Federal Register Notice: 89 FR 51554, Federal Register :: Networking and Information Technology Research and Development Request for Information on Digital Twins Research and Development, June 18, 2024.

Request for Information on the National Digital Twins R&D Strategic Plan

SLAC National Accelerator Laboratory

Digital Twins for Enhanced Modeling and Control of Light Sources and Neutron Sources [1]

SLAC National Accelerator Laboratory

Contributors: Auralee Edelen, Jana Thayer, Frederic Poitevin, Daniel Ratner, Ryan Herbst, Ryan Coffee, Quynh Nguyen, Paul Fouss, Matt Seaberg, Apurva Mehta, Michael Kagan

1 Introduction

Light sources and neutron sources play a key role in understanding the fundamental properties of complex matter at different time and length scales by capturing their structural and electromagnetic dynamics. These scientific facilities rely on some of the most complex machines humans build. For example, X-Ray Free Electron Lasers (XFELs) are driven by particle accelerators and produce highly coherent light for detailed imaging of samples, and their operation requires close integration of many subsystems: high-performance particle accelerators, sensitive magnetic undulators to produce X-rays, high-power X-ray optics, and sophisticated detectors and complex sample environments (e.g. synchronized pumping with ultrafast lasers). Comprehensively exploiting the capabilities of light sources and beamlines can lead to new scientific discoveries in a wide range of fields, such as biology, chemistry, physics, and material science. Increasingly more complex instruments and light source capabilities can enable unprecedented measurements to unravel the fundamental properties of matter. However, the scarcity of neutron and light sources relative to the large experimental demands leads to a shortage of allocated beam times. Consequently, there is a pressing need to develop real-time data analysis and experimental guiding capabilities in order to make efficient use of limited experimental time and maximize the scientific value of the collected data. Additionally, there is a need to reduce the large amount of time that is currently spent setting up facilities for delivery to different experiments.

Experiments at light sources can benefit greatly from **digital twin (DT)** technology, which can leverage prior measurements, known parameters, and theory to guide sampling strategies during an experiment and to produce unique scientific insights. DTs will be critical for streamlining operation of user facilities, which involves complex system control. Light sources are also ideal test beds for the development and deployment of DT technology. They are highly dynamic systems with many deliberate and unintended changes in conditions over time, they are made of multiple complex interacting sub-systems that need to operate in concert for optimal performance, they have physics simulations that can be readily leveraged and fused with measured data, they provide a more contained environment in which to explore DT concepts than many other applications (in contrast, for example, to DTs of global climate), and there are many light sources worldwide with shared designs, enabling the exploration of technology that is easily interoperable and exchangeable across systems. Experience with such test beds will be critical for developing reliable, sustainable, interoperable DT infrastructure that can be used across the numerous application areas of US national interest (climate, energy grid, etc.).

A prominent example of a complex light source is the one-of-a-kind, high repetition rate

[1] This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Digital Twins R&D Strategic Plan and associated documents without attribution. Submitted July 28, 2024.

1-MHz LCLS-II, which opens a new era for XFEL science and technology. This new superconducting linear accelerator will be delivering photons at unprecedentedly high repetition rate and brightness, with ultrafast time-resolutions to unravel new physical phenomena. The LCLS-II exploits complex instrumentation capabilities to enable unprecedented scientific measurements. To make full use of these capabilities, at SLAC we have begun developing and deploying parts of a facility DT ecosystem and identifying areas of urgent need for current and future R&D. In this response to the NSF request for information [1], we outline challenges for operation of light sources, the role DTs could play in improving operation of light sources and promoting scientific discovery, and R&D needs to bring DT technology to fruition for these systems.

1.1 Challenges and Needs for Operation of Light Sources

Many challenges complicate operation of a light source, including the high dimensionality of available settings and input parameters, numerous sources of uncertainty, and nonlinearity of system responses. These can make it extremely challenging to fulfill specific beam parameter requests within required tolerances for different applications. User experiments also require different experimental parameters (e.g., x-ray parameters, different samples, different acquisition and analysis modes) that need to be delivered on-demand. Because meeting these needs typically requires laborious hand-tuning by expert human operators, the range and number of science experiments that can be performed is significantly reduced. This is both due to the time it takes to set up individual experiments and the challenges in achieving and maintaining the required beam quality. Time-varying changes in the facility conditions (e.g. initial beam characteristics, RF cavity phase calibrations, ambient temperature) complicate the task further.

An increasing suite of experiments also require precise dynamic control over the beam during an experiment. For example, X-ray correlation spectroscopy requires the separation between two electron bunches to be smoothly scanned, which is highly challenging and necessitates joint tuning on accelerator and photon beamline settings to achieve extremely delicate final beam parameters. The demanding level of fine control required by new experiments requires monitoring and adjustment of settings across the entire facility to carefully control the beam evolution from the beam source to the experimental area. Finally, although physics simulations can aid experiment planning and operation, these are also challenging to construct in ways that provide sufficient accuracy relative to the real machine behavior, due to the many nonlinear phenomena and the high dimensionality of these systems (for example, LCLS-II at SLAC will have over 2M variables to monitor). Despite many advances in automated tuning based on artificial intelligence and machine learning (AIML), the majority of adjustments, complicated diagnostic analysis, and assessment of machine behavior still require human operators.

Another aspect of operation is "experiment steering," in which information relevant to the final science outcomes is provided on-the-fly during an experiment and used to suggest next steps in the experiment (e.g. new beam conditions to examine for a particular sample). These approaches can provide insight into how best to maximize information gain for a given experiment. This requires, for example, fitting of many physically relevant model parameters and determining which are most likely given the data. Techniques such as Bayesian Optimal Experiment Design (BOED) can then be used to suggest experimental inputs that would further reduce possible model uncertainty and aid better determination of the physics model. This process requires sophisticated analysis of user experiments (including both diag-

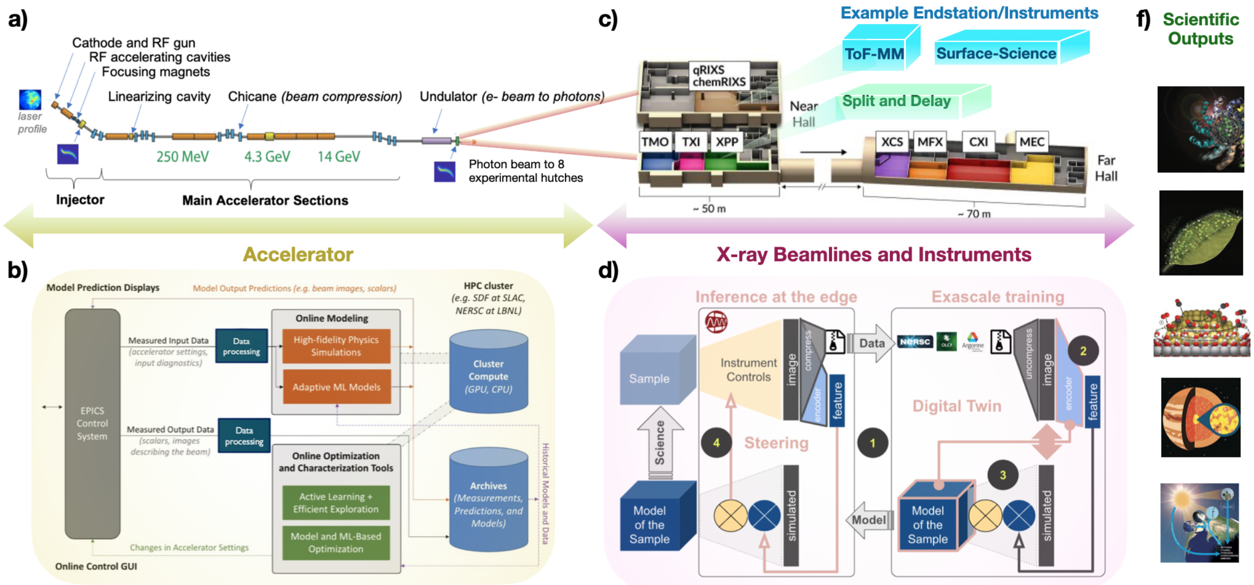

Figure 1: Layout of the LCLS accelerator and X-ray Free Electron Laser (XFEL), which delivers photon beams to myriad scientific users in biology, molecular dynamics, drug discovery, and chemistry. An outline of desired end-to-end alignment spanning from the accelerator to XFEL and scientific user experiments. Schematic diagrams of (a, b) the accelerator and associated DT and ML-based control infrastructure under development. (c) The x-ray free electron laser that delivers photons to different beamlines for user experiments, including gas phase, liquid, material, biological science. (d) Desired DT model and framework for an LCLS beamline and instrument. The flow chart describes the needs and challenges, including (i) framework for timely data and model transfer across data generating and data processing DOE facilities, (ii) large area detector image encoding, (iii) accurate and trustworthy differentiable simulators, (iv) leveraging information flow from sample parameters to instrument parameters for actionable experiment steering. (f) Scientific outcomes of experiments, which will directly benefit from developed end-to-end DTs and machine learning models.

nostic data processing and inclusion into physics modeling/analysis) to be tightly integrated with the optimization of (at minimum) the photon beamline setup. Incorporation of this information into start-to-end DTs to aid joint accelerator optimization and control, photon beamline optimization and control, and experiment analysis, would greatly reduce overall time-to-science, improve efficiency, and aid discovery.

Overall, there is an urgent need to decrease the overall time-to-science, ensure sufficiently high beam quality can be achieved and stably maintained for experiments, and enable new capabilities such as dynamic control over the beam. Coupled to this, there is a need to integrate on-the-fly analysis of user experiments to feed back into the direction of a given user experiment.

1.2 Role of Digital Twins in Meeting Light Source Needs

Accurate fast-to-execute system models that are kept updated as the light source enters new regions of parameter space could substantially improve online monitoring and control of the facility, as well as aid the ability to infer structures and molecular dynamics directly from the data, achieving the demanding performance for experiments that will fuel future discovery.

We take a DT to be a data-informed digital model of a complex physical system in all its aspects [2] that tracks the real inputs to the system and provides feedback to the system. As the RFI notes, the two-way interaction is a key element of a DT. They may leverage both physics information and artificial intelligence / machine learning (AI/ML).

In particle accelerators and light sources, a facility DT can be used to provide insight into global machine behavior during operation (including to identifying anomalous behavior and prompt faster intervention), provide estimates of beam behavior that are not able to be observed at sufficient speed or resolution with conventional methods (i.e. virtual diagnostics), be used directly in online model-informed optimization and control, and be leveraged in experiment steering. This includes optimization for experiment setup, continuous control for stable delivery and correction of unintended time-varying changes, and integration of live feedback from the scientific output of user experiments. Note that in addition, a DT can be copied at a certain point in time and used offline for experiment planning, design of new diagnostics or beam manipulation hardware, and prototyping of software and algorithms for analysis and control. Aside from a DT of the facility, a DT can also span the interaction between the incoming beam, a particular physical/material process under study (including unknown physics parameters), and measurement devices. DTs and ML models can also enable extraction of scientific understanding from new measurements. Combined with accelerated and parallel computation on GPUs, this can enable live-analysis to steer experiments and data collection. By linking facility DTs to DTs for specific experiments, active feedback to both the next steps in the experiment and to the facility controls can be linked to enable highly-efficient, precise experiment guiding.

SLAC has been developing and deploying technology to enable DTs of LCLS instruments, the LCLS-II accelerator, and other particle accelerators it stewards (e.g. FACET-II, MeV-UED, SSRL). These efforts span improvements in the speed of execution of physics models through the use of AI/ML, improvements to the overall accuracy of system models through coupling physics simulations and data, incorporation of deployed system models into improved online optimization of machine setup, development of software infrastructure and workflows for deployment of DTs and feedback control, and approaches for streaming ML-ready data. Work has also been ongoing to integrate new analysis techniques, including based on ML, into fast analysis of user data to aid experiment steering. Below we briefly describe several current and future applications and areas of need for DTs at LCLS-II, including open challenges.

1.2.1 Current Examples and Future Use Cases for LCLS-II

Accelerator Operation: As one example, open-source software developed at SLAC has enabled substantially easier construction of start-to-end simulations that use different physics codes for different parts of the accelerator [3, 4]. These tools are highly inter-operable and have, for example, been used with minimal additional effort to run online physics simulations for the LCLS-II injector and the FACET-II injector. These online physics simulations were used during LCLS-II injector commissioning to aid the intuition of physicists in the control room, which when coupled with automated tuning reached the highest beam quality (measured in terms of the beam emittance) then seen during commissioning. This infrastructure has now been expanded to start-to-end simulations of the accelerator and deployment on the local HPC cluster (the S3DF) using tools such as Prefect and Kubernetes. This enables physics simulations with relatively high fidelity to be computed based on live input in minutes for quasi real-time information in the control room. Surrogate models based on ML enable even greater execution speed. These models have then been used to provide priors to

Bayesian optimization and similar approaches to more efficiently tune the accelerator (e.g. see [5], and [6] for a survey of techniques). Substantial software development and infrastructure is still needed to develop and deploy the system for true facility-scale prediction and control.

Tuning of Complex Photon Instruments and Beamlines: Relevant challenges for XFELs and ultrafast lasers includes stochastic characteristics of input laser and x-ray parameters, constantly changing sample environment, massive data rates and quantity ($\sim$ TeraBytes), live-analysis, and fast feedback for experiment steering. Hence, accelerated computing and computation are required to extract scientific understanding on a reasonable timeline. Trained DTs can enable pre-alignment, optimization of crucial parameters, and auto-alignment during beam times to maintain optical conditions. Efficient and live analyses of the massive incoming data become crucial for steering relevant experiments and data collection. We have been developing DTs for the split and delay, ToF-MM, cyo-EM, and will expand to other complex spectrometers, the RIXS endstation, and the LCLS-II/-HE beamlines.

Split and Delay Lines: Particular challenges include the tuning and control of complex photon optics in a beamline, which consists of perfect crystals to focus and redirect beams. X-ray beams are sensitive to nanoradian-scale mirror misalignments and picometer-scale thermal deformation of crystal optics. The very narrow and highly non-linear acceptance of X-ray optics (especially diffractive single crystal optics), make for isolated, sharp optima in in otherwise vast and mostly featureless search space. As state-of-the-art light sources such as XFELs mature, the complexity of X-ray optical schemes tends to increase to push the boundaries of beam characteristics for enabling new scientific directions. Examples include split and delay systems that enable X-ray pump/X-ray probe and X-ray probe-probe measurements, tunable high resolution monochromators, nanofocusing/nanoimaging systems, pulse shapers, etc, some of which already exist and others that are under development. There are typically > 10 critical motion degrees of freedom for most of these optical system. Extreme sensitivity to changes in thermal loading from absorbed beam power, frequently deforms the system significantly from ideality, making tuning and drift correction by a human operator very challenging if not impossible.

RIXS and Momentum Microscope: The resonant inelastic X-ray scattering (RIXS) and time-of-flight momentum microscopy (ToF-MM) instruments/ end-stations are currently being commissioned and offered to the scientific communities. Developing DTs and ML models can tackle these challenges with the ongoing advancements and innovation on the instrumentation and beamline frontiers. For example, the end-to-end ML model will enable real-time optimization of electron optics to minimize time to alignment and optimization of ToF-MM, which is a momentum imaging electron spectrometer, while incorporating feedback from a DT of the detector systems to increase the feasibility of studying complex matter. ToF-MM consists of 500+ tuning parameters to enable direct momentum or spatial imaging to comprehensively capture the properties of matter. The vast parameter space presents a prohibitively difficult optimization problem during limited time during beam times for complex targets. Automated tuning leveraging a DT could help gather data to further develop analysis of new classes of materials and molecules via photo-induced processes, such as phase change, band transitions, electron-phonon coupling dynamics, and photoinduced isomerization.

Photon Experiment Steering: Analysis of imaging experiments typically entails solving ill-posed inverse problems through iterative methods that cycle through estimating parameters of a forward model mimicking the experiment and updating of the sample of interest. Traditional approaches to the first step suffer from lack of scalability to large datasets while the second step often imposes limits on the complexity of the models used to describe the experiment. Hence, researchers at SLAC have explored the ability of these generative modeling strategies to perform single shot ab initio processing of unfiltered cryoEM datasets, freeing the analysis process of typically tedious and error prone manual intervention, and the ability to tackle heterogeneous datasets that resisted traditional solvers and allowed the discovery of new sample states [7]. They demonstrated a general framework (Fig 1. d.) that can accommodate different representations of the sample and its dynamics by representing them directly at the atomic level [8–10]. They then successfully applied this model to X-ray coherent diffraction images of isolated particles [11]and are actively working on mapping the framework to all the imaging modalities encountered at LCLS, including ptychography [12], inelastic scattering [13] or crystallography.

2 Needs and Challenges in Digital Twin Technology for Light Sources and Neutron Sources

While substantial progress has been made toward development and deployment of DTs for light source facilities and experiments, much work remains to see DTs brought to fruition and put into day-to-day use, especially at the scale of an entire facility. This ranges from fundamental improvements in AI/ML techniques for modeling and analysis, to the need for reliable software ecosystems for deployment of DTs and handling of streaming facility data, to improved physics modeling, to integration with high performance computing (both local systems and the larger DOE HPC ecosystem). We describe these needs in further detail below, loosely following the structure of topics listed in the request for information.

2.1 AI and Digital Twins

AI/ML for Facility System Models and User Experiments: A major challenge in light source facilities and accelerators is the ability to obtain physics models with sufficient accuracy and execution speed to be leveraged in live prediction and control. The as-built system often differs from design due to numerous compounding sources of error and non-idealities that are typically not included in physics simulations (including time-varying behavior). Physics simulations are computationally expensive (e.g. it can take many minutes to hours for a single simulation). ML models can aid automatic determination of likely error sources (and associated uncertainties), even for high-dimensional systems; they can also serve as fast-executing surrogate models for physics simulations. In cases where no physics simulation is available, ML models can be trained on empirical data.

However, there are challenges in reliably using ML for modeling these systems. The time-varying nature means they are both deliberately and unintentionally brought outside of the statistical distribution of the training data, prompting a need for more generalizable modeling methods, improved uncertainty quantification, and the ability to adapt to changes over time (see "continual learning" below). These systems are also very high-dimensional, which in and of itself can be challenging for ML modeling. Further, because a major aim is to use ML system models in online experiment steering and control, it is desirable to have some interpretability in the model.

Light source facilities are also simultaneously data-rich and data-poor: there is usually extensive archived facility and user data, but it can be spread across many different distinct operating modes in ways that are not conducive to easily learning global system models. Physics simulations can be run on HPC systems, but are computationally expensive in ways that limit the ability to make sufficiently large data sets (e.g. tens of thousands of samples).

There is a need for sample-efficient methods for driving data gathering in simulation and measurement, as well as techniques for sorting through and cleaning large amounts of archive data (which may include ML-based approaches to data clustering and tagging).

Physics Informed ML and Differentiable Physics: Various approaches are being investigated to combine ML models and physics information to improve the sample-efficiency, generalizability, and interpretability of learned models. For example, multi-fidelity modeling can capture information from multiple different types of models (coarse analytic representations, detailed numerical physics simulations) and different data sources (e.g. slower detailed diagnostics and faster coarser or noisier diagnostics). Differentiable physics simulations can also be coupled to ML components to enable learning of high-dimensional unknowns in a way that is highly constrained against the physics (e.g. see [14, 15]).

Unfortunately, differentiable physics simulations are challenging to produce. Automatic differentiation (AD) systems are ubiquitous for training traditional AI/ML systems; however, the complexity of physical models can require reaching beyond this typical scope of AD systems. For example, some challenges require novel R&D for their use in differentiable simulators, such as the presence of discontinuities in physical models, especially discontinuities that depend on the parameters of interest, and require the development of novel methods for differentiation. Several strategies are in development for such challenges, such as smoothing, relaxing, or adding randomized perturbations to such systems, to enable differentiation [16–19]. Further R&D on differentiable simulations will greatly expand the our capabilities to develop accurate and more robust differentiable simulations for scientific facilities.

Continual Learning: As ML models are used on data that is outside of the statistical distribution of the training data, they become less accurate in their predictions. A major challenge is thus keeping these models updated over time and quantifying uncertainties in their predictions, alongside trying to develop methods for improved generalization to unseen conditions. For an accelerator setup, this could be due to putting the accelerator in a fundamentally different configuration of settings to enable new beam parameters, or due to slow unintended drift in parameters. The challenges are even more acute for photon beamlines, which are frequently reconfigured for different experiments and examine an ever-changing variety of different types of samples. Software infrastructure to support continual learning is needed in addition to fundamental AI/ML approaches. Constraining learning with physics information can help aid continual learning, as could the development of larger, more comprehensive community data sets, leading toward foundation models. Overall, this is a large area of need for R&D.

AI/ML on the Edge: The re-programmability of emerging EdgeAI accelerators makes it possible to deploy a custom ML inference model for each detector and experiment, and even for individual shifts. This flexibility allows EdgeAI accelerators to deliver actionable information within minutes of source or target conditions changing during active experiments. FPGA and edge devices, however, have severe resource constraints that limit the number of parameters to the order of tens of thousands (in contrast to millions of parameters). The resources available at the edge are also limited in terms of I/O and memory. Developing a methodology to take offline-developed algorithms and adapt them to the constraints of the ASIC or FPGA environment and the streaming environment will be important if DTs are deployed in this hardware. Readout of high rate instruments may also be distributed; any given ASIC/FPGA may only see a fraction of the data, so algorithms must adapt to this distributed data environment. This is a major area of needed R&D.

2.2 Adoption of Data and Model Management Best Practices

ML-Ready Data and Data Cleaning: The BES light sources and neutron sources could generate thousands of petabytes of data per year[20]. DTs have the potential to transform experiment operations at these facilities and significantly reduce the time to science. This will only be possible, however, if BES leverages the totality of the acquired and simulated data and provides resources to ensure the data are FAIR. This includes the acquisition and central registry of data and metadata, curation and access to high-value datasets, the ability to search through the data for key characteristics, and the capture of provenance and context. Tools will need to be built to enable the registry, indexing, and availability of these large and multi-modal datasets. Likewise, the AI/ML products generated also need to be curated, the provenance and context recorded, and the models rendered easily findable by users. In many cases, preparing the data for ingestion to a DT or AI/ML model is vital and time-consuming; data preparation could include reformatting the data, eliminating outliers and missing data, and ensuring that the metadata provided are accurate and have appropriate units.

On-the-fly Data Analysis, Data Logging: Information about the performance of the DT on data in-flight is vital to validate the performance of the DT, train new DTs, and for doing error analysis. A DT is never just one piece of code running in isolation, but a collection of codes running in, potentially, geographically separated environments. We must maintain the ability to benchmark the performance of the end-to-end system and log the results of each individual job.

Data Movement: At LCLS-II data rates and data volumes, required computational resources will scale to hundreds of GPUs. These resources are not locally available, so LCLS will be looking to the DOE Leadership Class Facilities for training. Once the data moves outside of the local lab, and in fact, even when the data must move through local networks within a lab, data movement is a key concern. New methods for doing parallel data transfer on tens to hundreds of files while they are being written will need to be developed in order to accommodate the models and use cases that we are envisioning. Data movement also requires some certainty that data integrity is maintained, the ability to buffer data when there are bottlenecks along the path, and the ability to do data transfer seamlessly with little human intervention. Low latency networking at Tb/s is needed to keep pace with the expected data rate from LCLS in 2028.

Best Practices for Handling of Scientific User Data in Light Source Facilities: In BES, user data belongs to the user. Although individual user groups may extract science from their own data, the scientific community is missing an opportunity to fully leverage the overall data sets for the national good. Collective use of datasets to train DTs will enable a host of experiments, especially as DTs become more sophisticated and adaptable/tunable to new circumstances. However, if data are used to train a common model for the good of all, then there is an additional burden on the trainers and maintainers of that model to publish which data was used for the training and to define appropriate constraints on the use cases for that model. The facility should be able to use data to provide generic algorithms for the general good. The data generated by the BES user facilities is a national resource and models and DTs developed from these data are powerful tools that should also be useable across user facilities. Such data use will significantly enhance facility performance, but data provenance and use will need to be tracked through the system. In a national environment where DTs and models are reused at different facilities, it will become necessary to develop a methodology to understand the applicability of a model to an environment or experiment, or to develop a workflow for rapidly adapting/retraining a model to the new modality. Policy changes coupled to infrastructure improvements will be necessary to make DOE data and models a national resource, available to all with a clear understanding of how to use and adapt models to new modalities. DTs also cannot be a black box to the users. Users must have the ability to validate the performance of the DTs against their specific use case. Without some ability to verify the performance in real time during an experiment, users will be reluctant to use tools provided by the facility no matter how well documented and vetted.

2.3 Software Ecosystems and Computing Infrastructure

For successful deployment, we need to have a robust data path (maintain integrity, low latency, high throughput, with a low-touch interface), the ability to deploy code on remote HPC (many different architectures, environments, and local policy implementations), the ability to control/messaging pathways between distributed infrastructures, and workflow orchestration tools to allow users to deploy desired DTs with the desired parameters (adapted to the data/situation), parallelization, and ability to scale to HPC.

Another aspect is visualization and user interfaces. Even in a fully-autonomous DT environment, it is vital to monitor data analysis progress and measurement results as well as the recommendations from autonomous data analysis during experiments. Depending on the specific aspects of the experiment, it may be advantageous to incorporate a human in the loop, enhancing human insight with autonomously generated information. Development of tools that allow users to monitor status and interact with live data as its being produced to understand the performance of the data pipeline will be necessary to assist in the trustworthy deployment of such models. Similarly, for facility control, human operators typically need to monitor and interpret numerous signals streaming from the facility; having visualization tools coupled to a DT will be essential for aiding monitoring and diagnosis of operational issues, as well as aid decision making for improving or maintaining beam quality being delivered to experiments.

It also essential to develop software ecosystems that can be used across different facilities. For example, accelerator systems (even beyond light source use-cases) share many similar component designs, control challenges, and types of diagnostics. Photon science experiments and beamlines similarly share many characteristics and types of diagnostics. The establishment of a multi-facility framework and/or modular ecosystem of tools that can handle fast data analysis, autonomous experiment steering, and facility monitoring and control is crucial for supporting advanced research.

Computing systems are important to facility operation, data interpretation, and scientific productivity. DTs require computing to run simulations, to analyze experiments, to train new models, to retrain or adapt existing models to new modalities, and coordinate end-to-end workflow execution. As LCLS-II and subsequent upgrades ramp up, the computing needs will exceed available local resources. LCLS plans to make use of DOE Leadership Class Facilities such as NERSC for all aspects of DT workflows. Analyzing data and retraining/adapting models on experimental timescales will require a data path of sufficient bandwidth and integrity, tools for automated network orchestration, as well as the ability to deploy complex workflows deployed across local edge, including ASIC, FPGA and other AI/ML accelerator resources, local compute, and remote HPC resources. Workflow orchestration tools are needed to automate data movement, analysis tasks, and control algorithms to translate actionable information to directives used to steer the accelerator, experiment, or beamline based on the results derived from DTs executing elsewhere.

Seamless connection to Leadership Class Facilities will be vital to successful deployment of DTs at the light sources. The High Performance Data Facility (HPDF) and Integrated Research Infrastructure (IRI) program will be integral components in the overall strategy to harness the tremendous data rates and volumes produced by the light sources and neutron sources. Interoperability and uniform deployment of code on the leadership class facilities and uniform methods for securely transporting data (with integrity and low latency) will be vital to the successful deployment and use of DTs to steer accelerators and experiments and analyze scientific data in the future.

2.4 Other Considerations:

International Collaborations on DTs: There are numerous light sources worldwide that share similar challenges in operation, modeling, and data handling. For example, the layout, component design, and operation of the normal-conducting accelerator for LCLS is very similar to the accelerator for SwissFEL. This presents many opportunities for shared use and development of technology for DTs, ranging from software frameworks for deployment to AIML techniques and even direct transfer of ML models between facilities. Funding support and incentives for shared software development are needed.

Long Term Research Investments: Aside from the topics described earlier, some examples of long-term research topics include foundation models for enabling cross-facility and cross-experiment DTs, methods to appropriately integrate quantum computing (e.g. in cases where it is beneficial to include quantum simulations for physical processes, or solve complex optimization problems suited to quantum computing), and fully integrated co-design of future light source facilities to aid designing machines with the highest efficiency, broadest experimental capabilities, and/or best performance across target experiments. Improvements in operation of present facilities through the use of DTs, alongside the software infrastructure required to create them, can help inform design of future ones.

Workforce Development: DTs can also be used in training and workforce development. Training of beamline scientists and accelerator operators can leverage offline copies of DTs. Facility-agnostic infrastructure for DTs could be used at smaller-scale accelerators, such as those at universities, exposing students to advanced computing workflows, AI/ML, and advanced control in addition to preparing them for careers at light sources.

Business Case Analysis: In 2015, it was estimated that approximately 400 hours were spent on manual tuning by accelerator operators for initial experiment setup at LCLS, corresponding to 9-10 additional user experiments and millions of dollars equivalent value (taking into account the operating budget). ML-based automated tuning algorithms informed by system models have shown substantial improvements in tuning speed (e.g. 20x faster electron injector tuning [21]). Leveraging a DT of the entire facility in online tuning would substantially reduce time-to-delivery.

Going beyond the efficiency of individual facilities, the similar needs and structures across facilities means that critical infrastructure for DTs can be shared across them. Rather than having individual facilities construct their own infrastructure (which is a very large undertaking), ensuring software can be shared and co-developed across facilities will ensure resources are not wasted on duplicative efforts. Joint developments across facilities (including modular, interoperable software developments and standards) will ensure modeling and control technology can be readily transferred between facilities.

3 References

[1] *Networking and Information Technology Research and Development Request for Information on Digital Twins Research and Development.* https://www.federalregister.gov/documents/2024/06/18/2024-13379/networking-and-information-technology-research-and-development-request-for-information-on-digital?utm_medium=email&utm_source=govdelivery.

[2] Louise Wright and Stuart Davidson. "How to tell the difference between a model and a digital twin". In: *Advanced Modeling and Simulation in Engineering Sciences* 7.1 (Mar. 2020), p. 13. ISSN: 2213-7467. DOI: 10.1186/s40323-020-00147-4. URL: https://doi.org/10.1186/s40323-020-00147-4.

[3] C. E. Mayes et al. "Lightsource unified modeling environment (LUME), a start-to-end simulation ecosystem". In: *Proc. of IPAC.* 2021, THPAB217. URL: www.lume.science.

[4] Christopher Mayes, Ryan Roussel, and Hugo Slepicka. *ChristopherMayes/Xopt: Xopt v0.5.0.* Version v0.5.0. Oct. 2021. DOI: 10.5281/zenodo.5559141. URL: https://doi.org/10.5281/zenodo.5559141.

[5] Tobias Boltz et al. "More Sample-Efficient Tuning of Particle Accelerators with Bayesian Optimization and Prior Mean Models". In: *arXiv preprint arXiv:2403.03225* (2024).

[6] Ryan Roussel et al. *Bayesian Optimization Algorithms for Accelerator Physics.* 2023. arXiv: 2312.05667 [physics.acc-ph].

[7] Axel Levy et al. "Revealing biomolecular structure and motion with neural ab initio cryo-EM reconstruction". In: *bioRxiv* (2024), pp. 2024–05.

[8] Youssef Nashed et al. "Heterogeneous reconstruction of deformable atomic models in Cryo-EM". In: *arXiv preprint arXiv:2209.15121* (2022).

[9] David A Klindt et al. "Towards interpretable Cryo-EM: disentangling latent spaces of molecular conformations". In: *Frontiers in Molecular Biosciences* 11 (2024), p. 1393564.

[10] Axel Levy et al. "Solving Inverse Problems in Protein Space Using Diffusion-Based Priors". In: *arXiv preprint arXiv:2406.04239* (2024).

[11] Jay Shenoy et al. "Scalable 3D Reconstruction From Single Particle X-Ray Diffraction Images Based on Online Machine Learning". In: *arXiv preprint arXiv:2312.14432* (2023).

[12] Oliver Hoidn, Aashwin Ananda Mishra, and Apurva Mehta. "Physics constrained unsupervised deep learning for rapid, high resolution scanning coherent diffraction reconstruction". In: *Scientific Reports* 13.1 (2023), p. 22789.

[13] Sathya R Chitturi et al. "Capturing dynamical correlations using implicit neural representations". In: *Nature Communications* 14.1 (2023), p. 5852.

[14] Ryan Roussel et al. *Efficient 6-dimensional phase space reconstruction from experimental measurements using generative machine learning.* arXiv:2404.10853 [physics]. May 2024. DOI: 10.48550/arXiv.2404.10853. URL: http://arxiv.org/abs/2404.10853 (visited on 07/10/2024).

[15] Jan Kaiser et al. “Bridging the gap between machine learning and particle accelerator physics with high-speed, differentiable simulations”. In: *Physical Review Accelerators and Beams* 27.5 (2024), p. 054601.

[16] Mathieu Blondel et al. *Fast Differentiable Sorting and Ranking*. 2020. arXiv: `2002.08871 [stat.ML]`. URL: `https://arxiv.org/abs/2002.08871`.

[17] Lawrence Stewart et al. *Differentiable Clustering with Perturbed Spanning Forests*. 2023. arXiv: `2305.16358 [cs.LG]`. URL: `https://arxiv.org/abs/2305.16358`.

[18] Felix Petersen et al. *Differentiable Top-k Classification Learning*. 2022. arXiv: `2206.07290 [cs.LG]`. URL: `https://arxiv.org/abs/2206.07290`.

[19] Felix Petersen et al. *Deep Differentiable Logic Gate Networks*. 2022. arXiv: `2210.08277 [cs.LG]`. URL: `https://arxiv.org/abs/2210.08277`.

[20] Nicholas Schwarz et al. “Enabling Scientific Discovery at Next-Generation Light Sources with Advanced AI and HPC”. In: *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI. 17th Smoky Mountains Computational Sciences and Engineering Conference, SMC 2020, Oak Ridge, TN, USA, August 26-28, 2020, Revised Selected Papers* (Aug. 26–28, 2020). Ed. by Jeffrey Nichols et al. Springer, Cham, 2020, pp. 145–156. DOI: `10.1007/978-3-030-63393-6`.

[21] Sara Ayoub Miskovich et al. “Multipoint-BAX: a new approach for efficiently tuning particle accelerator emittance via virtual objectives”. en. In: *Machine Learning: Science and Technology* 5.1 (Jan. 2024). Publisher: IOP Publishing, p. 015004. ISSN: 2632-2153. DOI: `10.1088/2632-2153/ad169f`. URL: `https://dx.doi.org/10.1088/2632-2153/ad169f` (visited on 05/10/2024).