport
Line (200 m)
Undulator (130 m)
Near Experiment Hall (NEH)
Far Experiment
Hall (FEH)

LCLS Instruments



LCLS has already had a significant impact on many areas of science, including:

- ➔ Resolving the structures of macromolecular protein complexes that were previously inaccessible
- ➔ Capturing bond formation in the transition-state of a chemical reaction
- ➔ Revealing the behavior of atoms and molecules in the presence of strong fields
- ➔ Probing extreme states of matter

LCLS Beam Parameters

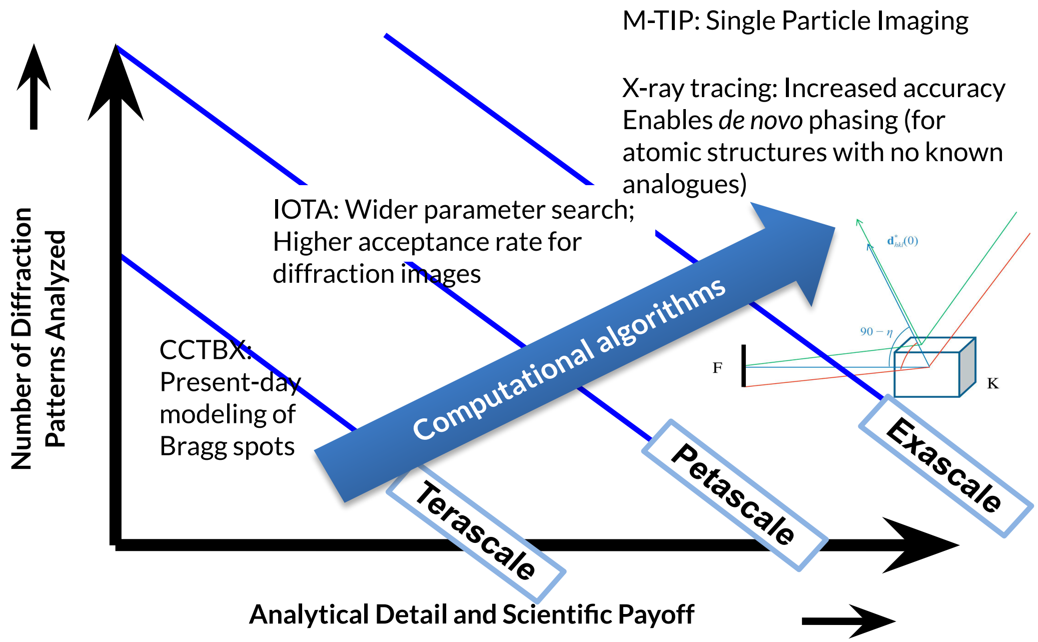
The upcoming superconducting LINAC will dramatically increase the pulse rate

Photon Beam Parameters	Symbol	NC linac HXU x-rays		NC linac SXU x-rays		SC linac SXU x-rays (Projected***)			Unit
		$\hbar\omega_{max}$	$\hbar\omega_{min}$	$\hbar\omega_{max}$	$\hbar\omega_{min}$	$\hbar\omega_{max}$	$\hbar\omega_{nominal}$	$\hbar\omega_{min}$	
Photon Energy	$\hbar\omega$	**25000**	**1000**	**5000**	**200**	**1300**	**800**	**200**	**eV**
Fundamental wavelength	λ_r	0.5	12.4	2.5	62.0	9.5	15.5	62.0	Å
Final linac e- energy	γmc^2	16.5	3.5	10.0	3.5		3.5 - 4.0		GeV
FEL 3-D gain length	L_G	4.0	1.0	2.5	1.0		TBD		m
Peak power	P	20	80	50	30	3	2.5 - 7	8	GW
Pulse duration range (FWHM)		10 – 50		10 – 250			20 - 40		fs
Nominal pulse duration (FWHM)	$\Delta\tau_f$	30		50			20		fs
Max Pulse Energy*	U	0.6	2.0	2.5	1.5	0.06	0.05 - 0.14	0.16	mJ
Photons per pulse*	$N\gamma$	0.15	14	3.1	47	0.28	0.4 - 1.1	5.0	10^{12}
Peak brightness*	$B_{pk,\,SASE}$	7800	425	2250	19	20	8.6 - 24	1.7	10^{30} §
Average brightness (120Hz for Cu-linac)*	$\langle B \rangle$	280	16	138	1.5	137 @ 33 kHz	57 - 161 @ 33 kHz	12 @ 33 kHz	10^{20} §
SASE bandwidth (FWHM)	$\Delta\omega/\omega$	30	2	10	2	4	3	3	eV
Photon source size (rms)	σ_s	8	20	16	46		TBD		µm
Photon far field divergence (FWHM)	$\Theta_{FWHM,x,\infty}$	1	12	3	25		TBD		µrad
Max. Beam Rate	ϕ_{FEL}	120		120			1,000 – 40,000**		Hz
Avg. x-ray beam power	P_x	0.07	0.24	0.30	0.18	2.0 @ 33 kHz	1.7-4.6 @ 33 kHz	5.3 @ 33 kHz	W
Linear Polarization (100%)	$\langle P \rangle$	Vertical		Horizontal			Horizontal		

Data Analytics for high repetition rate Free Electron Lasers

FEL data challenge:

- Ultrafast X-ray pulses from LCLS are used like flashes from a high-speed strobe light, producing stop-action movies of atoms and molecules
- Both data processing and scientific interpretation demand intensive computational analysis



LCLS-II will increase data throughput by three orders of magnitude by 2026, creating an exceptional scientific computing challenge

Material science experiments at an XFEL

Courtesy of Apurva Mehta (LCLS Dep Head for Material Science)

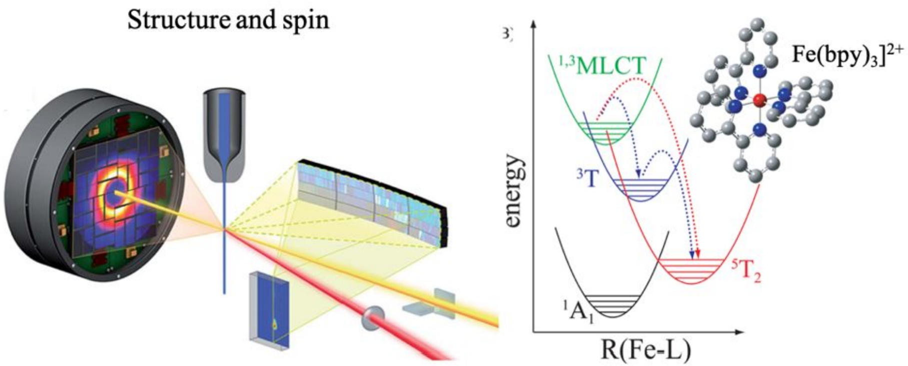
Scientific Thrust for Material Science in LCLS

Use X-ray techniques to study short lived excited states

- Study emergence, suppression and evolution of excited states using spectroscopic techniques, inelastic X-ray scattering, X-ray emission and absorption spectroscopy
- Transient structural phases
- Study changes in material morphology and defects

A selected examples:

- Magnon dynamics in cuprates
- Laser ablation
- Metal additive manufacturing

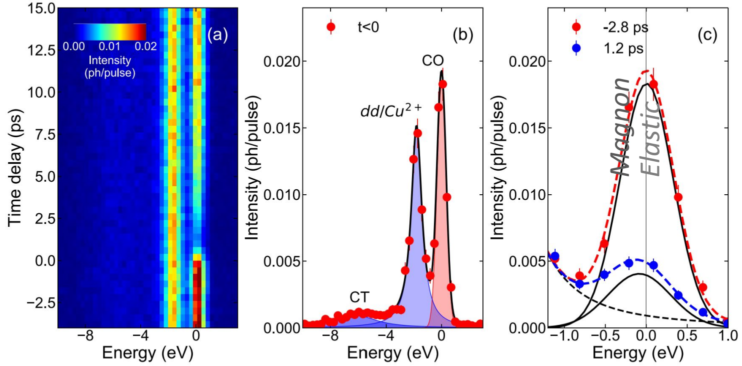


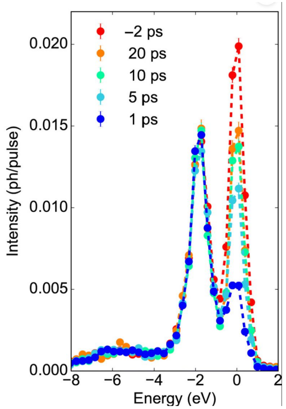
Potential energy surface diagram depicting excited state evolution in $[Fe(bpy)_3]^{2+}$ in solution following charge-transfer excitation
The involvement of the ^{3}T state had never been detected before LCLS

K.S. Kjær et al., *Chem. Sci.*, **10** 5749 (2019).
W. Zhang and K.J. Gaffney, *Acc. Chem. Res.*, **48** 1140 (2015).

Photo-excitation of the Stripe Phase

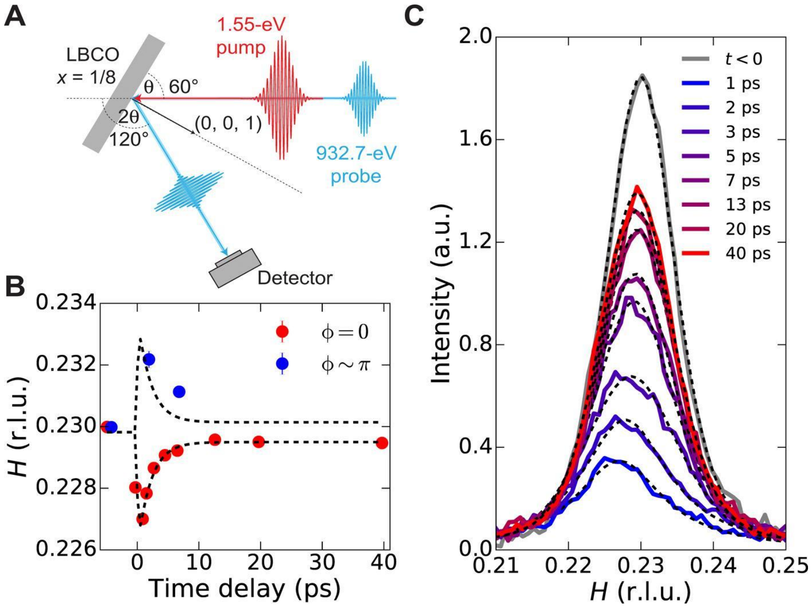
Understand the Origin/Stabilization of the Stripe Phase for LBCO - resolve paramagnon dynamics





Miao, et al., PNAS, 114 (47), 12430-12435 (2017)

LCLS can reveal the interactions among the fundamental excitations of solids such as phonons, magnons, and topological excitations with far more precision than has previously been possible.

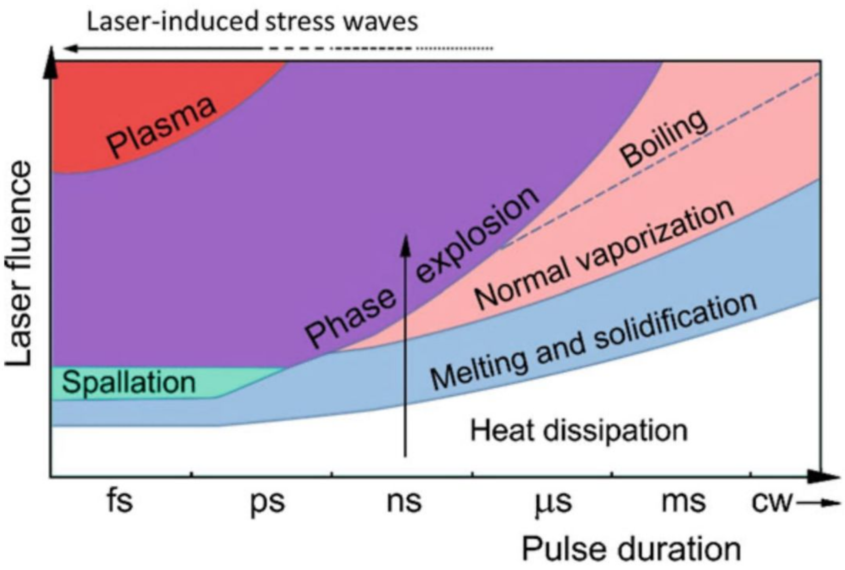


SXR 2017 (E res ~ 700 meV)

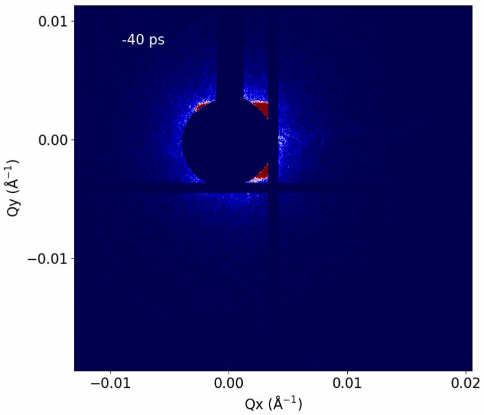
Miao, et al., PNAS, 114 (47), 12430-12435 (2017)

Understanding Laser Ablation

Pulsed laser machining - hard materials, delicate materials, complex and large aspect ratios

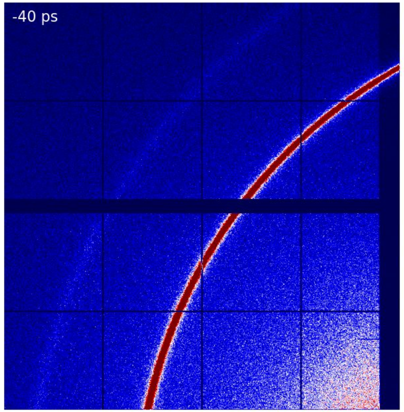


Shugaev, et al. In *Handbook of Laser Micro-and Nano-Engineering*



Small Angle Scattering

Particle formation



Wide Angle Scattering

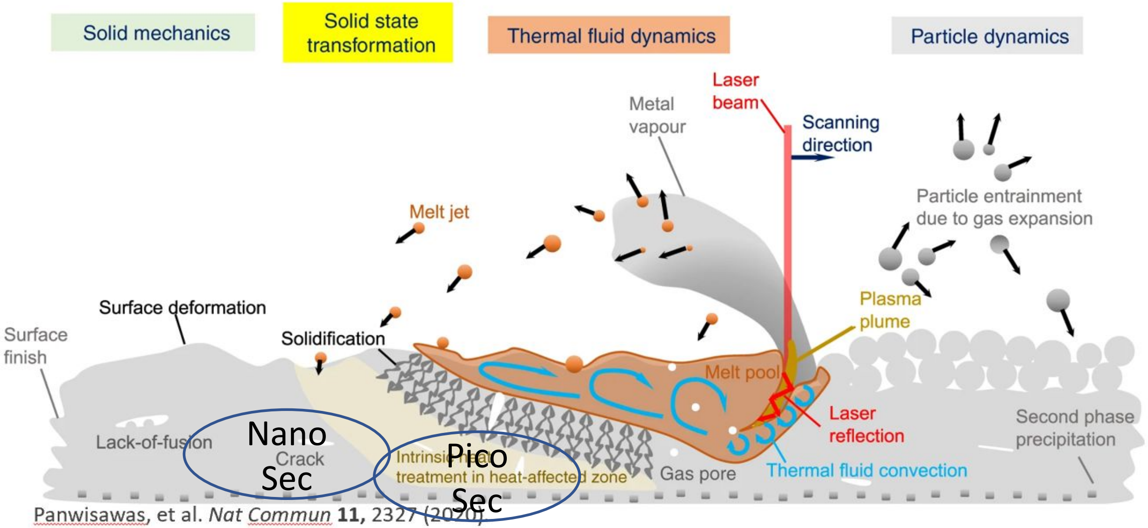
Changes in crystallinity

Courtesy of Yanwen Sun (unpublished)

Metal Additive Manufacturing

Hierarchy of time-scales

- Solidification dynamic – nucleation barriers
- Non-equilibrium metastable phase phases
- Novel microstructure
- Thermal transport
- Crack formation



Panwisawas, et al. *Nat Commun* **11,** 2327 (2020)

LCLS Science Impact Assessment Report on Theory and Simulation in condensed matter physics and materials science

Strong recognition of the role of simulations at LCLS

LCLS experiments have stimulated the development of advanced combined computational experimental methods that include artificial-intelligence-guided data analysis and strategies for comparing theory with vast arrays of experimental data.

The experimental work at LCLS has had a significant impact on the way researchers approach modelling and data visualization. For example, in the case of dynamic compression experiments, where the use of X-ray diffraction is ubiquitous, modelers now routinely use molecular dynamics or crystal plasticity codes to directly output simulated diffraction profiles, not only to compare with experimental data after the event, but also to inform the design of experiments before they are performed. Furthermore, and as alluded to in the section above, experimental data has informed the plasticity models, spanning the length scales.

… huge impact on atomic kinetics calculations that are widely used within theoretical laser-plasma community. In particular, the intense X-ray excitation that LCLS affords means that many standard codes, which had not included the relevant atomic configurations (or super-configurations) within them, do not accurately predict the observed spectra, and that has led to considerable efforts to create codes specifically designed for XFEL-matter interactions.

… observation of ionization thresholds at energies differing from standard theory has led to many groups now employing density functional theory to attempt to understand matter under extreme conditions.

Submitted by the LCLS Science Impact Assessment Committee members:

Antoinette (Toni) Taylor, Los Alamos National Laboratory (Chair)
Paul Evans, Department of Materials Science and Engineering, University of Wisconsin- Madison
Claudio Masciovecchio, FERMI FEL, Elettra Sincrotrone Trieste Laboratory
James McCusker, Department of Chemistry, Michigan State University
Janet Smith, Department of Biological Chemistry and Life Sciences, University of Michigan
Marc Vrakking, Max-Born-Institute and the Freie Universität Berlin
Justin Wark, Department of Physics, University of Oxford
Philippe Wernet, Department of Physics and Astronomy, Uppsala Universitet

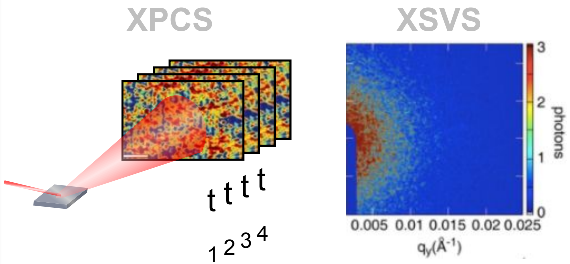
Complex workflows for FELs

Performance and automation

Linac Coherent Light Source: Workflows Benefit from Integration of Experimental Data with Simulations/Modeling

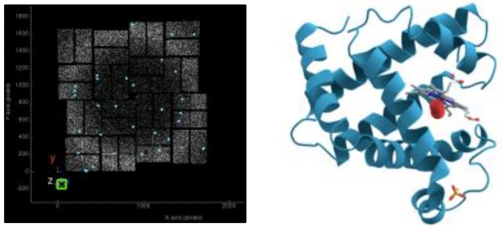
20+ Scientific Workflows with Unique Requirements

Coherent Scattering



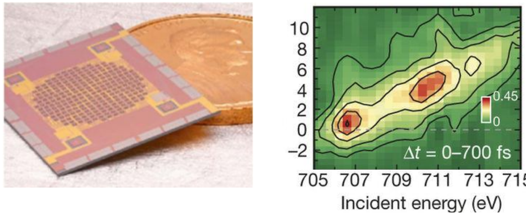
2022: 20 GB/s, 4 TF (reduction), 34 TF (analysis)
2026: 80 GB/s, 34 TF (reduction), 270 TF (analysis)

Nanocrystallography



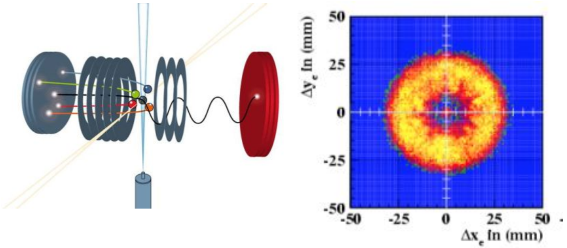
2023: 64 GB/s, 3 TF (reduction), 4 TF (analysis)
2026: 1.2 TB/s, 16 TF (reduction), 20 TF (analysis)

Resonant Inelastic Scattering



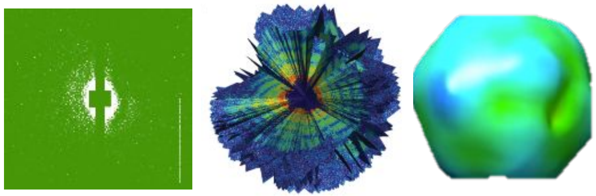
2023: 20 GB/s, 4 TF (reduction), 1 TF (analysis)
2026: 200 GB/s, 40 TF (reduction), 2 TF (analysis)

Coincidence Spectroscopy



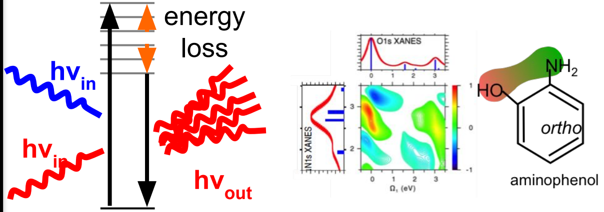
2021: 200 GB/s, <1TF (reduction), <1TF (analysis)

Coherent Imaging



2022: 64 GB/s, 3 TF (reduction), 270 TF (analysis)
2026: 1.2 TB/s, 16 TF (reduction), 1340 TF (analysis)

Nonlinear Spectroscopy

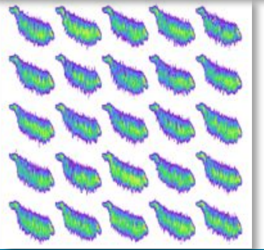


2023: 20 GB/s, 3 TF (reduction), <1 TF (analysis)
2026: 80 GB/s, 16 TF (reduction), <1 TF (analysis)

Example of Massive Throughput Workflow: Coherent Imaging at LCLS

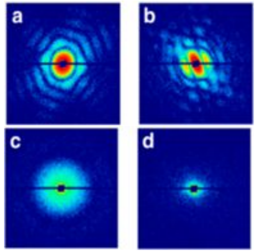
One workflow must encompass several areas and disciplines: integrated approach required

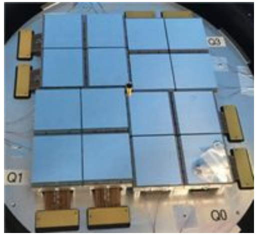
Accelerator Diagnostics



Machine Learning Optimization

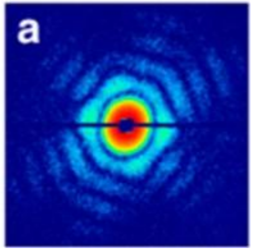
X-ray Images

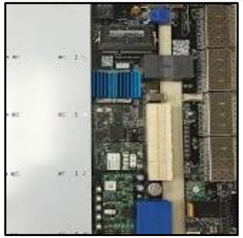




High Frame Rate Detector

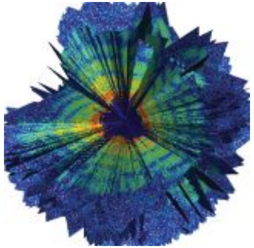
Instant Data Reduction

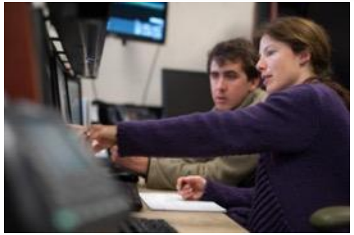




Edge Computing

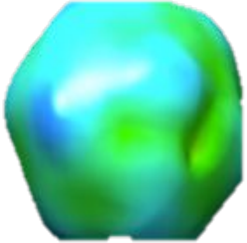
Assess Data Quality from Integrated Data

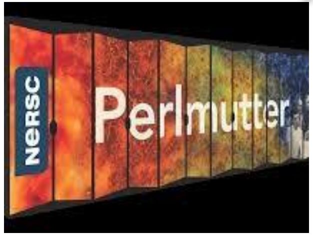




Local Fast-feedback Computing

Interpretation of Structural Dynamics

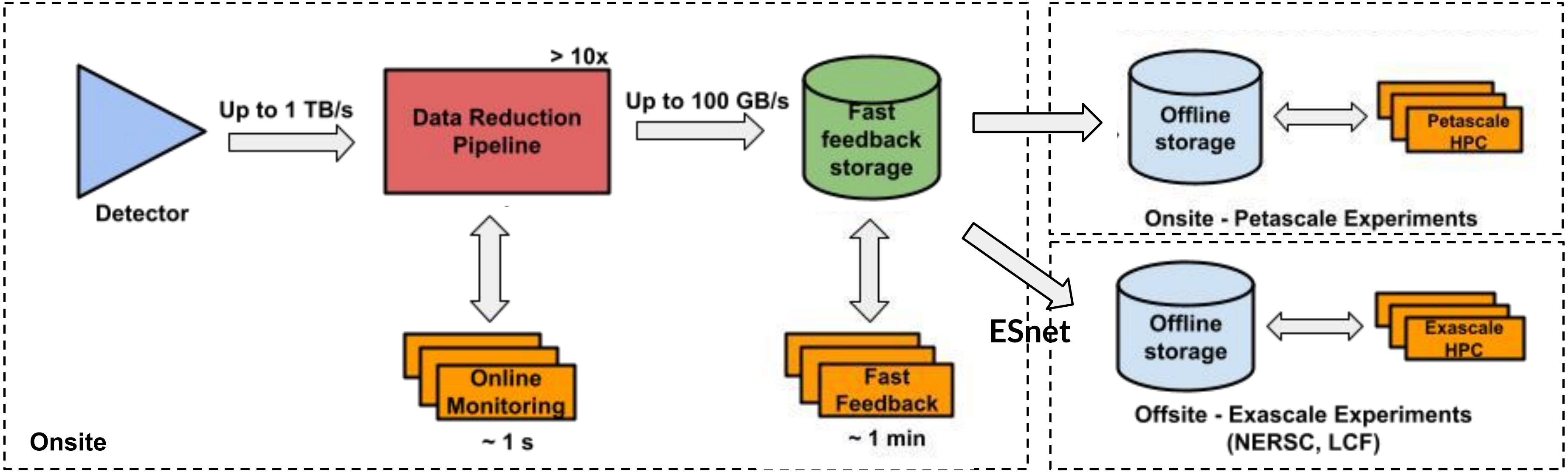




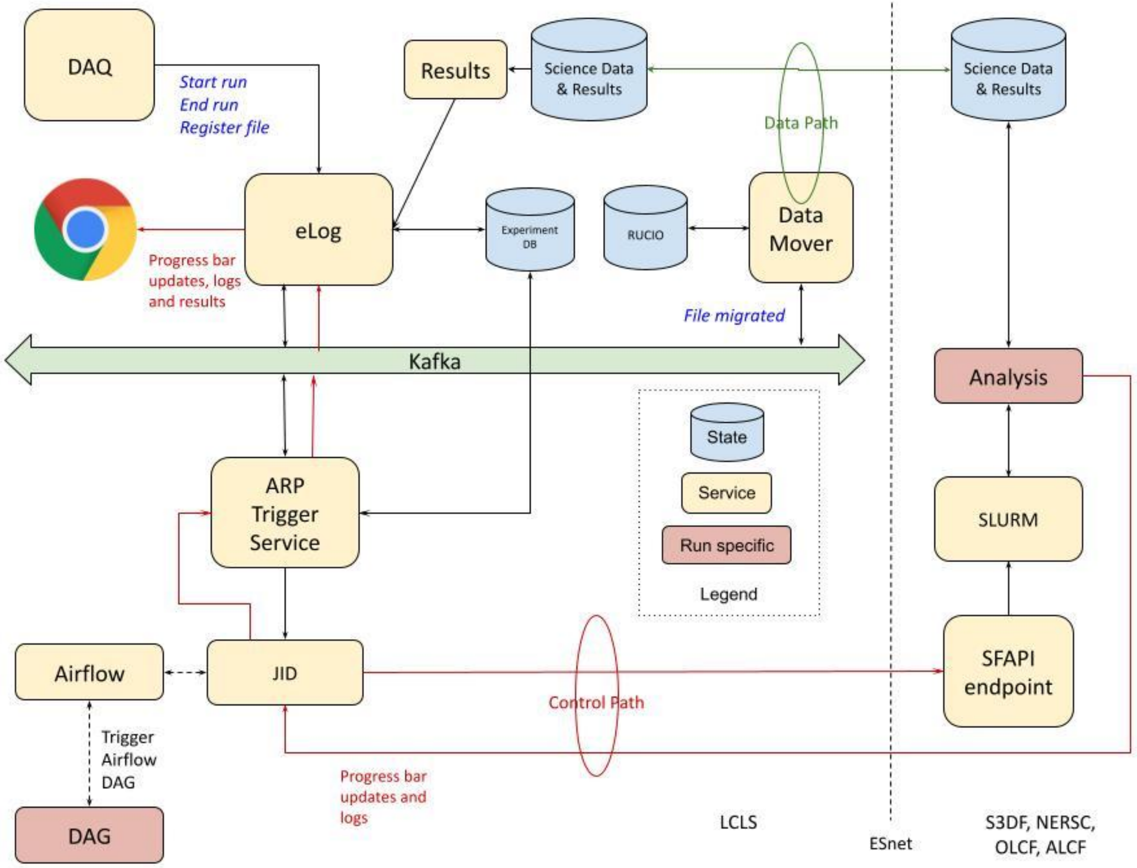
High End Computing

LCLS Data System, a scalable, adaptable system

Mix of automatic, on-demand, and user driven data flows - combination of onsite and offsite resources



Workflows for Automated Run Processing



Data Management for data transfer, metadata capture, and automatic run processing

- Uses superfacility API to connect to NERSC
- Streams data to remote computational resources
- Automatically launches remote workflows
- Returns results to experiments in a web browser
- Everything (airflow, etc) runs on LCLS side and funnels through SF API

Introduction to the digital twin approach

Integrating simulations with experiments in real-time

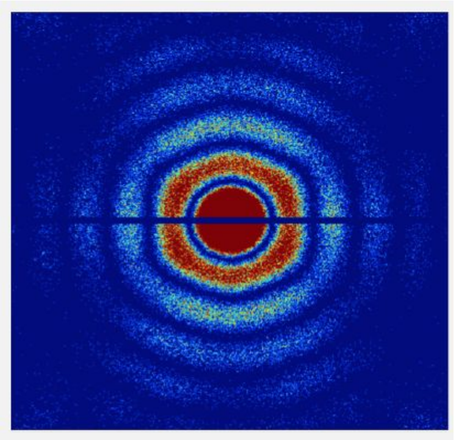
Role of Simulations at Light Sources

Up to now, not much but…

Even small improvements in use of beam can impact science outputs — especially at high rep rate facilities where sufficient data may be collected in a short time at the right conditions

Future light sources could use simulations to predict the behavior of new experiments, steer experiments, build smarter and faster experiments by, for example:

- Interpreting experimental data, testing experimental assumptions
- Providing information about parts of a system that are inaccessible (destructive measurement, slow measurement, or not directly measurable)
- Training ML algorithms or determine corrections and calibrations (e.g. knowing when to stop optimizing an experiment)
- Prototyping novel accelerator operation modes without using valuable machine time or to predict which parts of an experimental phase space are likely to yield the most useful data



Simulated diffraction pattern of a rice dwarf virus (RDV)

Challenges of Using a Digital Twin

A Digital Twin (ML/AI or model) can mimic the behavior of the accelerator (and help tune it more efficiently), a beamline (optics, detector) or a physical system (protein, material)

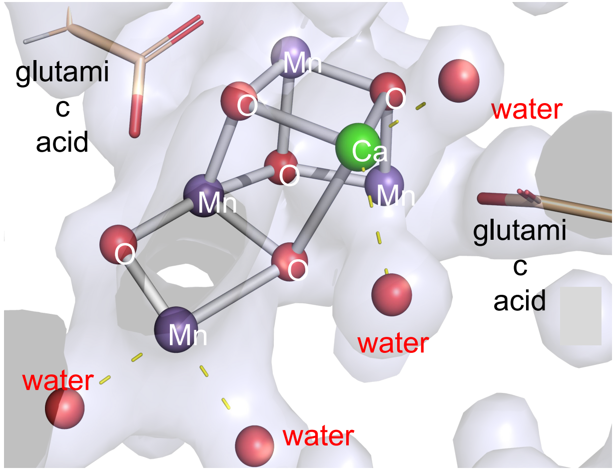
To be used effectively in the context of an experiment, several issues must be addressed:

- Timescale: accelerator tuning, experiment steering happen in ~real-time during experiments but timescales for experimental analysis and simulation often differ by orders of magnitude
- Phase space: Availability of experimental data, appropriate simulated data. Vast and multi-dimensional phase space to explore (e.g., temperature, chemical potential, ionic strength, solvent dielectric, chemical composition,...)
- Computational cost: Simulations that include nonlinear and collective effects are powerful tools, but they can be computationally expensive - need access to HPC, develop models to provide fast approximations to simulations, or both

Digital Twin Example 1: X-ray Tracing

Simulations aiding convergence and resolution in nanocrystallography reconstruction

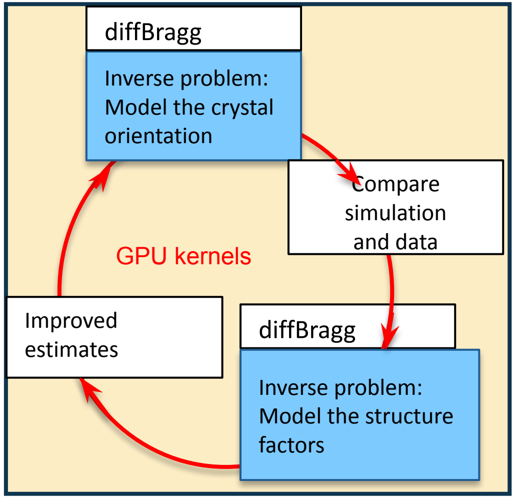
Science Goal: enable the time-domain "movie" of an enzymatic reaction

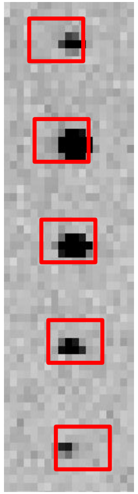


Photosystem II protein complex:

$2H_2O \rightarrow O_2 + 4H^+ + 4e^-$

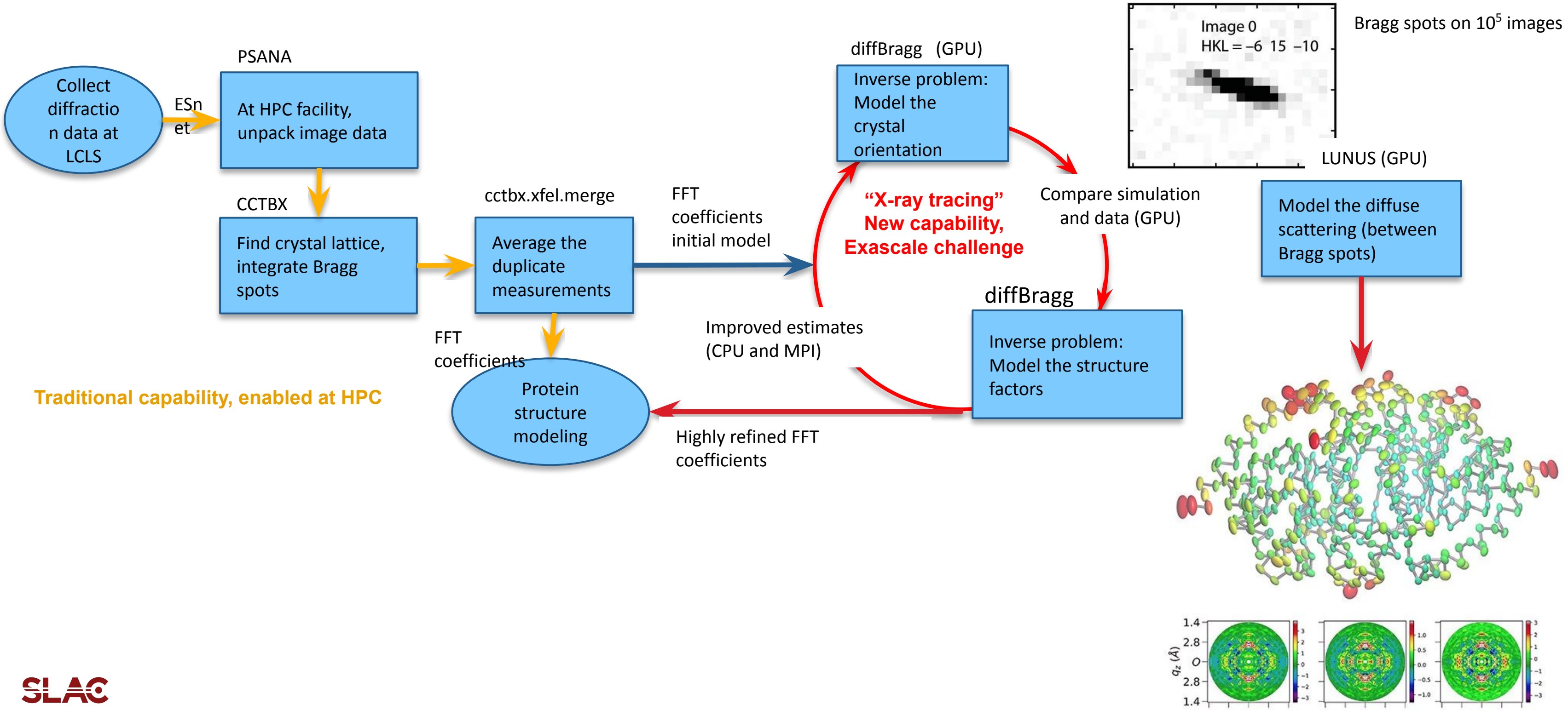
X-ray Tracing Approach: apply L-BFGS to estimate structural parameters from diffraction data - handle polychromatic X-rays and reflects imperfections in crystal





Digital Twin Example 1: X-ray Tracing (continued)

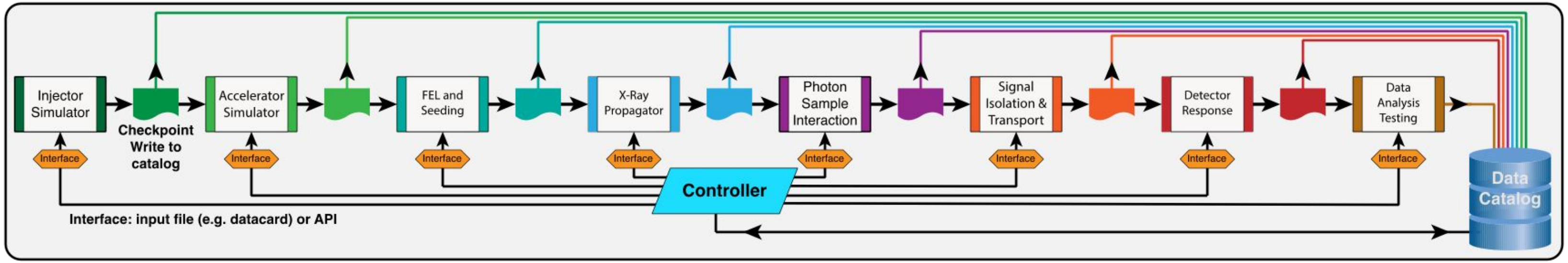
Pixel-level Bragg analysis: First XFEL modeling of spectral shape and crystal mosaic texture - goal is high resolution atomic detail



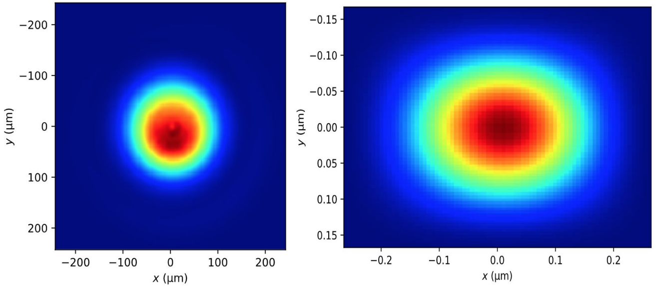
SLAC

Digital Twin Example 2: Accelerator Optimization

Start-to-end simulation for FELs: Light Source Unified Modeling Environment (LUME)



- How should accelerator be configured to produce the best pulses?
- How do (realistic) pulses propagate through the instrument and interact with optics?
- How does the X-ray pulse interact with the sample?
- What is the impact of detector performance?
- In a statistical measurement, how much data is enough?



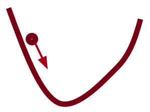
SASE wavefront before and after propagation

Digital Twin Example 2: Accelerator Optimization (continued)

Tuning approaches leverage different amounts of data / previous knowledge suitable under different circumstances

less ← assumed knowledge of machine → more

Model-Free Optimization

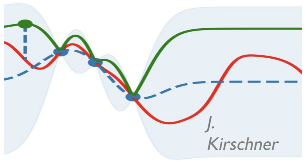


Observe performance change after a setting adjustment

→ estimate direction or apply heuristics toward improvement

gradient descent
Simplex ES

Model-guided Optimization

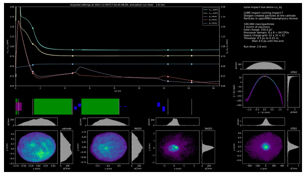


Update a model at each step

→ use model to help select the next point

Bayesian optimization
reinforcement learning

Global Modeling + Feed-forward



Make fast system model

→ provide initial guess (i.e. warm start) for settings or fast compensation

ML system models +
inverse models

Accelerator tuning is aimed at combining the strengths of different approaches. General strategy: start with sample-efficient methods that do well on new systems, then build up to more data-intensive and heavily model-informed approaches.

SLAC

Conclusions

The role of simulations at FELs is rising

As FEL science is getting more mature, making good use of beamtime is becoming critical, for both facility and users

Simulations can optimize data taking by reducing the time needed to acquire scientific results and improving the quality of data collection through

- Better reconstruction algorithms
- Sampling strategies to adaptively narrow the space of probable solutions instead of blindly search through all possible combinations without leveraging the information being acquired
- Identification of sources of error between simulations and measurements (UQ) via fast-executing models

If you have ideas of possible engagements with LCLS, please reach out

(contact: Apurva Mehta, LCLS Dep Head for Material Science, mehta@slac.stanford.edu)

How a Free Electron Laser Works