2022 Review of Data-Driven Plasma Science

Rushil Anirudh, Rick Archibald, M. Salman Asif, Markus M. Becker, Sadruddin Benkadda, Peer-Timo Bremer, Rick H. S. Budé, C. S. Chang, Lei Chen, R. M. Churchill, Jonathan Citrin, Jim A. Gaffney, Ana Gainaru, Walter Gekelman, Tom Gibbs, Satoshi Hamaguchi, Christian Hill, Kelli Humbird, Sören Jalas, Satoru Kawaguchi, Gon-Ho Kim, Manuel Kirchen, Scott Klasky, John L. Kline, Karl Krushelnick, Bogdan Kustowski, Giovanni Lapenta, Wenting Li, Tammy Ma, Nigel J. Mason, Ali Mesbah, Craig Michoski, Todd Munson, Izumi Murakami, Habib N. Najm, K. Erik J. Olofsson, Seolhye Park, J. Luc Peterson, Michael Probst, David Pugmire, Brian Sammuli, Kapil Sawlani, Alexander Scheinker, David P. Schissel, Rob J. Shalloo, Jun Shinagawa, Jaegu Seong, Brian K. Spears, Jonathan Tennyson, Jayaraman Thiagarajan, Catalin M. Ticoş, Jan Trieschmann, Jan van Dijk, Brian Van Essen, Peter Ventzek, Haimin Wang, Jason T. L. Wang, Zhehui Wang, Kristian Wende, Xueqiao Xu, Hiroshi Yamada, Tatsuya Yokoyama, and Xinhua Zhang

Manuscript received 28 May 2022; revised 15 February 2023 and 25 March 2023; accepted 1 April 2023. Date of publication 9 August 2023; date of current version 22 August 2023. This work was supported in part by the U.S. Department of Energy, the Office of Science, and the Office of Advanced Scientific Computing Research, SciDAC Program, through the FASTMath Institute, Oak Ridge National Laboratory, under Contract DE-AC05-00OR22725; in part by the United States Department of Energy (DOE)/High Energy Physics (HEP) under Grant DE-SC0016804; in part by the U.S. DOE through the Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344; in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 776262 (AIDA, www.aida-space.eu); in part by the European Union via the Euratom Research and Training Programme under Grant Agreement 101052200—EUROfusion; in part by the JSPS Core-to-Core Program under Grant JPJSCCA2019002; and in part by the National Research Council of Science and Technology (NST) grant by the Korea Government (MSIT) under Grant CRC-20-01-NFRI. The work of Markus M. Becker was supported in part by the German Federal Ministry of Education and Research (BMBF) under Grant 16FDM005 and Grant 16QK03A. The work of M. Salman Asif was supported in part by U.S. Air Force Office of Scientific Research Award FA9550-21-1-0330. The work of Sadruddin Benkadda and Satoshi Hamaguchi was supported by the CNRS International Research Project (IRP) FJ-IPL. The work of Rick H. S. Budé and Jan van Dijk was supported in part by the PPS-Contribution Research and Innovation of the Ministry of Economic Affairs and Climate Policy (The Netherlands) and in part by ASML. The work of C. S. Chang and R. M. Churchill was supported by the U.S. DOE Office of Fusion Energy Science and the Office of Advanced Computing Research under Contract DE-AC02-09CH11466 to Princeton University on behalf of the Princeton Plasma Physics Laboratory, the SciDAC Partnership Center for High-Performance Boundary Plasma Simulation, and the Theory Department. The work of Satoshi Hamaguchi was supported in part by the Japan Society of Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (S) 15H05736 and (A) 21H04453, in part by the JSPS Core-to-Core Program under Grant JPJSCCA2019002, and in part by the Osaka University International Joint Research Promotion Programs. The work of Todd Munson was supported in part by the U.S. Department of Energy, the Office of Science, and the Office of Advanced Scientific Computing Research, SciDAC Program, through the FASTMath Institute, Argonne National Laboratory, under Contract DE-AC02-06CH11357. The work of Habib N. Najm was supported by the U.S. Department of Energy, the Office of Science, and the Office of Advanced Scientific Computing Research, SciDAC Program. The work of Jan Trieschmann was supported in part by the Deutsche Forschungsgemeinschaft (DFG), German Research Foundation, under Project 138690629—TRR 87 and Project 434434223—SFB 1461; and in part by Tokyo Electron Technology Solutions Ltd. The work of Zhehui Wang was supported in part by the U.S. Department of Energy through the Los Alamos National Laboratory. The work of Kristian Wende was supported by the German Ministry of Education and Research (BMBF) under Grant 03Z22DN12 and Grant 03Z22D511. The review of this article was arranged by Senior Editor E. Schamiloglu. *(Corresponding author: Satoshi Hamaguchi.)*

Please see the Acknowledgment section of this article for the author affiliations.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPS.2023.3268170.

Digital Object Identifier 10.1109/TPS.2023.3268170

***Abstract*— Data-driven science and technology offer transformative tools and methods to science. This review article highlights the latest development and progress in the interdisciplinary field of data-driven plasma science (DDPS), i.e., plasma science whose progress is driven strongly by data and data analyses. Plasma is considered to be the most ubiquitous form of observable matter in the universe. Data associated with plasmas can, therefore, cover extremely large spatial and temporal scales, and often provide essential information for other scientific disciplines. Thanks to the latest technological developments, plasma experiments, observations, and computation now produce a large amount of data that can no longer be analyzed or interpreted manually. This trend now necessitates a highly sophisticated use of high-performance computers for data analyses, making artificial intelligence and machine learning vital components of DDPS. This article contains seven primary sections, in addition to the introduction and summary. Following an overview of fundamental data-driven science, five other sections cover widely studied topics of plasma science and technologies, i.e., basic plasma physics and laboratory experiments, magnetic confinement fusion, inertial confinement fusion and high-energy-density physics, space and astronomical plasmas, and plasma technologies for industrial and other applications. The final section before the summary discusses plasma-related databases that could significantly contribute to DDPS. Each primary section starts with a brief introduction to the topic, discusses the state-of-the-art developments in the use of data and/or data-scientific approaches, and presents the summary and outlook. Despite the recent impressive signs of progress, the DDPS is still in its infancy. This article attempts to offer a broad perspective on the development of this field and identify where further innovations are required.**

***Index Terms*— Artificial intelligence, data-driven plasma science, machine learning, nuclear fusion, plasma control, plasma diagnostics, plasma processing, plasma simulation.**

NOMENCLATURE

AI	Artificial intelligence.
AIDA	Artificial intelligence data analysis.
AI/ML	Artificial intelligence/machine learning.
AM	Atomic and molecular.
AML	Adaptive machine learning.
ANN	Artificial neural network.
APC	Advanced process control.
AWAKE	Advanced proton-driven plasma wakefield acceleration experiment.

BCA	Binary collision approximation.
BE	Boltzmann equation.
CCD	Charge-coupled device.
CMEs	Coronal mass ejections.
CMOS	Complementary metal–oxide–semiconductor.
CNN	Convolutional neural network.
CRP	Coordinated research project.
CS-MRI	Compressed sensing magnetic resonance imaging.
CSP	Computational singular perturbation.
CT	Computed tomography.
CTS	Collective Thomson scattering.
CV	Cross-validation.
DCGAN	Deep convolutional generative adversarial network.
DDPS	Data-driven plasma science.
DFT	Density functional theory.
DM	Diffusion map.
DNN	Deep neural network.
DRM	Dimension reduction method.
DT	Deuterium–tritium.
EEDF	Electron energy distribution function.
EFFIS	End-to-end framework for fusion integrated simulation.
EES	Equipment engineering system.
EM	Electromagnetic.
EUV	Extreme ultraviolet.
EVDF	Electron velocity distribution function.
FACET-II	Facility for Advanced Accelerator Experimental Tests.
FCN	Fully convolutional network.
FDC	Fault detection and classification.
FDP	Fusion data platform.
FEL	Free electron laser.
FGM	Flamelet generated manifold.
FOV	Field of view.
GAN	Generative adversarial network.
GAP	Gaussian approximation potential.
GB	Gigabyte(s).
GP	Gaussian process.
GPR	Gaussian process regression.
GSA	Global sensitivity analysis.
GST	Goode Solar Telescope.
HEDP	High-energy density plasma.
HLLE	Hessian locally linear embedding.
HMI	Helioseismic and magnetic imager.
HPC	High-performance computing.
HTS	High-throughput screening.
IAEA	International Atomic Energy Agency.
ICA	Independent component analysis.
ICF	Inertial confinement fusion.
IDA	Integrated data analysis.
IFE	Inertial fusion energy.
ILDM	Intrinsic low-dimensional manifold.
IRSA	Infrared Science Archive.
ITER	International Thermonuclear Experimental Reactor.
JSPS	Japan Society of Promotion of Science.
KPCA	Kernel principal components analysis.
LAPD	Large plasma device.
LCLS	Linac coherent light source.
LDA	Linear discriminant analysis.
LLNL	Lawrence Livermore National Laboratory.
LOS	Line-of-sight.
LPS	Longitudinal phase space.
LSST	Large Synoptic Survey Telescope.
LSTM	Long short-term memory.
LWFA	Laser wakefield accelerator.
MB	Megabyte(s).
MCS	Monte Carlo simulation.
MCF	Magnetic confinement fusion.
MCMC	Markov chain Monte Carlo.
MCU	Microcontroller unit.
MD	Molecular dynamics.
MDS	Metadata schema.
ME	Milne–Eddington.
MGI	Materials Gnome Initiative.
MHD	Magnetohydrodynamics.
ML	Machine learning.
MLP	Multilayer perceptron.
MOS	Mean opinion score.
MPC	Model predictive control.
MR	Magnetic resonance.
MRI	Magnetic resonance imaging.
NERSC	National Energy Research Scientific Computing Center.
NIF	National Ignition Facility.
NIRIS	Near InfraRed Imaging Spectropolarimeter.
NN	Neural network.
NPE	Neural posterior estimation.
OES	Optical emission spectroscopy.
OLED	Organic light-emitting diode.
PC	Polynomial chaos or principal component.
PCA	Principal component analysis.
PCR	Principal component regression.
PDE	Partial differential equation.
PES	Potential energy hypersurface.
PI	Plasma information.
PI-VM	Plasma information-based virtual metrology.
PIML	Physics-informed machine learning.
PINN	Physics-informed neural network.
PIV	Particle imaging velocimetry.
PMT	Photomultiplier tube.
PSI	Plasma–surface interaction.
PTV	Particle tracking velocimetry.
PWFA	Plasma wakefield acceleration.
QDB	Quantemol chemistry database.
QoI	Quantity of interest.
QSSA	Quasi-steady-state approximation.
R&D	Research and development.
ResNet	Residual network.
RF	Radio frequency.
RL	Reinforcement learning.
SciDAC	Scientific Discovery through Advanced Computing.
SDO	Solar Dynamics Observatory.

SR	Super-resolution.
SVM	Support vector machine.
SWx	Space weather.
TB	Terabyte(s).
TOX	Thermal oxide.
UQ	Uncertainty quantification.
VAMDC	Virtual Atomic and Molecular Data Center.
VCG	Vickrey–Clarke–Groves.
VDF	Velocity distribution function.
VESPA	Virtual European Solar and Planetary Access.
VM	Virtual metrology.
VUV	Vacuum ultraviolet.

I. Introduction

PLASMA science, like other branches of natural science, such as particle and high-energy physics, condensed matter physics, physics of fluids, nuclear physics, atomic and optical physics, astrophysics, cosmology, material science, biology, and chemistry, is founded on experiments and observations. Experimental data-driven activities through the collection of experimental data, analysis of the data, reduction of the data to knowledge and comparison of experimental data with theory, and computational and statistical models play a central role in plasma science and technology. In recent years, DDPS and technology are going through a renaissance, picking up new meanings, and revealing unexplored directions because of the advances in data science and technology both inside and outside the domain of plasma research and applications.

One of the most widely known, and possibly most successful, examples of data-driven science applied to conventional scientific disciplines is the MGI [1]. Similar projects also took place around the world around the same time. In this project, the search for new functional materials was assisted by data-driven approaches, rather than the experience and intuition of engineers in materials science, and the efficiency of discovery of new materials is said to have been significantly improved. In the project, not only the existing material data were fully exploited by newly developed ML techniques and AI but also efficient methods to collect a large amount of data in relatively short periods, which is called "HTS," were also developed. In general, in materials science, shortage or the lack of data is often the problem for efficient material discovery, so the development of HTS techniques, especially those fit the latest ML and AI techniques, played a key role in the success of the MGI project.

Similarly, the search for the best plasma conditions for specific applications, such as nuclear fusion and semiconductor device manufacturing, is often one of the most essential R&D activities in plasma science and technologies. Therefore, similar approaches developed in the MGI may also be useful in this field. Especially, systematic collection, classification, and improved accessibility of data for reuse may also be crucial in promoting data-driven approaches to problem-solving in plasma science and technologies.

Data-driven science is sometimes called the fourth paradigm of discovery [2]. The previous three paradigms are empirical or experimental (Galileo Galilei), theoretical (Issac Newton), and computational (it may be hard to credit a single person for this) according to a classification by Jim Gray in his talk at the Computer Science and Telecommunications Board of the National Research Council in 2007. Data-driven science is fundamentally different from the previous three paradigms and, thus, transformational. Most notably, it could take human intelligence out of the discovery process and make fully automated scientific discovery possible through AI. It has been predicted by Frank Wilczek that such a transition could take about 100 years [3], [4].

We may recognize several pillars in DDPS: availability of big data (come in different forms), availability of a large number of advanced algorithms and methods, including theoretical-driven algorithms, such as a finite element solver and statistical-driven algorithms, and availability of the inexpensive computational platform. We have summarized the current status of data-driven research activities in plasma science and technologies in this review article. The article is organized in the following manner. In Section II, fundamental data science is briefly reviewed, especially in light of applications to plasma science and technologies in general. In Section III, examples of data-driven approaches for the analyses of basic plasma physics and laboratory experiments are discussed. In Section IV, an overview of data-driven analyses in MCF research is presented. Another large field of high-power/high-energy plasma physics is ICF, whose latest data-driven analyses are presented in Section V. In space and astronomical plasmas, a large amount of observational data has been accumulated over many decades, and data-analytic techniques have been extensively studied. The latest development of such research activities is summarized in Section VI. Plasma technologies are also widely used in industries, and cost-effective development is always of interest to the industries. Some latest developments of data-driven approaches to R&D in industries and related academic problems are highlighted in Section VII. Section VIII discusses the current status of various databases that may be of interest to the plasma community. A final summary is given in Section IX.
[Zhehui Wang and Satoshi Hamaguchi]

II. Fundamental Data Science

A. Introduction

This section provides a brief overview of the present status and future direction of the fundamental analysis methods for scientific data, on which the present and future plasma data science rely. The transition from data to information is given in the data reduction, dimensionality reduction, and feature extraction processes. Data reduction and compression are mixed concepts that comprise data selection and feature extraction/reduction [5], [6], [7]. We discuss data reduction and compression methods that operate on diverse architectures and also on streaming data, capable of high compression rates while preserving targeted QoIs. We describe dimensional reduction and sparse modeling techniques that promote a scientific understanding of high-dimensional data and reduce the analysis and storage costs of the data. We note here that the feature extraction or 1-D CNN may not involve a dimensionality reduction and can be fully automatic, thus avoiding the use of any data model or heuristic [8]. We cover ML enhancements

to modeling and simulation that can be utilized to accelerate simulations and provide accurate and robust closure models. In addition, we discuss other fundamental ML methods throughout this section. Intrinsic to the analysis methods and tools presented in this section is the hardware used to execute these methods. We discuss the mapping of ML models to large-scale computing systems touching upon the techniques used to integrate hardware capabilities with the numerical methods presented in this section. We focus on advancements in workflow automation, which is necessary to store, move, and process the complex scientific data produced at leadership experimental and computing facilities. We overview the explosion of theory, algorithms, and tools that have been developed over recent decades in UQ. Advancements in visualization and data understanding are described, which can be used by domain experts to facilitate knowledge and discovery from scientific data. This section ends with ML control theories that are applicable to highly nonlinear and multivariable plasma dynamics, and can take into account the safety-critical plasma applications.
[C. S. Chang]

B. Data Reduction/Compression

Experimental, observational, and computational facilities are facing a crisis because of the large increase in data being produced at these facilities. New technologies allow more data to be captured at higher rates, which increases data volumes and velocities, and necessitates the need for streaming reduction techniques. Hence, there is a crucial need for fast reduction techniques that must work on diverse architectures and stream data across processes in complex workflows, ensuring that short- and long-term events can be captured and analyzed in the reduction process.

There are several cross-cutting challenges that are not specialized to a particular application and can be thought of as reduction motifs that can work for a variety of applications and can be further customized and tuned for different scientific instruments. The first motif is for reducing "noisy" data when the signal-to-noise ratio is low and where computational signatures are often needed to extract the signal. The second motif is for high-dimensional data, often illustrated in plasma physics applications, which often simulate 6-D physics. The third motif is for nonuniform and unstructured data, often produced by MHD codes. The fourth motif is for reducing data as it streams, which can be from a live experiment or from an exascale simulation. Finally, the fifth motif is to ensure that simple and complex QoIs from downstream processing have a user-specified uncertainty to ensure trustworthy data used for later postprocessing. Many of these motifs can be put together to illustrate new scientific challenges, for example, when large simulations that search for features, events, and anomalies and produce QoIs that could be combined with in situ ML and AI workflows to produce reduced order models in addition to a complete data model repository. In all of these cases, there is a crucial need for fast reduction techniques that must work on diverse architectures and stream data across processes in complex experimental workflows, ensuring that short-term events can be captured and analyzed in the reduction process.

Both compression and analysis share a common goal: to extract science from the raw data that involve extracting the essential structure and key features of the phenomenon under study while ignoring or discarding the noise and data that have little or no impact on the QoIs. It is important to understand how reduction methods affect the specific QoIs used in the analysis so that reduction does not alter the results of the analysis. Lossless methods have unfortunately been generally unable to achieve the high compression ratios needed to handle the large quantities of data generated by facilities, often reducing data by less than 15%. This means that we have to look at lossy methods, which bring the fidelity of the reduced data into question. Fidelity and reduction are directly in competition with each other, and so it is important to consider what exactly is required of a reduced dataset in order for it to serve as a scientifically useful surrogate.

Lossy compressors should be flexible with regard to the structure of the data, generalize to arbitrarily high dimensions, and allow control of errors both in the original degrees of freedom and in downstream QoIs. Compressing data in the same high-dimensional space where it is defined can make more of the data's spatial correlations visible to the compression algorithm, resulting in higher compression ratios. Similarly, compression algorithms should make use of as much of the data's spatial structure as possible. Compressing nonuniform or unstructured data as though it was defined on uniform grid risks obscuring redundancies and patterns in the data, resulting in lower compression ratios. Another design goal is the control of errors incurred by compression algorithms. Often, scientists are concerned with the change to the QoIs from the compressed data, hoping to make sure that all of the features in the QoIs are preserved to a high enough accuracy. The mathematics required to relate errors in the raw data to errors in QoIs is nontrivial, especially for QoIs that are nonlinear and/or obtained by complex postprocessing. Empirical approaches can provide estimates for, but not guaranteed bounds on, QoI errors by extrapolating from previously encountered datasets and QoIs. We can often look at feature extraction as a method to understand specific QoIs, which must be controlled during the reduction, but, sometimes, we need strict mathematical control of more than the features, such as requiring the magnetic field has zero divergences; this strict control of the error during reduction must be mathematically guaranteed.

Reduction algorithms need to be efficient as well, meaning that they need to use a minimal set of computational resources (time to solution, memory, network, and computational) and should be able to reduce the time to solution in application workflows. Reduction algorithms must further satisfy the following set of requirements: 1) ability to quantify the uncertainty of errors in the raw-data and the derived QoIs; 2) ability to be efficient in its use of computational resources; and 3) ability to work with high-dimensional data on structured and unstructured meshes, which sets the overarching requirements so that scientists can both trust and efficiently use the communities data reduction algorithms.

A 2018 survey by Li et al. [9] organized data reduction techniques for scientific data into five categories: truly lossless,

near-lossless, lossy, mesh reduction, and derived representations. Lossless compression includes techniques, such as entropy-based coders (such as Huffman coding [10], which is used by bzip2 [11], and arithmetic coding [12]) and dictionary-based coders (such as LZ77 [13] and LZ78 [14], which have inspired many variants, such as those used in DEFLATE [15], gzip [16], and zlib [17]). That said, lossless compression often achieves only modest reductions, for example, fpzip achieved a 3.7X reduction on a simulation of a Rayleigh–Taylor instability by the Miranda simulation code [18]. Near-lossless compression refers to rounding errors that occur during reconstruction from transforms; their reduction capabilities are often similar to lossless compression. Lossy compression includes techniques such as truncation, quantization, predictive coding schemes, and transform-based compression schemes. The Li survey points to many instances of lossy compression packages applied to scientific data, including MLOC [19], fpzip [20], ISABELA [21], SZ [22], VAPOR [23], JPEG2000 [24], and zfp [25]. Several of these are included in SDRBench [26], as well as some additional packages that have emerged in recent years: DCTZ [27], MGARD [28], and TTHRESH [29]. Mesh reduction techniques include decimation (surveyed by Weiss and De Floriani [30]), multiresolution techniques, subsetting (such as with querying with FastBit [31]), and temporal sampling (i.e., triggers [32], [33], [34]). Derived representation techniques use alternate representations of the data, typically statistical in nature, but also including approaches such as topological features [35] and imagery [36]. Recent work using ML to reduce scientific data [37], [38], [39] could also be considered a derived representation.

1) R&D Necessary for the Future: The provision of realistic numerical bounds is essential if the scientist is to have confidence in applying data reduction. In order for existing and future reduction algorithms to be "trustworthy," it will be necessary for algorithms to come with some kind of certificate or guarantee on the fidelity of the reduced data to the original data. This may take the form of rigorous mathematical bounds on the loss incurred measured in a norm that is relevant to the application. More generally, future research in data reduction procedures should ideally aim to provide the scientist with the capability to specify a set of application-dependent QoIs, which should be preserved to a user-specified tolerance. The reduction procedure should have the flexibility to effectively reduce the data while maintaining the set of QoIs to the level specified by the user and providing realistic bounds on the actual loss incurred. Certificates of this type are essential for the "trustworthiness" of the reduction routines.

In order for a reduction algorithm to be "effective," it must be capable of providing meaningful levels of reduction while incurring a level of loss that is acceptable to the user. Achieving a balance between these two competing criteria encapsulates the essential difficulty in developing effective data reduction algorithms and constitutes a major challenge for future research in data reduction. Nevertheless, a reduction algorithm is only effective if it is applicable to the types of data of interest. Many existing reduction algorithms are effective at reducing structured data, such as uniformly spaced data and data specified on tensor product grids. However, the performance or applicability of the algorithms to more general data formats, including unstructured grids and particle data, is less well-understood. Research into understanding whether or not, and how, existing approaches can be extended to more general types of data will be needed if effective algorithms are to be developed.
[Scott Klasky]

C. Dimensional Reduction and Sparse Modeling

Often in physics, a change of basis can greatly simplify the analysis of measured data and promote scientific understanding. This is the driving force behind dimensional reduction methods, remapping high-dimensional data to a compact representation in low-dimensional space that preserves information, increases understanding, and reduces analysis and storage costs. We will briefly review the dominate methods in this field of research, providing insight into these methods and their best uses.

At a high level, we split DRMs into two major categories, based upon if they have a nonlinear component of the method. We sample the space of DRMs by describing three linear and nonlinear DRMs. Specifically, we discuss the popular choices of PCA, LDA, and ICA for linear DRMs and KPCA, DMs, and ML for nonlinear DRMs.

One unique driving force in dimension reduction is sparsity in modeling and data. To describe the area of sparse modeling, we will focus on one method that utilizes most of the key mathematical tools of the sparse modeling field. Sparse phenomena can occur in scientific data; for example, a sparse set of vibrational modes in a material. Sparse spectral techniques are designed to find sparsity and use this to create dimension reduction transformations. We will discuss Laplacian eigenmaps and HLLE.

Even in this short review, we have touched on many viable DRMs. The key to using these methods is understanding the type of data that they were designed to analysis and employing them accordingly. When we introduce each method, we make sure to include a clear statement for the type of data that these methods are designed to operate on.

PCA has a long history in data analysis [40] and performs one type of transformation very effectively. PCA performs an orthogonal basis transformation where dimensions are ordered to represent as much of the data's variation as possible. Thus, often, most of the data's variation can be represented in a low number of dimensions.

Data variation is just one way of generating a basis set for data. If the data have class information, then this information may provide hints to an effective DRM. LDA uses class information to create dimension reduction by determining the best linear transformation that maximizes class separation while minimizing the scatter within a class.

It can be seen that the major difference between these linear DRMs is the procedure that is used to create a basis set to project onto. In the case of ICA [41], data are equated to a linear transformation of maximally independent components for some definition of independence. The two most common

definitions of independence are the minimization of mutual information and the maximization of non-Gaussianity.

Introducing nonlinear methods into DRM can produce more sophisticated methods. The KPCA method [42] uses kernel functions to generate representations of data. Some examples of these functions include Laplace, square, exponential, and Mátern kernels, to name a few. We refer the interested reader to [43] for more information. The key difference between PCA is that KPCA computes the principal eigenvectors of the kernel matrix, rather than those of the covariance matrix of the data. When kernels are used to transform data, the resultant analysis is done in a reproducing kernel Hilbert space, which can be a high-dimensional space, possibly infinite, compared to the original data space. KPCA is a more effective DRM compared to PCA if few eigenvalues are needed to capture the data information in the reproducing kernel Hilbert space than the original space.

Once common assumption in dimension reduction is that data reside on a manifold in high-dimensional space. Determining this manifold creates the ultimate representation for dimension reduction. DMs [44] use methods from dynamical systems and statistics to approximate data on manifolds in high-dimensional space. DM uses multiple Markov random walks on the graph of the data, measuring the so-called diffusion distance. Here, data manifolds can be determined by integrating over all paths through which the graph creates an isomap of the manifold, where short-circuiting helps discover the diffusion path and, hence, the manifold.

Sparse patterns in high-dimensional data can be identified using sparse spectral methods. These methods setup sparse (generalized) eigen problem that is capable of low-dimensional reductions while retaining the local structure of the data.

Laplacian eigenmaps find a low-dimensional data representation and preserve local properties of the manifold [45]. The Laplacian matrix is constructed using a weight function, where the distance between the nearest neighbors is used as the input. Minimization of a cost function based on the graph ensures that points close to each other on the manifold are mapped close to each other in the low-dimensional space, preserving local distances.

HLLE [46] is another type of sparse spectral method that learns manifolds in high-dimensional data, maintaining the balance between local geometric information and overfitting. Hessian is calculated at every datapoint and is used to create a localized parameterization of the manifold.

ML has become a dominate force in the field of dimension reduction and sparse modeling [47]. As described in this section, each method carried with it a set of assumptions on the form of data. ML has the flexibility to learn unique transforms beyond the data assumptions of this section. The limitations of ML are centered around the fact that these methods will only perform transformations that they are trained to execute. [Rick Archibald]

D. ML-Enhanced Modeling and Simulation

Computationally expensive operators in the system of equations being used to simulate the plasma can significantly impact our ability to sufficiently simulate the plasma. By replacing these expensive operators with less expensive surrogate models, we can improve the overall performance of the simulation. These surrogate models can be learned, for example, by applying ML techniques that use a corpus of data to train a DNN model. Alternative surrogate models can be obtained via techniques such as kernel regression, GPR, support vector regression, and dictionary learning. The resulting surrogate model is then used as a replacement for the operator in the simulation. However, the surrogate needs to conserve relevant physical quantities, such as mass and energy, for the resulting simulation that includes the surrogate model to be stable and meaningful. These physics-informed surrogates, such as PINNs [48] and physics-informed GPR [49], offer the potential, however, to be much faster to evaluate than the original operator while providing a sufficient approximation. As an example, the Fokker–Planck–Landau collision operator has a computational cost that grows at a quadratic rate as the number of species increases and needs to be evaluated many times when forming the right-hand side of the system of equations. ML surrogates have been successfully developed for this operator [50]. In other settings, data-driven ML models have been developed to estimate closures for plasma fluid models [51], [52] and fluid turbulence models [53] and for coupled simulations [54]. The amount of available data and the methods applied, including the network architecture for DNN models, all have an impact on the quality of the resulting surrogate models.

An assumption typically made when developing surrogates is the availability of a large corpus of data. In this data-rich regime, DNNs with many parameters and layers can be trained, often resulting in good surrogate models, as the information content of the data often exceeds the number of parameters. When only a small amount of data is available, greater care must be taken in the network architecture and regularization techniques to produce a reasonable surrogate model, especially when the amount of data available is less than the number of parameters in the surrogate model or the information content of the data is insufficient. Sparse NNs that are not fully connected between layers, for example, can be beneficial in the data-poor regime, and hyperparameter optimization methods can be applied to search for a sparse network architecture for the ML surrogate model. Regularization terms that promote sparsity can also be used with regression-based methods. Embedded physics knowledge can also reduce the amount of data required to train the surrogate model [48], as the equations reduce the number of degrees of freedom.

There are two basic approaches for incorporating physics knowledge into ML surrogate models: changing the architecture of the network or changing the training problem to penalize deviations from the physics constraints. Both approaches have benefits and limitations. These approaches have analogs in other regression-based methods, such as physics-informed GPR.

Changing the network architecture by adding projection layers guarantees that the ML surrogate used in the simulation will conserve the relevant physical quantities. Whenever a physical quantity that needs to be conserved is added, the

network has to be updated and the model retrained. Moreover, when the training data only approximately conserve the quantities, as is often the case with real and simulated measurements, the trained surrogate model may not satisfy the conservation constraints exactly. As a premium is placed on the satisfaction of the conservation constraints, the trained surrogate model may not represent the training data, as well as one would like.

Adding the deviation from the physics constraints to the loss function used during training can result in a good surrogate model that conserves the physical quantities. This approach offers flexibility in making a tradeoff between an accurate representation of the training data and the conservation error by adjusting the penalty weights. Conserving additional quantities, such as fluxes, amounts to adding new terms to the loss function and retraining. As the conservation constraints are not preserved exactly, one may not be able to completely rely on the trained ML surrogate model in the simulation. Moreover, choosing the best weights for the terms in the loss function and the formulation of the (scaled) conservation errors can greatly impact the quality of the trained surrogate model. Methods that dynamically adjust the penalties, using, for example, an augmented Lagrangian formulation, have been developed that can circumvent some of the challenges in determining the best weights.

The surrogate model that is integrated into the simulation may not always be a good approximation to the true operator, particularly under conditions not well represented in the training set or when rare events occur. By developing metrics to identify when the surrogate model is inadequate, such that when the output has a large conservation error, one can store the inputs and the simulation can continue by reverting to the computationally expensive operator when it is available and computationally tractable and store the correct outputs. The stored input/output data can be added to the corpus of data and a retraining strategy employed to produce an updated surrogate model. A complete workflow where the surrogate model is retrained while the simulation is running, with the improved surrogate model fed back to the simulation, can be pursued. [Todd Munson]

E. Mapping ML Models to Large-Scale Computing Systems

One challenge with achieving performance for simulation models of fusion science problems is sparsity and irregular data formats. This is a problem in multiple domains as well, and the simple explanation is that nature is connected in such a way that the elements that affect one another are not contiguous, and the connections are not always organized in an orderly fashion. The result is that, when the data are arranged in a matrix, there are large gaps where the array elements are zero, and the columns that are nonzero are often jagged. This, in turn, presents a problem for digital computers that want to calculate ordered sets of data, and obviously calculating an operation with zero is not very efficient. This issue is exacerbated when the natural phenomenon being modeled is inherently noisy or turbulent. In this case, the arrays are highly irregular, and the methods for modeling them accurately are challenging due to the high level of dimensionality.

The general challenge with these approaches is that most of the numerical algorithms used in science require full 64-bit precision, so two full 64-bit operands must be moved to generate a multiply or divide, and with a sparse array, these are often multiply or add zero.

There is a large body of work on sparse methods to help alleviate this problem, but they all introduce some level of overhead in their attempt to reorder the matrix or compact the elements. Another related issue is that many of the best-in-class algorithms use operations that are vector by vector or matrix by vector, and these often cannot fully utilize all the computational infrastructures.

Advanced ML algorithms have emerged as a mechanism to deal with some of these constraints. The ML algorithms, such as DNNs, can be classified as universal function approximators. So rather than start with the equations of state that define the natural phenomenon, the function approximators are trained from data. There are physics-informed methods that will use the equations of state as part of the model development, but the ML methods do not directly approximate the equations of state but use them to govern the loss or provide input to the training process.

In general, this approach has been shown to offer a number of advantages for developing models for complex natural phenomena in multiple domains. The overarching benefit is that the training of the model and the resulting inference are more efficient as they do not require full precision, and the algorithms used to train and execute the resulting inference can be posed as matrix-by-matrix operations where a digital computer can be much more efficient.

This allows the hardware to be more efficiently used or perhaps more efficient hardware to be efficiently used. The training process is generally more expensive but is done far less often than the inference, and both are more efficient than classical methods to model the same natural phenomenon. The matrices for the ML methods can be sparse, but they can generally use matrix-by-matrix operations and reduced precision. Thus, the overall improvement in time to solution and resource consumption is multiple orders of magnitude relative to classical methods.

The dramatic improvement in time to solution and resource consumption opens the door for the ML methods to be used both for data center and experimental use case settings. [Tom Gibbs]

F. Workflow Automation

The rate and size of data generated by cutting-edge experimental science facilities and large-scale simulations on current HPC systems are forcing scientists to move toward the creation of autonomous experiments and HPC simulations. However, efficiently moving, storing, and processing large amounts of data away from the point of origin present an incredible challenge. ML approaches are being used to learn insights from I/O patterns; in-memory computing, in situ analysis, data staging, and data streaming methods are being explored to transfer data between coupled workflows. However, many challenges remain to offer scientists the tools that they need to efficiently automate their workflows. Modern scientific workflows are

often collaborative in nature, consisting of multiple heterogeneous and coupled processes that must dynamically interact and exchange data on the fly during or after execution. This dynamic nature adds another layer of difficulty in managing these massive datasets.

Steering experiments in near real time is becoming critical, demanding further automation of the experimental scientific workflow. For example, scientists often run simulations and experiment analyses on different (sometimes geographically distributed) computing resources to simulate different components of the same physical phenomena. These codes need to interact with each other, and often, they must exchange data with analytical or visualization processes in near real time to help scientists understand the simulation results in a timely fashion. As a result, the need to automate these efforts has grown. Research in this field requires moving large amounts of data from the point of origin (e.g., simulations, experiments, and instruments) to HPC facilities that can perform reduction, analysis, and visualization.

A workflow management system capable of automating the execution of complex large-scale workflows must provide scientists with the ability to optimize their experiments for the maximized acceleration of the scientific process and use the experimental observations efficiently. There are currently several limitations and challenges that the science community faces in constructing resilient, distributed workflows for making NRT decisions. There is limited support for the semiautonomous, resilient execution of a workflow in NRT. Resiliency is constrained to addressing failures through general task restarts, where restarts may not be done in NRT, and policies are not implemented for tasks in order of their priority in a workflow. Support for dynamic control is limited to a set of basic actions that a user may take at runtime. The ability for a user to query or analyze workflow execution and steer a workflow using monitoring data is largely absent.

Automating workflows and coupling experiments with workflows and automating the data movement for real-time analysis and visualization are new research areas moved forward by the needs of current large-scale applications. In order to achieve this goal, several projects have made progress in resource allocation across multiple machines, data streaming over large areas, resiliency, security, and so on. Once such effort is the NERSC Superfacility [55] project that aims to provide an ecosystem of connected facilities and software for the NERSC computing center. Its main focus is on providing a vision for making resource reservations using an API that can be used to connect to the center's HPC systems. The project does not address resiliency concerns for distributed workflows and does not allow near real-time decision-making in their process.

There is currently a long list of workflow management systems and tools used by the scientific community focusing on different aspects of workflows. Examples include Pegasus [56], Kepler [57], RADICAL-Pilot [58], and others. A common theme across existing workflow management systems is the focus on execution patterns and optimizing computational throughput, dynamic support constrained to task restarts but almost no support for real-time data delivery, monitoring, and workflow steering. The EFFIS [59] framework, initially designed to loosely couple multiple fusion codes running on HPC resources, is a workflow management system that uses a combination of enabling technologies, including ADIOS [60], Kepler, and eSimMon [61], a web-based dashboard. EFFIS is built upon the Cheetah-Savanna [62] suite of workflow tools and provides an API for composing and executing codesign studies for online data analysis on different supercomputers. It supports both the execution of strongly coupled workflows on HPC resources [63] and the execution of data streaming from the fusion KSTAR experimental facility to NERSC [64]. EFFIS is being successfully used by applications at the Oak Ridge National Laboratory. However, it is only a first step toward providing a workflow infrastructure capable of efficiently coupling complex geographically distributed workflows.

An important aspect of workflow automation is managing distributed resources and dynamically controlling a running workflow. Scheduling schemes for supporting real-time jobs, along with traditional batch jobs on HPC systems, have been evaluated by the community in several studies [65], [66]. Current approaches include the usage of basic manual intervention and preprogrammed scripts to control a workflow dynamically. Challenges around live monitoring, analysis, and control of running workflows still remain open issues. In addition, runtime control in current solutions is limited to restarting failed tasks. There is little support for more resilient and policy-based execution of a distributed workflow.

The current workflow system does not provide native support for scientific data management middleware to tune data delivery. Most workflow systems cannot interact with streaming scientific data management frameworks. They support staging data as files across resources using tools such as GridFTP, but they do not provide low-level tuning of data streams for low-latency data delivery. Data objects are seen as black boxes, and support is provided only to the stage and persist them; there are no ways to switch between file- and stream-based options.

1) R&D Necessary for the Future: Coupling experiments with workflows containing simulations, surrogate models, analysis, and visualization codes requires geographically distributed resources: large-scale systems, edge devices at facilities, and computers at home institutions of the science team members. The science team must be able to discover and provision the resources required to execute their workflows and monitor the data generated by the workflow transparently. Since these workflows can be executed for making near real-time decisions, some components are critical to ensure that vital information is delivered in a timely fashion. Workflow management systems need to be able to provide rich monitoring and provenance information for a running workflow, an interface to steer the workflow dynamically, a resource management layer for elastic resource provisioning, and a policy-driven design for constructing resilient workflows.

Uninterrupted availability of data needs to be guaranteed. In order to support coupling of experiments with surrogate models across a distributed set of computational resources, it is essential to build a workflow infrastructure capable of

resiliently executing simulations with analysis and visualization, and transferring the information in NRT. Resiliency policies centered on priority-based redundant computations are needed to ensure that a set of surrogate models and analysis services is coupled in near real-time fashion in order to make timely decisions and control experiments.

The ability to dynamically control and steer workflows needs to be provided. Runtime validation of experiments and simulations via dynamically spawning models and analysis tasks is needed for more efficient usage of experimental resources. Consequently, this requires computational resources to grow elastically when they are needed. A command-and-control interface is needed so that scientists can make informed decisions on the experiment in near real time.

It is imperative to have a data streaming and management system capable of moving workflows and data efficiently through distributed resources. Streaming methods impose several different challenges: the data sources could be many: large-scale experiments, such as the Large Hadron Collider, or the results of large-scale simulations; the data might need to be processed in real time and streamed directly to the data analysis processes completely bypassing the file systems; and the data producer and the data consumer are often independent programs running on different nodes or systems geographically distributed. Ultimately, these tools and services will allow the sharing of machine data between the experimental analysis and computational simulations, which will allow scientists to steer their analysis using AI-/ML-based surrogate models, helping them better use their time on the experiments.
[Ana Gainaru]

G. Uncertainty Quantification

Recent decades have seen increasing awareness of the role that UQ can play in science and engineering. The goal of UQ in this context is to enable predictive computations of physical systems by providing means for quantification of uncertainty in computations, including estimation of uncertainty in model inputs, parameters, and structure, and the forward propagation of these uncertainties to model output predictions. UQ encompasses developments in applied mathematics and statistics, including fundamental theory, numerical algorithms, and software tools, for the assessment of uncertainty in computational models and their predictions [67], [68], [69], [70], [71], [72]. It is relevant across the board in the modeling of physical systems, including two essential elements, namely, the inverse UQ problem, relevant in learning model parameters/inputs from experimental, observational, or computational data [73], [74], [75] and the forward UQ problem involving the propagation of uncertainty from parameters/inputs to outputs of computational models [67], [76], [77], [78]. A range of other UQ activities builds on these fundamental components, including hypothesis testing, model selection and validation, and optimal experimental design, as well as robust optimization and control under uncertainty. In the following, we present brief highlights of the state of the art and challenges in UQ, focusing on the probabilistic UQ framework.

The primary goal of the inverse UQ problem is model calibration/fitting or parameter estimation, accounting for data noise/uncertainties and model error, to arrive at learned uncertain model parameters/inputs. The probabilistic framework, and thus the statistical inverse problem setting, provides improved conditioning of the often notoriously ill-conditioned inverse problem. This is due to the change from the original inverse problem goal of estimation of the best-fit value (which may be nonunique and surrounded by many local minima) to the estimation of a distribution on the regions in parameter space with a high probability measure. Furthermore, specifically in the Bayesian inference context, the use of priors provides for additional regularization that is often indispensable. In the Bayesian context, the goal of the inverse UQ problem is the estimation of a posterior density on model parameters/inputs. The challenges of model complexity, computational cost, and high dimensionality have always been considerable obstacles to the application of statistical inversion in large-scale computational models of physical systems. This is particularly true when, as is often the case, multiple challenges are present simultaneously. The forward UQ problem involves the propagation of uncertainty, e.g., specified with a given density, from model inputs to outputs. While not plagued with ill-conditioning as in the inverse problem, the forward UQ problem is similarly challenged with model complexity, cost, and high dimensionality.

In order to facilitate inverse and forward UQ in relevant problems, considerable effort has targeted the development of surrogate models that, when fit to represent the dependence of computational model observables or QoIs on parameters of interest, can be substituted for the original model. Surrogate models have been built employing a wide array of technologies. One approach, employing expansions in orthogonal basis functions, particularly PC [67], [77], [78], [79], [80], [81], [82] expansions, has been a considerable focus in forward UQ, the result of which is also precisely the surrogate needed for the inverse problem [83], [84]. PC constructions have been fit using generalized sparse quadrature and regression methods, often relying on sparsification via compressive sensing [85], [86], [87], [88] when high-dimensional. Other surrogate constructions have employed interpolants [89], [90], [91], [92], [93], low-rank tensors [94], [95], [96], GPs [97], [98], [99], and NNs [48], [100], [101]. Moreover, multilevel/multifidelity methods have emerged as essential means to facilitate surrogate model constructions in high-cost computational models [102], [103], [104], [105], allowing the use of model computations at varying degrees of fidelity/resolution, and hence cost, to achieve requisite surrogate accuracy at much reduced cost.

In the inverse UQ context, MCMC methods are generally used to solve the Bayesian inverse problem by providing samples from the posterior density on model parameters, thereby enabling estimation of the posterior density or moments thereof [75], [106], [107], [108], [109]. Advanced MCMC methods have been developed to deal with the complexity and high dimensionality of posterior distributions [110], [111], [112], [113]. Furthermore, approximate Bayesian computation (ABC) methods have been developed to tackle expensive/intractable models/likelihoods [114], [115], [116]. Developments in this area have also pursued the design of

reduced representations and distance metrics that address the complexity of the model response, particularly in dynamical systems, to provide tractable dynamical observables and likelihood loss functions that are smooth in parameters of interest while capturing essential dynamical features [117], [118]. Addressing high dimensionality in Bayesian inference has also led to advances in the design of MCMC methods for infinite-dimensional problems [119] and in identifying lower dimensional subspaces where data are, in fact, informative [120], [121] and where MCMC random sampling can be focused.

These and other developments have been documented in reviews/books [67], [68], [70], [122], [123] and deployed in open-source tools [124], [125], [126]. Resulting capabilities have enabled the use of UQ methods in complex problems of physical relevance, including, e.g., transport in porous media [127], [128], seismic sensing [129], fluid dynamics [77], [130], [131], plasma physics [132], [133], [134], [135], [136], [137], chemistry [138], [139], reacting flow [140], and materials [141], [142].

Despite these achievements, numerous open challenges remain in the practical use of UQ methods. High dimensionality remains a universal challenge, particularly when combined with model cost and complexity. The identification of a sufficiently low-dimensional subspace of "important" parameters, e.g., using GSA, is crucial for facilitating UQ in practical problems. Often, however, high dimensionality is an inherent, irreducible challenge, e.g., when dealing with models where there is no such lower dimensional important space. Examples include problems where nonsmooth observables/QoIs are of interest, e.g., detailed turbulent motions or material fracture. They also include DNN models, whose practical utility relies on their expressiveness that comes with the exceedingly large number of weight parameters. The DNN setting is also highly challenged by the remarkable complexity of the loss surface and the lack of informed priors. Even where there is a low-dimensional subspace of important parameters, however, GSA is in-itself a challenge with expensive models, particularly when dealing with models having discontinuities/bifurcations. Such artifacts are also generally problematic with surrogate construction. Other challenges include the estimation of probabilities of rare events and the design of smooth, sufficiently informative, observables in dynamical systems. These challenges render subsequent "outer loop" UQ activities doubly difficult. This includes, e.g., model marginal-likelihood/evidence estimation for model selection purposes, Bayesian optimal experimental design, and optimization/control under uncertainty.
[Habib N. Najm]

H. Visualization and Data Understanding

One of the primary challenges facing scientists is extracting understanding from the large amounts of data produced by simulations, experiments, and observational facilities. The use of data across the entire lifetime ranging from real time to post hoc analysis is complex and varied, typically requiring a collaborative effort across multiple teams of scientists. The rapid growth in the rate and size of data generated at these facilities makes gaining understanding even more difficult.

This section highlights examples of simulation and experimental plasma data collection and visualization. A comprehensive review paper on the subject would be lengthy and in all probability obsolete in several years. In the 1970s, plasma datasets of 10 MB seemed enormous. Computers and visualization software were in their infancy. Now, TB datasets are becoming common. More than 50% of the human brain is devoted to visual processing [143], so the use of visualization to understand such large and complex data is essential.

There is a wide range of software available for analysis and visualization of scientific data. Many packages (e.g., Python, Interactive Data Language (IDL), MATLAB, and Mathematica) have easy-to-use visualization routines built in. A number of tools for 3-D visualization are also available. Tools such as VisIt [144] and ParaView [145] have a broad range of capabilities. These tools can be used for post hoc processing and in situ processing using LibSim [146] and Catalyst [147]. A service-based approach to analysis and visualization [148] builds upon a hardware-portable visualization toolkit (VTK-m) [149] and a data model (Fides) [150] that can be integrated into automated workflows, such as EFFIS. The serviced-based approach supports both post hoc and in situ processing, and provides the flexibility for workflow systems to schedule analysis and visualization tasks as they are needed. To support collaboration between teams of scientists, a web-based dashboard, eSimMon [151], allows scientists to see near-real-time visualizations of different variables and QoIs.

As the complexity of data processing and visualization has grown, scientists have relied on complex workflows to orchestrate the collection and/or generation data, as well as the processing and movement throughout the lifetime of a scientific campaign. To be truly useful, analysis and visualization tasks must be able to access data in a variety of different ways in these complex workflows. These range from traditional post hoc visualization where data are accessed from disk to in situ visualization where the data are accessed as it is being generated. Visualization tasks must be able to integrate in a robust manner into these autonomous workflows and be dynamically controlled.

Fusion science requires a number of different techniques for gaining an understanding of the data being generated. These include both 2-D and 3-D visualization. In simulations, because of the nature of fusion science, a large amount of analysis can be done using 2-D slices through the simulation mesh. These 2-D slices can be colored by different quantities and provide a summary view of the fusion around the tokamak. An example would be the averaged value in the toroidal direction at each point in the plane. The time evolution of these visualizations can provide valuable insight into the behavior of the simulations. The use of 3-D visualization can be used to illustrate the changing complexity of the physics in the toroidal direction. These provide valuable insight into how features in the plasma vary toroidally. Another important type of visualization is that of the particles used in the simulations. Typically, there are a very large number of particles, and only particular types of particles are of interest. These include

particles that become trapped in the plasma or travel to particular regions in the plasma.

One critical aspect of visualization involves derived quantities. These include regions of relatively high energy (called blobs) that can develop in the plasma and move around as the plasma evolves. A more complex example is the generation of Poincare plots. Poincare plots are used for the analysis of the magnetic field in the plasma. These are created by advecting a large number of particles around the tokamak and marking the punctures that each particle makes with a plane normal to the toroidal direction. These puncture patterns provide valuable information about the evolution of the plasma.

The visualization of experimental data typically involves 1-D and 2-D visualizations that evolve over time. The 1-D visualizations are typically time-varying curve plots. One common source of 2-D data is cameras that operate at very high speeds and are growing in resolution. Feature detection and tracking are one key type of visualization performed for 2-D data.

As a concrete example for illustration, we discuss an experiment involving colliding magnetic flux ropes in a strongly magnetized background plasma [152]. The flux ropes were kink unstable and designed to collide periodically, at the kink frequency. When the ropes collided, magnetic field line reconnection occurred somewhere in the plasma. The process of reconnection results in the annihilation of a small portion of the magnetic field. The magnetic energy is converted into heat, flow, and waves. One outstanding question in this experiment, and in general, is where in the large volume of plasma does this occur?

The experiments were carried out in the LAPD at the University of California at Los Angeles (UCLA) [153]. This, coupled to computer-controlled probe drives, allows the collection of volumetric datasets. The problem required the acquisition of a large amount of data. Three-axis magnetic pickup coils measured the magnetic field from which the vector magnetic field and plasma currents are derived. Other quantities measured with different sensors were the plasma flow, electron temperature, plasma density, and plasma potential. They were measured at over 42 000 spatial locations and 7000 timesteps ($\delta t = 0.32$ μs. The measurement volume was 30 cm on a side in ($\delta x = \delta y = 0.5$ cm) on 15 planes transverse to the background magnetic field ($B_{0z} = 330$ G). The planes were 64 cm apart in z or the axial direction of the cylindrical geometry of the experiment. 1-D data in this experiment are not enough to get the true picture of what occurred. At any given location, the magnetic field oscillates at the kink frequency (5.2 kHz) and varies smoothly in the transverse direction. 2-D data are far more helpful but can sometimes be misleading. Fig. 1 is a vector map of the transverse (B_x–B_y) magnetic field on a plane $z = 512$ cm from the start of the ropes and at an instant of time when the flux ropes collide. The plasma current density is superposed as a color map. The small transverse field near the center is close to the point of collision, where a small pink dot is drawn. From Fig. 1, one could guess that the location at which reconnection occurs is somewhere near the pink dot, but it is not that simple. A 3-D picture, constructed from volumetric data, is given in Fig. 2. Important topological quantities that shed light on the reconnection location are the quasi-separatrix layer, magnetic twist (the rotation of a field line around its neighbors), and the winding number, which is a measure of entanglement of field lines [154]. Most are displayed in Fig. 2. The magnetic twist is given by

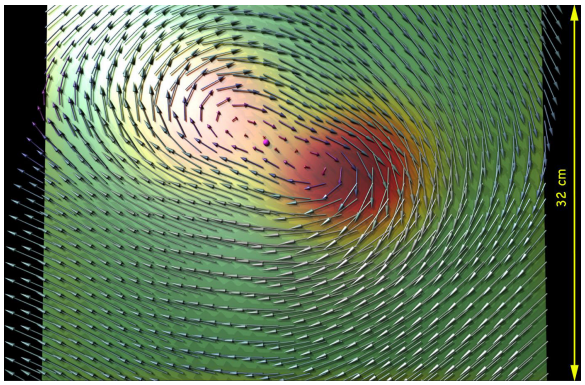

Fig. 1. Vector plot of the transverse magnetic field at $z = 512$ cm and $t = 5.673$ ms. The axial component, B_z, is suppressed. The background colors correspond to the current density on the same plane. The maximum value is 3.0 A/cm^2. The largest arrow corresponds to a magnetic field of 16 Gauss.

$$T(\vec{r}, t) = \int_{\gamma(\vec{r})} \frac{\vec{J} \cdot \vec{B}}{B^2} ds \tag{1}$$

where ds is a line element for integration along a fieldline γ. The twist calculated along the field lines is shown as white sparkles in Fig. 2. The winding number Θ (larger sparkles correspond to larger Θ) is a measure of the entanglement of field lines. First, one must calculate the winding angle Θ

$$\Theta(\vec{r}_0, \vec{r}, z, t) = a \tan\left(\frac{\gamma_y(\vec{r}, z, t) - \tilde{\gamma}_y(\vec{r}_0, z, t)}{\gamma_x(\vec{r}, z, t) - \tilde{\gamma}_x(\vec{r}_0, z, t)}\right). \tag{2}$$

Here, γ is the vector of x, y coordinates for a test field line anchored at $\vec{r}_0$ which passes through successive domains D (regions that the field lines pass through) transverse to the background magnetic field. There are 2800 $\vec{r}$ locations on a plane for all the other field lines γ, and z is the plane in question for which Θ is evaluated. Once Θ is calculated, another test field line is chosen, and the calculation is repeated for every γ in the plane. To measure the average entanglement of γ with the rest of the field, we integrate Θ over all field lines at positions $\vec{r}$. The winding number L is given by

$$L(\vec{r}_0, z, t) = \frac{1}{2\pi} \int_{D_0(t)} \left[\Theta(\vec{r}_0, \vec{r}, z, t) - \Theta(\vec{r}_0, \vec{r}, 0, t)\right] dA. \tag{3}$$

The winding number is shown as red sparkles in Fig. 2. For most field lines, the winding number begins to grow at about $z = 5$ m, and it is the largest near the axis of the machine. It was established that the reconnection occurred in the region where the twist became small and the winding number became large [154]. To confirm this, one must study what is displayed in Fig. 2 over many viewing angles. This is possible with existing software packages. The upshot of the analysis is that these topological quantities and one not mentioned, the

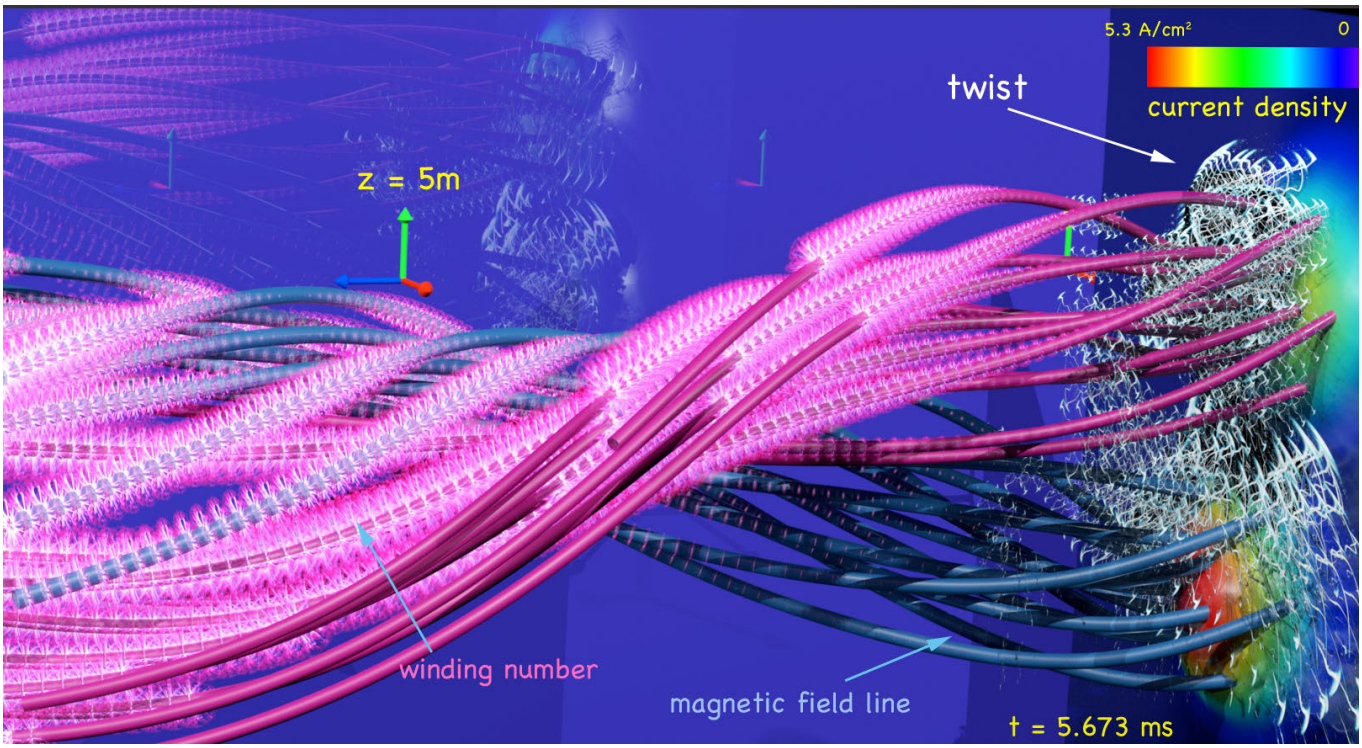

Fig. 2. Data from an LAPD experiment on magnetic field line reconnection that occurs when two magnetic flux ropes collide. Isosurfaces of the current in the current channels are shown on the right-hand side. The data plane is at $z = 0$. The maximum current density J (5.3 A/cm^2) is colored red. A colormap is provided at the top. The data plane at $z = 0$, on the right-hand side, is 30 cm on a side; however, the axial distance (z) spans 9 m. The current permeates the volume. The magnetic field **B** was measured at 48 000 locations. This was used to generate field lines shown as red and blue tubes. The magnetic twist is depicted as white sparkles and is the largest in the first 2 m. The winding number (L) along the field lines is depicted as red sparkles. The ropes start at $z = 0$, and the first transverse plane upon which data are acquired is at $z = 64$ cm.

quasi separatrix layer, were used to identify additional 3-D volumes in which reconnection occurred. When there was no reconnection, these quantities vanished. To belabor a point, these quantities could not be derived without fully 3-D, time-dependent data.

As useful as visualization has been, challenges remain to meet the future needs of fusion scientists. The emergence of computing ecosystems that couple experiments, simulations and surrogate models, and reliance on streaming data will require significant work for visualization to continue as a critical aid to gaining understanding from data. This increased complexity will require the use of automated workflows to compose and orchestrate the set of tasks that are required to do the science. The resources available to perform the analysis and visualization will dynamically change over the course of a scientific campaign. In addition, the resource requirements for visualization will vary as well. Visualization algorithms have different scaling characteristics depending on how they are run. Recent work has explored cost models for the placement of tasks using different placement strategies [155], [156], [157]. The rapid increase, size, and rate of data are also requiring the use of data reduction techniques. Analysis and visualization tasks must be able to adapt to the uncertainties introduced by data reduction. In order to provide trustable visualization from reduced data, the uncertainty must be conveyed to the scientists. The uncertainty will come from the raw data and the algorithm that is producing the visualization. If the uncertainties are too high, the algorithms must be smart enough to request additional data that will lower the errors and/or use different algorithms with higher accuracy. This of course can require additional resources in order to compute accurate results. As such, there must be an integration with automated workflows in order to ensure that enough data are used with the proper algorithm running on the right amount of resources. Solutions to these challenges will require visualization to integrate well into the controlling workflows. This includes abstractions for access to data, the ability to be composable, and schemas to describe the underlying streamed data. Platform portability will be required for placement across a wide range of computing devices. Performance models will be required so that visualization tasks can be placed on the proper resources with access to the proper accuracy of data. The development and use of smart dashboards, where scientists can see the current status of a simulation or experiment, will make it possible for teams to efficiently collaborate. These dashboards should be customizable by each scientist. AI can be used in these dashboards to learn the interaction patterns of scientists to ensure that the most relevant visualizations are displayed, features of interest are highlighted, and anomalies are highlighted.

Advances in the cloud and web-based visualization are making it easier to interact with published results [158] One day, moving images may be feasible in publications, or users may navigate through 3-D data in publishers' repository in real time; 3-D displays, both traditional and wearable, have become inexpensive. The development of large, high-quality

projectors has been driven by the movie industry. The use of 3-D in scientific meetings, however, is rare [159]. High-quality projectors (necessary for large audiences) are expensive to ship and rent. They come with a small team of operators and require special screens that reflect light without changing its polarization. As with television, shutter glasses are required for every member of the audience. The gaming community is making a push for virtual reality through the use of wearable devices. One day, these may find use in scientific visualization.

It is possible to embed holograms in scientific publications as was done for a cover of National Geographic [160]. They are expensive to produce, especially if the image quality is high, but we should not rule out their future use. There is speculation that images using organic LEDs could be embedded in the paper, which would enable publications to have moving color images. Finally, one may look to science fiction to imagine what future visualization systems might be. Characters in a book by Gibson [161] have chips implanted in their brain that can make telephone calls and project 3-D images in space before them. The chips are controlled by small movements of their tongues on the upper palate of their mouths.

[Walter Gekelman and David Pugmire]

I. ML Control Theory

Historically, control of plasma processes was based on statistical process control approaches, which are open loop in nature and are merely suitable for monitoring the process performance. Recent years have witnessed a growing interest in model-based feedback control approaches for confinement fusion reactors and low-temperature plasma processes to cope with intrinsic variabilities of plasmas and exogenous process disturbances. To this end, MPC [162], which relies on real-time optimization and is the prime methodology for constrained control, has emerged as a promising advanced control technology for plasma processes (e.g., [163] and [164]). This stems from the ability of MPC to handle the highly nonlinear and multivariable plasma dynamics and to explicitly account for constraints on process variables, which is crucial for safety-critical plasma applications. However, the conventional MPC paradigm follows the strict separation of a design phase, which mainly involves model development and controller tuning using offline data, and a closed-loop implementation phase, during which the controller remains largely intact. Such a controller design strategy can limit the MPC performance for plasma processes whose complex dynamics can span over multiple length and time scales. Recent advances in the field of ML, along with enhanced computational, sensing, and communication capabilities, have created ample opportunities for safe learning-based control of the hard-to-model behavior of plasma processes at exceedingly fast sampling rates.

Data-driven methods can aid in the design of MPC approaches for plasma processes in two primary ways.

1) Learning the System Dynamics: The performance of MPC is heavily dependent on using a suitable and sufficiently accurate model representation of the system dynamics. ML has shown great success in deriving data-driven, multivariable representations of complex system dynamics that are amenable to real-time optimization and control. Data-driven models can embed varying degrees of physics-based knowledge of a process. In the absence of theoretical plasma models, control-oriented models can be readily learned from data that are collected offline [165], [166]. Alternatively, when theoretical plasma models are available, surrogate modeling, in which dynamic models are trained based on high-fidelity simulation data [167], [168], has proven useful for deriving computationally efficient models suitable for control. Yet, an emerging approach to learning-based MPC is to combine a prior model (data-driven or physics-based), which represents our available system knowledge, with a learning-based model that is adapted in real time [169], [170], [171]. Such a learning-based modeling scheme is particularly useful for capturing the hard-to-model and time-varying nature of the plasma behavior when it cannot be captured a priori via offline data or high-fidelity simulation data. To this end, GPR has proven especially useful for not only learning the unmodeled system dynamics but also characterizing the uncertainty of model predictions, which can be incorporated into the MPC design to robustify control actions with respect to uncertainties. In particular, GPR models of system uncertainty, which are dependent on the system states and inputs and, thus, can be reevaluated online at every measurement sampling step, can be especially useful for MPC of plasma processes under variable conditions [171], [172].

2) Learning the MPC Law: Another important research direction in ML for MPC focuses on learning the control law, as opposed to a prediction model. MPC relies on online solutions of often a nonlinear optimization problem that can be computationally prohibitive for real-time control of fast sampling systems. This can especially be the case when sophisticated process models are used for MPC or when the control application involves fast measurement sampling frequencies on the order of KHz to MHz or even possibly faster. ML has proven useful for developing so-called *approximate* MPC approaches that learn a cheap-to-evaluate, an explicit expression for the MPC law using data generated from an offline solution of an MPC problem [173] (see Fig. 3). A variety of function approximators, ranging from polynomials to DNNs, have shown promise for approximating optimization-based control laws with surrogates that can be evaluated on fast sampling times. The resulting low-complexity controllers typically exhibit a limited memory footprint, which makes them particularly suitable for implementations via resource-limited (i.e., low power and memory) embedded control systems [174]. ML can also be used to learn other components of an MPC formulation, such as the control cost function, directly from data, as discussed in [169] and the references therein.

A largely open area of research in learning-based MPC is how to confer an *active learning* mechanism to a controller to simultaneously explore and exploit the system dynamics toward actively mitigating the model uncertainty. To this end, there has been significant interest in leveraging RL [175] and Bayesian optimization [176] methodologies to design

TABLE I
EVOLVING APPROACHES TO AND ROLES OF EXPERIMENTAL DATA IN PLASMA PHYSICS. M = MANUAL; CA = COMPUTER-AIDED AUTOMATION. BIG DATASETS, AIDED BY ALGORITHMS SUCH AS NNS, COULD BE USED FOR PREDICTIONS, OPTIMIZATION, AND CONTROL OF EXPERIMENTS. BIG DATA CAN ALSO BE USED FOR MODEL DEVELOPMENT AND POTENTIALLY FOR SELF-CONSISTENT NEW DATA INTERPRETATION, WHILE THE EQUIVALENT ROLES OF SMALL DATA ARE LIMITED OR IMPOSSIBLE FOR SIMILAR APPLICATIONS OR WORKFLOWS

Application/workflow	Small data	Big data
Acquisition	M/CA	CA
Processing/analysis	M/CA	CA
Interpretation	incomplete	self-sufficient
Prediction	No/limited	Yes
Optimization	No/limited	Yes
Control	Maybe	Yes

learning-based controllers. These methodologies will allow us to combine learning and feedback policy design into a unified framework that provides a "self-optimizing" feature via systematically balancing learning (i.e., exploration) and feedback control (i.e., exploitation) of an uncertain system [177]. Another crucial consideration in learning-based control is to ensure safe learning of the unknown and hard-to-model process behavior. In particular, it is imperative to guarantee the safe operation of safety-critical plasma processes despite uncertainties in models and variabilities in the process itself. In general, safety guarantees for learning-based controllers can be established by decoupling optimization of the control objective function and requirements of constraint satisfaction [169]. Nonetheless, safe learning-based control, particularly for controllers with an active learning mechanism, remains an open and active area of research.
[Ali Mesbah]

III. BASIC PLASMA PHYSICS AND LABORATORY EXPERIMENTS

A. Introduction

Recent advances in data science and data methods, together with the continued decline in the cost of computing hardware and data acquisition instruments, are not only reinforcing the traditional roles of experimental and observational data but also rapidly changing how the data are used in interpretation, prediction, and optimization problems, as illustrated in Table I. As the datasets get bigger, computer-aided automation is essential in the workflow of acquisition, data processing, and analysis. Big data, aided by data-driven algorithms, also offer new methods for interpretation, prediction, and control of plasma experiments, which is otherwise difficult if feasible when only small datasets are available. More comprehensive discussions on big data, data analytics, and infrastructure needs can be found in Sections II and VIII and the literature outside the field of plasma physics.

Interpretation, prediction, optimization, and control are too sophisticated for traditional computer programs, which can only repeat the preprogrammed tasks or workflows without glitches. ML methods, such as deep learning and kernel learning, bring several important features that are missing from traditional computer programs. Multilayer NNs, which mimic the way the network of neurons in the human brain processes information, for example, have the ability to learn without being explicitly programmed. The same NN architecture can also be reprogrammed or "retrained" for different datasets or multitasks. In other words, an NN developed for material science, biology, or even outside natural science can be adopted to solve plasma physics problems [178]. As another example, NNs provide a new tool for the fast solution of repetitive nonlinear curve fitting problems encountered in experimental data fitting [179]. NNs continue to grow in size and architectural variety from thousands to more than one billion of simple computational units or the "artificial neurons." In comparison, there are about 100 billion neurons in the human brain.

A growing number of libraries and predesigned NN architectures are now available through open sources, such as GitHub. ML process in NNs such as CNN and LSTM is equivalent to iterative tuning a large number of "weight" parameters associated with neurons or nodes. The initial network configurations and states, such as the connections between different neurons, are called hyperparameters. There are many hyperparameters for an NN, including the number of layers; the orders of neuron connections; the number of nodes in a layer (which intrinsically determines the connections); the dimensions of the kernels in convolutional layers; the selection of nonlinear truncation functions such as sigmoids, rectified linear units (ReLUs), and licking ReLUs; orders of pooling operations and their type, with or without loops; and so on. The initial values for the parameters of the connections, or the weights, are usually set at random within a certain range, by following heuristic rules that are well known as "Xavier initiation" [180]. CV for hyperparameter selection is also supplemented by heuristics in practice, depending on the different ML approaches, such as SVMs [181], Bayesian learning, or GPs [182].

Laboratory plasmas also provide rich experimental datasets and data varieties to test and develop ML methods or a plasma-trained "artificial brain" that can potentially benefit other scientific fields. Laboratory plasmas are extremely diverse, ranging from microplasmas produced by short-pulse lasers to table-top experiments and to the nearly 30-m-tall ITER experiment. In spite of the difference in plasma density, temperature and data-driven methods are generic for interpretation, prediction, and control problems. A generic approach, which is independent of the hardware details, to construct a data-enhanced instrument, is illustrated in Fig. 4. In the following, we shall highlight the applications of different data methods in laboratory experiments as illustrative examples.
[Zhehui Wang]

B. Spectroscopy, Imaging, and Tomography

Optical, UV, and X-ray spectroscopies are widely used for plasma density (n_e), electron temperature (T_e), ion temperature (T_i), neutral atom temperature (T_g), and impurity measurements. Passive spectroscopy using plasma self-emission is sometimes preferable over Langmuir probes for reasons

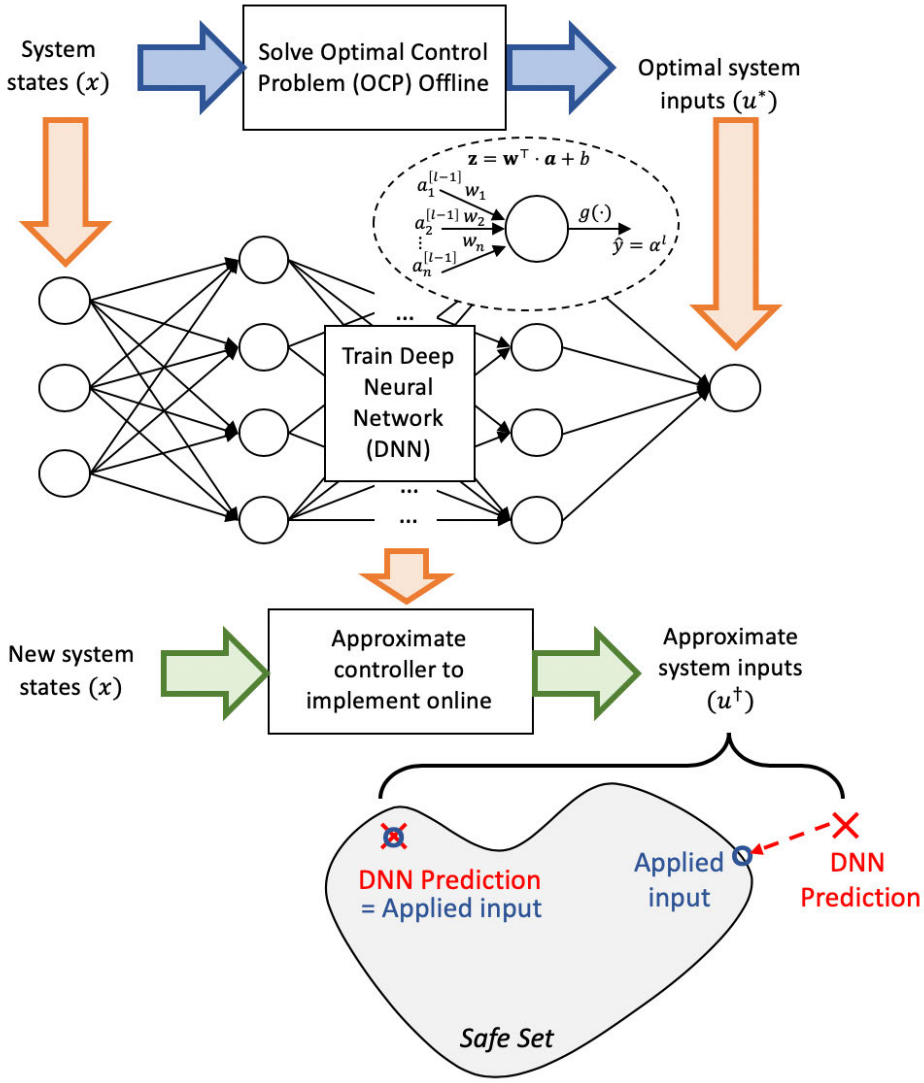

Fig. 3. DNNs can effectively approximate optimization-based control laws with a cheap-to-evaluate explicit control law that has low memory requirements. To guarantee satisfaction of safety-critical constraints of plasma processes in the presence of approximation errors and system uncertainties, the control inputs computed by the NN can be projected onto a safe input set that is constructed using the notion of robust invariant sets. Safe NN-based controllers can play a pivotal role in the control of fast-sampling plasma processes using resource-limited embedded control hardware [173].

such as no perturbation to plasmas and free of complex plasma–material surface interaction. However, data analysis to retrieve T_e information may be more complicated than a probe measurement, which measures the local T_e and n_e. For ion temperature T_i, Doppler broadening of ion spectral lines width may be used. For T_e measurement, one common approach is to use intensity ratios of multiple emission lines, which may be measured using several line-filtered PMTs [183], an array of photodetectors, or a spectrometer with an imaging camera. For local thermal equilibrium with a temperature T_e, if a pair of emission lines originate from the same ground state of an atom or an ion, and the excited states are mostly empty, the line ratios give the value that is proportional to $e^{-\Delta E/kT_e}$. Similar physics-motivated analytical and empirical formulas can be derived for other diagnostics, including Langmuir probes [184], which usually form the basis of diagnostic data interpretation. NNs can replace such formulas and represent a much more complicated correlation between measurements and physical quantities, such as n_e, T_e, T_i, and T_g, as illustrated for T_e in Fig. 4.

Multichord spectroscopy, and similarly multichord interferometers, reflectometers, and bolometers, can be used to obtain the 2-D profile distribution of the plasma emissions through inversion algorithms. NNs have been implemented to reconstruct electron temperature T_e profiles from multienergy soft-X-ray arrays and other plasma diagnostics with fast time resolution [185]. By training a three-layer fully connected feedforward NN to match fast (>10 kHz) X-ray data with T_e profiles from Thomson scattering, the multienergy soft-X-ray diagnostic can be used to produce T_e profiles with fast time resolution. The typical network input nodes for soft X-ray signals were up to 20. The number of output nodes for T_e was comparable to the number of inputs. The number of hidden layer nodes was about 40. A sigmoid activation function, in the form of the logistic function $f(x) = 1/(1+e^{-x})$, was used to sum up the inputs of each hidden node. Adding spectroscopic

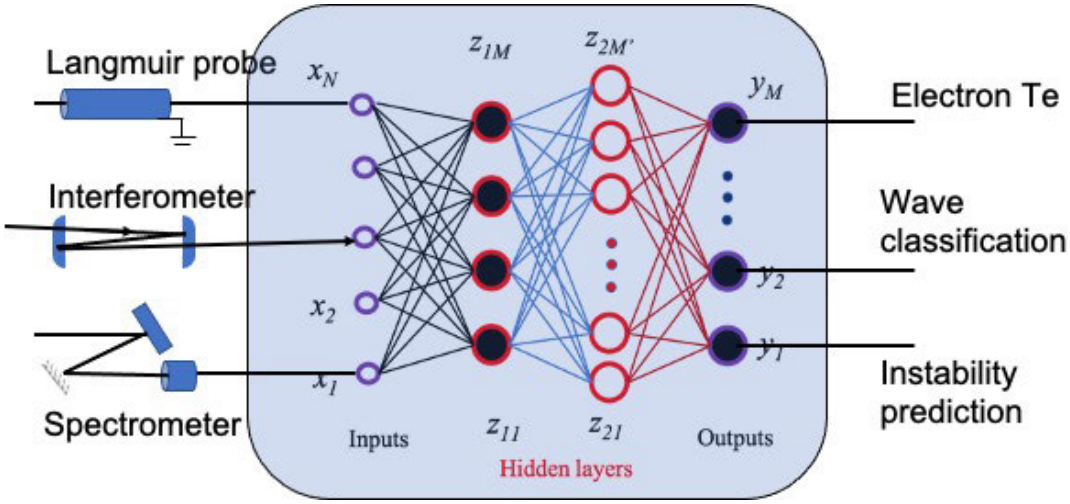

Fig. 4. Generic approach, which is independent of the hardware details, to construct the data-enhanced instrument. A Langmuir probe, a single-channel interferometer, and one LOS of spectroscopy are shown as examples. The output of such a synthetic instrument, such as electron temperature (T_e) measurement, or a binary prediction about the onset of instability, is significantly enhanced by the NN, including noise reduction and frequency retuning. A virtual instrument with multiple inputs from different diagnostics to an NN is also possible through multistream data fusion as illustrated.

data as inputs was found to decrease the root mean square (rms) error of the temperature predictions by as much as 50%. Deep learning for a multichord bolometer was reported [186]. A so-called upconvolutional network, a variant of CNN that takes 50 1-D inputs and generates 2-D (120 × 120) outputs, was used. After training on JET data, the network provides accurate reconstructions with an average pixel error as low as 2%.

A feedforward fully connected NN has been implemented to measure the electron temperature directly from the EUV/VUV emission spectra (photon wavelength in the range of 50–160 nm) of the divertor region of the DIII-D tokamak plasma [187]. The plasma temperature was below 100 eV in the region. The best-performing NN had 12 hidden layers of 12 neurons that were sandwiched between a 1000-element input vector (the spectra) and the single output node (T_e). Each neuron in the model used an exponential linear unit (ELU) activation function with the exception of the final output neuron, which did not have an activation function so that it could take on any value. An Adam optimizer was used to calculate the changes to the model weights. The Python model construction and training were handled with TensorFlow. The full dataset consisted of 1865 input (spectrum time slices) and output pairs, of which 25% are reserved for evaluation. The rest of 75% were further split to a 3:1 ratio for training and training assessments.

In addition to tunability and higher emission intensities, laser-based spectroscopy, such as Thomson scattering, and charged particle spectroscopy, such as charge exchange recombination spectroscopy (CHERS), can overcome the limitations of passive spectroscopy using plasma emission. For example, passive spectroscopy gives line-averaged information along the LOS. Laser and particle beam techniques can localize the plasma temperature and density in space and time. A three-layer (one hidden layer with eight nodes) NN was used to calculate the electron temperature in Thomson scattering diagnostics, replacing the traditional χ^2 method [188]. One advantage of the NN was to speed up the data processing time by almost 20 times over the χ^2 method. NN has also been used to speed up the analysis of CTS data [189]. As a result of scattering by fluctuations in the electron density, electric field, magnetic field, and current density, CTS has been used to diagnose ion temperature and fast ion velocity distribution [190]. Recovery of T_i and T_e from CTS usually requires time-consuming simulations to produce synthetic spectra from a set of input parameters, including T_e and T_i. A feedforward ANN with three hidden layers was implemented with scikit-learn [189]. The T_i mapping error was less than 5%.
[Zhehui Wang]

C. Sparse Measurement and Noise

The problem of image reconstruction and 3-D tomography from a limited number of measurements or sparse measurements commonly arises in plasma experiments and in computational, medical, and scientific imaging [191], [192], [193]. Even by using multichord configurations and detector arrays, measurement of plasma such as through spectroscopy is intrinsically sparse. The number of chords is limited by the real estate and viewing ports around a plasma device. The number of photons recorded is limited by the plasma emissivity or the laser power in terms of the scattering experiments and the signal integration time. The detectors have a finite spatial and temporal resolution. The electronics have a finite bandwidth and a finite sampling rate. Even though the state-of-the-art oscilloscopes have now an impressive bandwidth of tens of GHz, it may still be insufficient for example in ultrafast plasma experiments when sub-ns resolution may be desired for hundreds of or more channels. In short, Shannon's information theory, which requires that the sampling rate or the Nyquist rate must be at least twice the bandwidth of the signal, can be too restrictive for experiments. In other words, the Shannon–Nyquist sampling theory considers the worst case scenario and requires the number of samples to

be large enough to recover an arbitrary signal in the signal space [194], [195]. Similarly, the mathematical formulation of the inversion algorithms, such as the Radon inverse transform, also assumes a large number of projections, which may not be practical in experiments. In situations when the Nyquist rates are achievable, the volume of data generated may be too large and can result in transmission, storage, and processing challenges.

The ubiquitous presence of noise can further complicate 2-D profile reconstructions from multichord line-integrated measurements for both traditional and ML methods [196]. A related problem in tomography is to use 2-D projections to reconstruct 3-D volumes [197]. Noise is probably the hardest type of signal to reproduce based on physics and first principles because of its seemingly random nature. The noise of different origins is present along the full chain of signal generation, propagation, and registration (digitization). For spectroscopy measurements, the limited amount of light can appear as Poisson noise. Background noise is spectral-dependent and experiment-specific, i.e. the background light for optical spectroscopy is different from soft X-ray spectroscopy due to, for example, different geometries of the setup, FOV or the solid angle of the light collection, surface reflectivity, and the light path setup that may be susceptible to atmospheric turbulence. Electronic noise, which has different sources by itself, is also not avoidable. Yet, noise removal and reduction are needed for any measurements, especially when the measurements are photon starved. ML methods for noise classification, denoising, and even noise modeling are of growing interest [198], [199].

The compressed sensing or compressive sampling principle has emerged as a powerful framework for data acquisition, detector designs, and signal processing, including inversion problems [192], [193], [200]. The compressed sensing spectral imaging system was reported for plasma OES [201]. A single PMT detector and a variable encoding mask (a digital micromirror device) are designed and implemented for the measurement of molecular and ion vibrational temperature. In other examples, compressed sensing was used to decompose emission spectra from an extended plasma source, such as the Sun [202]. A combination of compressed sensing and ML led to dimensionality reduction, so the flow properties, such as the Reynolds number, pressure, and flow field, can be obtained from sparse pressure measurements [203]. More recently, a compressed sensing framework is implemented in variational autoencoder (AE) and GANs [204]. The method can use five to ten times fewer measurements than the least absolute shrinkage and selection operator (LASSO) for the same accuracy. A canonical imaging system can be represented as

$$y = Ax + \eta \tag{4}$$

where $y \in \mathbb{R}^m$ represents images from the measurements, $x \in \mathbb{R}^n$ represents the unknown scene to be reconstructed, A represents an imaging formation operator, and η denotes noise in the measurements. The problem of reconstructing x from y is underdetermined if $m < n$, and we need to use some prior knowledge about the scene signal structure. Classical signal priors exploit sparse and low-rank structures in images and videos for their reconstruction [205], [206], [207], [208], [209], [210], [211], [212], [213], [214], [215], [216], [217]. However, the natural images exhibit far richer nonlinear structures than sparsity alone.

A recent trend is to use data-driven methods, mainly based on deep learning and NNs, to perform image reconstruction. Deep learning-based methods can be broadly divided into the following categories:

1) end-to-end networks that are trained to map the sensor measurements onto the desired images [218];
2) learned NNs that are used as denoisers plug-and-play priors during the recovery process [219];
3) trained generative networks that are used as priors for natural images [220]; and
4) untrained networks that are learned while performing image reconstruction [221], [222].

In the following, we highlight techniques that use *pretrained* or *untrained* networks within an optimization algorithm in order to leverage the information from both the data-acquisition model and the learned prior. This is a rapidly evolving research area with a number of recent theoretical and practical developments [223], [224], [225], [226], [227], [228].

Generative Models as Image Priors: Deep NN-based generative models have emerged as useful image priors in recent years. In a nutshell, a deep generative model represents a function $G(\cdot)$ that maps a low-dimensional, latent vector z into a high-dimensional image as $x = G(z)$ [220], [229]. The weights of the generative network and the distribution of the latent vectors can either be learned using training images or the generative network can be learned while solving the image recovery problem.

Let us denote a generative model as

$$x = G_\gamma(z) \equiv g_{\gamma_L} \circ g_{\gamma_{L-1}} \circ \cdots \circ g_{\gamma_1}(z). \tag{5}$$

$G_\gamma(z)$ denotes the overall function for the deep network with L layers that maps a low-dimensional (latent) code $z \in \mathbb{R}^k$ into an image $x \in \mathbb{R}^n$ and $\gamma = \{\gamma_1, \ldots, \gamma_L\}$ represents all the trainable parameters of the deep network. $G_\gamma(\cdot)$ as given in (5) can be viewed as a cascade of L functions g_{γ_l} for $l = 1, \ldots, L$, each of which represents a mapping between the input and the output of the respective layer. Fig. 5(a) illustrates a generative network based on DCGAN architecture that is usually used as an image prior [229], [230]. Some other commonly used generator architectures include U-net [231] and deep decoder [222], as shown in Fig. 5(b). To recover an image using generative models as image priors, we can either use a trained or an untrained network. We briefly discuss both approaches in the following.

Trained Networks as Image Priors: A number of papers have recently explored the idea of replacing the classical (hand-picked) signal priors with deep generative priors to solve inverse problems [220], [221], [232], [233]. Recovery of an image using a trained generative model ($G_\gamma(\cdot)$) can be formulated as the recovery of the latent code (z). To learn a latent representation of an image with respect to a generator, we often need to solve a nonlinear problem [234], [235], [236],

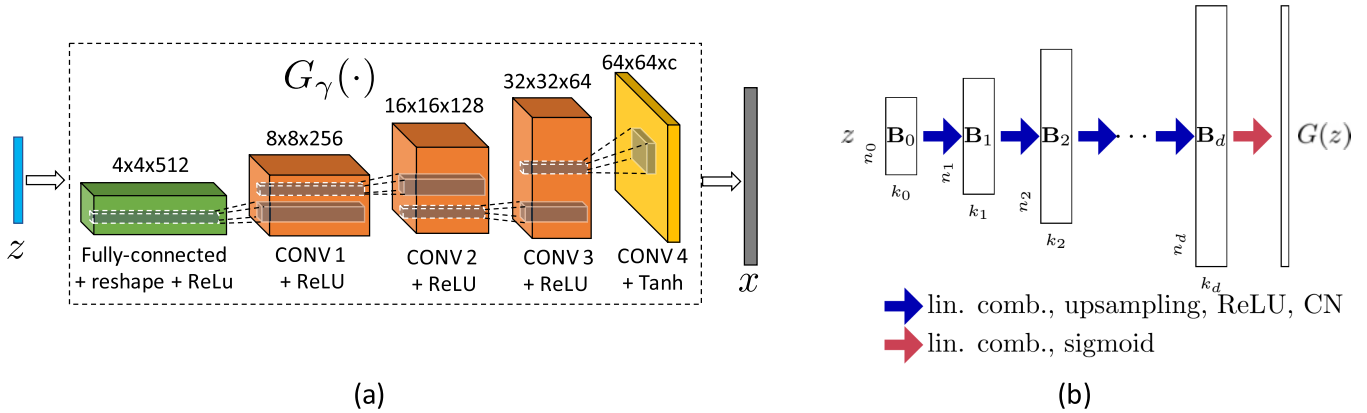

Fig. 5. Examples of NNs commonly used as image priors. (a) DCGAN [230] architecture that maps a low-dimensional latent vector z into an image as $x = G(z)$ (generative model). (b) Deep decoder [222] architecture that uses an untrained network as an image prior.

[237]. Given a pretrained generator G_γ, measurements y, and the measurement operator A, we can solve the following optimization problem to recover the low-dimensional latent code:

$$\underset{z}{\text{minimize}} \ \|y - AG_\gamma(z)\|_2^2. \tag{6}$$

The reconstructed image can be computed as $x = G_\gamma(z)$, where z denotes the solution of the problem in (6). We can solve (6) using a gradient descent-based method that iteratively updates z to minimize the objective function. The gradient of the objective function in (6) with respect to z can be computed using backpropagation. This approach is employed in [220], [234], and [238]. An alternative approach is to solve the following (nonlinear) projection-based method [239], [240]:

$$\underset{z,x}{\text{minimize}} \ \|y - Ax\|_2^2 \quad \text{s. t. } x = G_\gamma(z) \tag{7}$$

where we alternately update x via gradient descent and project the estimate onto the range of the generator $G_\gamma(\cdot)$.

Untrained Networks as Image Priors: Trained networks serve as good image priors, but they require a large number of training samples, which limits their use in settings with limited data. Furthermore, trained generators can only correctly recover images that are close to the training samples. In recent years, a number of methods have shown that untrained networks can also be used as image priors [221], [222]. The deep image prior method in [221] first showed that an overparameterized network can be trained to generate natural images by early stopping. This observation led to the use of untrained generative models as image priors for solving different inverse problems [221], [222], [229], [233], [241], [242]. A number of theoretical results have also appeared recently that provide conditions under which an untrained network can solve different inverse problems [241], [243], [244]. In practice, untrained networks perform almost as good as trained generative networks when the test data lie in the range of the trained generators. Untrained networks perform better than trained networks when the test data do not fall within the range of the trained networks.

Untrained generative prior is free from limitations as we use random weights to initialize the network and update the weights as we go along. However, it is natural to question the theoretical validity of such priors.

[M. Salman Asif and Zhehui Wang]

D. Synthetic Instruments and Data

Multiphysics simulation tools are now available to design and simulate plasma experiments, up to full-scale experiments in realistic geometries [245]. Such tools have been adopted for the modeling of plasma instruments and data interpretation. The multiphysics model for an instrument is sometimes called a synthetic diagnostic [246], [247], [248]. In parallel to hardware-based instruments for the diagnosis of a real plasma, a synthetic diagnostic can be regarded as a numerical instrument for the diagnosis of a numerical model of a plasma, as illustrated in Fig. 6. Due to the complexity of the plasmas and instrumental responses, synthetic diagnostics are indispensable for the quantitative interpretation of the experimental data from a physical instrument and for comparison of the experimental data with plasma simulations [247]. As mentioned above, spectroscopy, tomography, interferometry, and others, such as electron cyclotron emission imaging (ECEi) and millimeter-wave imaging reflectometry, are usually LOS or volume integrated and time integrated, while plasma simulations usually give physical quantities, such as temperature and density as a function of position and time. The synthetic data generated from a synthetic instrument can be flexibly converted into both experimental and simulation formats. Another function of synthetic diagnostics is to quantify uncertainties and sensitivities of the instrument to different plasma conditions and noise, with applications in improving instrument design. In the case of synchrotron emission from runaway electrons, geometric effects are shown to significantly influence the synchrotron spectrum. A simplified emission model that does not include detection physics can lead to incorrect interpretation of the measurements [249]. A third function of a synthetic diagnostic is for experimental control and plasma parameter optimization [250].

In addition to synthetic data generation and "data fusion" between experiments and simulations, synthetic imaging has been proposed to replace hardware or components, such as focusing optics in experiment [251]. Not only that synthetic

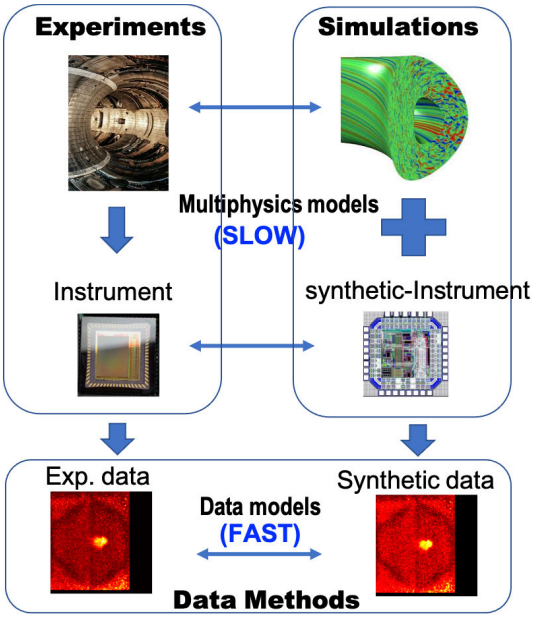

Fig. 6. Data methods including ML motivate the development of synthetic instruments and allow direct and faster interpretation of complex experimental data. The ability to extract information faster from measurements is particularly important for real-time control of plasma experiments.

imaging is simpler but also that, on many occasions, the optics may not be available or difficult to implement. In hard X-ray imaging, for example, the focusing optics are difficult to fabricate due to the small refractive index difference from the vacuum for essentially all materials and the sub-nm X-ray wavelength. For microwave imaging, the wavelengths are several centimeters, which makes the focusing optics very large. There are plenty of examples outside plasma physics. Computational X-ray imaging, including lensless X-ray imaging, has been reported [252]. Synthetic aperture microwave imaging has been used for imaging laboratory plasmas [253].

An emerging framework for synthetic data generation is GANs [254]. GANs demonstrated that deep learning could discover hierarchical probability distributions of data [255], which is common for experimental plasma physics and other branches of science. In this framework, generative models are trained in an adversarial process: a discriminative model that learns to determine whether a sample generated by a generative model is from the data distribution. The modules that correspond to the generative models and discriminative models are generators and discriminators, respectively. Adversarial nets [254] implemented both generator and discriminator as MLPs and demonstrated their applicability to generating images of datasets, such as MNIST [256] and CIFAR-10 [257]. GANs have the advantage that Markov chains are never needed, only backprop is used to obtain gradients, no inference is needed during learning, and a wide variety of functions can be incorporated into the model.

However, GANs have been known to be unstable to train, and generators often produce nonsensical outputs. DCGANs [230] addressed this issue by implementing both generator and discriminator as deep CNNs. The visualization of the convolutional filters learned by DCGANs empirically showed the connections between the filters and specific objects. This was convincing evidence that DCGANs could learn a hierarchy of representations from object parts. It follows that convolutional GANs are a promising approach to generating images with complex structures.

In addition to the issue of training stability, the unconditioned generative models of GANs can cause difficulties in controlling the modes of data being generated. This is because many interesting problems are more naturally thought of as probabilistic one-to-many mapping. For example, an image can have multiple tags. The conditional GANs (cGANs) [258] addressed this issue by using conditional probabilistic generative models. This approach allows GANs to be conditioned on class labels, some parts of data, or even data from different domains. Preliminary results of conditional adversarial nets on image tag generations demonstrated the potential of this approach on multimodal learning.

By following the conditional and convolutional approaches, various GANs were developed for cross-domain image synthesis. Those conditional and convolutional GANs tailored their generators, discriminators, and loss functions for specific applications. Image-to-image translation, for instance, is a problem that is involved in many image processing, graphics, and vision problems. One of the data-driven image-to-image translation approaches is to learn mappings between paired input and output images by using GANs. For example, in [259], a U-Net-based generator [231] was used to learn image-to-image mappings, and a Markovian discriminator called PatchGAN was proposed. This work demonstrated that the proposed GANs could synthesize photos from label maps, reconstruct objects from edge maps, and colorize images. Paired training data are, however, not easy to acquire in practice. CycleGANs [260] achieved image-to-image translation on unpaired data by using a cycle-consistency loss function. It has been proven that cycle-consistency is an upper bound of the conditional entropy. Qualitative results of CycleGANs were presented on several tasks where paired training data did not exist, for example, collection style transfer, object transfiguration, season transfer, and photo enhancement.

In parallel with the studies of conditional and convolutional GANs, unconditional and convolutional GANs were studied in applications that involved intradomain image synthesis. Image SR, for instance, is about how to recover the finer texture details when images are super-resolved at large upscaling factors. SRGANs [261] employed a deep ResNet [262] with skip-connections, SRResNet, as its generator. As the objective of SRGANs was to achieve photorealistic single-image SR, the authors proposed a perceptual loss function that consisted of an adversarial loss and a content loss. They also introduced an MOS, which evaluated the qualities of reconstructions by humans. They found out that the SRResNet without the adversarial component sets a new state of the art on public benchmark datasets when evaluated with the widely used PSNR measure, whereas the SRResNet with the adversarial component, i.e. SRGANs, was the best in terms of MOS. More recently, SinGAN [263] achieved the unconditional generation of synthetic images by using only one training image. This was achieved by adopting a multiscale approach: the pyramid representation. This work demonstrated that a pyramid of fully

convolutional GANs could learn the generative model of the complex structures of a single natural image.

While most existing studies on GANs concentrate on natural images, they have inspired studies of GANs on nonnatural data, such as medical images. In [264], an FCN [265] was used to learn mappings from MR images to CT. Experimental results showed that this method was accurate and robust for predicting CT images from MR images. Using GANs to accelerate CS-MRI reconstruction is another example. CS-MRI needs only a small fraction of data to generate full reconstruction. However, this method suffers from long running time due to the extra computational overhead for dictionary training and sparse coding. RefineGAN [266] was built upon ResNet and GANs, with a novel cycle-consistency loss function, so that it shifted the time-consuming process from the reconstruction phase to the training (preprocessing) phase. RefineGAN achieved state-of-the-art CS-MRI reconstructions in terms of running time and image quality.

Inspired by the previous research and applications of GANs, multiphysics simulation is potentially another area that can use GANs for acceleration. Fig. 7 shows an example of generating synthetic experimental images from a single experimental image by using SinGAN. However, this is just an initial attempt to show the potential of using GANs to accelerate multiphysics simulations. The generation of experimental images is different from the generation of nonphysics images in terms of their underlying physical laws. For this reason, physics-informed methods are necessary for the generation of synthetic data that are sensical to actual physical processes. It has been shown that a physics-informed GAN [267] can approximate the generation of stochastic processes so that it can solve stochastic problems.

Some of the main bottlenecks in developing and deployment of a synthetic instrument are good physics models for different components of an instrument, the slow process in carrying out multiphysics simulations, especially for high-fidelity models [268]. Compared with classical computational methods, such as finite difference and finite elements, the ML method can significantly accelerate the simulation for instrumentation applications. A recent work that combines a CNN and a traditional direct computational method has shown 40- to 80-fold computational speedups [53]. Data methods offer a new way to combine simulations and experimental data [269]. [Xinhua Zhang, John Kline, and Zhehui Wang]

E. High-Rep-Rate Laser Experiments

The use of high-intensity laser pulses as drivers for the next generation of accelerators has received considerable attention over the past decade, and demonstrations of multi-GeV electron acceleration [270], [271], 100-MeV ions [272], and energetic positron beams [273] have been performed. Beam quality and control are the approaches that are needed for applications such as X-ray and neutron production, as well as for IFE drivers. However, the main disadvantage of laser sources is the relatively low rep rate and stability of the drivers. For example, applications for an LWFA or a laser-driven neutron source would be dramatically enhanced if the laser driver rep rate could be increased to 10 Hz or more. For IFE, such rep rates are also necessary.

In an LWFA, the laser pulse drives the relativistic plasma wave via the ponderomotive force, which depends on laser intensity, pulse shape, and spectral content. In general, all of these parameters are constantly evolving throughout the acceleration process. Although it is possible to obtain simple expressions for the dependence of electron beams produced by an LWFA with regard to plasma density and laser intensity for an unchanging laser pulse, in reality, the evolution of laser parameters makes analytical treatment less tractable. Furthermore, there are a large number of input parameters that must be tuned to optimize the accelerator performance. The usual approach to optimization and "ML" is to perform a series of single variable scans in the neighborhood of the expected optimal settings. These scans are challenging, as the input parameters are often coupled and the highly sensitive response of the system can lead to large shot-to-shot variations in output. Moreover, due to the nonlinear evolution of the LWFA, altering one input can affect the optimal values of all the other input parameters. Hence, sequential 1-D optimizations do not reach the true optimum unless initiated there. A full N-D scan would be prohibitively time-consuming for $N > 2$, and so more intelligent search procedures are required [274]. At the University of Michigan, we have implemented such optimization using genetic algorithms acting on the actuators of a deformable mirror that controls the laser focal spot characteristics.

ML techniques are ideal for these kinds of problems. Consequently, it is possible to use genetic algorithms, Bayesian optimization, and other methods; using the spatial phase of the laser to optimize a keV electron source (see Figs. 8 and 9), and subsequently using both spectral and spatial phase to optimize multi-MeV sources [275]. In these cases, only some of the laser parameters were controlled preventing full optimization of the LWFA, which relies on the complex interplay between the laser and the plasma. Furthermore, these optimizations often do not incorporate experimental errors and fluctuations, and can be therefore prone to distortion by statistical outliers. For the extension of these techniques to ICF experiments at a high rep rate, fully automated laser pulse optimization at high power and energy is needed in addition to control of laser pointing, which adds a fluctuating component to the laser pulse. In performing such optimizations, the algorithms will need to build a surrogate model of the parameter space, including the uncertainty arising from the sparsity of the data, fluctuations, and measurement variances.

Consequently, it is clear that the work to develop feedback control of high-power high-rep-rate laser pulses with respect to focal spot shape, temporal pulse shape, spectral control, and laser pointing will be required simultaneously. In work up to now, the performance of LWFA has been dramatically improved—using deformable mirrors and control of the laser pulse shape through the applied phase (Dazzler) [275]. Extensions of this work to the higher laser energies needed for fusion will enable the use of feedback techniques of the pulses needed for reproducible direct drive implosions at a high rep rate. Use

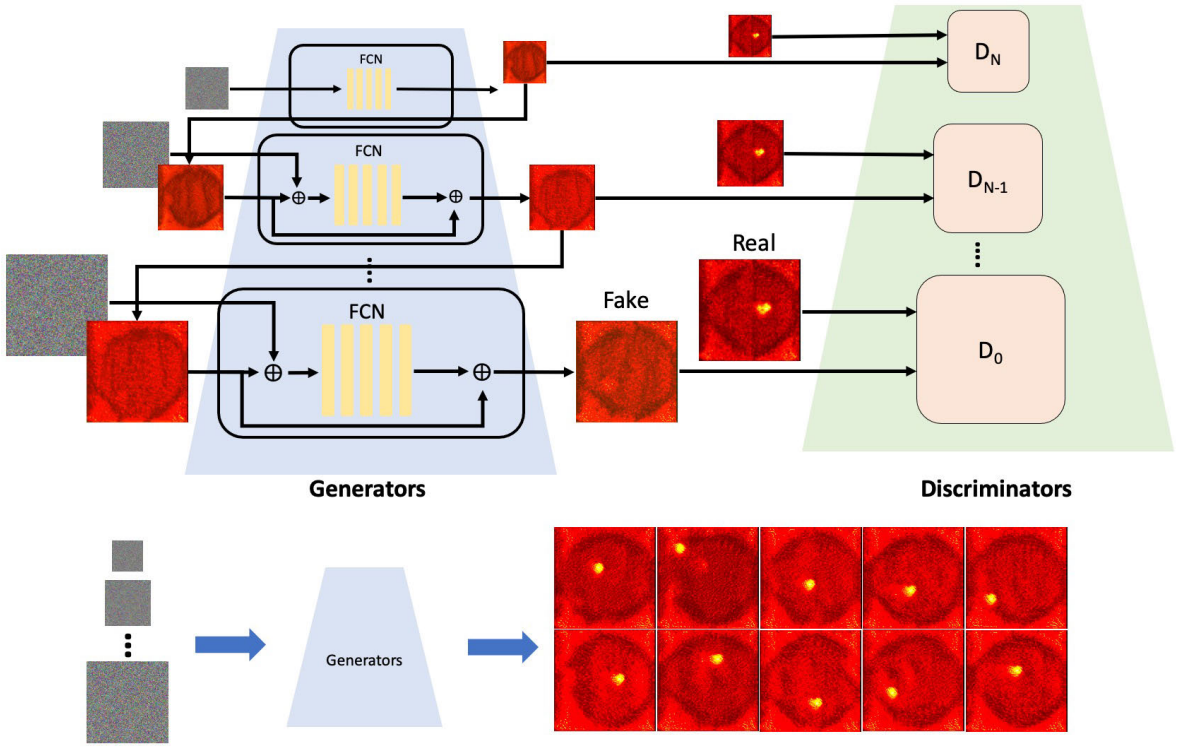

Fig. 7. SinGan consists of two pyramids of generators and discriminators at different scales. Each generator is an FCN. D_i, where $i = 0, 1, \ldots, N$, is the discriminator. Given images with random values, a trained generator of SinGAN can produce a set of synthetic images. The image used in this example is an ICF experimental image [197].

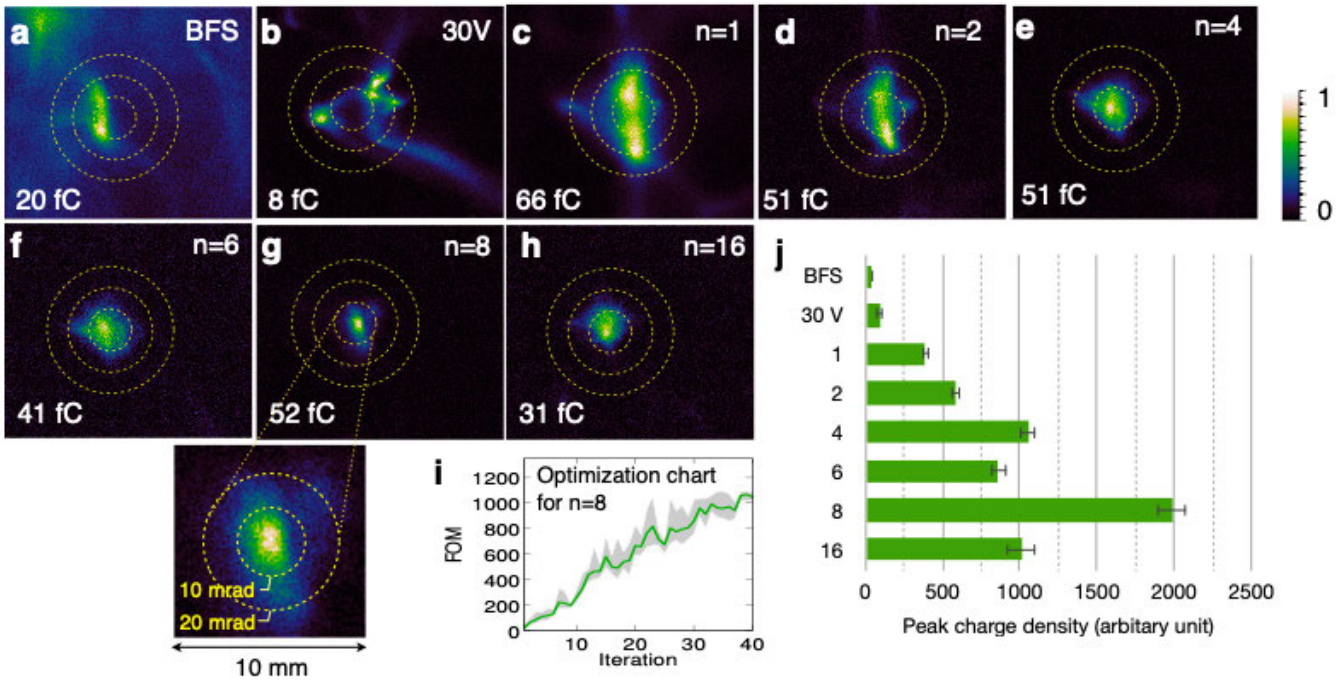

Fig. 8. Optimization of the electron spatial profile from an LWFA. Electron beam profile image integrated over 50 shots (100-ms exposure time) with a deformable mirror configuration. (a) Corrected for the best focal spot (BFS) and (b) 30 V on all actuators. (c)–(h) Single-shot electron beam profiles after genetic algorithm optimization using different weighting parameters, n. (i) Convergence of the genetic algorithm with $n = 8$. The shaded gray area represents the range of the ten best children in each iteration, and the solid green curve is the average. (j) Comparison of the peak charge density in a single-shot electron image. Contours shown are for 20, 40, and 60 mrad, centered on the beam centroid.

of adaptive optics in combination with temporal pulse shape control and pointing stabilization can potentially optimize all aspects of the laser focal spot for enhancing absorption and reducing instabilities during laser plasma interactions. While optimization and ML at 10 Hz work more "slowly" than that at kHz rep rates, demonstration of the viability of the

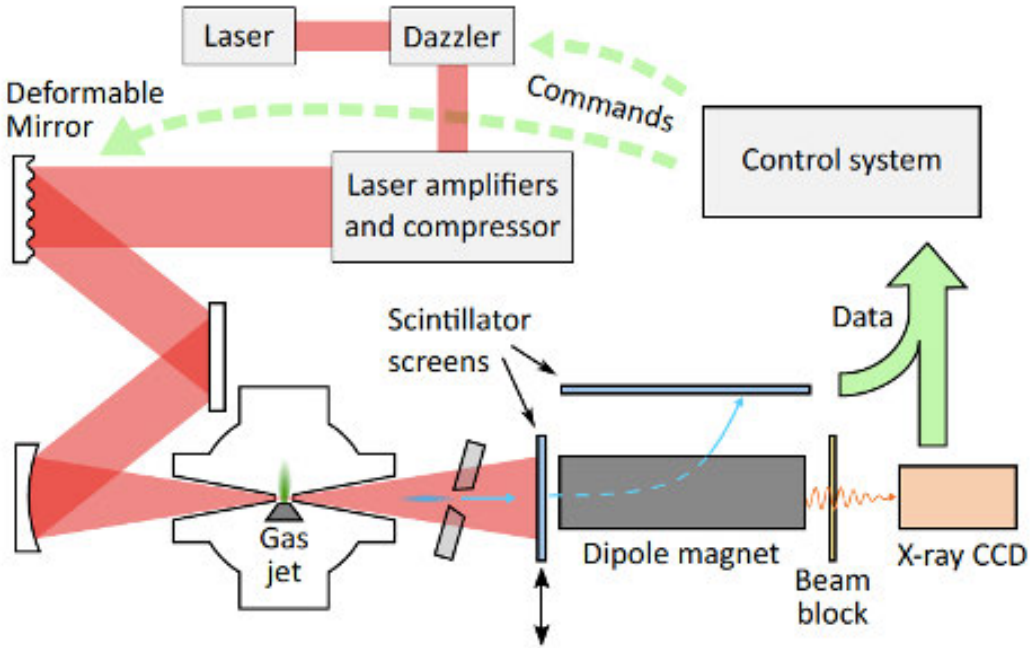

Fig. 9. Schematic of the experimental setup for ML. Pictured are the laser system, Dazzler, deformable mirror, inner chamber, gas jet, and diagnostics.

technology at these higher energies will be possible in the near term.
[Karl Krushelnick]

F. Charged Particle Beams

Beam-driven PWFA can achieve the same energy gain in a single meter, for which conventional accelerators require several kilometers but has not yet achieved the same beam quality (in terms of metrics such as energy spread and transverse emittance) as conventional accelerators. PWFA requires extremely intense, high current, and sometimes extremely short charged particle bunches with complex beam dynamics and phase space manipulations [276], [277]. The bunches required for the PWFA process must be extremely short (~3 fs) to achieve the extremely high peak currents (20–200 kA) with bunches having a few nC of current, making them very challenging to control. The PWFA process is extremely sensitive to the detailed longitudinal current profiles of these bunches and requires precise control over these profiles. However, the dynamics of extremely short and intense charged particle beams are difficult to control and quickly/accurately model due to collective effects, such as space charge forces and wakefields. Furthermore, diagnostics are extremely limited for such high current, high energy, and short electron bunches.

For example, the FACET-II at the SLAC National Accelerator Laboratory is being designed to provide custom-tailored current profiles for various experiments with bunch lengths as low as (1 μm or ~3 fs) [278], [279]. Another example is the AWAKE that uses transversely focused (~200 μm), high intensity (2.5–3.1 $\times$ 10^{11}), and high energy (400 GeV) protons from CERN's Super Proton Synchrotron (SPS) accelerator to create a 10-m-long plasma and wakefields into which ~18.8 MeV electron bunches with charge ~656 pC are then injected for acceleration up to energies of 2 GeV [280].

PWFAs are driven by kilometer-long accelerators that are composed of thousands of interacting EM components including RF accelerating cavities and magnets. The performance of all of these components is susceptible to drift, e.g., such as thermal drifts. There is also uncertainty in and time variation of the electron distribution coming off of the photo cathode and entering the accelerator. Traditional model-based control and diagnostics approaches are severely limited by such uncertainties and time variation of both the accelerated beam's phase space distribution and the accelerator's components, as well as misalignments, thermal cycling, and collective effects, such as space charge forces, wakefields, and coherent synchrotron radiation emitted by extremely short high-current bunches. Adaptive feedback and ML methods have the potential to aid in developing more advanced controls and diagnostics for complex accelerator facilities.

Static Systems: For simulation studies or for small accelerators whose properties do not change significantly over time, surrogate models are very useful examples of ML applications in the accelerator community. NN-based surrogate models can be trained to quickly map between accelerator parameters and beam properties, providing faster estimates than possible with computationally expensive physics models. Surrogate models can also be used to generate datasets for ML training and optimization studies [281], [282], [283], [284], [285], [286], [287].

An effort has also been made toward developing ML-based accelerator controllers using Bayesian and GP approaches for accelerator tuning [275], [288], [289], [290], [291], [292], including various applications at the Large Hadron Collider for optics corrections and detecting faulty beam position monitors [293], [294], [295], [296], and PC expansion-based surrogate models for UQ [297]. RL tools have also been developed for online accelerator optimization [298], [299], [300], [301].

Time-Varying Systems: An open problem in ML is the development of tools for quickly time-varying systems and systems with distribution shifts. If a system quickly changes with time, it is no longer accurately represented by the data that were used to trail the ML model. Therefore, the accuracy of the ML methods for accelerators will quickly degrade for systems that change with time, for which previously collected training data are no longer accurate.

Transfer Learning for Slowly Changing Systems: For systems that change very slowly with time and for which gathering large amounts of new data is feasible without interrupting operations, it is possible to utilize transfer learning techniques in which a network is modified to be accurate for a new dataset by taking advantage of some learned feature extraction capabilities and fine-tuning others for the particular problem of interest [302].

The most common transfer learning technique is retraining. For a particle accelerator, a retraining approach may start by using large amounts of simulation-based data to train ML models and then "freeze" most of the weights in the layers that have learned the high-level features of the physical systems for which they were trained and then fine-tune only a few layers, such as input layers that must handle real data rather than simulation-based data as inputs, by using much smaller experimental datasets. Another approach to transfer learning is domain transform in which a much smaller NN, such as a U-Net approach, is developed using a small amount of experimental data and is used as the input layer of our trained NN; the U-Net encodes and decodes data to translate between experimental and simulation domains [303]. These transfer learning techniques are not limited to NNs. For example, they can be applied to GP-based algorithms in which the prior and parameter correlations are first estimated via simulation studies and then fine-tuned with experimental data.

Such transfer learning techniques have been demonstrated to be very successful on a wide range of systems with recent applications including cross-modal implementations [304], and both retraining and domain transform were recently demonstrated for mapping electron backscatter diffraction patterns to crystal orientations in which simulation-based data were first used, and then, many orders of magnitude fewer experimental datasets were successfully used for transfer learning to make the networks accurate for experimental data [305].

Adaptive ML for Time-Varying Systems: For most accelerator applications, repetitive retraining is not feasible because detailed beam measurements are time-consuming and invasive procedures that interrupt regular operations. Furthermore, for quickly changing systems continuous retraining may be required forever chasing the changes. For such quickly time-varying systems, adaptive feedback techniques exist, which are model-independent and can automatically compensate for unmodeled disturbances and system changes. Recently, novel adaptive feedback algorithms have been developed, which are able to tune large groups of parameters simultaneously based only on noisy scalar measurements with analytic proofs of convergence and analytically known guarantees on parameter update rates, which makes them especially well-suited for particle accelerator problems [306].

Adaptive methods can be applied online in real time for drifting accelerator systems. For example, these methods have now been applied to automatically and quickly maximize the output power of FEL light at both the LCLS and the European X-ray free-electron laser (XFEL), and are able to compensate for unmodeled time variation in real time while optimizing 105 parameters simultaneously [308]. Adaptive methods have also been demonstrated for real-time online multiobjective optimization of the electron beam line at AWAKE at CERN for simultaneous emittance growth minimization and trajectory control [309]. These methods have also been demonstrated at FACET to provide noninvasive LPS diagnostics to predict and actively track time-varying TCAV measurements as both accelerator components and initial beam distributions drift with time [310]. Adaptive methods can also be applied for online RL in which optimal feedback control policies are learned directly from data to learn optimal feedback control policies that are parametrized by a set of basis functions whose coefficients are adaptively tuned online [311].

Adaptive methods are usually local feedback-based and can become stuck in local minima. An active area of research is the combination of ML and adaptive feedback in an AML approach, which combines the robustness of model-independent algorithms with the global learning power of ML tools, such as NNs. For example, at the LCLS FEL at the SLAC National Accelerator Laboratory, an NN was combined with adaptive feedback for fast automatic LPS tuning, quickly guiding the system to a neighborhood of the global optimum, and allowing the system to adaptively zoom in on and track the time-varying optimal conditions for fast automatic LPS control of the electron beam [312]. This general AML method has also been utilized for 3-D coherent diffraction imaging for accurate reconstructions of 3-D electron densities by combining adaptive feedback with 3-D CNN [313].

Novel AML methods are being developed, which utilize adaptive feedback to tune the low-dimensional latent space of encoder–decoder type CNNs based on real-time measurements and for online adjustment of inverse models that can provide a realistic estimate of the accelerator's input beam's phase space distribution based only on downstream diagnostics [314], [315]. Such AML tools have the potential to enable truly autonomous accelerator controls and diagnostics so that they can continuously respond to unmodeled changes and disturbances in real time and, thereby, keep the accelerator performance (beam energy and energy spread, beam loss, phase space quality, and so on) at a global optimal, not allowing it to drift as things change with time.

In a recent example of adaptive latent space, tuning a noninvasive diagnostic for the FACET-II beam line was studied in which a CNN was trained to map inputs of 2-D (x, y) electron beam images and vectors of seven accelerator parameters to 75 phase space distributions that were all 15 unique 2-D projections of the charged particle beam's 6-D phase space at five different accelerator locations. The input images were 128×128 pixels, and so, combined with the input vector, the total input has a dimensionality of 16391. These high-dimensional inputs were reduced down to a 2-D latent

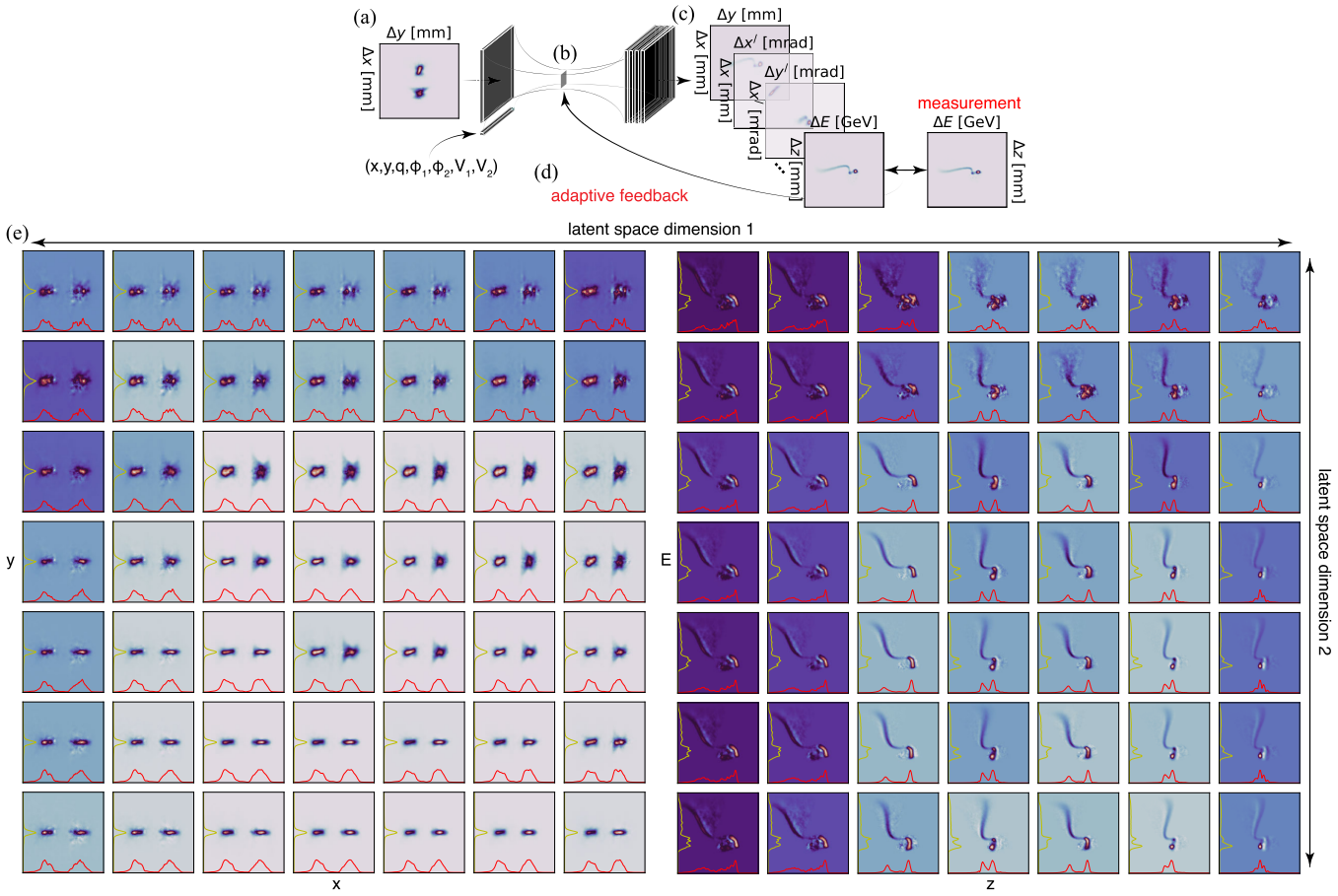

Fig. 10. (a) Encoder–decoder CNN setup is shown, which takes an image of an electron beam's (x, y) phase space distribution as an input together with a vector of accelerator parameters. The high-dimensional inputs are squeezed down to (b) 2-D latent space, from which 75 2-D distributions are then generated, which are all 15 2-D projections of the beam's 6-D phase space at (c) five different particle accelerator locations. Some of the projections, such as the (z, E) LPS distributions, can be compared to TCAV-based measurements to guide adaptive feedback, which takes place in the low-dimensional latent space to compensate for unknown changes (d) in both the accelerator parameters and in the initial beam distribution. (e) Variation of the (x', y') and (z, E) 2-D phase space projections is shown as one moves through the 2-D latent space learned by the network and adaptively tuned [307].

space from which the output beam distributions were then generated. By forcing the CNN to generate a large number of phase space projections simultaneously, the network was forced to learn correlations between various phase space coordinates. In order to utilize the encoder–decoder as a noninvasive diagnostic, it was then demonstrated that, by just comparing the predicted (z, E) projections to their TCAV-based measurements, and adaptively tuning the latent space in order to make them match, all of the other 2-D projections of the beam's 6-D phase space could be predicted and tracked even as both the input beam and accelerator parameters changed with time [307]. The setup for the adaptive encoder–decoder latent space tuning approach is shown in Fig. 10.

[Alexander Scheinker]

G. Control and Optimization of Plasma Accelerator Experiments

Plasma accelerators exploit the strong EM fields supported by plasmas to generate relativistic electron and ion beams. In a plasma-based electron accelerator, an ultrashort driver, either an intense laser pulse [316] or high-current particle beam [317], excites a trailing wakefield as it propagates through an underdense plasma (Fig. 11). Relativistic ion beams can be produced in laser–plasma interactions through the use of near-critical or overdense plasma sources [318]. The accelerating fields in these devices can reach hundreds of GV/m—more than three orders of magnitude higher than available in conventional RF accelerators—allowing for the production of multi-GeV electron beams over centimeter-scale lengths or multi-MeV ion beams in lengths on the order of tens of micrometers.

Plasma-based electron accelerators offer a route to drastically reduce the size and cost of brilliant light sources. In this domain, they have demonstrated the production of synchrotron-like X-ray beams [319] and FEL gain [320]. Furthermore, the technology offers a promising compact alternative to future high-energy colliders based on conventional technology [321]. Compact ion accelerators might find application in medical treatment, material science, or ICF technology [321].

However, while the future of plasma-based accelerators is extremely promising, they are not yet devices in a state of technological readiness where they could be used in place of today's RF accelerators. Some of the critical challenges

in making this transition are improving the control and optimization of the acceleration process, and reliably and robustly automating the accelerator operation.

As with any nonlinear system, small changes to the input parameters can constitute a significant shift in the behavior of the interaction. Plasma accelerators are no exception. In these devices, the relativistic interaction of the intense laser or particle beam with the plasma represents a strongly coupled system that dynamically evolves throughout the acceleration process. Add to this the shot-to-shot fluctuations in driver and plasma source parameters, as well as uncertainty and noise in the experimental diagnostics, and the task of manually controlling and optimizing the multidimensional parameter space of these machines becomes onerous.

One route to improving the performance of plasma accelerators while simultaneously adding automation and advanced diagnostic capability is through the application of ML and data science. Here, key experimental controls and diagnostics of the plasma accelerator are given to an ML algorithm to exploit their unique capabilities in multidimensional optimization, pattern recognition, and predictive analytics.

State-of-the-Art: Plasma-based accelerators have recently started to adopt the use of several different supervised ML techniques for the control and optimization of the electron, ion, and X-ray beams that they produce.

Several key experiments confirmed the fundamental feasibility of applying ML techniques for the real-time optimization of plasma-based acceleration of electrons [274], [323], [324], [325] and ions [326], [327]. These experiments utilized genetic algorithms to control specific aspects of the experiment, such as the spatial or spectral phase of the driving laser, and, in some cases, demonstrated optima with order-of-magnitude improvements over manual system optimization or found significant improvements with unexpected driver properties.

A key drawback of the genetic algorithm approach was the inability to incorporate experimental uncertainty and shot-to-shot variations in experimental parameters. Recently, Bayesian optimization based on GPR has been explored for the control of plasma accelerators due to its ability to incorporate uncertainty into the optimization process. This, coupled with the simultaneous tuning of multiple facets of the experimental arrangement, has enabled significant control over the form and parameters of the electron beam phase space [275], [328]. It has additionally allowed for the optimization of specific parameter regimes, such as stable operation, which is of paramount importance for the long-term development of plasma accelerators [328].

In addition to the optimization of the specific experimental outputs, the data generated through the long-term operation of these devices can be combined with ML and data science techniques to provide insight into the underlying phenomena.

For example, surrogate models can provide a cheap-to-evaluate, continuous, and noise-free abstraction of the complex plasma interaction allowing for an investigation into the underlying parameter dependencies and how they influence the achieved optima. It has been demonstrated that the GP models generated during Bayesian optimization can naturally serve such a purpose [275].

ANNs are also gaining traction as tools for exploring complex experimental datasets. For example, they have found use in explaining and quantifying the influence of drive laser fluctuations on electron beam quality [329]. Such knowledge is vital to improving the shot-to-shot stability of these machines.

In a similar fashion, several different supervised learning techniques have been applied in a predictive capacity to compare their performance in determining the charge generated in a laser–plasma accelerator as a function of changes to the laser wavefront [330].

In the context of plasma-based ion acceleration, it has further been shown that surrogate models can replace costly simulations, based on training NNs with comparably sparse sets of particle-in-cell simulations [331], [332].

Current and Future Challenges: Over the last two decades, important proof of principle experiments has shown that plasma-based acceleration is a technology that, in principle, can provide competitive beam parameters for accelerator applications, such as brilliant light sources. However, due to the limitations of the driver technology and the experimental nature of the setups, the findings of these experiments were often based on a small amount of data or even just single events.

Today, building on the results of these early experiments, the field is making significant progress in improving the reliability of the acceleration process to allow for stable long-term operation [333], [334]. In addition, promising progress has been made in using low-energy high-repetition-rate drivers [335], [336], and high-power high-repetition-rate laser drivers are foreseeable in the near future.

This progress in both stability and data availability has been a key enabler for the recent advances in the ML and data-driven methods listed above. Consequently, with the current trajectory of the field, ML and data-driven research demonstrates great potential but also faces key challenges. These include the aggregation of data at a high repetition rate, comprehensive diagnostics of the relevant parameters, and, finally, the development of algorithms that can handle the large data throughput.

Therefore, with the transition toward production machines, plasma accelerators will naturally adopt more and more concepts that are currently being established in the field of conventional accelerators [312], [337]. This is expected to be especially prevalent in the case of beam-driven plasma accelerators that, by their nature, operate in very close synergy with conventional machines.

Among these concepts are complex virtual diagnostics [284], [285], [338], [339] that allow noninvasive measurements of beam properties that would otherwise require destructive diagnostics, such as fluorescent screens. For this, ML models, typically NNs, are trained to predict the outcome of an invasive diagnostic from machine parameters that can be measured noninvasively.

For Bayesian optimization, it has been shown that domain knowledge can be used in physics-informed GPs [289], [340] to increase the speed and robustness of the optimizer. Furthermore, methods for efficient multiobjective optimization have

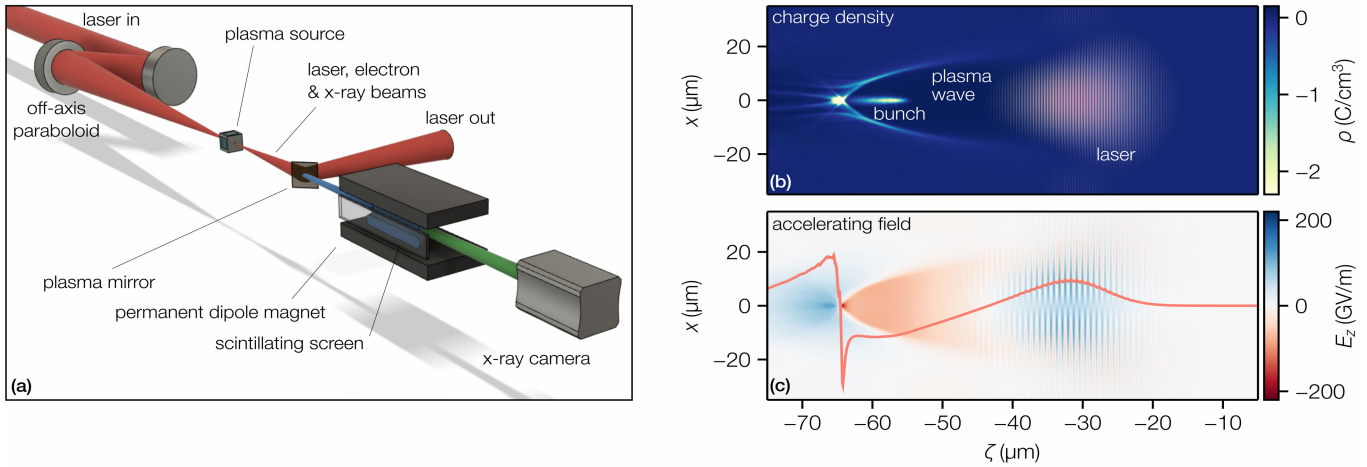

Fig. 11. (a) Conceptual layout of a laser-driven plasma accelerator adapted from [275]. (b) Snapshot of a particle-in-cell simulation, performed using FBPIC [322], showing the plasma wave driven in the wake of an intense laser pulse (traveling from left to right) along with an injected electron bunch. (c) Accelerating electric field generated by the separation of charge shown in (b).

been explored to find optimal machine states given competing optimization goals [282], [341].

Moreover, RL agents [300] that are either trained on the experiments themselves or on surrogate models resembling these promise to be a useful tool when confronted with dynamic conditions that tend to be a challenge for other optimization methods.

Concluding Remarks: Plasma accelerator technology is currently in a transition period, moving from single experiments to study fundamental concepts toward robust machines fit for applications in future light sources, high-energy colliders, and beyond. The increase in the quality and quantity of data has brought with it a commensurate uptake in ML and data science techniques for experimental control, optimization, and data analysis. It is foreseen that, in the future, the use of these techniques will rapidly expand.

Plasma accelerators offer a unique and timely testing ground to translate lessons learned in the control and optimization of high-repetition-rate *big physics* machines, such as conventional particle accelerators, to the laser–plasma community at large. As such, there is a significant advantage to be gained through close collaboration between members of all facets of laboratory plasma physics research.

[Sören Jalas, Manuel Kirchen, and Rob J. Shalloo]

H. Dusty and Complex Plasmas

Complex plasmas or dusty plasmas consist of nanometer to micrometer-sized dust particles immersed in a partially ionized plasma environment [342]. All plasmas, whether they are in a laboratory or natural environment, such as Earth's ionosphere, interplanetary solar wind, the interstellar medium in the Milky way, or intergalactic medium farther away, are dusty to a degree due to the ubiquitous interactions and mixing of plasmas with condensed matter [343], [344], [345]. Supernovas or massive star explosions are a source of dust, or "dust factories," which contribute to the cosmic dust population and have been studied, for example, by the Spitzer Space Telescope [346]. The discovery of the plasma crystals or Coulomb crystals of dust in low-temperature plasmas in the 1990s by multiple groups was a major milestone in laboratory dusty plasma physics research. In laboratory plasmas, these microparticles and nanoparticles usually attain a negative charge due to the higher mobility of electrons. The highly charged particles interact with one another electrostatically and exhibit collective behaviors, such as crystallization, melting, demixing, self-excitation of waves, and turbulence (see [347], [348], and references therein). Difference forces, including neutral-gas drag force, ion drag force, thermophoretic force, and Earth's gravity, can also affect the dynamics of the individual dust motion and the collective multiple-particle dynamics. Experiments such as the PK-3 Plus laboratory onboard the International Space Station (ISS) have been used to isolate the effects of Earth's gravity [349]. Tesla-strong magnetic fields have also been applied in the laboratory to examine the effects of magnetization [350]. The processes of self-organization and phase transition can be observed on the single particle level using laser scattering and imaging cameras, such as CCDs. Together with the table-top experimental footprint and modest hardware cost, dusty and complex plasma experiments are highly accessible to data science.

Leveraging the fact that individual dust particles can be detected together with a cloud of dust, tracking individual dust and collective dust motion is an important and unique experimental technique in dusty plasma research. Dust tracking and imaging (see Fig. 12), coupled with theory and dust dynamic simulations (a cousin to molecular dynamic simulations), are used to examine a broad range of problems such as the dynamics of dust charging and motion; dust crystal-liquid phase transition; nonthermal and statistical physics; discovery of new phases of dust clusters, such as glass phase and supercooled dust liquids, nucleation and dust growth, dust acoustic waves and instabilities, nonlinear physics, formation of 2-D and 3-D dust structures, and anisotropic dust clusters under microgravity; ac electric field; cryogenic temperature; charged-particle beams; and shock wave conditions. For example, electrorheological complex plasmas can evolve into a string phase when an external ac electric field is applied [351]. Fluid demixing

and crystallization can be examined with a mixture of two or more types of microparticles. Dust acoustic waves have been extensively studied theoretically and experimentally [352], [353]. Dust acoustic wave turbulence, when coherent dust motion oscillations change to a turbulent state of motion with many harmonic modes, was also reported [354]. More recently, through novel multidimensional empirical mode decomposition based on the Hilbert–Huang transforms, 3-D dust acoustic wave turbulence has been decomposed into a zoo of interacting multiscale acoustic vortices, exhibiting attraction, repulsion, entanglement, bunching, and synchronization, in the 2-D + 1-D spatiotemporal space [355].

TB datasets are available from dusty plasma experiments through particle tracking and imaging [356]. Dusty plasma movies have been recorded at about 1500–5000 frame length, at the rates between 100 and 500 frames/s and each image size of a few MB per frame [357]. For an experimental campaign consisting of a few hundred runs, more than 1.5 million movie frames or more than 1 TB of raw data becomes available [358]. Automated particle tracking through ML is emerging as necessary to process a large number of images and to extract the particle trajectories [356]. Particle tracking and PIV techniques have wider applications than plasma physics. In addition to the traditional probabilistic algorithms, new PTV and PIV algorithms based on U-Net, CNN, and PIML [359] are emerging. Other examples of ML applications may be found in the phase transitions in the dust cloud [360], the correlation of current–voltage (I–V) characteristics given by a Langmuir probe with the main plasma parameters [361], to identify the boundary layer between mixed regions of dust particles with different diameters [362], and the response of a single dust particle levitated in the plasma sheath, to a nonlinear excitation frequency [363].

[Zhehui Wang and Catalin M. Ticoş]

I. Physics and ML

Prior to the recent introduction of ML models, physics-based hypothesis-driven models are the most powerful tools for natural sciences, including plasma physics. ML has now been used in many scientific domains with few exceptions [364], [365], [366], [367], [368], [369], [370], [371], [372]. ML as a new scientific tool is as generic as traditional physics-based hypothesis-driven methods and allows broad implementations by different scientific domains and subfields. Automated data processing through ML has led to the acceleration of every aspect of the scientific activities or "scientific workflows," from observations and experimental data taking to hypothesis generation, model construction, model execution through computation, and model validation and predition [373].

Some plasma problems parallel their counterparts in other scientific domains, which may justify the use of similar ML algorithms. Understanding plasma waves and instabilities in plasma physics poses similar challenges as in understanding diseases in biology [369]. Plasma flow and turbulence, which resemble charge-neutral fluids, are also further enriched in structures due to the EM interactions [374]. New phases of matter, including quantum phases of matter, are expected in high-energy-density plasma experiments due to the extremely high pressure that can be created [375]. Plasma–material surface interactions are encountered in both low- and high-temperature plasmas. Plasma–material interface engineering poses one of the most significant challenges for both fusion energy and plasma technology applications. The computational complexity is comparable to and may even exceed quantum DFT calculations for materials. A comprehensive physics-based description of this multiphase system requires an integrated approach to plasma physics, material science, and their interactions. The length scales involved range from sub-nm to above 1 m in the largest laboratory plasma apparatus. The temporal scale spans 1 fs to the order of a second. Hundreds of controllable parameters may be needed in the search for the best recipe for generating and controlling a plasma, making plasma optimization problems high dimensional. Automation through ML is necessary for model reduction and to accelerate the plasma physics workflows for more accurate predictions, more reliable controls, and more accessible optimization.

One latest trend is to combine ML with physics toward PIML. A combination of a deep learning architecture and high-dimensional datasets has shown to be more effective than earlier ML methods, such as SVMs, small MLPs, random forests, and gradient-boosted trees [376]. High-dimensional data came from multiple plasma apparatuses and different experimental conditions from about 9000 experiments. Physics consideration guided the selection of more than a dozen features, including plasma density, plasma temperature, and so on, as the NN inputs. Physics-motivated dimensionless combinations of the raw measurements were used for input data normalization. Reliable predictions with 82% or better accuracy have been demonstrated on another plasma from the one on which the NN was trained. Construction of Grad–Shafranov equilibria is usually the first step in understanding and control of magnetically confined plasmas. A five-layer fully connected DNN was reported for solving the Grad–Shafranov equation constrained with measured magnetic signals in real time [377]. The computing time was approximately 1 ms on a personal computer, potentially allowing applications in real-time plasma control. An encoder–decoder NN model of tokamak discharge is developed based on the experimental dataset alone [165], without a direct reference to a physics constraint, such as the Grad–Shafranov equation. Electron density, stored energy, and loop voltage were reproduced with close to 90% fidelity to experimental data from a series of actuator signals using the NN. The method provides an alternative to the physical-driven method for plasma modeling, experimental planning, and model validation. Variations of experimental plasma conditions are usually captured by statistical models. The stored energy of a plasma E_{tot}, for example, may be a function of input power (I_p), plasma geometry (Δ), magnetic field (B), ion species (Z_k), impurity (n_i), and so on. The statistical mean of E_{tot}, $\bar{E}_{\text{tot}}$, may be given by

$$\bar{E}_{\text{tot}} = \sum_j E_j P_j \left(I_p, \Delta, B, Z_k, n_i, \ldots,\right) \tag{8}$$

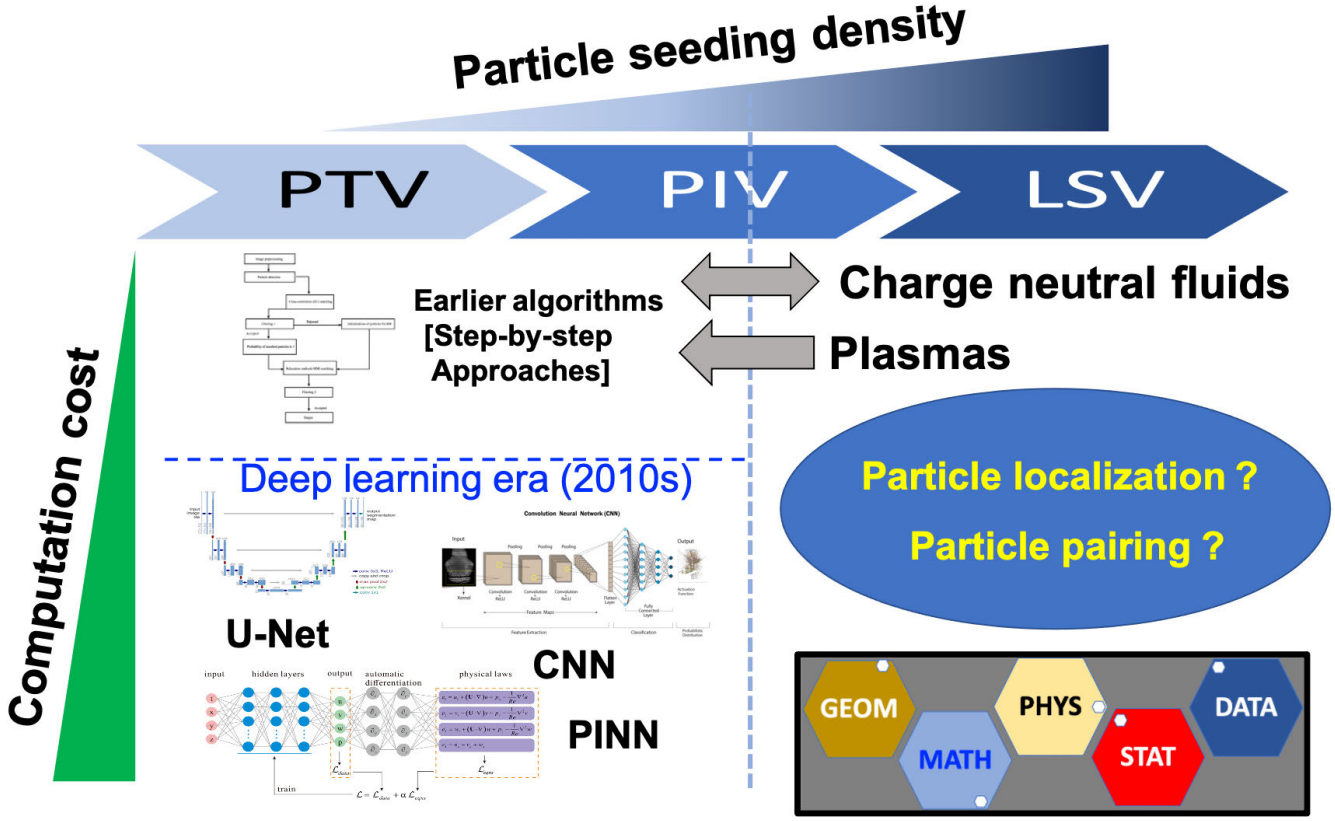

Fig. 12. As the particle seeding density increases, three particle imaging and tracking methods, PTV, PIV, and laser speckle velocimetry (LSV), have been developed for charge-neutral fluids, such as water and gas flows. For plasmas, PTV and PIV are more common. The two central questions for PTV and PIV algorithms to address are how to localize seeding particles from an image and how to pair up the same particles from different images to form particle trajectories. In the deep learning era (the 2010s), a growing number of PTV and PIV algorithms, such as U-Net, CNNs, and PINN, are being introduced to neural fluid and plasmas. The new algorithms have higher computational costs compared with earlier algorithms that process images step by step and more manually. While most such NNs are trained by large datasets, they can also take into account geometry, mathematical, physical, and statistical constraints.

where the probability function P_j corresponds to the energy content E_j. The statistical variance, ΔE_{tot}^2, is given by

$$\Delta E_{\text{tot}}^2 = \sum_j \left(E_j - \bar{E}_{\text{tot}}\right)^2 P_j\left(I_p, \Delta, B, Z_k, n_i, \ldots,\right). \quad (9)$$

To construct explicit probabilities P_j as a function of I_p, Δ and others present substantial challenges for theory but are important to experiments and controls. ML can be used to obtain implicit correlations between E_{tot} with input power I_p and so on. Meanwhile, there may be even features of plasmas that are hard to be captured by explicit physics model [378].

Even with the use of physics-motivated quantities and features, such as electron temperature, plasma density as NN inputs, successful scientific applications of deep learning for feature extraction, pattern recognition, classification, denoising, and nonlinear regression, statistical inference can still be perceived as a "black-box" magic [302], [379]. One may recognize similarly that modern computer codes are also quite complicated and not necessarily transparent to understanding. Code validation, therefore, has been an important part of the code development process. This apparent separation of the power of ML and AI from understanding through the fundamental laws of physics or corollary laws is convenient but not satisfying. The fundamental laws of physics are universally applicable to physics, chemistry, biology, geology, astronomy, and cosmology, to atoms, molecules, and bulk materials, and to different phases of matter, such as gases, fluids, solids, plasmas, and Bose–Einstein condensates. The difficulty of ab initio models is only that mechanical applications of these laws lead to equations much too complicated to be soluble [380]. Other difficulties include incomplete initial and boundary conditions, random noise, and errors that may accumulate with time and the number of elementary calculations. Yet, another difficulty is that data are sparse. Limited by instrumentation or numerical resolution, data and information sparsity increases as the length scale and time step decrease. These difficulties with the first-principle methods have given rise to corollary or empirical laws, such as quasi-linear theory, Kolmogorov turbulence scaling, Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy, adiabaticity of charged particle motion, and many others in plasmas. The corollary laws are approximations of the fundamental laws. They are not intended to be universal and are expected to be broken down. However, the corollary laws are effective methods for understanding complex phenomena and, meanwhile, are traceable to the fundamental laws. One open question is whether ML can be used to derive corollary laws, as a step toward the recovery of the fundamental physics laws behind the data. Another related question is whether such corollary laws and fundamental laws are as important to machine intelligence as they are to human intelligence.

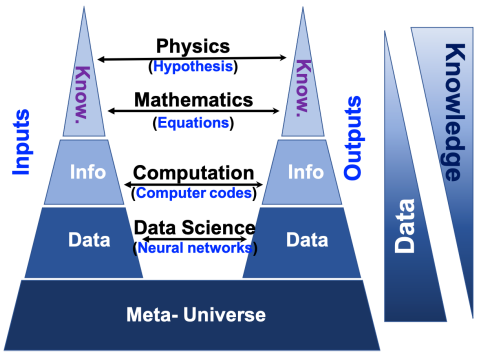

Fig. 13. Universe is the ultimate source of all scientific data, which collectively may be called "Meta Universe." A growing number of methods and tools are used to correlate the data, information, and knowledge, shown in hierarchical order as a subset of the Meta Universe.

Applications of ML in physics and its subfields pave the way toward a more satisfactory union between the two, namely, interpretable ML models based on physics and vice versa, and the discovery of new physics aided by ML. A theory of AI may still be a long way to go [381]. Interpretation of the ML-based algorithms may lead to even more powerful algorithms for plasma control [382]. The fundamental laws of physics are incomplete. With the growing evidence for dark matter and dark energy, and the ongoing effort to reconcile general relativity with quantum physics, there is apparently room for the discovery of fundamental physics through data science. In high-energy particle physics, pattern recognition and machine classification have found applications in data reduction, i.e., searching for extremely rare events that may hint at new physics beyond the existing frame work of quantum chromodynamics [383]. ML to recover hidden physics models could be extended to plasma physics [384], [385].

Physics can give rise to new concepts in ML and data science, such as physics-enabled ML and PIML [386], [387]. Quantum ML is emerging, which could transform both ML hardware and software [388]. The tensor network structure of quantum mechanics has inspired ML methods for classification [389]. One approach to PIML as discussed above is by using physics-motivated quantities or features as inputs and outputs for ML. Therefore, NNs can be trained to emulate corollary laws, such as empirical scaling relations that are widely used in plasma physics. Another approach is to use computer simulations to produce training data for NNs, which can then be used for nonlinear regression and prediction [390]. A recent approach to PIML has introduced differential-equation-based loss functions for NN training. Statistical physics may be used for UQ.

There are also physics concepts that may not be captured by differential equations. One class of such concepts is the principle of symmetry [391], which includes reflection or mirror symmetry, translational symmetry, and rotational symmetry. Galilean invariance is the hypothetic symmetry for different inertial frames. According to Noether's theorem, symmetry gives rise to conservation laws in physics. Momentum conservation is the consequence of translational symmetry. Energy conservation is derived from time-invariant symmetry. Mass conservation is the other familiar example. The probability, probability density of an electron or an ion distribution function, and the intensity of light on a sensor need to be positive. These symmetry, invariants, and the positiveness of many physical quantities may be used to regularize the parameter space of the inputs and outputs of an NN or the loss functions. It has already been recognized that image representations by NN such as CNNs should be invariant due to the translational and rotational symmetries [392]. Using kernel-based interpolation to tractably tie parameters, CNN has been generalized to deep symmetry networks [393]. By taking into account the spherical geometry of an object, spherical CNNs have been found to be more computationally efficient and accurate for 3-D model recognition [394]. There are also specific symmetries in plasmas related to the toroidal geometry of a plasma, periodicity. The concept of collective variables [395], when there is no obvious symmetry, might be useful for turbulent plasma feature extraction. Further exploration of these additional physics concepts for ML algorithms would become fruitful and rewarding in the near future.
[Wenting Li and Zhehui Wang]

J. Challenges and Outlook

Rapid advances in computing hardware, architecture, and data acquisition instruments present challenges and opportunities for plasma physics and science at large. One challenge lies in the fact that manual and even semimanual data mining methods face increasing difficulty in extracting new information and knowledge from large and multidimensional datasets, see Fig. 13 for data reduction and knowledge extraction hierarchy. Data science and ML offer transformative tools for laboratory plasma experiments and the physics of plasmas in the big-data era. The classical physics framework, which includes Newton's laws and Maxwell's equations, is the canonical pathway to understand plasmas and guide the designs of plasma experiments and inventions of plasma technologies. Many problems in plasmas rise from the complexity derived from a large number of particles (on the order of 1 mole in some laboratory plasmas) and their interactions with EM fields and material surfaces. The combination of accumulative computational errors, insufficient knowledge of the initial condition, boundary condition and perturbations, and the long computing time even by using state-of-the-art computers renders the canonical pathway ineffective if possible for reliable predictions and optimization problems in plasmas. There are also NP-hard problems in plasma physics, which may be difficult for both ML and traditional computation. Enabled by heterogenous multidimensional datasets, including experimental and observational data, simulation data, and other metadata, data science and ML have been successfully or can be used to accelerate all aspects of plasma research or the "scientific data flows," i.e., from observational and experimental data taking to hypothesis generation, model construction, modeling, and model validation. Despite their practical prowess and simplicity, ML methods for plasmas and other scientific domains are not completely understood at this time. Seeking a better union between the established knowledge framework of plasma physics and emerging information science is an exciting new frontier for data-driven

plasma physics and laboratory experiments. New results may be anticipated such as in the data-driven discovery of new plasma physics, the development of scientific ML algorithms that will be broadly applicable to problems beyond plasma physics, and a quantitative understanding of uncertainties for more effective predictions and optimization, paving the way toward automated plasma knowledge discovery and novel technologies.
[Zhehui Wang]

IV. Magnetic Confinement Fusion

A. Introduction

For the successful realization of the safe, unlimited, and carbon-free magnetically confinement fusion energy, the nonlinear nonlocal behaviors of ∼150 million °C plasma in the strong magnetic field need to be understood and predicted. There has been, and will be, a vast amount of experimental and computational data available, which may be used to build surrogate models and digital twins. Since a thermonuclear magnetic fusion device is extremely costly and takes tens of years to build, digital twins and surrogate models can be highly valued tools for scientific advancement. Fast surrogate models are also valuable for real-time workflow and control of the on-going long-pulse experiments and improvement of next experiments.

With the rapid advancement of computing power, extreme-scale simulations are supporting magnetic fusion energy (MFE) research by solving fundamental equations. However, the turn-around time for extreme-scale computational study is still too long for near-real-time input to the experimental studies. Data from such simulations can be used, together with experimental data, to raise the fidelity of the simpler models. Moreover, AI/ML can be used to replace computationally expensive kernels to accelerate extreme-scale simulations and enable physics discovery online from the big simulation data and compress the output data without sacrificing the important physics features.

Data-driven science in magnetic fusion research is only at the beginning stage. However, many useful developments have been reported, with some of them already in use in experiments and modeling. Topics covered here many not be highly comprehensive but will at least be representative.
[C. S. Chang]

B. Data-Driven Physics Models

Data-driven models have become increasingly popular in the scientific literature in recent years. One of the basic ideas motivating data-driven modeling is to utilize data from experimental systems (e.g., such as diagnostics, control systems, reactor consumables, and/or maintenance schedules) and simulation models of various fidelity to derive additional predictive models that are either physically informed or, at the other extreme, entirely empirical. Physically informed data-driven models are often either enriched versions of first-principle theoretical physics models (e.g., MHD and gyrokinetcs [396]), or they can be extracted models from data that constrain themselves to prescribed physical laws or conditions.

These techniques are distinct from a surrogate model generation for the acceleration of multiphysics modeling, which relies on model data for their training sets. This is discussed in Section IV-F.

Generally, a benefit of data-driven modeling is that the "validation" of data-driven models against experimental data is, in some sense, baked into the model itself. In other words, because the experimental data are used to train the model, the model validates against that data naturally, removing many of the concerns regarding whether the observed experimental phenomenon corresponds with (or validates against) the model itself. The primary concerns that tend to remain to include the following open questions: 1) whether these models can extrapolate well to different physics contexts (e.g., different machines or plasma configurations); 2) how dependent these models become on the underlying engineered hardware that drives some of the physics observed in the experiments (e.g., the specific engineering design and performance impact a specific divertor, cryostat, and so on may have on the resultant model system); and 3) whether these models are too "blackbox–like" to extract meaningful physical insight/understanding from. These common considerations are illustrated in Fig. 14.

While data-driven methods have been utilized in many contexts, for many purposes—such as for identifying error estimates in sophisticated validation studies using traditional physics simulations models [397], [398], as well as being used in semiempirical methods [399], stabilization analysis [400], the development of plasma stability control techniques [401], discharge control systems [402], deep statistical inference models on experimental data [403], [404], [405], and feedback control schemes [406]—many of these techniques are frequently considered more empirical than physics-based.

As a consequence, efforts have been undertaken to find physics-informed data-driven techniques that are capable of mitigating some of the limitations of these more empirical approaches. For example, PINNs [405], or PDEs solved and enriched using DNNs [407], have been recently developed and explored. These models have generally been used to solve traditional initial-boundary value problems in physics-based PDEs (e.g., multicomponent reactive MHD) but with the added benefits of: 1) significantly improved numerical regularity features; 2) the ability to readily incorporate large datasets into the "training regime"; and 3) the ability to simultaneously solve for solutions over an entire parameter sweep (e.g., over not only (t, x) but over (t, x, γ)). The major drawback of using these methods for solving numerical PDEs, however, is the slower overall runtime per forward solve that can render them impractical for high-dimensional systems (e.g., gyrokinetics), such as those necessary for understanding the plasma physics that drives magnetic fusion reactors [407].

In addition, some data-driven physics models can be applied simultaneously to numerical regimes alongside experimental data, leading to models that are automatically "discovered" from within the data [408], [409], [410], while remaining consistent with the observed data as well. Again, these discovered models (discussed more in Section IV-D) can be either largely empirical [411] or additionally constrained to

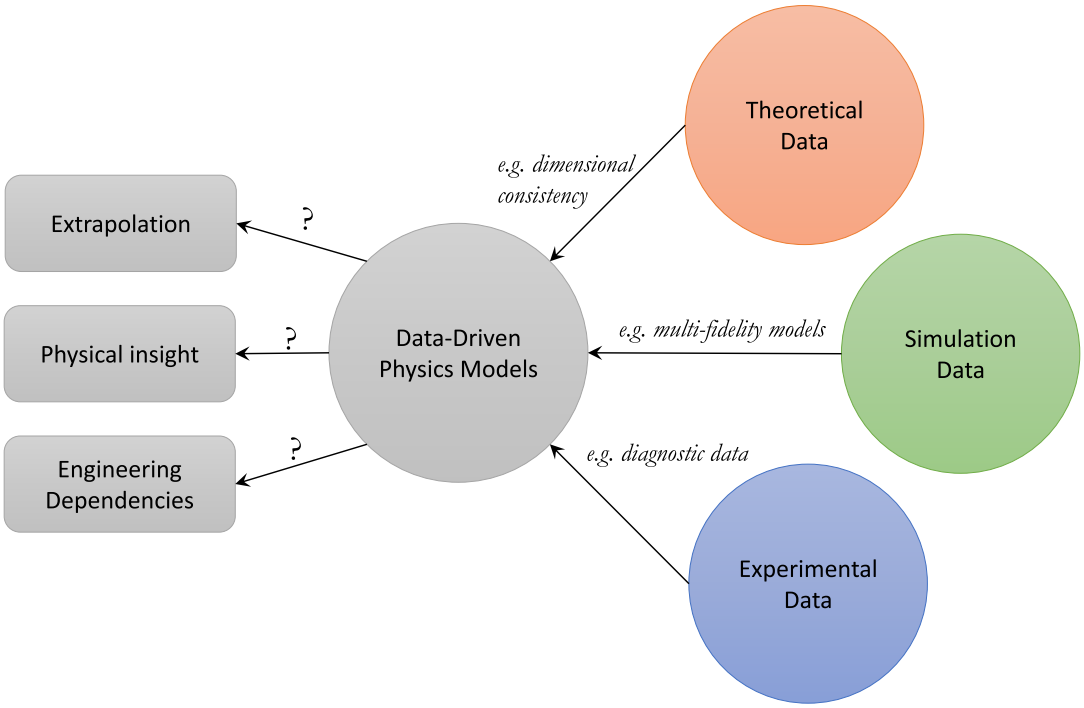

Fig. 14. Relation of data-driven physics models to experimental, simulation, and theoretical data streams.

be physically consistent with theoretical considerations [168] or simulation-based considerations (e.g., high-fidelity model predictions) [407], [412]. It is generally thought that, as the amount of both experimental and simulation data increases, data-driven physics models may become increasingly important for being able to predict and model experimental behaviors while simultaneously connecting the gained insights from these systems to traditional and first principle ways of understanding plasma physics.
[Craig Michoski and Jonathan Citrin]

C. Optimizing Experimental Workflows With Data-Driven Methods

The experimental campaign planning processes in MCF are currently not explicitly computer-aided or otherwise enhanced with optimization, ML, and related machinery. The typical chain of events leading up to experimental scheduling and execution starts with the open submission of proposals, followed by expert discussions in topical groups, and, finally, a selection by committee. It appears highly challenging to formalize this planning process toward a more quantitative exploration–exploitation mechanism, but it may be worthwhile attempting to do so. Since the ultimate purpose of the MCF device is to reliably maintain a high-performing MHD-instability-free fusion grade plasma, and several metrics to characterize such plasmas are available, it follows that the campaign planning mechanism could, and also arguably should, somehow consider those metrics algorithmically in order to optimize the progress toward this purpose.

Explicit human-in-the-loop computer-aided decision support in MCF has been attempted in more focused MCF devices [413], in ICF optimization enabled by data assimilation [414], and also in other process optimizations in experimental physics [289] with seemingly excellent results. The integration of such systems into large tokamak user facilities is a novel area, which is underexplored. Such systems may require original ideas to effectively allocate experimental resources for multiuser multiple-objective exploration and exploitation.

Practically implementing these types of policies in campaign planning may require a shift of the focus of discussions from what topical areas to prioritize next to what metrics to explore and exploit next, and let sanctioned algorithms automatically generate candidate experiments, which can be further discussed and iterated. Classical experimental design response surface methods [415], [416], standard Bayesian optimization [176], and mechanism design [417], [418] can all be envisioned as part of a toolset to build MCF planning decision support systems. In an abstract sense, any planning system for the experimental workflow is a mechanism that uses past data collected, plus external information including predictive simulation data, to propose where data should be collected next.

Mechanism design (not well known in physical sciences) could even be retrofitted onto existing user facilities planning processes. To introduce the idea, here follows a naive example of the optimization of collective valuation. The prototypical optimal social choice mechanism that incentivizes participants to provide truthful inputs is the VCG mechanism [418]. In the context of collaborative planning on a user facility, it could be used as follows. Based on the initial community input, management comes up with a shortlist of allocation options that are compliant with resource and contractual constraints and other programmatic boundary conditions. The user facility participants then submit the number of hours that they would be willing to work to realize each option. The VCG

mechanism selects the option that maximizes the collectively most desirable option (collective eagerness to work on its realization). Crucially, the VCG mechanism uses a formula to charge each participant (extra hours asked to work) such that each participant is best off providing their private true valuation (number of hours actually willing to put up for each option) to the mechanism. Presumably, it also holds that the participants' true valuation is positively correlated with their belief in the likelihood of making actual physics progress.

Improvement to programmatic decision support using data-based methods combined with designed value revelation mechanisms is an interesting direction for future research. User facilities are in this sense arenas where groups of tax-funded agents compete for access to a machine that can (should) convert their labor into a public good (research output that benefits all, not only the resumés of particular individuals) [419]. Revised incentive structures and transparent mapping of performance metrics across operational spaces may enhance this public-good aspect.
[Erik Olofsson]

D. Diagnostics and Fusion Data Streams

In fusion energy plasmas, many disparate diagnostic instruments are simultaneously used in order to cover the multiple physics phenomena covering a range of spatiotemporal scales. In addition, fusion experiments, such as ITER, will run longer pulses, with the goal of eventually running a reactor continuously. The confluence of these facts leads to large, complex datasets with phenomena manifest over long sequences. Fusion scientists have a range of data analysis timescales, from real-time processing for plasma control to between-shot quick processing of data to give insight to adjustments for next shots and to longer term deep analysis for science discovery. Diagnostic data analysis has always been fundamental to progress in MCF energy, and many current and emerging applications of ML are aiding scientists in these many tasks in making sense of diagnostic data.

ML is being applied to interpreting observed experimental data and extracting from it physical parameters of interest (e.g., electron temperature from line-integrated spectrometer measurements). Traditionally, this statistical inference of physics parameters from diagnostic data has been performed under the umbrella of "IDA" [420], performing Bayesian analysis leveraging potentially multiple diagnostics. Recent trends are integrating ML in the form of NNs to accelerate the IDA process, which usually either relies on analytic likelihoods, or resorts to slow, sequential MCMC samplers. NNs have been trained to do approximate Bayesian inference, replicating a Bayesian model, which is used to extract electron temperature from a lithium-ion beam emission spectroscopy (Li-BES) diagnostic on the JET tokamak [421]. The benefit of using an NN is that, now, the inference of electron temperature (with uncertainties) can be performed in microseconds versus the tens of minutes typically required for a single experimental time slice, enabling use in between shots or real-time control.

Similar techniques are being applied when the forward model relating physics parameters to observed diagnostic data is a more formal simulator, making the likelihood intractable. Simulation-based inference technique of NPE uses normalizing flow models [422] (built with NNs) to create flexible surrogates, performing the Bayesian inference to infer physics parameters consistent with the simulator, but again producing results in milliseconds. An example application used the fluid plasma and neutral edge transport code UEDGE, which takes in anomalous transport coefficients and produces plasma kinetic profiles of density and temperature. NPE was used to train a normalizing flow model on 10 000 UEDGE simulations, producing an NN that could then take in profiles of electron/ion density and temperature from diagnostics at the midplane and the outer divertor, and infer the corresponding anomalous transport coefficients that are consistent with UEDGE [423].

Various works are using other methods for speeding up and broadening the analysis that can be done with experimental diagnostic data for physics parameter extraction. For example, a simple feedforward NN was trained to extract electron temperature from a database of measured spectra from an EUV/VUV spectrometer based on the measurements of electron temperature from Thomson scattering diagnostics [187]. The above works and techniques aim to improve our physical understanding of fusion plasmas by leveraging ML to extract physics from experimental diagnostics.

Recent trends have focused on various ways to accelerate the identification of plasma modes or other events directly from diagnostic data using supervised learning. These applications are for aiding the researcher in identifying items of interest but also for the inclusion of real-time control algorithms. Reservoir computing, a dynamical ML model that trains quickly, has been applied successfully to the prediction of Alfven eigenmodes in the DIII-D tokamak [424]. NNs have also been used for very rare and difficult signals to fine, such as solitary bursts before Edge Localized Modes on KSTAR [425]. Also, convolutional NNs with dilated convolutions have found utility in working with long sequences for diagnostics with high sampling rates like the ECEi diagnostic at DIII-D [426].

Large-scale data analysis for experimental diagnostics can be accelerated using data science and networking techniques to stream the data from the experiment to large, remote HPC centers. By working with data streams and leveraging the large HPC compute resources, better and more data analysis can be performed, which can better inform fusion scientists between plasma shots on the best way to optimize the next shot [427]. A demonstration of this used the streaming framework DELTA [428] to stream ECEI diagnostic data from the KSTAR tokamak in South Korea to the NERSC HPC center in the USA and complete spectral analysis of all channel pairs using multiple CPUs on the Cori supercomputer. The entire streaming and analysis were completed in 10 min compared to the 10 h that sequential analysis would take. This opens the door for a range of large-scale parallel analyses, modeling, and simulation to further enhance the information scientists can extract from diagnostic data.
[R. M. Churchill]

E. Prediction of Tokamak Disruption

Disruption, which is an abrupt termination event of tokamak discharge, is one of the biggest issues in fusion energy development [429]. Magnetic and thermal energy as high as GJ is released in a very short time of the order of milliseconds at this event. Consequently, disruption causes harmful damage to tokamak through excessive thermal load on the wall, magnetic force, and run-away electrons. Therefore, prediction, avoidance, and mitigation of disruption are prerequisites for a tokamak fusion reactor.

Extensive works for disruptions have been done since the early stage of fusion research [430], and intensive works targeting the operation of ITER are being implemented by international collaborating efforts [431], [432]. Although understanding of the physical process of disruption has been deepened by the MHD theory and simulation, the prediction capability of disruption still remains limited. Since disruption is highly nonlinear dynamics with the complex interaction of different physical processes [433], it is essentially difficult to predict disruption by the framework of time-dependent differential equations defined a priori. Instead, data-driven approaches based on a posteriori observation are anticipated to give an induced model reliable for practical use [376]. In this section, the development of data-driven models for disruption prediction of tokamak plasmas is reviewed.

Deep-learning algorithm for multimachine disruption prediction has been proposed and achieved high predicting accuracy across multiple tokamaks [434]. This means that device-independent representations of disruptive characteristics have been identified. Simultaneously, this work has shown that nondisruptive property is device dependent, and only the use of existing tokamaks is still not enough to predict disruption in a new tokamak. It is also noted that synthetic data from numerical simulation do contribute to the improvement of prediction capability.

The approaches of interpretable ML models, which are contrast methodology of deep learning [434], NN [376], and generative topographic mapping [435], are attracting interests because not only they improve prediction capability but also their resultant expression enables exploration of underlying disruption physics. Physics validation of the model/hypothesis would secure the limitation of generalization performance. Also, these approaches have high potential compatibility with actuators for disruption avoidance and mitigation. Random forest algorithms and sparse modeling via exhaustive search and SVM are referred to as examples.

The random forest algorithm has been applied to the prediction of disruption is a variety of tokamaks, such as DIII-D [436], JET [437], Alcator C-Mod [438], and EAST [439], and it has been successfully integrated with the real-time plasma control system on DIII-D and EAST. Disruptivity is the final probability of disruption and is characterized by the average result of decision trees to classify disruption/nondisruption from training. It should be noted that this approach can quantify the relative contributions of the various input data signals to disruption. Disruptivity is expressed in the decomposition formula of the sum of each feature contribution and bias of the intrinsic value of the sample mean in the classification scheme. Since the decision paths in random forest trees provide measures of the explainability of input data, the effectiveness of new input data is easily assessed. For example, peaking factors of plasma parameters, such as temperature, density, and radiation, are proved to enable earlier prediction. In other words, the selection of input parameters based on hypothesis and physical insight is essential for the improvement of a predictor.

Not limited to disruption prediction, the selection of input parameters is an essential issue for ML. The exhaustive search, which exploits the inherent sparseness in all high-dimensional data to extract the maximum amount of information from the data, selects key parameters subject to the SVM classifier for disruption. With regard to high-β disruption in JT-60U, four physical parameters have been extracted as key parameters to describe the boundary between the disruptive and the nondisruptive zones [440]. Then, it has been found that disruption frequency can be expressed as the distance from the boundary in multidimensional space. Consequently, the disruption likelihood has been quantified in terms of probability based on this boundary expression. Fig. 15 shows the contour plot of the disruption likelihood on the plane of the normalized pressure β_N and the function of residual extracted parameters. It is noted that the boundary function is expressed in a power law, so as to be compatible with physics discussion.

Careful deliberation of the expression of the disruptivity/disruption likelihood, which is derived from ML, could lead to the elucidation of the underlying physics behind disruptions. A data-driven approach to the prediction of tokamak disruption is inevitable for the plasma control system and the device protection system in ITER, as well as the next demonstration fusion reactor.

[Tatsuya Yokoyama and Hiroshi Yamada]

F. Surrogate Models of Fusion Plasma

A challenge in multiphysics simulation of MCF systems [441] is to achieve high physics fidelity at a computational burden that is compatible with the desired use case. This is particularly acute for many-query applications, such as sensitivity studies, UQ, scenario optimization, and reactor design. Fast simulations can also be applied in control-oriented simulators, where high accuracy is critical for powerful new techniques, such as controller design through RL [442].

Carrying out regression of the individual physics models that comprise the multiphysics suite, using supervised learning methods, can circumvent the conflicting constraints of model speed and accuracy. The ML-learned surrogate model then provides faster (often by orders of magnitude) multiphysics simulation when applied as drop-in replacements for the original models. The computational cost is relegated to the training set generation phase, facilitated by HPC resources. See Fig. 16 for a conceptual overview. In principle, physics models that are too slow for routine direct application in multiphysics simulation can also be incorporated in such a manner, as long as there is sufficient computing resources for generating the required training set. This idea is compelling since the ML-surrogate has the potential to then be both faster

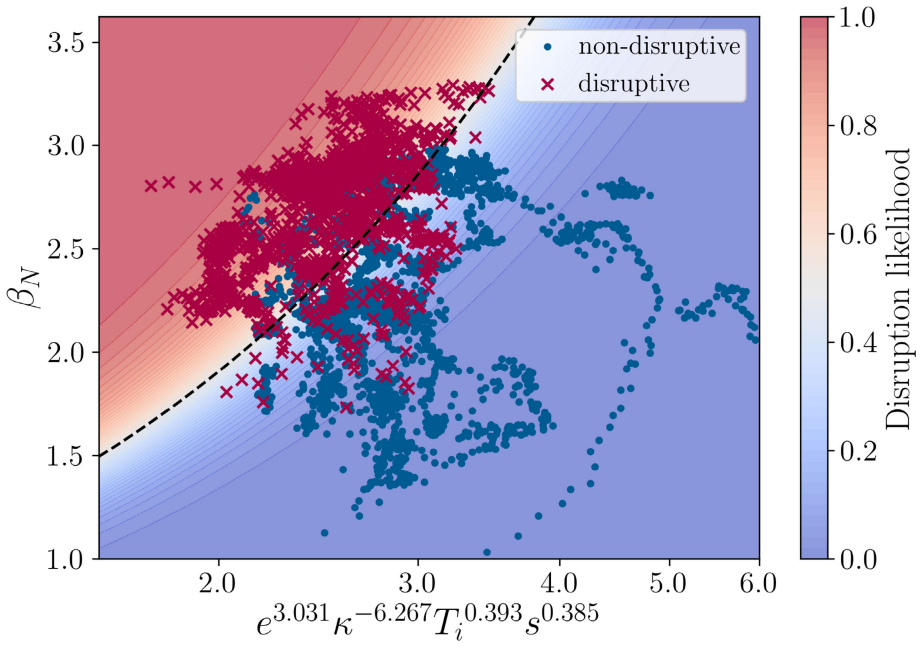

Fig. 15. Contour plot of disruption likelihood. Here, β_N, κ, T_i, s, and e are normalized beta, plasma elongation, ion temperature, magnetic shear, and Napier's constant.

and more accurate than present-day multiphysics modeling capabilities.

To date, multiple surrogate models have been developed for fast MCF modeling applications, primarily (but not exclusively) applying feedforward NN architectures. A nonexhaustive list of examples is summarized in the following.

1) *NUBEAM Monte Carlo Neutral Beam Heating Code [443]:* Principle component analysis was applied to reduce the dimensionality of the 1-D input and output profiles, and extensively validated on DIII-D [444] and NSTX-U [443].
2) *Turbulent Transport Models:* The QuaLiKiz-neural-network [445], [446] utilizes prior knowledge of the physical input–output mapping structure for determining physics-informed constraints of network topology and optimization cost functions that improve model fidelity. Applications include JET tritium ramp-up optimization [446] and ITER scenario optimization [447]. Similar work was carried out for TGLF [448] with applications for scenario optimization and control [449], as well as the multimode model [450]. A surrogate of the higher fidelity turbulent transport model GKW has been developed for JT-60U parameters [396].
3) EPED NN for pedestal predictions and core-pedestal coupling workflows [448].
4) 3-D MHD equilibrium calculations for stellarator optimization applications [451].
5) MHD instability calculations, as part of a disruption predictor stack [452].
6) Surrogate formula for divertor heat-load width built from combined experimental and gyrokinetic simulation data [412].

Further extension of these techniques to incorporate all components of the MCF multiphysics simulation stack provides a pathway toward fast and accurate interpretation of present-day experiments, scenario design and optimization (including intershot), and control-oriented modeling. Future devices, such as ITER, will require the availability of such a pulse design simulator to increase shot efficiency and reduce risks.

A common challenge in constructing the surrogate models is on the data generation side, particularly for high-fidelity physics models with a higher computational burden. It is critical to establish robust high-volume computation workflows, automated data validation and filtering pipelines, and selective sampling techniques. The NN outputs also need UQ to establish trust zones and flag when the surrogate model is extrapolating. At the simplest level, this is achievable by assessing the variance of an ensemble of identically trained models. Ideally, the model UQ should be coupled to an active learning pipeline, whereby the model training set can expand when new parameter space is encountered.
[Jonathan Citrin]

G. Magnetic Fusion Energy Data Challenges and Solutions

Data access patterns for ML workflows are fundamentally very different than traditional access patterns for magnetic fusion experimental or simulation data. Conventional repositories of experimental data have been designed for small-scale human consumption in the control room and are mostly aimed at simultaneous visualization of small amounts of data gated by the visual/mental response time of the human operator. In significant contrast, ML access patterns are driven by algorithms that can potentially read and use vast amounts of data, requiring substantially more computational resources for data loading and processing. In addition, issues associated with data curation, such as data discovery, cleaning, normalization, and labeling, are all critical components of successful fusion ML studies.

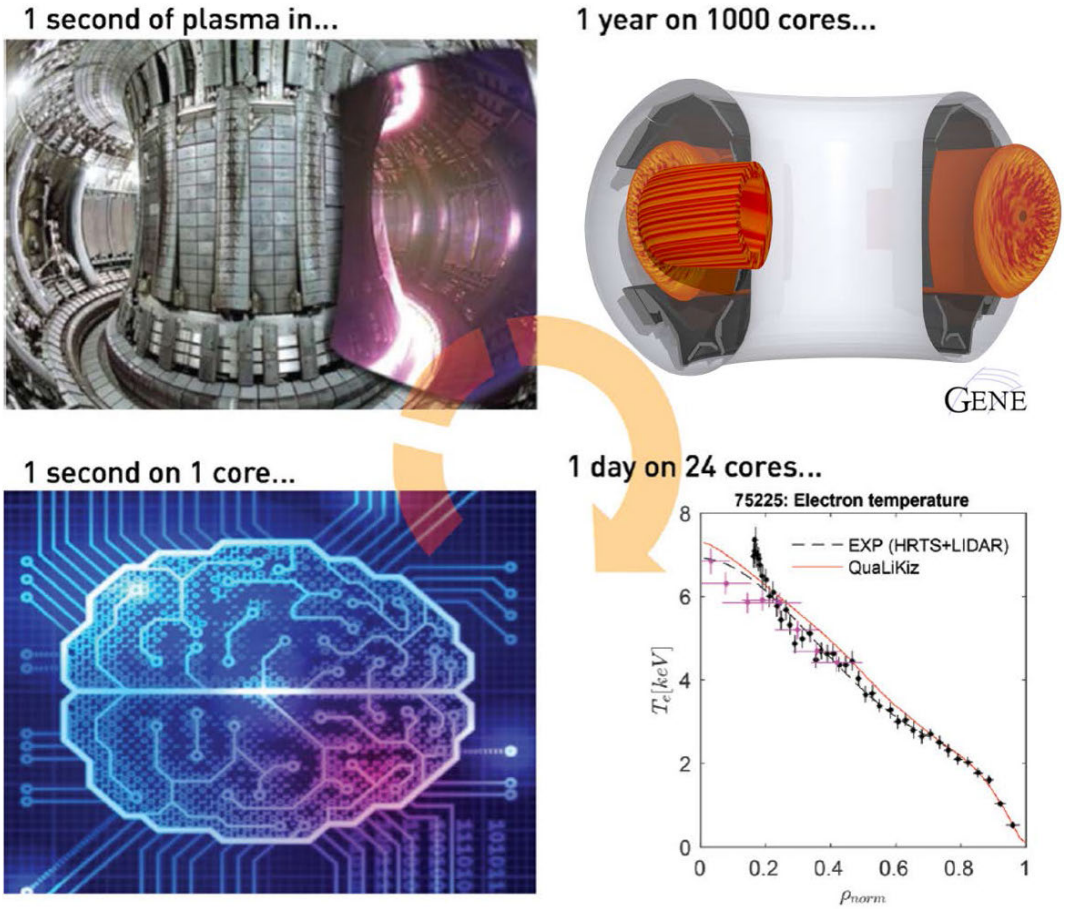

Fig. 16. Hierarchy of models applied toward fast and accurate multiphysics simulation, with the example of tokamak core turbulence. Routinely modeling a tokamak scenario is prohibitively expensive with high-fidelity nonlinear gyrokinetics (upper right). However, the high-fidelity model verifies and validates reduced-order-models (lower right), which are then applied to generate training sets for ML-surrogates (lower left plot) for fast simulation.

These issues were outlined in the Report of the Workshop on Advancing Fusion with ML [453], which highlighted several limitations of the conventional data repositories; shortcomings that need to be addressed to fully harness the transformational potential that ML could provide in many areas of fusion energy. In particular, the report supports the idea of a community-wide FDP targeted at ML research. The core idea of such an FDP is to provide an integrated environment for ML and data exploration studies, supported by a common interface. Data must be staged and supported with sufficient metadata to support rapid, iterative ML workflows, an example of which is illustrated in Fig. 17. ML studies typically integrate a large number of software tools, so a significant library of tools must also be included to support such workflows. Examples of support tools include data visualization, dimensionality reduction, and rapid data space analysis tools, along with the tools needed to actually conduct ML training, testing, and inference.

The DIII-D data archive provides representative examples of both the size and variety of data used by the fusion ML community. It currently consists of ~0.4 pb of data accumulated over decades of operation. It contains both raw, unprocessed signal data that are stored in the GA-implemented PTDATA system [454] and processed data (such as equilibrium reconstructions) that are stored in MDSplus [455]. The data contain a wide array of dimensionalities, ranging from scalars to images, and signals with sample rates spanning multiple orders of magnitude. Historically, the most typical access pattern for these data has been experimental scientists analyzing on the order of ten [$O(10)$] shots with $O(10)$ signals per shot, with the I/O and processing capabilities of the archive system sized accordingly. The access patterns required for ML applications have proven to be significantly more resource intensive. A typical ML study conducted using DIII-D experimental data might be able to take advantage of up to the scale of $O(10^5)$ discharges. In recent years, DIII-D has sought to deal with this need for large-scale data access by deploying a scaled-up data access and processing system. This system includes a complete copy of the DIII-D experimental archives on a BeeGFS parallel file system [456], along with the TokSearch [457] framework for parallel data processing, allowing for multiple order of magnitude data processing throughput improvements for typical ML use cases.

Data discovery relies to a large extent on the ability to perform expressive queries for metadata. For example, a plasma disruption study needs expert-labeled annotations indicating both the time of occurrence and the type of disruption. The DIII-D experimental data system is integrated with a Microsoft SQL Server [458] relational database that records $O(100)$ metadata fields across $O(10)$ tables for each shot, including

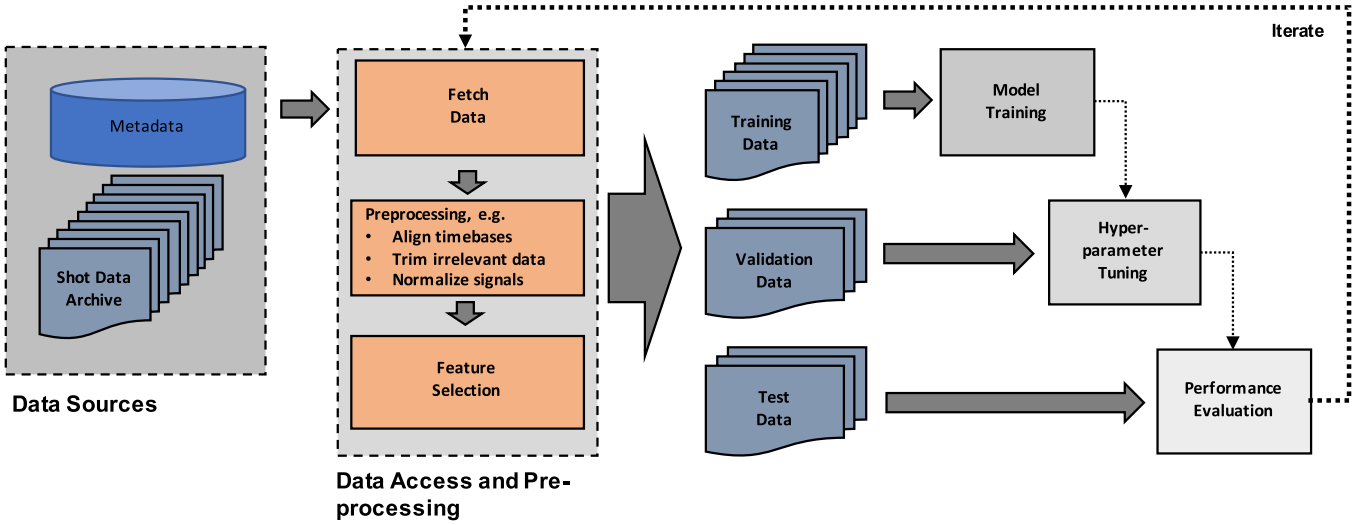

Fig. 17. Typical supervised learning workflow for data-driven MFE studies. Data exploration, access, and preprocessing are conducted iteratively in conjunction with ML modeling. An FDP would facilitate the rapid execution of this loop.

a schema for recording disruption information. A typical ML application will often gather a preliminary list of shots to process by first querying the relational database. As a simple example, one might be interested in shots from a particular date range, or one might have search criteria related to shot length, shot start time, experimental logbook entries, maximum plasma current, and so on, all of which can be queried using standard SQL. However, it is worth noting that, while the approach taken by DIII-D might effectively utilize one set of tools, there has not been a community-wide effort at standardization, particularly with regard to metadata management, an issue that a dedicated FDP would address. Such standardization would also facilitate increased engagement with domain experts who could more easily provide the annotations needed for classification studies.

Magnetic fusion data are fairly unique in their variety and scope. A single ML study might utilize the following:

1) 0-D scalar time series (e.g., magnetics);
2) 1-D profile time series (e.g., current profile);
3) 2-D grid data time series (e.g., equilibrium reconstructions);
4) image time series (e.g., infrared camera data).

Each of these items may be stored in a different file format, and each may have one or more associated metadata elements. Such breadth and depth of data underline the need for a community-wide effort toward standardization, which, given the critical importance of data quality and availability for ML, will have a dramatic impact on the ability of the community to execute data-driven studies.
[Brian Sammuli and David P. Schissel]

H. Data Science for Extreme-Scale Simulation

Global nonlinear simulation using fundamental kinetic equations in the whole plasma volume including realistic divertor geometry requires extreme-scale simulations. The soon-to-arrive exascale computers will be great tools, but the size of the filesystem capacity is relatively small compared to the compute node memory. This brings up the necessity for online data analysis and data reduction/compression before being written out to the filesystem. The online data analysis can be done in the simulation codes at every timestep if the analysis routine is well parallelized. However, there are analysis routines that may not be easily parallelized. In this case, the simulation data can be offloaded to some analysis nodes using the asynchronous remote direct memory access (RDMA) or one-sided message passing interface (MPI) data transfer at every timestep. Thus, the computing does not slow down, while the data are analyzed in the analysis nodes. AI/ML can be used in the analysis nodes for efficient visualization and scientific discovery. Reduction and compression of the analysis data can also be performed in the analysis nodes.

Data-driven AI/ML can accelerate extreme-scale simulations by replacing some compute-intensive kernels with AI/ML inference routines. The Fokker–Planck collision operation is an example [50]. Preconditioners and PDE solvers can be good candidates. However, some difficulty lies in the accuracy and physics property conservation in the data-driven routines, e.g., L2 error, mass conservation, momentum conservation, energy conservation, and viscosity conservation. If we aim for 1% error at the end of 1000 timesteps' simulation, a data-driven routine must have $<10^{-5}$ relative error to avoid accumulation in the possible "drifting error." This level of error bound in AI/ML is not easy and requires support from fundamental AI/ML scientists.

Data-driven AI/ML can also perform other functions to help the extreme-scale simulations in real time: by detecting and mitigating possible load imbalance, by detecting and suppressing known numerical instabilities, by utilizing UQ techniques to request simulation steering into needed input/output parameter space and to execute autonomous validation tasks using preloaded experimental data in the independent data analysis nodes, by combining simulation-experimental data to help construct predictive surrogate models (see Section IV-B), and so on. Fig. 18 depicts a summary of the data science topics for the extreme-scale kinetic magnetic fusion simulations.
[C. S. Chang]

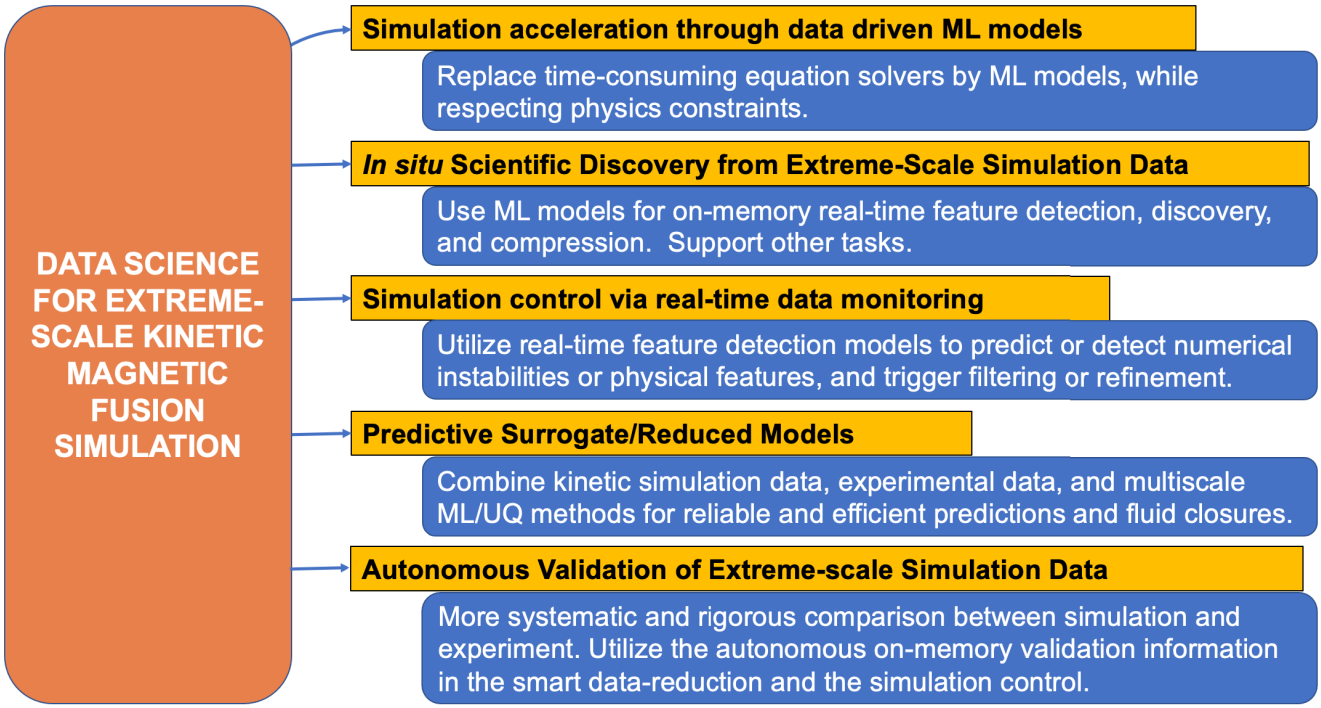

Fig. 18. Data science topics for extreme-scale kinetic magnetic fusion simulation.

I. Challenges and Outlook

As for other application science areas, there are numerous challenges in utilizing data-driven sciences in MCF research. Besides the challenges and outlook listed in each of the above subareas, an important aspect to keep in mind in discussing the challenges and outlook is that MCF is different from many other scientific projects in that it is an international-scale mission-oriented project. This means that global collaborations among geographically separated large-scale laboratory facilities and between laboratory experiments and high-performance computations are key to success.

A vast amount of data produced (and stored) in different format at different experimental facilities and by different simulation codes over decades of time span (see Section IV-G) may require building a community-wide federated database and workflow system [459], which is based on the meta-database management system and which honors individual institution's and code's data format transparently and maps multiple autonomous database systems into a single federated database via wide area network without the need for centralized data mirroring.

The inference codes can be placed on or nearby the collaborative experiments, such as ITER or future prototype reactors. However, their learning should be performed on remote HPCs, with frequent RL for timely updates of the inference codes, using streaming data to cope with observational variance. To achieve this, a global management system is needed over a wide area network for efficient workflow (see Fig. 19).

The continuous accumulation of data to be generated by ITER, or future fusion test reactors, will reach to be enormous (tens of exabytes over the lifetime of the ITER experiment). Historically even in Today's tokamak experiments, once the experimental data hit the permanent storage tape, they are seldom utilized for scientific discovery. It is a challenge but desirable that the streaming data out of the various experimental diagnostics are to be organized according to the features and reduced/compressed without the loss of the features on the way to the permanent storage. In this process, a special request can be sent to the simulation communities, together with the feature-preserved reduced data, for the timely study of the observed experimental phenomena and feedback for the design of improved experimental scenarios, as mentioned in Section IV-D. Various AI/ML techniques are expected to be a highly valuable tool in accomplishing this, including workflow framework building. All the data science techniques discussed in Section II and in this section can be utilized in this workflow framework at various stages.

[C. S. Chang]

V. Inertial Confinement Fusion and High-Energy-Density Physics

A. Introduction

The field of HEDP is typically defined as plasma physics with energy densities $>10^{11}$ J/m^3, equivalent to pressures $>10^6$ bar. HEDP research covers a broad range of systems from strongly coupled "warm dense" matter, through laboratory astrophysics and ICF, to ultraintense laser–plasma interactions, and more. While these subfields probe a zoo of physics phenomena, they are all underpinned by the twin pillars of experimentation and simulation. The difficulties in reaching the conditions of interest in an experiment, collecting high-quality data, and modeling the results mean that both pillars rely on the largest experimental and computing resources available worldwide.

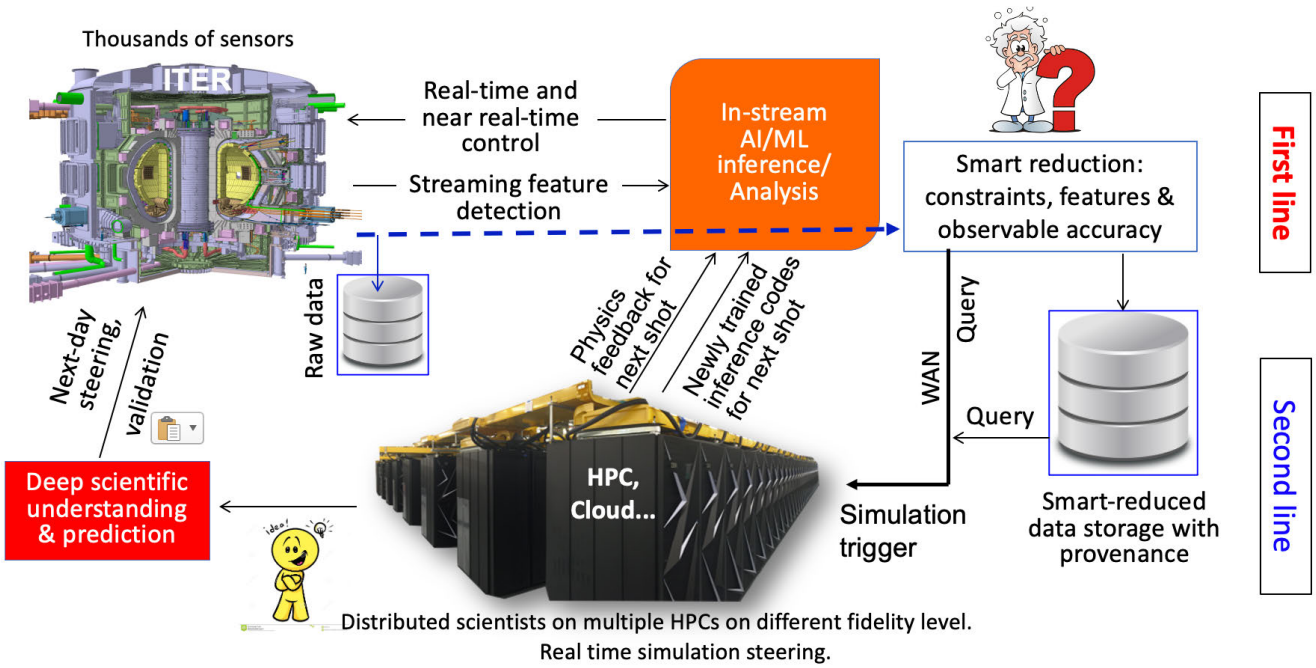

Fig. 19. Schematic of the data movement workflow, showing the first line data science region at the experimental site and the second line of data science and HPC studies at remote sites.

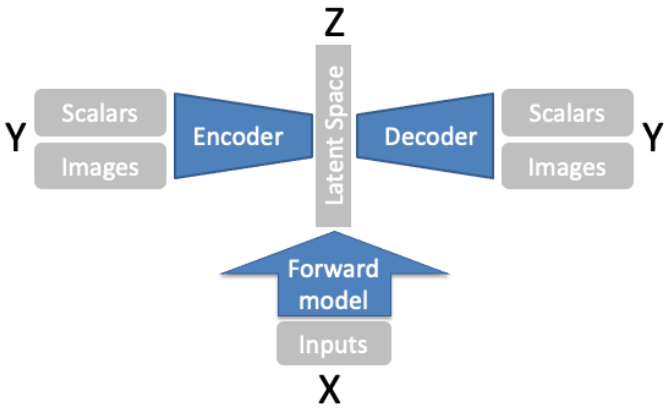

Fig. 20. NN architecture predicting multimodal outputs Y of the simulations. The outputs are first compressed using an AE into a latent space Z, and then, a forward model is trained to predict these compressed outputs from the inputs X.

We envision data-driven methods as a cross-cutting third pillar that both improve HEDP experiments and simulations, and sit at the interface between the two. Data-driven methods provide an opportunity to efficiently feature our complex datasets, reliably combine information from simulations and experiments, and accelerate the rate at which simulations and experiments can be performed. As a result, HEDP and ICF problems are quickly becoming an important driver of data-driven methods for science. In the remainder of this section, we will describe some aspects of the research in these areas.
[Brian K. Spears]

B. Representation Learning for Multimodal Data

Predictive models in plasma physics are used to set our expectations about future experiments with varying designs; explore, optimize, and automate new designs; and infer important physics parameters that cannot be accurately measured or simulated, thereby allowing for an improved understanding of the experiments. While expensive simulations can generate a variety of diagnostic data types, in many applications, simulations need to be replaced with fast-to-evaluate predictive surrogates, which have traditionally been fit to only a handful of scalar diagnostic outputs. This approach ignores rich observational and simulated data, such as high spatial and temporal resolution X-ray and neutron images, or neutron yields recorded at multiple azimuths around the burning plasma. These nonscalar detectors are routinely deployed during experiments at nuclear fusion facilities, such as NIF, Omega, and Z, providing additional and more detailed information about processes operating within the plasma. Including these multimodal data can help break degeneracies in the scalar-only models and reduce model uncertainty.

Combining multimodal data poses a challenge as it requires the model to predict thousands of variables in each image or array, and these variables are typically correlated both within and across data modalities. Rather than training the model to predict raw data, ideally, one would like to find a representation of these data in terms of a set of independent variables corresponding to the key physics parameters controlling the experiment. Unfortunately, standard compression techniques cannot detect correlations between different data modalities and collapse an arbitrary combination of data arrays into a set of decorrelated variables. Recent advances in deep learning, however, provide tools for building data-driven representations that both compress and decorrelate multimodal data making them suitable for inclusion in the predictive models. At the same time, the computing power at the national laboratories has grown to the point, where a sufficiently large number of

expensive, radiation hydrodynamics ICF simulations can now be run to train data-hungry deep learning models [460].

Equipped with more powerful supercomputers and deep-learning tools, researchers at LLNL have designed a new deep learning architecture to include multimodal data and build more robust predictive surrogates of ICF simulations [461]. In this architecture, simulation outputs Y, consisting of images and scalars, are embedded by an AE into a low-dimensional manifold Z (see Fig. 20). The AE consists of two NNs: an encoder $E{:}Y \rightarrow Z$ and a decoder $D{:}Z \rightarrow Y$. To reduce statistical dependencies between the compressed latent variables, a Wasserstein AE is used instead of a standard AE. Adding the adversarial training strategy, in addition to the standard L-2 norm minimization, causes the AE predictions to look like training samples, enforcing consistency with the physics relations built into the simulation. The second part of the architecture is the forward model $F{:}X \rightarrow Z$ connecting the input design space X with the diagnostic outputs compressed by the AE Z. The robustness of this model is improved by imposing a cyclic consistency regularization to penalize predictions that are inconsistent with the pseudoinverse network, which is trained simultaneously with the forward model.

While the advanced features of this architecture allow the model to predict multimodal simulation outputs nearly perfectly, physicists need to know whether the model also preserves physics relations learned from the simulation. One such relation was investigated by Anirudh et al. [462]. Using the approximation of Planck's law, the brightness of images from four energy bands was converted into the electron temperature and compared with the ion temperature—one of the scalar diagnostics. These two temperatures are strongly correlated in the simulation outputs. The correlation is very well preserved in the predictions of the NN model for the validation samples even though this correlation was not imposed as a constraint during the AE training.

In summary, representation learning enables the inclusion of diverse types of diagnostic data in the training of accurate, scalable, and predictive surrogates of the simulations. Advanced deep learning techniques allow for building representations that are better at preserving physics relations between predicted diagnostics than standard NNs.
[Bogdan Kustowski and Rushil Anirudh]

C. Transfer Learning for Simulation and Experiment

Standard computer simulations for indirect drive ICF, without platform-specific corrections, often show discrepancies with experiments. In the ICF community, a new approach to calibrating simulations to experimental data has been shown to create models that can predict the outcome of ICF experiments better than simulations alone.

This approach leverages an ML technique called "transfer learning" to merge simulation data and experimental results into a common model. Transfer learning is when an NN trained on a large dataset to solve a given task is partially retrained to solve a different, but related task, for which little data are available. For example, an NN trained on the ImageNet dataset to label random objects (such as cars, trees, and cats) can be modified by retraining just a few layers of the NN to label very specific images, such as the type of aircraft in a photograph, which has a significantly smaller training dataset.

In ICF, transfer learning is used to take simulation-based NNs and partially retrain them on sparse sets of experimental data, creating a model that is more predictive of experiments than simulation alone.

Two approaches to transfer learning for ICF have been published in recent years, one which learns an NN mapping from design input parameters (such as target geometry and laser pulse) to experimental outputs, and one which transforms simulation outputs to experimental outputs via a transfer learned AE.

The input-to-output mapping approach was first demonstrated by predicting the outcome of direct drive ICF experiments at the Omega Laser Facility. An NN trained on 30 000 1-D LILAC ICF simulations was partially retrained on 19 experiments that spanned to the same design space as the simulations. The model predicted the subsequent four experiments with significantly higher accuracy than the LILAC simulations alone; this is shown in Fig. 21 (left).

The AE-based transfer learning technique was developed to overcome challenges associated with indirect drive ICF—the expense of integrated hohlraum simulations and the sparsity of indirect drive ICF data. An AE trained on a large database of capsule-only simulations learns to encode ICF outputs (such as yield, temperature, and density) into a latent space and decode back to the outputs. The model is transfer learned with pairs of integrated hohlraum preshot simulation outputs and corresponding experimental measurements for a database of 50 ICF experiments carried out at the NIF. The resulting model produces an accurate mapping from preshot simulation predictions to expected experimental measurements; resulting predictions from this model are shown in Fig. 21 (right).

A key benefit of each approach to transfer learning is the ability to immediately update the model after each experiment by retraining the network with the new data. This means that the models get more accurate over time, providing a powerful new tool for future design exploration by providing empirically realistic sensitivities to design parameters. Furthermore, the transfer learned models can guide us toward high-performing designs more efficiently than simulations alone.

While transfer learning techniques described above predict scalar diagnostic data, multimodal can be incorporated into predictive models using representation learning. Matching additional data types to better inform the model is particularly important in transfer-learned models because they are retrained on only a handful of experimental samples. ML literature, however, does not explain how to apply transfer learning in multimodal architectures with AEs, such as the one discussed in Section V-B. Multiple retraining options have been tested at LLNL and discussed in [463]. Using synthetic ICF data, the authors demonstrated that retraining the decoder part of the NN architecture allows for correcting systematic biases in important characteristics of X-ray images, such as the hot spot size, shape, and brightness. Such correction is possible even when only a handful of synthetic experiments are available, as in the case of real ICF experiments. Ongoing research aims

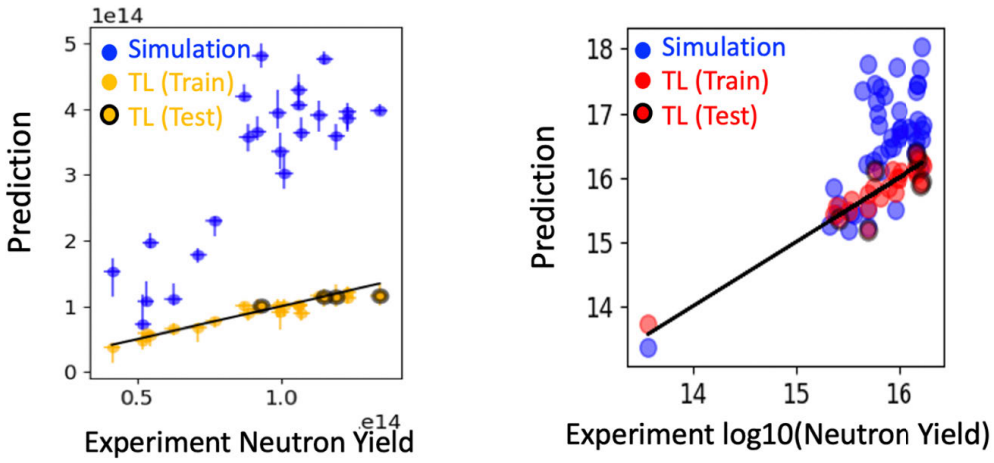

Fig. 21. Left: actual versus predicted values of the neutron yield for Omega ICF experiments based on simulations only (blue) and the transfer learned NN (yellow). Right: actual versus predicted values of the neutron yield for indirect drive ICF experiments from the NIF. The transfer-learned AE predictions (red) are significantly more accurate than the preshot simulation predictions (blue).

to improve this method to handle larger, and more realistic, biases between simulations are real experiments.

Because transfer learning has shown promise at correcting simulated images to match synthetic experiments [464], it will be natural to apply this method to multifidelity simulations. An initial model could be trained on a large number of 1-D radiation-hydrodynamics simulations and then elevated to match a smaller number of expensive, 2-D simulations, potentially eliminating the need to run thousands of them to train the model from scratch.

[Kelli Humbird and Bogdan Kustowski]

D. Uncertainty Quantification and Bayesian Inference

Quantifying uncertainty presents huge challenges in studies of ICF and HEDP systems that stem from the complexity of both experiments and physics models. Experimental observations are sparse, difficult to diagnose, and limited in the range of parameter space that they can access; as a result, they provide limited information, and there is usually some amount of extrapolation to regions where predictions are needed or new physics may be learned. A proper accounting of how much information that we have about a system of interest is, fundamentally, a question of uncertainty, and this puts UQ at the forefront of ICF and HEDP research. In recent years, data-driven methods have been pushing the boundaries of what is possible resulting in more reliable estimates of uncertainty and, hopefully, more predictive computer models.

From a data science perspective, experimental datasets are rarely complete enough to make purely experimental-data-driven approaches feasible. Instead, the usual approach is to use the available data to make point checks of physics models (*benchmarking*) or to fit a handful of parameters to observations (*tuning*). The tuned and benchmarked physics model can then be used to make predictions at a new point of interest, with a limited (or no) understanding of the uncertainty in the prediction. Recently, ICF and HEDP researchers have started to formalize the process by treating physics as a second source of information and build data-driven models that are in some way informed by both sources. A variety of approaches have been explored, for example, using simplified physics models [465], [466], [467], [468], by using physics considerations to limit the size of the design space [469], [470], or by attempting to combine data from disparate but physically related experiments [471]. Other important efforts aim to pose the benchmarking and tuning of large-scale multiphysics simulations as a Bayesian inference [414], [472].

The Bayesian approach has the advantage that the results automatically capture uncertainties in a statistically consistent manner, while methods that use multiphysics simulations are our best representation of current physics understanding, making the results interpretable. However, using simulations in a Bayesian inference framework requires huge computational resources since large numbers of simulation runs, each requiring hundreds of CPU hours to complete, are required. Overcoming this computational barrier has relied on the use of surrogate models [461], [473], [474] that aim to replace the simulation with a cheaper approximation. A large set of simulations is run—requiring tens of millions of CPU hours—and then used to train an approximate interpolator that maps simulation inputs to predicted observables. The key point is that the generation of training data is a massively parallel operation that can leverage leadership-class high-performance computing facilities and software tools, while including simulations in the inference directly requires the samples to be run serially. With a good choice of surrogate [475], a high-fidelity analysis that would be impossible with the simulation itself can be run in a few hours, opening the door for thorough and realistic UQ studies. The process of building surrogate models for large simulations has motivated many of the developments described elsewhere in this article and has made ICF and HEDP datasets [476] a key driver of developments in scientific ML.

A recent application of the Bayesian approach aimed to interpret results from a series of so-called "BigFoot" ICF implosions at the NIF [477], [478] (see Fig. 22). This work used 100 000 2-D HYDRA simulations [479], in a Latin hypercube design over eight input dimensions, to train a novel cycle-consistent DNN surrogate [461]. The DNN was trained

in an approximate Bayesian manner [480], giving uncertainties in the surrogate prediction, which were calibrated by tuning the prior on DNN weights [481]. The trained and calibrated surrogate was used in an MCMC inference of probability distributions over the 8-D input space in order to match a set of experimental observables for NIF shot N180128. Comparing the observed quantities with posterior predictive values from the inference [see Fig. 22(a)] shows a match to multivariate experimental data that would be extremely difficult to achieve with the simulation in the loop, and the Bayesian approach provides a meaningful measure of the quality of the fit in the form of predicted errorbars. Since the analysis includes high-fidelity physics, the fits can be easily interpreted as modifications to radiation drive and degradations [see Fig. 22(b)]. Finally, the use of a DNN surrogate allows for the inclusion of nonscalar data, such as X-ray images [see Fig. 22(c)], which suggests a path toward future analyses that can use all of the information collected in an experiment (i.e., without first projecting nonscalar observations to scalar features).
[Jim A. Gaffney and Jayaraman Thiagarajan]

E. High-Performance Computing and Simulation Acceleration

The vast amounts of data generated by these simulations and required for training some of these models can create a substantial demand for scalable training algorithms and leadership-class HPC resources. In developing models for these multimodal datasets, we have created new techniques for composing data-, model-, and ensemble-level parallelisms and working with 100M sample datasets with 1.5B scalar fields and 1.2B images [482]. Using the LTFB algorithm developed by Jacobs et al. [482], [483] enabled the entirety of a supercomputer, such as Sierra, to be used when training a single model architecture and was able to produce a single instance of a well-converged model. Some of the techniques that have been developed are a coupled, tournament, training algorithm that intertwines the training of a set of model instances to produce a single, best model that has been trained on a sufficient portion of the training data to generalize across a held-out tournament and validation datasets. In addition, we developed a scalable, in-memory data store, and data ingestion algorithm that is able to fetch a massive, distributed data set efficiently and use only a single pass over the data for the entire training regime. Finally, we have developed methods for both model- and data-parallel training of each individual instance of the NN architecture and optimized it for the IBM Power9 + Nvidia Volta architecture of the Sierra system. Building upon these capabilities enabled us to produce a demonstration on training a generative molecular model on 1.6B small samples, which was selected as a finalist for the 2020 Gordon Bell Special Prize for COVID-19 research [484]. These algorithms have been implemented in the Livermore big ANN (LBANN) scalable deep learning toolkit, which is open-source and is being optimized for the next generation of leadership-class computing systems, Fugaku, Frontier, and EL Capitan.

In addition to the optimization of deep learning training for HPC systems, we are also exploring the integration of next-generation AI accelerators and hardware platforms. Specifically, we have integrated two stream dataflow architectures, the Cerebras CS-1 and SambaNova SN10-8, into two of our HPC systems, Lassen and Corona, respectively. Using these systems, we have started to evaluate these accelerators that may be able to serve in a Cognitive Simulation workflow, offloading data-driven, in-the-loop, surrogate models from traditional graphics processing unit (GPU)-accelerated compute nodes.
[Brian Van Essen]

F. Design Exploration and Optimization

A key challenge for ICF is the relative lack of experimental data. Leadership class experimental facilities may only be able to execute a few experiments per week, with single campaigns consisting of perhaps dozens of experiments. As such, a major challenge is how to design and optimize an experiment for a desired outcome (such as high nuclear yield), with only very few opportunities to experimentally test that design. Historically, the community has heavily leveraged high-fidelity full system numeric simulations to first design experiments in silico. Then only after searching for a likely-to-be effective design numerically is a candidate design fielded and tested in an experiment. Numerical simulations, therefore, play a crucial role in the design and optimization of ICF experiments.

However, the digital design of full-system experiments brings with it another set of challenges. First, while ICF drivers and targets facilitate great flexibility, this flexibility comes with a cost: the design space is extremely large. For instance, laser pulses can change their time- and space-dependent power distribution. An ICF capsule needs to define its ablator layer thicknesses and material compositions. In the case of indirect drive, a hohlraum's material and geometry also need to be defined. Furthermore, the tolerances on ICF designs can be very tight, requiring micrometer precision. In all, to fully define an ICF experiment can easily require setting a few dozen independent parameters. The setting of these parameters has historically been done by subject matter experts who leverage physics knowledge and intuition to smartly find new designs. A major advance would be to move from this labor-intensive manual process toward automatically discovered and rigorously optimal designs.

Mathematically optimizing functions of several dozen parameters would not be a challenge, except that the simulations are expensive. A full indirect-drive coupled hohlraum-capsule simulation can cost a few node days, and the simplest simulation that treats just a capsule with low-fidelity physics models can still take a few core minutes. Mathematically, this means that the objective function is very expensive to calculate. Since navigating high-dimensional spaces requires many function evaluations,[1] mathematical optimization and design exploration for ICF seem to be prohibitively expensive: the search space is too large and the simulations are too costly.

[1]For instance, a simple gradient-based optimization algorithm would need to run at least as many simulations as the size of the search space to calculate a finite-difference approximation of the derivate.

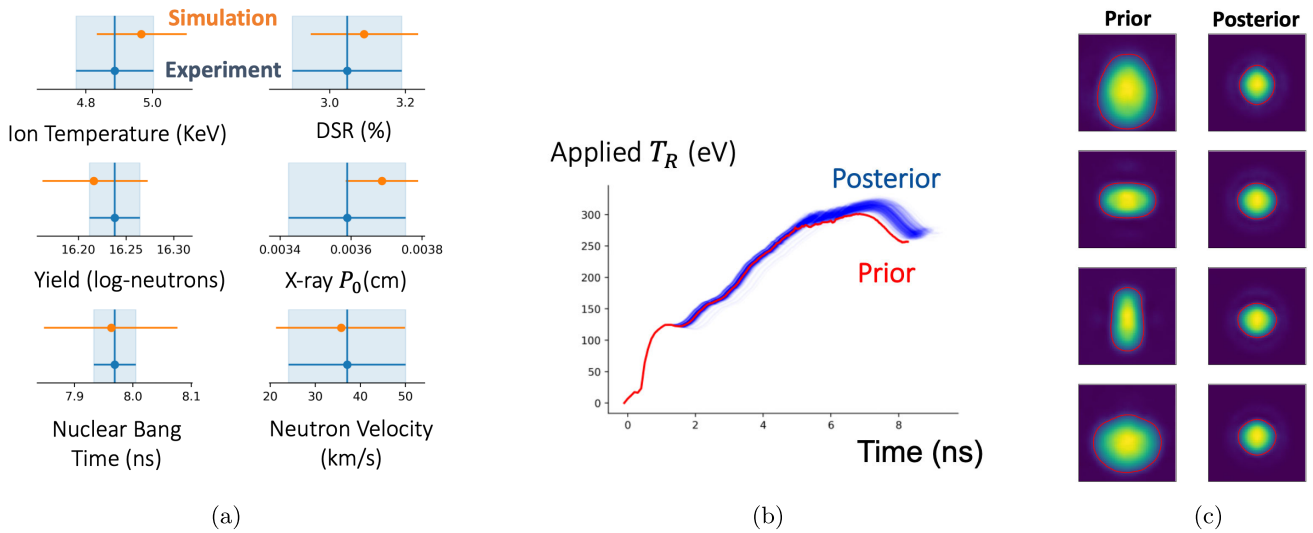

Fig. 22. Results of Bayesian inference of inputs to a high-fidelity multiphysics simulation based on experimental data from the NIF (shot N180128). Inference was enabled by a DNN surrogate trained on 100k expensive (∼10 node-hours/simulation) simulations. (a) Quality of match to experimental observables; the Bayesian approach gives an unprecedented quality of fit and provides uncertainties in the match. (b) Posterior predictive radiation temperature driving the implosion, T_R, demonstrates the inherent interpretability of this approach. (c) Prior and posterior equatorial X-ray images for the shot, which are enabled by our use of DNNs that are highly effective for nonscalar data.

However, recent advances in ML and computational hardware are beginning to usher in a new era of optimal digital design for ICF. Peterson et al. [485] leveraged high-frequency ensemble computing and surrogate modeling to discover a digitally optimized design. The computational workflow to do this was rather complex since it had to automatically mange and coordinate the execution and postprocessing of several thousand concurrently running independent HPC simulations. To do so, the authors developed and deployed cloud computing workflow technology on the Trinity supercomputer at the Los Alamos National Laboratory not to run a large high-fidelity model, but rather to run several thousand lower fidelity models. In all, they were able to execute 60 000 simulations, which spanned a nine-parameter capsule design space, enough to adequately train a random forest regression model. Once trained, the surrogate model was fast enough to embed into a global optimization algorithm. The authors also introduced the idea of "robust design," whereby the design parameters themselves could be uncertain (for instance, due to finite manufacturability precision or tight engineering tolerances). Instead of maximizing the nuclear yield, they maximized the probability that the simulation achieved some threshold yield, given the variability about the desired target design. After finding a predicted location for a new optimally robust design, the authors then double-checked the result by running new simulations. Interestingly enough, these new simulations suggested a new kind of physics regime for ICF, defined by asymmetric capsule implosions filled with instability-suppressing vortical "zonal" shear flow. Zonal flows, while being common in magnetic fusion, had previously been unseen in ICF, and their discovery would not have been possible without an automated optimal design framework.

Automated design optimization has also yielded more intuitive results, as in [486]. This work avoided the gradient-in-high-spaces problem not with a surrogate model but via a genetic algorithm for use in ICF capsule optimization. Good performing simulations had their capsule layer thicknesses and material compositions "bred" together in an iterative fashion, with the fittest candidates surviving to breed in subsequent generations. Within a few dozen generations, the best design that emerged appeared as a canonical ICF target, with low-density DT gas surrounded by high-density DT ice encased in an ablator layer. In this example, an automated optimal design was able to navigate a high-dimensional space and settle on a design template not-unlike one that human subject matter experts have learned via decades of study.

While automated design in ICF has shown some early success in being able to discover both intuitive and nonintuitive designs, the solutions that they discover are inherently limited by the simulator used. That is, if a model disagrees with an experiment, the optimal model-based design may be of little interest since it may or not reflect reality. In this case, it could be possible to use techniques such as transfer learning [269], [487] to postprocess raw simulation data during the optimization process. That is, the optimization cost function evaluates not the output of the simulation, but rather the output of an ML model that adjusts the simulation output to better match what might occur in an experiment. A similar technique that used statistical linear regression to modify simulation outputs drove an experimental campaign on Omega to record yields [469].

Given the early numerical and experimental success of deploying automated and optimal design exploration, its use for ICF is likely to grow. Table II summarizes some of the key challenges and opportunities as the field progresses. Surrogate-based and gradient-free optimization can be enhanced with Bayesian optimization [488] techniques that use surrogate model uncertainty to balance exploration and exploitation (provided that surrogate models produce uncertainties that increase in unexplored areas). These iterative

TABLE II
CHALLENGES AND OPPORTUNITIES FOR AUTOMATED DESIGN OPTIMIZATION AND EXPLORATION FOR ICF

Challenge	Opportunity
Relatively few experiments	Model-based design
Costly simulations	Surrogate-enhanced optimization, multi-fidelity optimization
High-dimensional design spaces	Bayesian optimization, gradient-free and agent-based optimization
Tight engineering tolerances	Stochastic optimization, robust design
Complex simulation pipelines	Next-generation hardware; advanced workflow software
Simulation-experiment discrepancy	Transfer-learning of simulation data to match experiments

techniques, in contrast to the single-pass or human-in-the-loop surrogate-based optimization, however, become increasingly complex since simulation postprocessing, surrogate model training, and optimization and simulation launching must be automatic. Such heterogeneous, dynamic, high-frequency computing is less common in a traditional HPC environment than it is in data science. However, the melding of AI and scientific computing is a broad trend, and next-generation computer hardware and software will likely see a continued merger of ML and traditional HPC technologies [489], making the infrastructure needed for automated design more common. [J. Luc Peterson]

G. Self-Driving Experimental Facilities

The sections above have introduced various technologies from data representations to design optimizations that address a number of important challenges in plasma science and scientific ML in general. Here, we show how the combination of these techniques can tackle an even broader challenge to develop self-driving experimental facilities. One of the dominant trends in large-scale experiments, manufacturing, and even computational sciences is the rapid increase in automation. Whether it is particle physics, 3-D printing, or managing massively parallel workflows, the underlying processes are too complex, and decisions need to be made too quickly for humans to be directly in control. In the context of plasma science, we are particularly interested in high-repetition laser experiments. The state of the art in laser experiments used to involve one shot per hour or even per day, which provides ample time for an initial analysis and to adjust experimental parameters on the fly. Effectively, this created a manual, expert-driven optimization loop with each experiment hand-selected and curated. Current systems allow multiple shots a second and soon may reach frequencies of tens or even hundreds of hertz. In this new regime, we can no longer optimize individual experiments but need to preplan entire sequences or even shot days. This invariably can lead to thousands of experiments being wasted as the preset plan proves less interesting than expected or through mistakes only discovered after the fact. If not addressed, these challenges could easily negate many of the benefits that the more frequent experiments provide. Instead, combining the various technologies introduced above, we are developing the fully automated and integrated control loop for laser experiments shown in Fig. 23. The overarching goal is to adjust the various laser controls, i.e., power, pulse shape, and so on, denoted as the input parameters X to optimize some scientific objective, such as maximizing electron temperature achieved in the experiment.

To build this system, we start with a large ensemble of simulations (see Section V-E) designed to mimic a planned experiment as best as possible given the constraints on computing resources and physics knowledge. This results in a large set of outputs representing synthetic diagnostics and internal states of the system (only observable in the simulations). Subsequently, we use representation learning (see Section V-B) to entangle all available multimodal output data (of the simulation) into a latent representation (Z in Fig. 23), which is then used to build a multimodal forward modal predicting the mapping to the full outputs. Similarly, we build an inverse model, and in fact, typically, these models are linked to ensure internal consistency [461]. We then start the experiment using (a set of) inputs initially assumed to provide high-quality outputs. Each shot records a set of experimental diagnostics assumed to be a subset of the simulated diagnostics from the simulations. Using manifold projections and ideas from transfer learning (see Section V-C), we then search the data representation for the Z whose corresponding outputs in the forward model best represent the experiments, taking into account experimental noise, distribution shifts, and so on. This ultimately leads to a set of what we call "enhanced diagnostics" that include not only the measures experimental diagnostics but also unobservable internal state information estimated through the mapping of the forward model. The enhanced diagnostics are then provided to the current inverse model to estimate simulation-based input parameters whose outputs are expected to match the observed results. Using the inverse model and the observed differences between the current simulation-driven forward model and the experiment, we can then exploit the design optimization techniques of Section V-F to compute a new set of inputs aimed at optimizing the objective. Once connected, this chain represents a closed-loop optimization approach in which the knowledge encapsulated in a large ensemble of preshot simulations is used to autonomously drive high-repetition laser experiments. Going forward, the next step is to include self-learning models as well and to use the observed discrepancies in both outputs and estimated inputs to improve both forward and inverse models on the fly. In the limit of sufficient experimental data, this will provide the means to incrementally modify the initial simulation-based model to create a fully experimentally informed one.
[Tammy Ma and Peer-Timo Bremer]

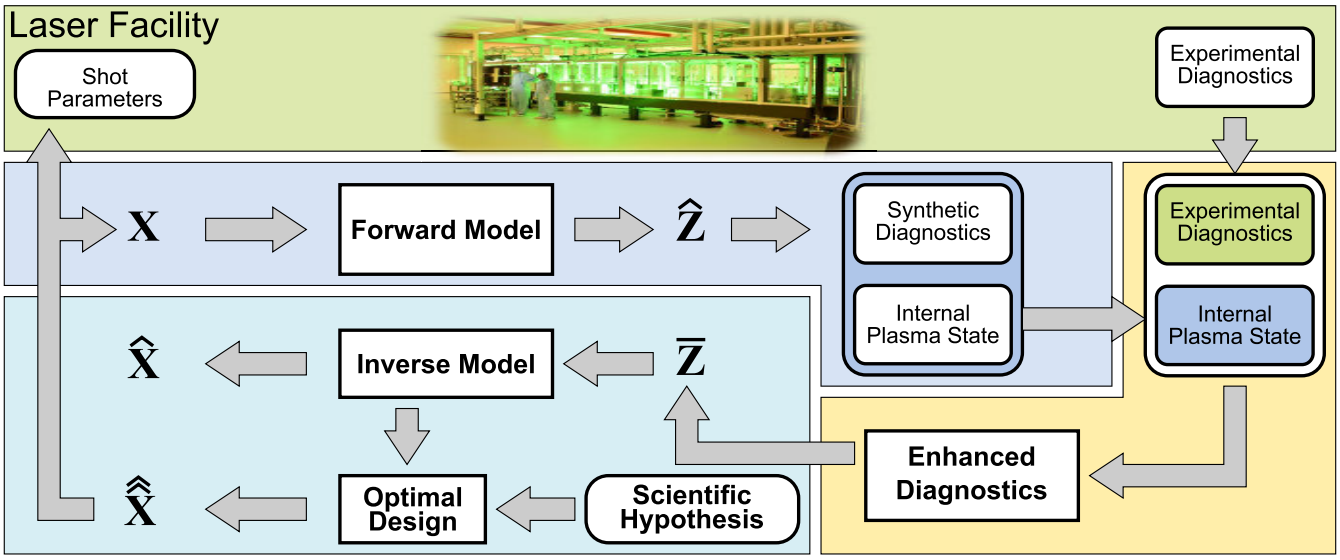

Fig. 23. Schematic of the envisioned control loop integrating simulation-based models with enhanced diagnostics and optimal design. For each shot taken at the facility (green box), the corresponding forward model produces the expected outcomes in the form of both synthetic diagnostics and predicted internal states. The expected internal state is then combined with the experimental diagnostics (yellow box) and used as the initial condition to estimate a corrected internal state Z. The estimated state is then used directly in the inverse model to potentially improve the surrogate but also informs the next shot by considering the current objective and the results of the previous shot to suggest new shot parameters.

H. Challenges and Outlook

We have described several elements of ongoing research that aim to make data-driven methods the third pillar of HEDP and ICF research, alongside large-scale experimentation and simulation. While each of these elements is ongoing work, the ultimate aim of the HEDP and ICF community is to tightly couple them into a continuous, iterative process of scientific discovery; high-fidelity simulations inform the design of experiments, and the resulting data are used to update physics models and propose new experiments at very high throughput. Many of the components of this vision are already in place, and ICF and HEDP research is pushing the remaining pieces forward.

[Brian K. Spears]

VI. Space and Astronomical Plasmas

A. Introduction

Besides the mysterious dark matter and dark energy, the observable universe is known to consist mostly of plasmas and EM fields [490], [491]. The mass of the solar system, which hosts an average size star, is dominated by the solar plasma confined to Sun's gravity. The solar, terrestrial, such as auroras, and extraterrestrial plasmas, such as solar wind and intergalactic clouds, are too large to fit in laboratory experiments. In other words, these natural plasmas would generate orders more data if they were subject to similar measurement schemes in the laboratory. These natural plasmas do share common physical mechanisms and processes, such as energy and mass transport on the meter size and smaller scales with laboratory plasmas, which can be probed and measured in a controlled setting. With the recent detection of gravitational waves, a golden age of astrophysics including astrophysical plasma physics has arrived. The growing number of satellite and ground instruments can generate unprecedented amounts of observation data from the RF to the gamma-ray region of the EM spectrum, and a lot more will become available through for example the LSST on the ground and the James Webb telescope in space. Within the solar system, space instruments can probe the solar, Earth-bound, and lunar plasmas with unprecedented spatiotemporal resolution through particle detectors, electric probes, magnetic probes, and concerted measurements from different satellites. On the largest spatial and temporal scales of the universe, these data provide information to address open questions and constrain theoretical models regarding the origin, the current state and structures, and the future fate of the universe. On the galactic scale, new phases of matter, such as double-pulsar systems [492], provide unique laboratories and observational data for the reconciliation of quantum theory and relativity, and open up new regimes of relativistic and quantum plasmas that only may exist inside nucleus or matter under extreme pressure [493]. On the solar scale, the data present opportunities for SWx forecasting and protection of the growing number of space assets. On the terrestrial scale, atmospheric plasmas, such as lightning, provide opportunities to understand climate change and other environmental issues.

The explosive growth of observational data is expected to continue on all length scales from cosmology to terrestrial plasmas. In addition to new windows of observation, such as Laser Interferometer Gravitational-Wave Observatory (LIGO), large digital sky surveys across the EM spectrum are a predominant source of observational data [494]. For example, between June 1997 and February 2001, the Two Micron All Sky Survey (2MASS) collected 25.4 TB of raw imaging data covering 99.998% of the celestial sphere in the near-infrared J (1.25 μm), H (1.65 μm), and Ks (2.16 μm) bandpasses [495]. As of 2019, the IRSA alone provides access to more than 1 petabyte of data consisting of roughly 1 trillion astronomical measurements, which span wavelengths from

1 μm to 10 mm and include all-sky coverage in 24 bands. The IRSA dataset will soon exceed 100 times the data size of the Library of Congress. The Sloan Digital Sky Survey (SDSS) telescope produces 200 GB of data every night. The new LSST captures 6-GB images at 3 GB/s with its 3.2 billion-pixel camera and will generate about 15 TB of raw image data every night. The Cassini mission collected over 600 GB of scientific data from 2004 to 2017 [496]. Big data have given rise to the interdisciplinary field of astroinformatics and astrostatistics. The importance of automatic data mining has been recognized by astronomers, cosmologists, astro and space plasma physicists, statisticians, and computer scientists alike in recent years [497], which not surprisingly coincides with the advances in novel NN structures, such as deep learning [255], [302], [498]. Even though ML and AI may not completely replace human intelligence in the foreseeable future, such revolutionary tools may lower the barriers for scientists from other fields and even hobbyists alike to contribute to data analysis and new knowledge mining, through the distributed open-source platforms, such as SpaceML [499].
[Zhehui Wang]

B. Space and Ground Instruments

First, the increasing volume and varieties of observational data from space and astrophysical plasmas are the results of the growing number of ground and satellite-based instruments. Examples of the EM instruments are summarized in Fig. 24. Ground-based instruments are limited to optical and radio wavelengths due to the absorption of Earth's atmosphere. Satellite instruments overcome this limitation and can also stay far away from human-generated backgrounds, such as lighting.

Second, benefiting from the advances in microelectronics, such as CCD and CMOS technologies, which have been characterized by continued reduction of feature size (currently down to nanometers as in the cell phones) or Moore's law, each instrument has more data capacity due to more pixels or channels, and each channel or pixel can have higher data acquisition rate and more data storage. The LSST CCD camera has a pixel size of 10 μm. Scientific CMOS imagers have been gradually replacing CCD imagers because of their low noise, small pixel format (around 1 μm), and high quantum efficiency (above 90%) [500]. Microelectronics further allows higher data yield instruments with lower weight, power consumption, or more compact size, and therefore, a greater number of instruments or channels can fit onto the same payload of a satellite. In addition to continuous improvements in instrumentation hardware, space instruments become more accessible due to the continued decline of the launch cost to the low Earth orbit, from about \$100k/kg in the 1980s to \$1–10k/kg in the 2020s.

Third, advances in detector materials and optics have also given rise to new capabilities in collecting more data and more efficiently. Astrophotonics is a relatively young field that leverages novel photonic components and integration for astronomical instrumentation. Integrated photonic technology is an extension of integrated electronics technology from processing electrons to photons. The integrated photonic circuits provide reduced size, weight, and power that are critical for compact instrumentation, especially for space-based systems [501]. Astrophotonic solutions are already becoming an integral part of existing instruments. Examples include photonic lanterns, complex Bragg gratings, spectrographs, frequency combs, and interferometry on a chip [502]. Astrophotonics also enables the next generation of large telescopes, such as the Extremely Large Telescope (39 m).

Finally, despite the advances in instruments and data processing hardware, the sheer volume of data from space and extraterrestrial plasmas, which is essentially infinite, requires intelligent data reduction strategies. Traditionally, such strategies come from human intuition, theory, and simulations. These established methods and scientific routines are useful in planning a measurement and designing the satellite orbits for the measurement, but are not enough for space-based measurement especially for in situ measurements. Plasmas within the solar system allow in situ measurements similar to laboratory experiments. The Parker solar probe has been flying into Sun's atmosphere since 2018. Equipped with six remote-sensing instruments and four sets of in situ instruments, the Solar Orbiter spacecraft has been collecting data since 2020. The Parker probe and the Solar Orbiter will not be the last ones of their kind since they can generate data that are essential to better understand the solar corona heating, the solar wind acceleration, the 11-year cycle of the solar magnetic activities, and space dust, paving the way toward more reliable SWx forecasting. The Parker Solar Probe is planned for two dozen flybys to Sun's corona with a temperature up to 1371 °C. Planning an in situ space plasma measurement ahead of time is like planning a trip, which is difficult due to the indeterministic nature of the SWx, the counterpart of the weather on Earth.

One emerging trend is to use machine intelligence for onboard data processing and reduction. Machine intelligence has already been routinely used for orbit maneuvers of individual spacecrafts, coordinated positioning of large satellite constellations, satellite communications, rendezvous, sample collection, and returns. Onboard classification of images by a $10 \times 10 \times 10$ cm^3 CubeSat in Earth orbit using a random forest classifier was reported [503]. The classifier was trained on the ground prior to launch using test imagery from a high-altitude balloon flight. The CubeSat used a nonradiation-hardened commercial Atmel AT91SAM9 processor (210 MHz) that costs about \$40. Another example is ϕ-sat-1, which has an AI chip to downselect image data before transmitting them down to the Earth. The use of state-of-the-art ML methods, such as deep learning onboard, has so far been limited by the satellite computation hardware and available power [504]. Deep learning algorithms, such as CNN and U-Net, are being adapted to fit onboard space applications. An ultralight CNN called CubeSatNet was described for image classification for eventual implementation on a 1U CubeSat [505]. CubeSatNet had the highest F1 score compared to trained SVM, the deep belief network (DBN), and AE models. The trained model, with an accuracy of around 90%, was slightly above 100 kB in size and can fit the memory size of an ARM Cortex MCU. A flight demonstration of various CNNs using TensorFlow graphs for image processing was

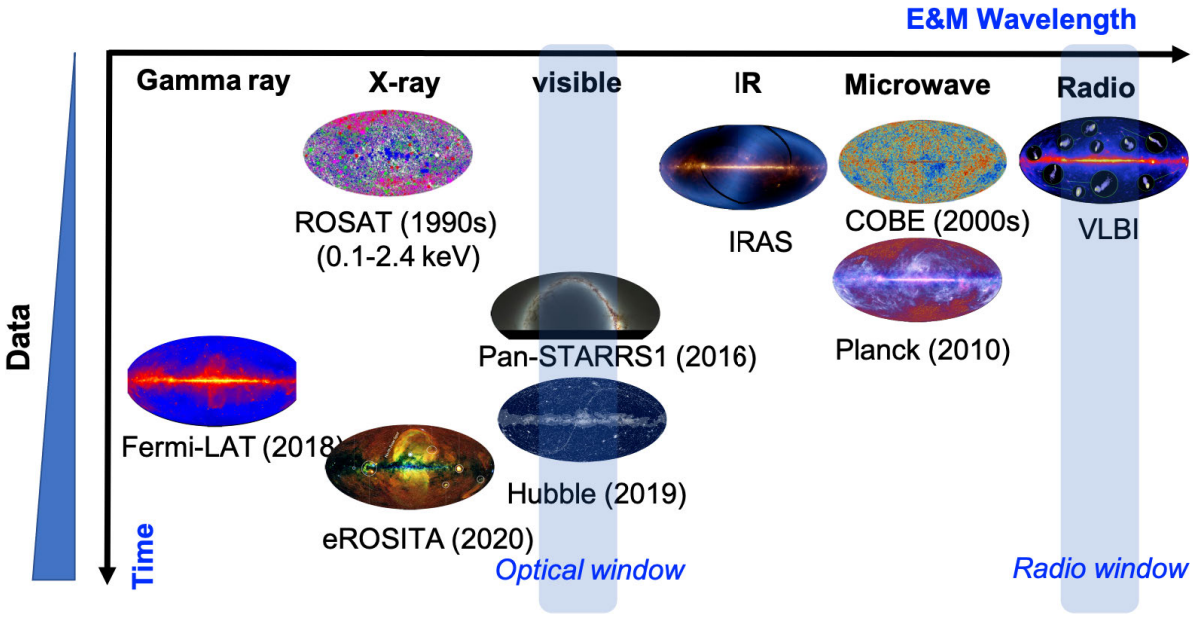

Fig. 24. Examples of the full-sky surveys of the universe using the EM waves. Space-based instruments allow the use of the full EM spectrum. Newer instruments also can produce significantly more data than their predecessors. All these factors combined contribute to the rapid growth in data size and variety.

described [506]. The constellation of satellites has also given rise to hive learning.

Using machine intelligence to enhance instrumentation performance and improve data quality does not have to limit to data reduction, including high-dimensional data reduction. Signal degradation by noise or systematics is a common problem, especially for low signal-to-noise scenarios, such as exoplanet search by measuring light curves [507]. In addition to intrinsic instrument and detector electronic noise, statistical noise from the small flux of photons, external noise, or systematics may include instrumentation jitter [508], stray star light, and cosmic ray background. An ensemble of Bayesian NN called plan-net produced more accurate inferences than a random forest approach [509]. The improvements in accuracy and uncertainties led to higher resolution spectra and physical properties of the atmosphere. Improvements in instrumental resolution can also require more sophisticated models for data interpretation. An unsupervised learning model called ExoGAN [510], which combined a GAN with semantic image inpainting [511], has reduced data processing time from many hours to minutes or faster, with a factor of several times in speed improvement. ExoGAN could also be retrained for other instruments. Image inpainting belongs to a class of methods for filling in missing or damaged regions in images. Inpainting can, therefore, also be used to restore images corrupted by instrument artifacts, remove undesirable objects, such as bright stars and their halos, and preprocess the Fourier or wavelet transforms [512]. Some space instruments may be too large to fit into a launch vehicle. CNN has been used to create a virtual "super instrument" for monitoring extreme UV solar spectral irradiance [513]. The virtual VUV instrument has now been in use as part of a Frontier Development Laboratory project for forecasting ionospheric disturbances and filling in the missing data from broken sensors.

[Zhehui Wang]

C. Space Weather Prediction

Due to the tremendous physical scale, high temperature, strong and dynamic magnetic and velocity fields, and its proximity to Earth, the Sun is regarded as an ideal plasma lab. In addition, the Sun is the source of SWx, which is defined by the transients in the space environment traveling from the Sun, through the heliosphere, to Earth. In the recent decade, the difficult task of understanding and predicting violent solar eruptions and their terrestrial impacts has become a strategic national priority, as it affects the life of humans, including communication, transportation, power supplies, national defense, space travel, and more. Its importance is highlighted by the Promoting Research and Observations of Space Weather to Improve the Forecasting of Tomorrow Act (PL 116-181) passed by the U.S. Congress in 2020. Advances in SWx research and forecasting have been made in recent years thanks to a great diversity of observations from state-of-the-art instrumentation from both ground and space. However, due to increasing spatial and temporal resolutions, researchers are facing tremendous challenges in handling massive amounts of data, especially for operational near-real-time utilization. For example, the flagship solar physics mission, SDO, produces multiple TBs of data daily. This task becomes more demanding as new facilities probe the rapid dynamics of physical processes at some of the fundamental scales. Two important areas of using ML tools to address these challenges are given in the following, which can benefit solar and SWx physics significantly.

Extracting Information Efficiently From Large Volumes of Data in Near Real Time: A required step of understanding magnetic field evolution prior to the onset of solar eruptions is to derive high-resolution vector magnetic and velocity fields quickly with high precision from spectroscopic observations. Scientists routinely use standard methods, such as the ME Stokes inversion to deduce the three components of vector

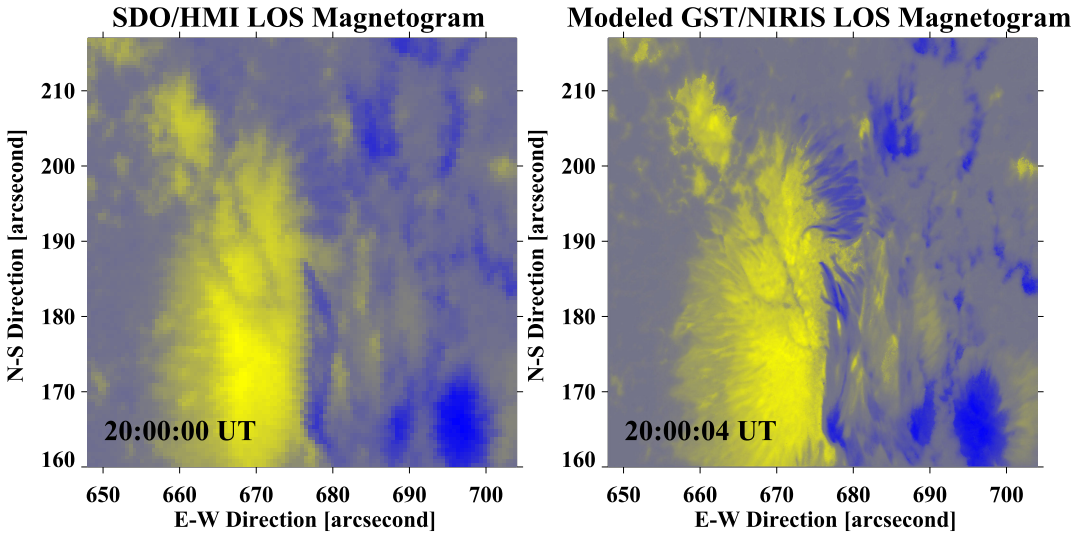

Fig. 25. Left: SDO/HMI measurements used as independent reference data obtained from the ME Stokes inversion tool developed by the HMI team. Right: inverted GST/NIRIS LOS magnetic field strengths derived by our CNN model for the same time (20:00 UT on June 25, 2015) and the same FOV. Magnetic structures look similar, while the GST-inverted magnetic map has about five times better spatial resolution (modified from [514]).

magnetic fields, Doppler shifts, and other plasma parameters. However, such inversion attempts do not always produce physically meaningful results, especially when Stokes profiles are complicated. Furthermore, the ME inversion for large datasets can be quite time-consuming. A Stokes profile can be modeled as waves, and a CNN is suited for capturing spatial information of the waves [514]. Fig. 25 (left) presents some results obtained from the CNN model. The Stokes inversion appears to be quite successful: the ML method is ten times faster than the ME technique with much-reduced noise [514]. Another example of information extraction using ML is SolarUnet [515] that identifies and tracks solar magnetic flux elements or features. The method consists of a data preprocessing component, a deep learning model implemented as a U-shaped CNN for fast and accurate image segmentation, and a postprocessing component that prepares tracking results. This method can be extended to identify and track various other solar and geospace features in large volumes of data.

Predicting Solar Eruptions and SWx Effects Using ML: The solar and SWx community targets predicting solar eruptions and SWx effects, namely, flares, CMEs, solar energetic particles (SEPs), and geomagnetic storms in near real time. The predictions use near real-time ML-processed data, some of which are described above. The predictions can be implemented from both empirical and physical aspects, which are complementary. Physical prediction relies on advanced physical modeling. For empirical prediction, ML becomes vitally important. For example, researchers utilize multiple magnetic parameters for flare prediction, including kernel-based regression analysis [516], ordinal logistic regression combined with SVMs [517], [518], [519], the random forest algorithm [520], ensemble learning methods [521], and LSTM networks [522], [523]. Liu et al. [524] demonstrated the feasibility of using recurrent NNs (RNNs) to predict CMEs. In addition, we noted the success of using CNNs in predicting geomagnetic storms [525]. This research can be advanced in two directions: 1) applying DNN models to perform multiclass prediction including the use of rich spatial–temporal information from ML processed time series of 2-D and 3-D images instead of derived magnetic parameters used in the previous studies and 2) adopting a combination of NNs and statistical methods that innovate on top of off-the-shelf ML algorithms to accommodate the complexity of flaring mechanisms. The second direction will not only benefit SWx prediction but also introduce novel methodological and theoretical challenges to the foundations of data science.

[Haimin Wang and Jason T. L. Wang]

D. Transfer Learning to Improve Historic Data

Modern solar observations provide unprecedented spatial resolution, sensitivity, and wavelength coverage. Solar and SWx research often rely on analysis of large examples of eruptions in the past. Therefore, it is important to use advanced ML methods to improve these historic data. Here, we present two examples in this direction.

Kim et al. [526] generated farside solar magnetograms from the Solar Terrestrial Relations Observatory (STEREO)/EUV Imager (EUVI) 304-Å images using a deep learning model based on cGANs. This opens an avenue of research to train an ML model using one kind of data and apply it to the other kind through transfer learning. For example, in the past, Halpha, CaK, and white-light data are available for over 100 years, while vector magnetograms are routinely available for ten years. The method above demonstrates the feasibility of creating vector magnetograms, which are extremely important for SWx research, from historic data.

The second example is related to the resolution improvement of historic data. The new observations can achieve a spatial resolution of around 100 km, while historic data had a resolution of no better than 1000 km. There is a need to improve the resolution of existing data to disclose the dynamic physics of solar active regions. Such a study has been

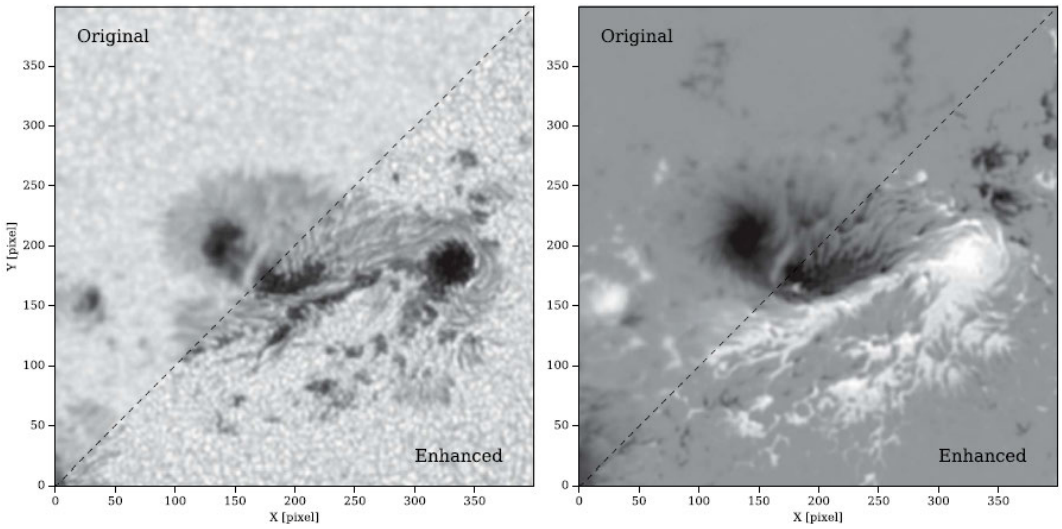

Fig. 26. Example of a CNN applied to (left) intensity and (right) magnetogram of the same region [527]. The FOV is divided into two halves. The upper half shows the original HMI image, without applying the NN. The lower half shows the enhanced image obtained by applying the NN to the original image. The original image was resampled to have the same scale as the network output. Figure credit: [527]. Reproduced with permission ©ESO.

demonstrated by using the Hinode-HMI/SDO data pairs with a CNN in [527], as shown in Fig. 26. Hinode's resolution is five times better than that of SDO. Future work can be extended to improve the resolution even further using observations from large-aperture telescopes.
[Haimin Wang and Jason T. L. Wang]

E. Surrogate Models of Fluid Closures Using ML

Many space plasmas can be described by a fundamental kinetic equation for microscopic descriptions or a set of fluid moment equations for macroscopic statistical descriptions. The traditional tradeoff of solving a set of fluid equations instead of a kinetic equation is generic accuracy verses practical computability. Direct simulation of physical processes on a kinetic level is still prohibitively expensive. Any system of moment equations suffers from the "closure problem": accurately capturing the behavior of an infinite-dimensional kinetic physical system via a few simplified equations. The problem arises when deriving fluid equations through the chains of moment equations for kinetic theories. The resulting lower order moment equations always contain a higher order moment. To truncate the moment hierarchy, a proper closure is, thus, required to approximate this higher order moment from existing lower order moments for microscopic descriptions, which is conventionally constructed by phenomenological constitutive relations.

Fluid moment closure hierarchies for kinetic theories are relevant to a wide range of scientific areas of research, including fluid dynamics, plasma physics, neuroscience, radiative transfer equation, and so on. In plasma physics, the widely used the Spitzer–Harm closure [528] and, similarly, the Braginskii closure [529] consider a strongly collisional plasma and predict heat flux $q \propto \nabla T$, both of which lack kinetic effects and start to break down when the particle mean-free-path approaches the characteristic length scale (i.e., in weakly collisional regime). Well-known closure models, such as the Landau-fluid closure model (or specifically the Hammett and Perkins model [530]), can efficiently incorporate certain kinetic effects within fluid models, such as wave–particle resonances. The Landau-fluid closure describes the nonlocal kinetic response of the heat flux to a temperature profile that has significant spatial variations on length scales that are smaller than the microscopic collisional mean free path. Over the years, the Landau-fluid closure has been extended to collisional [531], [532], [533], magnetized plasmas [534], and with dynamic perturbation [535], [536]. However, implementing Landau-fluid closures to high-performance fluid codes is numerically challenging as they are usually complex functions with both frequency and wave vectors in the Fourier space [531], [535], [536].

Riding on the rapid development of ML [255], ML moment closures for accurate and efficient fluid moment simulations have made significant progress recently. The fidelity of the ML surrogate models has been progressively increasing with the aim of reducing the computational cost and capturing the macroscale behavior of the system but uses only the microscale model to achieve efficiently integrated multiscale simulations, ranging from learning some complex moment closure functions [51], [52], [537], the learned multimode (LMM) closure from kinetic simulation data [538], and learning the calculation of the fivefold integral collision operator in the BE [539] to learning uniformly accurate surrogate hydrodynamic models for kinetic equations [540]. The ML moment closures have been used for accurate and efficient simulation of polydisperse evaporating sprays [541], [542], for the radiative transfer equation [543], and for the moment system of the BE [544]. Miller et al. [50] pursue encoder–decoder NN for solving the nonlinear Fokker–Planck–Landau collision operator in XGC. In [545], the surrogate models have been trained for integrated simulations for the calculation of the core turbulent transport fluxes and the pedestal structure.

Two novel applications of ML techniques to Landau-fluid closures in plasma physics were recently published [51], [537]. In these new studies, the researchers explored how well three

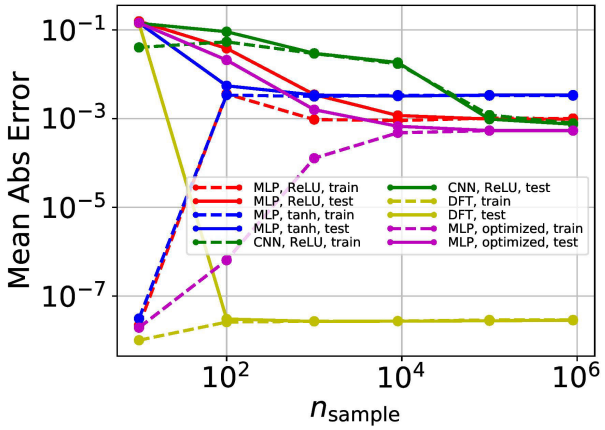

Fig. 27. Mean absolute error versus the number of training samples. Dashed and solid lines denote training and testing errors, respectively; red, blue, and green lines represent MLP with ReLU, MLP with tanh, and CNN with ReLU as the combinations of network and activation function, respectively. The yellow line represents discrete Fourier transform results, while the purple line represents the result of an optimized Bayesian model. Reproduced from [51] with the permission of AIP Publishing.

different types of NNs could reproduce the kinetic Landau-fluid closure. The three networks employed were MLP, CNN, and two-layer discrete Fourier transform. They found that, with appropriate tuning and optimization, all three types of NNs were able to accurately predict the closure, while other existing simplified closure models could not yield the same accuracy at an equivalent computational speed. According to the publication [51], Fig. 27 illustrates the Mean Absolute Error (MAE) of three distinct neural network types, i.e., MLP, CNN, and DFT. These neural networks employ different activation functions and undergo training using varying sample sizes (n_{sample}) of training data. These studies reveal that achieving optimal network performance requires a minimum size for a training dataset. It is also noteworthy that MLP necessitates a minimum number of neurons in its hidden layers, equivalent to the degrees of freedom in Fourier space, despite the input data being fed into the configuration space. Among the three models examined, DFT exhibits superior performance for clean data, possibly attributed to the presence of a simple Fourier expression for the Hammett-Perkins closure. However, it proves to be the least robust when confronted with input noise. Using this new approach, fluid simulations enabled by deep learning, with complicated spatiotemporal closure functions predicted by the NNs, were, for the first time, shown to give the correct Landau damping rate for a wide range of length scales. These results offer a promising pathway to capturing complex phenomena associated with microscopic physics, which is still computationally efficient and accurate when applied at the macroscale.
[Xueqiao Xu]

F. Magnetic Reconnection

The amount of observational data and simulated data relative to space plasma physics is growing exponentially: as more computational power becomes available, on the ground or in situ in space, more data are being generated. This ever-growing data availability is now met with a growing use of ML tools able to consider amounts of data that even a large team of human researchers could not process.

Generally speaking, ML for data processing can be subdivided into two types.

First, *supervised ML* tools are designed to replace tedious well-known steps of processing with automatic tools. The typical scenario is that of taking a large but manageable set of cases and processing them the human way labeling each dataset according to our understanding of it. The machine can then learn from this dataset and replace the humans for the task. The archetypal case is that of image recognition, one of the greatest successes in ML. Since 2015, tests have shown ML tools surpass humans in the accuracy of image recognition [546]. In terms of speed, there is obviously no contest. The same techniques can be applied to analyze scientific data transferring human skills to the machine. However, this approach presumes that we already know how to analyze the data, and we simply want to transfer this knowledge to a machine.

The second approach is that of *unsupervised ML*. In this case, different methods of ML are applied in ways where the machine learns on its own how to treat the data. The central idea is to deploy or design a method where the machine arrives at a reduced description of a complex dataset, and then, the human scientist investigates the reduced description attempting to make sense of it in light of our understanding of nature. The archetypal example is that of classification. The machine can sort all cases into a number of classes, where the number can be preset or can be an unknown of the process itself. In the end, the challenge for the researcher is to understand what the meaning of the different classes is. With this challenge comes the opportunity to discover something new and unexpected.

The research is progressing at unparalleled speed in both categories of ML tools. Fortunately, the meagerly funded space science community can benefit from the general growth in ML tools developed for other applications. Some specific tools have been developed to make the progress in ML available to the community of space scientists. The aim is to vulgarize the more esoteric aspects of ML and make them accessible to scientists whose background is in space and not in computer science. We mention here the project AIDA that has precisely this goal: www.aida-space.eu. The AIDA project takes some of the state-of-the-art ML tools and applies them to typical use cases common in space science. Each use case is documented detailing how to use ML tools in a step-by-step process that is aimed at training non-ML experts.

Space science has a peculiar constraint unique to its nature: much of the data generated in space cannot be transferred to the ground due to the limited telemetry. In other applications, the data can always be stored or at least processed; in space, the memory onboard cannot be always transferred in its entirety; and only a portion can be downloaded to Earth. This limitation opens a new opportunity for ML deployment in space so that the data can be processed on board, and only the outcome of the analysis needs then to be transferred to the ground. This is a pioneering new possibility, and great challenges need to be overcome because the processors used in space are much less powerful than those used on the ground due to the intense space radiation environment. ML tools are highly computing demanding making their deployment in space a great topic of research.

Finding Reconnection: Reviewing the explosively growing area of space applications of ML is beyond our scope here and is a futile exercise as many new developments will be published, while this article is being processed. We focus instead on a few examples that provide a view of the type of activities that ML can take over. We focus then on only one well-known but very difficult task: identifying reconnection regions.

Reconnection [547] is a process that converts magnetic energy into kinetic energy. Its characteristic feature is the breaking and reconnecting of magnetic field lines, giving it its name. Recognizing reconnection is not as simple as 2-D cartoons might seem to imply.

In 2-D, one can consider the problem more or less solved in terms of making a definite determination of where reconnection happens: based on the out-of-plane vector potential, the null points of the in plane magnetic field can be characterized using the Hessian matrix as o-points and x-points giving an unequivocal answer [548]. However, there are two problems. First, finding nulls and computing Hessian matrices require complete spatial information, something that we have only in simulations, but we do not in experimental data. Often in space, we know quantities only at one location (or a handful, in the case of multispacecraft missions). Second, nature is 3-D, and there is no equally rock-solid definition of reconnection in 3-D. There are situations where experts might argue endlessly on whether there is or is not reconnection.

This provides a unique opportunity to apply ML. Let us then review how different data feeds can be used to find reconnection.

Traditionally, reconnection is identified by using a proxy. A review of the different reconnection indicators is provided in [549]. The simplest is finding high-speed jets. Of course, many processes can lead to high-speed jets, and only the expert can combine the analysis of different quantities and arrive at the conclusion that reconnection is really taking place. A more recent discovery is that reconnection is associated with peculiar electron velocity distributions that present croissant-shaped features called *crescents* [550]. An especially convenient way to identify a possible reconnection site is the local measure of the so-called Lorentz indicator based on computing the speed of a frame transformation that eliminates the local magnetic field [551]. An example of this indicator is shown in Fig. 28 where many 3-D reconnection sites are identified in a turbulent region.

This accumulated expertise provides a great opportunity for creating human-labeled datasets to use as training for a supervised ML tool. However, the intrinsic complexity even for a human to decide what is and is not reconnection gives unsupervised ML tools the opportunity for new discoveries. In the following, we explore some methods recently published to identify reconnection with ML.

Identifying Reconnection From Velocity Distributions: VDFs, $f(v_1, v_2, v_3)$, are provided as 3-D datasets by instruments that measure the count of collected particles in situ. Kinetic simulation can provide a synthetic version of the same information. This information can be in different forms: energy angles, 3-D velocity bins, or polynomial expansions (e.g., Hermite and spherical harmonics). The first aspect of this type of data is the overwhelming size. As an example,

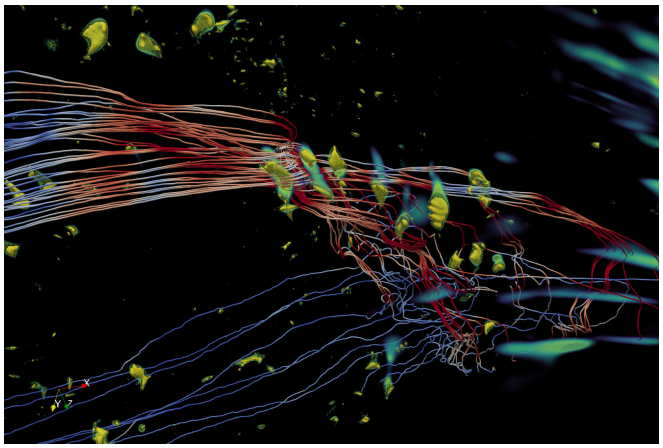

Fig. 28. Indicator defined in [551] identifies many reconnection sites visible in the picture as ghostly yellow-green areas. A group of electron flowlines is shown passing one of these reconnection sites and encountering also others. The flow lines are colored by the intensity of the local electric field that transfers energy between the magnetic field and the electrons, accelerating them and creating a turbulent flow.

a typical modern particle in cell simulation produces TB of distribution data for each time step. At each time step, there are millions or billions of such distributions to analyze. A recent mission, magnetospheric multiscale (MMS) produces in burst mode one distribution every 30 ms (though not all can be transferred to the ground due to limited telemetry). A survey of the literature shows that these distributions are rarely used in their full 3-D complexity, and usually, only very few 2-D reductions of specific instants are studied. The choice is guided by analyzing other quantities that suggest what distribution to study. There is no systematic analysis of all data taken; it is humanly impossible, but not impossible for ML.

Supervised ML can be trained to recognize features like the crescent using a human-labeled dataset, an application of the widely used image recognition software. However, shapes in VDF are more in the imagination of the viewer than an objective feature. VDF, especially in observed data, is highly noisy and structured. A promising approach is to use unsupervised ML. The complexity of a VDF can be classified using clustering methods.

The Gaussian mixture model (GMM) represents a distribution using a superposition of overlapping Gaussian distributions [552]. With this approach, Dupuis et al. [553] showed that reconnection can be associated with a high number of Gaussian beams, with different classes of distributions capable of identifying the inflow and the outflow region of reconnection. The method automatically determines the number of Gaussian beams using information theory criteria that make the best compromise between the efficiency of description (that requires as small a number of beams as possible) and accuracy (that is always higher, and the more beams are used). The method also determines the properties (mean and variance) of each beam. From this ML analysis, physical meaningful quantities can be determined. Especially useful is the determination of the "intrinsic" thermal spread of each beam in the mixture and the "pseudo" thermal speed due, instead, to the relative speed between the different beams in the mixture [554]. Dupuis et al. [553] showed that these physical quantities can be used to determine the electron and ion diffusion region around a reconnection event.

Another approach to the unsupervised ML analysis of distribution functions uses the subdivision of the VDF in nonoverlapping beams or arbitrary shapes. The k-means method [555] can be applied for this task leading to the identification of different populations with different physical origins [556].

Identifying Reconnection in Spatial Data: The quintessential example is the 2-D image: in the case of reconnection, this is a 2-D view of the EM field component or of a plasma moment. It is obvious that this type of data can benefit from the methods developed for image processing.

CNNs [557] can be trained using expert-labeled images. Hu et al. [558] report the example of a dataset of 2000 cases labeled using the expert community via zooniverse.org. The project can be accessed via http://aida-space.eu/reconnection where a tutorial on how to identify reconnection sites is provided. The project is public and unbiased experts helped with labeling. Once the labeled dataset is available, the CNN can be trained to recognize reconnection.

Unsupervised ML can detect reconnection based on spatial information by using the clustering of pixels in the spatial data. Reconnection is identified using the physics properties of the resulting classes. Sisti et al. [559] use the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [560] and k-means [555] to identify current layers with a sufficiently large aspect ratio to flag reconnection.

In principle, 3-D datasets and 1-D fly-through datasets can be treated using similar methods, and future research will likely investigate this possibility.

[Giovanni Lapenta]

G. Challenges and Outlook

Besides the mysterious dark matter and dark energy, the observable universe is known to consist mostly of plasmas. The explosive growth of observational and simulation data is expected to continue from the cosmological scale to terrestrial-size plasmas, which can supply data not accessible to laboratory experiments. The essentially unlimited and heterogeneous data, sophisticated multiphysics models, and, lately, the universal data mining tools, such as deep learning, have ushered in the new *precision* epoch in cosmology, astrophysics, space, and terrestrial science, including plasma physics. On the one hand, the wealth of data allows detailed tests of the existing physics-based models, including the underlying fundamental physics, such as quantum mechanics and general relativity, and fine-tuning of ad hoc parameters in some models. On the other end, such data permits systematic searches for new physics motivated by dark matter, dark energy, neutrino mass, high-energy cosmic rays, and quantum information centered around the blackholes. On the applications front, data science has opened doors to real-time predictions of solar CME and SWx forecast. Data science has already given rise to new disciplines, such as astroinformatics and astrostatistics; it may also provide a generic framework to better integrate plasma-driven physics into the existing models when plasma effects have so far been left out, for example, of the standard model of cosmology. Data science and ML have been successfully or can be used to accelerate all aspects of "scientific data flows" in astronomy and astrophysics, i.e., from enhanced instrumentation and data acquisition to automated feature extraction and classification, hypothesis generation, model construction, modeling, and model validation. Despite their practical prowess and simplicity, ML methods are not fully understood at this time. Seeking a better union between the established knowledge framework of physics-driven models with data-driven models is an exciting new frontier. New results may be anticipated such as in the solving the outstanding problems as mentioned above, development of scientific ML algorithms that will be broadly applicable, and quantitative understanding of uncertainties for more effective predictions and optimization, paving the way toward automated space and astro plasma observations, discoveries, and novel space technologies.

[Zhehui Wang]

VII. Plasma Technologies for Industrial Applications

A. Introduction

Plasma technologies are widely used in industries [561], [562]. One of the largest industrial applications of plasmas is plasma processing for semiconductor devices and other related microelectronics devices, such as displays and sensors. Especially for the latest and most advanced semiconductor devices, the device dimensions (i.e., typical sizes of transistors) are now approaching the atomic size. Therefore, the further miniaturization of a single device is now facing its physical limit and can no longer be expected as a means to pack more devices in a single chip. Instead, further improvement of device performance must be achieved by other means, such as the use of complex 3-D device structures and new materials.

Mass production of such complex devices with atomic-scale accuracy poses enormous challenges in their manufacturing technologies. Plasma etching and plasma-enhanced deposition processes [563] need not only to improve their accuracy in spatial dimensions but also to handle nonconventional materials, such as ferromagnetic metals for magnetoresistive random access memories (MRAMs) and perovskite-type oxides for resistive random access memories (ReRAMs), just to name a few. However, in most cases, the interactions between the newly introduced material surfaces and conventional or newly introduced gaseous species of plasma processing are not well understood, which makes the process development highly challenging and costly. Furthermore, having a variety of choices for surface materials and process conditions increases the complexity of process development, and the exhaustive search for process optimization by experiments becomes prohibitively expensive. One of the possible means to tackle these challenges is to use ML to predict gas phase and surface reactions of plasma processing based on the existing knowledge of such systems.

Other technological applications of plasmas that have attracted much attention from the plasma community recently are those for medicine, agriculture, biology, and environmental protection [564], [565]. Although practical applications

of these technologies at the industrial level are yet to be seen, some of them are considered to be game-changing innovations. As in plasma applications for semiconductor technologies, gas phase and surface chemical reactions play critical roles in plasma processes in these fields, and the exhaustive search for optimizing their process conditions by experiments can be prohibitively expensive as well. This reasoning applies similar to thin film deposition of hard and functional coatings and plasma-assisted catalysis [566]. Diagnostics and modeling are crucially challenged by intrinsic multiscale and multiphysics phenomena, including yet-to-be-revealed nonequilibrium plasma-surface chemical reactions. Moreover, the exploration and discovery of novel plasma and solid phase materials systems, e.g., for energy-efficient gas conversion and synthesis, are a severe limitation. Systematic collection of data and the use of data-driven approaches to make full use of such data are expected to enhance the efficiency of process development and promote (or even enable) transitional changes in these fields.

In this section, we present how such data-driven approaches are used in the semiconductor and microelectronics industries in the following three subsections. From a more academic point of view, examples of data-driven approaches are then presented as new tools to analyze PSIs, plasma simulations, plasma chemistry, and plasma medicine in the subsequent subsections. The final subsection briefly summarizes the challenges and outlook in industrial applications of technological plasmas.

[Satoshi Hamaguchi and Jan Trieschmann]

B. Data-Driven Approaches for Plasma-Assisted Manufacturing in the Semiconductor Industry

With the explosive growth in data creation, estimated to surpass 180 zb by 2025 due to the increasing popularity of the Internet of Things (IoT), there is an unprecedented demand for the storage and processing of large volumes of data. Today's data-centric world increasingly relies on semiconductor manufacturing to fabricate chips with integrated circuits that can realize the data storage and computational capabilities required for harnessing data and AI. Of the hundreds of steps used to fabricate a chip, nearly half use plasma processing. This is because nonequilibrium plasmas offer several benefits over thermal processing, including lower energy barriers to promote surface adsorption, resulting in the reduction of high-temperature requirement for certain materials; ion acceleration toward the wafer due to sheath physics, resulting in directional behavior; and enhanced surface reactions due to presence of neutrals and ions. These plasma effects will lead to better film uniformity, conformality, and roughness control with atomic layer processing. As the semiconductor industry continues to innovate by building chips with smaller feature sizes, the cost to design such chips and the cost to equip fabrication facilities with state-of-the-art process tools required for making these chips have increased dramatically (see Fig. 29). In addition to the cost, the time taken to complete a chip has increased, as more process steps are required to achieve the desired results [567]. Thus, it is paramount to accelerate the design and development time of the plasma reactors, optimize the processes used to create the desired features while improving the efficiency of engineering staff, and provide adaptive control to mitigate uncertainty at the chamber level, the tool level, and the fleet level. Smart manufacturing practices and advances in sensing capabilities and product metrology have created unprecedented opportunities for the semiconductor fabrication equipment industry to improve yield, efficiency, and speed to solutions using data-driven approaches.

Applications of data-driven approaches for plasma-assisted processes in semiconductor manufacturing can be categorized in three interrelated areas (see Fig. 30): design and production of plasma processes, optimization of plasma processes and engineering efficiency enhancements, and adaptive process control and operation. As the complexity of shrinking technology nodes has increased, Moore's law has not been followed in recent years, i.e., the cost reduction per bit in the case of NAND memory has decreased [568]. To overcome challenges in shrinking technology nodes, equipment makers look to build new processes capable of handling new materials at a much more accelerated timescale in order to meet the ever-challenging demands for new applications. Process design optimization using surrogate models has received increasing attention in the semiconductor industry to facilitate design as well as testing what-if scenarios in a resource-efficient manner. In these approaches, cheap-to-evaluate models based on physics-based simulations are used to construct nonlinear relationships between various design parameters [569]. Surrogate models are also becoming increasingly important for constructing the digital twin of a system [570], which allows for developing process design and optimization solutions based on fast surrogate evaluations under different operating conditions. For the design of parts, especially with additive manufacturing that has gained popularity for quick prototyping and making complex designs possible, generative design approaches have proven useful by combining computational design, simulation, optimization, and data visualization. To achieve the most optimal design, an initial design is "evolved" under multiple constraints. Such methods can allow process engineers to analyze various tradeoffs in the design by determining the Pareto-optimal solution under multiple constraints [571]. Another area of growing importance is material identification and characterization for process design. As new chemistries are introduced in the process reactor and different plasma regimes explored, there is a need for new materials in the system to withstand more challenging conditions, such as corrosion, crack, warpage, and thermal creep. Material informatics create new opportunities to select the correct material for the given application and minimize the extensive evaluation cost of materials that may not work in the given conditions. As the industry continues to shrink the technology nodes, equipment makers must constantly add more process knobs to meet the stringent specifications for the layer under consideration, such as deposition or etch rates, uniformity metrics, critical dimensions at desired locations, and other properties (e.g., stress and refractive index). In addition, there are other requirements set at the system level,

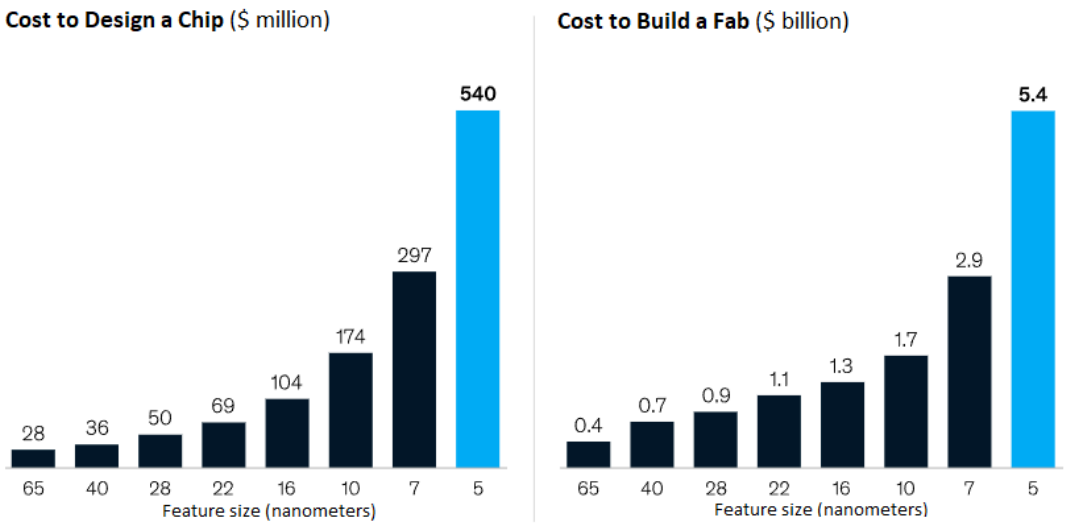

Fig. 29. Cost to design state-of-the-art chips and build the semiconductor fab equipped with the latest process modules to fabricate these chips has grown dramatically with the scaling of the nodes [567].

such as defectivity, sustainability, throughput, and various cost constraints. Data-driven approaches have shown promise for speeding up process development by optimizing the recipe setpoints for the ideal film, as well as improving the efficiency of process engineering given the vast design space for recipe optimization. In order to assess the outcome of a process, automated image analysis capabilities are developed to measure dimensions of interest [572] and to improve the quality of the image [573]. Recipe optimization is performed not only based on current data collected but also prior knowledge developed using ML algorithms [574], [575]. To accommodate for upstream film variations and variations in tools, real-time analysis methodologies for endpoint detection are developed [576]. Data-driven approaches are used to characterize defects automatically by assigning classes to wafer map patters, morphology, and chemical spectra [577], [578], as well as detecting and triggering autoclean routine to improve productivity by minimizing the failures caused by these defects. In addition to process challenges at the unit process level, data-driven approaches can be used for optimizing the entire process flow, allowing engineers to study the sensitivity of a particular layer and build appropriate tradeoffs to achieve their desired product [579]. With the proliferation of more sensors in semiconductor manufacturing equipment, new opportunities have also been created for APC, including operation analytics for online equipment health monitoring to enable predictive and prescriptive maintenance of processing tools [580]; soft sensing and VM for enhanced process monitoring and fleet matching for yield improvement; FDC for timely diagnosis of potential process anomalies [580]; feedback control strategies, such as predictive control and run-to-run control for accommodating process-to-process variability, high product mixes, and process dynamics; and predictive scheduling for improving the overall fab productivity by minimizing idle tool time [581].

A fundamental requirement for the success of data-driven approaches for the design, optimization, and control of plasma-assisted processes in semiconductor manufacturing lies in the interpretability of the data-driven models. As the number of process tuning knobs increases to meet challenging demands for scaling needs in the industry resulting in over 10^{23} possible permutations of recipes and the continued demand to match system states across a fleet of tools with more than 10^{100} possible states, quantum computing can play a transformative role in the years to come to facilitate AI applications involving complex high-dimensional data or discrete/combinatorial optimization. To this end, there is a need for further advances in data management, better algorithms, resilience in cyber–physical systems, and innovation in advancing compute and storage of data. Other emerging applications of AI include automated visual inspections of parts, supply chain optimization, and augmenting human capabilities through concepts of extended reality. The field of Industry 4.0 is just beginning for the semiconductor industry and will rapidly grow with the goals of accelerating the time-to-market of new processes and products, as well as the relentless drive for greater productivity and yield.

[Kapil Sawlani and Ali Mesbah]

C. Plasma Information-Based Virtual Metrology

The necessity of a high-value process strategy for the semiconductor- and OLED display-manufacturing industries, which requires ultrafine plasma process technology, is ever-increasing to achieve an increase in device production throughput. To manage the process results efficiently in this ultrafine-scale plasma process, an automated control system, such as FDC and APC logics, is needed. It requires the development of a VM model, which directs the process control. The prediction accuracy of the VM is a crucial component of the performance of the FDC or APC system [582]. The VM was developed from classical chemical processes to predict process results based on the statistical analysis of monitored sensor data. According to Cheng et al. [583], VM is a method for estimating the manufacturing quality of a process tool based on data sensed from the process tool and without physical metrology operations. Therefore, the

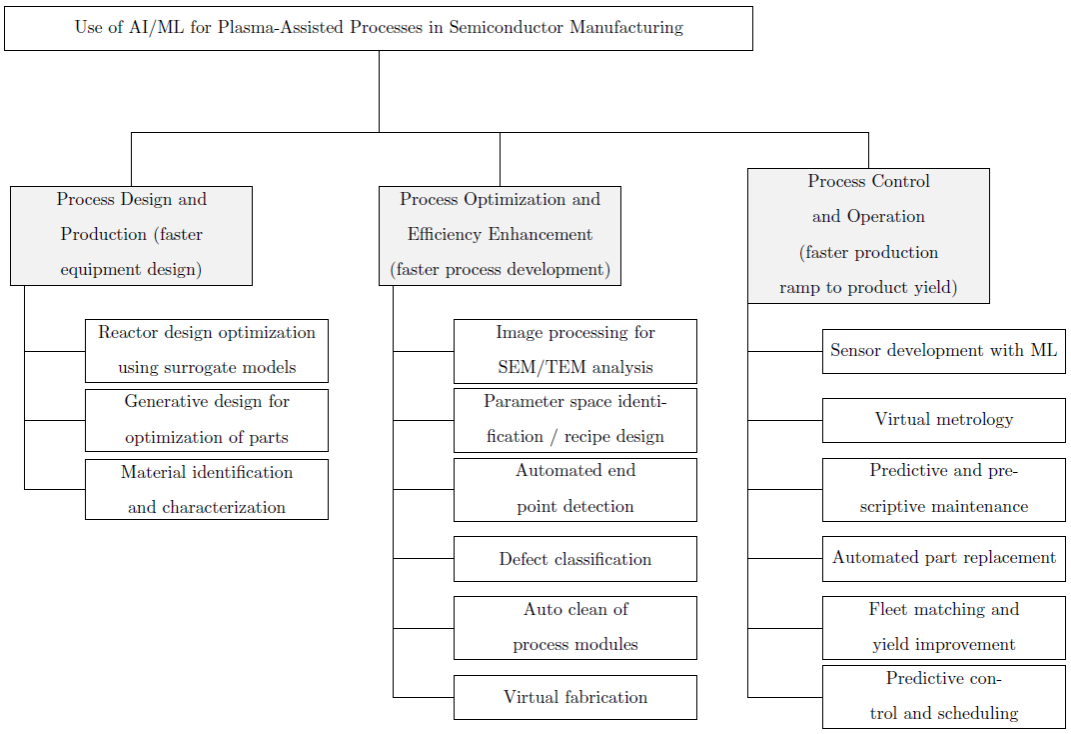

Fig. 30. Overview of the applications of AI/ML for the design, development, and operation of plasma-assisted processes for semiconductor manufacturing, toward accelerating the time-to-market of new processes and products for the consumer electronics industry via smart manufacturing practices.

development of the VM for plasma processes was likewise initiated from statistical approaches. Development of the statistically established VM began with correlation analysis of the variables with process results. To this end, various ML models are applied to VM modeling [584], [585]. However, this statistical-method-based VM has shown unsatisfactory prediction accuracy when applied to numerous cases of plasma-aided processes [586].

To develop high-performance VM models, the efficient containment of the "good information" representing parameters, that is, the parameters representing the process plasma state, is needed rather than the direct application of the ML methodologies. These parameters should efficiently mediate between state variables monitored from the sensors and performance variables, and the specificity of a plasma-assisted process mechanism should be considered [582], [586]. Lieberman discussed the importance of the reactions in the plasma volume, sheath, and target surface in terms of the progress of the process reactions, such as etching, deposition, sputtering, and ashing [561]. These overall reactions are strongly correlated with each other and governed by the properties of the process plasma. Therefore, to develop the VM for plasma-assisted processes, the process PI, including parameters representing the reaction properties in the plasma based on the volume-sheath-surface reaction mechanism, is required. To attain this concept of the VM for plasma-assisted processes, new parameters called "PI" were introduced, which are applicable as powerful variables in 2015. They have been used to predict various process performances, such as etch rate, deposition rate, defect particles, etching profile, deposited thin film quality, and spatial uniformity of the processed results. They have been applied to the control and management of the OLED mass production lines last six years [586], [587], [588], [589], [590], [591], [592].

Fig. 31 compares the predicted etch rates for the C_4F_8-based plasma-assisted silicon oxide etching process with the measured etch rates. To test the performance of fundamental ML methodology, a PCR-based VM to predict the etch rate was modeled. 79 EES sensing variables from the power, pressure, gas, chiller, heater, and exhaust system and 1670 parameters from the OES intensities were combined into the PCs and were regressed, as shown in Fig. 31(a). The correlation coefficient between the measured etch rate and VM result was $R^2 = 38.8\%$. By adopting the PI parameter of b-factor measured by the OES data as a PC into the PCR-based VM model (PCR_b), the correlation coefficient between the measured etch rate and VM result was $R^2 = 57.2\%$, as shown in Fig. 31(b). Here, the b-factor is the shape factor in the generalized form of the EEDF, $f(\varepsilon) \sim \exp(-c\varepsilon^b)$ with the coefficient c and electron energy, ε [593]. The distribution shape varies from the well-known high-energy tail developed Maxwellian distribution with $b = 1$ to the curtailed Druyvesteyn distribution with $b = 2$ in general [593], [594]. Finally, by adopting presheath potential and surface passivation representing PI parameters synthesized from the monitored OES and EES data, the prediction performance of the VM was enhanced to $R^2 = 96.9\%$, as shown in Fig. 31(c). These results imply that selecting the variables according to the reaction mechanisms in the process plasma is important to achieve the performance of the VM for plasma-assisted process monitoring. PI-VM modeling, especially includes the characteristics of the EEDFs, can be an efficient method to include the information about

the process state into the VM model and is useful to obtain high-performance of the VM applicable to the real field [586].

Developed PI-VM algorithms were applied to the mass production line of the OLED display manufacturing to solve four kinds of problems that occurred in the real field: The defect particle caused process fault prediction [587], root cause analysis of the high-aspect-ratio contact (HARC) etching process faults [588], the management of the mass production discontinuities with a proper application of the in situ dry cleaning (ISD) [589], and the microuniformity problems in the process results [590]. These PI-VM models optimized for each issue have shown enough prediction accuracy to apply for the long-periodic mass production running. Therefore, by applying the PI-VM models to the control of the OLED display manufacturing processes, overall production yields were relevantly progressed in the last six years. Especially, the mass production management referring to the discontinuity qualifying PI-VM, effective application of ISDs, and yield loss was successfully suppressed by about 25% for 42 process chambers in the fab.

[Seolhye Park, Jaegu Seong, and Gon-Ho Kim]

D. Data Management in Manufacturing

The semiconductor device technology is now far below the 10-nm critical dimension in manufacturing with its sights set on 2 nm. Successful device scaling, historically driven by lithographic patterning, is now driven by plasma etch. The tight process control afforded by modern plasma sources has enabled scaling. Moving past 3 nm requires even tighter levels of control. Tighter control translates to atomistic control. The "smart manufacturing" initiative is a means to meet this end. Its purpose is to enable adaptability in plasma process tools [595] facilitating reliable and accurate APC systems delivering nanometer-scale precision. APC in the form of run-to-run control enables continuous process tuning. Process output parameters are monitored by metrology tools potentially including VM models. Adjustments to process tuning knobs are dictated by control models' responses to measured deviations between the process outputs and control limits [596]. The control model itself is a barrier to achieving accuracy and reliability. The relationship between control parameters and the surface processes to be controlled is complex even for nominally simple plasma processes. Numerical plasma models require significant computational resources, making them difficult to use in a control context directly. The accuracy and reliability of theory-based numerical models are also an issue. Despite decades of progress, numerical models are difficult to validate. AI/ML technologies enable high-accuracy VM model prediction by capturing variations originating from complex plasma and surface reaction phenomena without reliance on physical assumptions [597], [598], [599]. While AI/ML algorithms are attractive options for implementing high-accuracy VM models, there are some disadvantages. They require a large number of training datasets and lack the inference capability needed to link the predicted variations to their root causes. Plasma diagnostics paired with appropriate sensor technologies can reduce the advanced data processing load while maintaining or even improving model accuracy. This is done through direct extraction of variables that should correlate with target metrics via theory or model [591], [600], [601], [602], a methodology termed "data quality improvement." Data quality improvement relies heavily on the selection of appropriate in situ sensors, which, in turn, requires specific plasma domain knowledge. Fig. 32 shows how plasma domain knowledge is incorporated to improve data quality for building VM or classification models.

1) *Type 1 Domain Knowledge—What to Measure:* Plasma variables such as the ion flux, neutral flux, and deposition rate defined by phenomenological surface reaction models or interpretation of postprocess profile formation using theoretical mechanisms [603], [604], [605].
2) *Type 2 Domain Knowledge—How to Measure:* Noninvasive in situ sensors to measure plasma variables that are derived from type 1 domain knowledge.

Preprocessing measured data with the interpretive functions afforded by type 1 and 2 domain knowledge is key. Preprocessing not only involves the conversion of raw data into plasma variables but also is important for error removal [606]. OES provides a good example of useful preprocessing. OES intensities most often vary during production runs due to varied transmittance through the view window caused by film deposits. The intensity variation is independent of the plasma condition, hence registered as an error. One way to reduce error is to normalize the OES spectra by the OES intensity at the chosen wavelength [607]. Both sensor and preprocessing method selection (i.e., data quality) can be evaluated by benchmarking VM model performance with versus without the studied sensor data added to other default sensor datasets. The following example illustrates one such evaluation. TOX flat wafer etching rates were varied by installing various combinations of new and used chamber parts. OES and RF sensor data were collected during the etching of TOX wafers. VM models were constructed to predict TOX etching rates using exhaustive least-squares regression with preprocessed OES and RF sensor data. The number of terms in the VM models was limited to below 4 to enhance the sensitivity to data quality. The impact of RF sensor data on the VM model performance was evaluated using CV scores calculated as an average of R2 values from each fold of the fivefold CV. Fig. 33 shows CV scores of all VM models generated from exhaustive least-squares regression with two datasets—OES only (OESonly) and OES with RF sensor data (OESwRF). del_RF represents models that include RF sensor data, i.e., del_RF = OESwRF − OESonly. As can be seen, significantly improved high-performance VM models were generated with RF sensor data. The results illustrate overall data quality improvement by adding RF sensor data with preprocessing.

Successful development and deployment of APC to meet the tight control limits demanded by sub-10-nm technology plasma processes require AI/ML to be augmented by improved data quality. Data quality improvement with domain-knowledge-aided preprocessing was illustrated in this article for the simple example of TOX etch. The availability of noninvasive in situ sensors for plasma and surface parameters is an issue. Therefore, the concerted development of these sensors will be an area of emphasis for the industry. An area of particular importance for sensor development is drift-free molecular species measurement during production

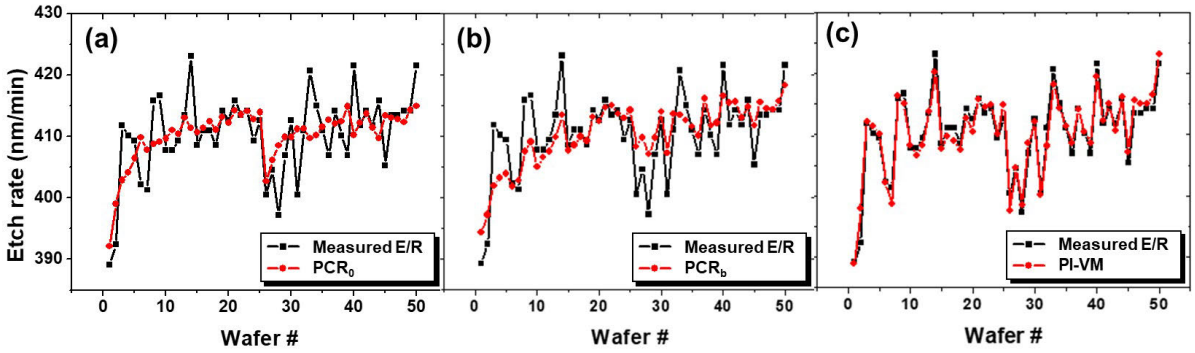

Fig. 31. Comparison of the measured etch rate and predicted etch rates of 50 wafers with (a) basic PCR model, PCR_0, (b) with the adoption of b-factor, PCR_b, and (c) fully PI variables adopted PI-VM.

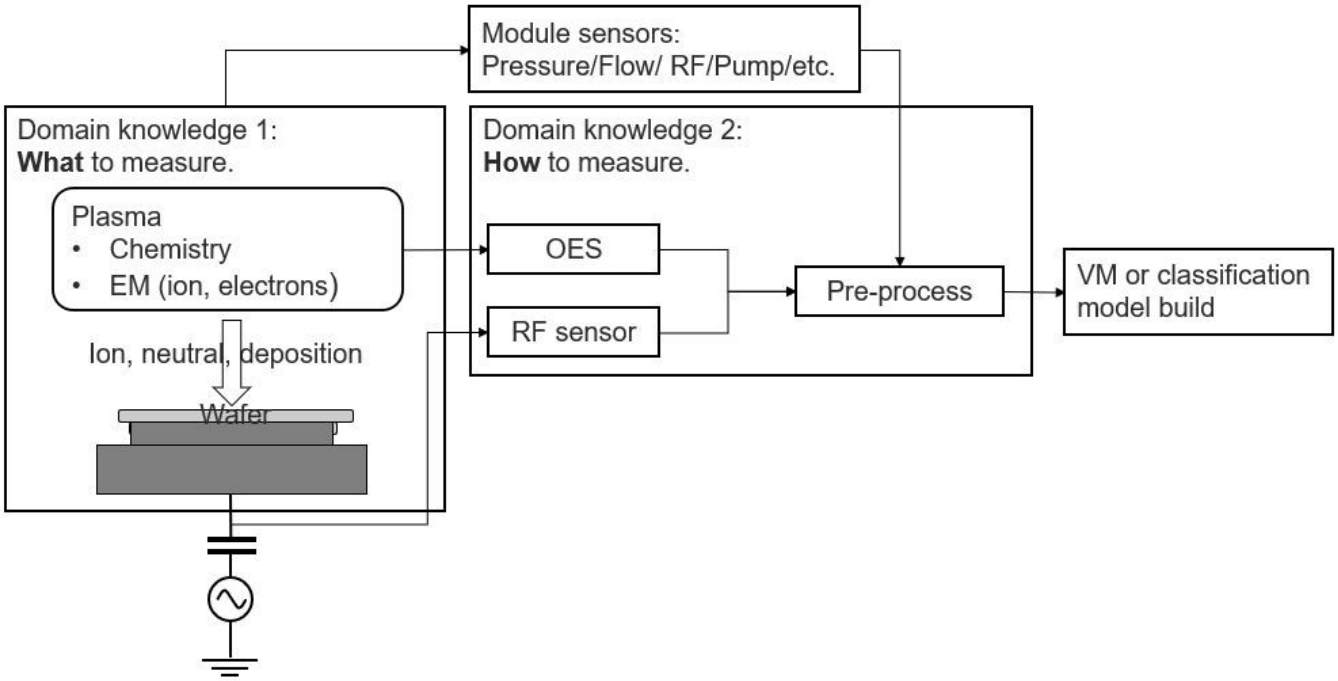

Fig. 32. Schematic of a generic plasma process tool and in situ sensors.

runs. The ubiquitousness of pulsing in plasma processing poses additional challenges and opportunities. Faster data rates are needed for in situ sensors to be able to characterize the important aspects of complex pulse trains.
[Jun Shinagawa and Peter Ventzek]

E. Data-Driven Analysis and Multiscale Modeling of Plasma–Surface Interactions

The majority of technological (and fusion) plasmas are subject to interactions with bounding surfaces. It is essential for plasma processing but is typically considered inevitable in fusion devices with harsh plasma environmental conditions. The role of PSI is generally bidirectional: 1) particles from the plasma volume may cause modification of surface material (e.g., etching/deposition, chemical reactions, structure, and phase transition) and 2) the surface may influence the plasma volume through particles emanating from the walls due to related physical phenomena (e.g., sputtering, chemical reactions, and secondary electron emission). This feedback implies that PSI cannot be considered independent but consistently coupled. It requires a bidirectional relation following 1) and 2) between plasma and surface conditions at multiple time and length scales.

Several data-driven approaches have taken PSI into account macroscopically for plasma process control. They used PI-VM for plasma etching with experimental data sources [589], as well as MPC for atmospheric pressure plasma dose delivery [608] or reactive magnetron sputtering close to mode transition [609]. In contrast, theoretical multiscale analyses of technological plasmas have been restricted to classical modeling and simulation (e.g., combining MD, BCA, and kinetic Monte Carlo models at the atomic level [610] or unidirectional coupling the reactor scale to the feature scale in complex capacitive RF plasmas [611], list not exhaustive).

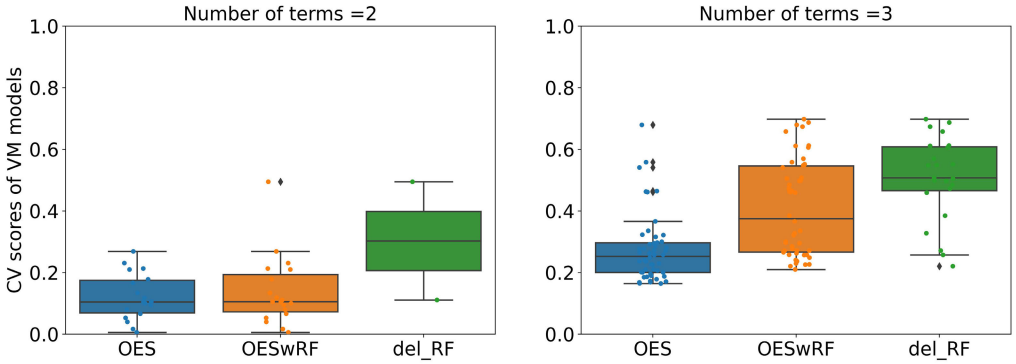

Fig. 33. CV scores of the VM models built with OES only (blue) and OES with RF sensor (orange) datasets. The VM models that include RF sensor data were grouped into del_RF (green). Each data point represents the CV score of VM models generated from the exhaustive least-squares regressions. The CV scores of VM models were significantly improved when RF sensor data were included, indicating the improvement of data quality with RF sensor data.

So far, the focus has been on route (1) toward the surface. The physical complexity and the computational expenses of atomic-level PSI models restrict the return route (2) toward the plasma. If considered, the latter is often reduced to simple analytical approximations. This may be a severe limitation when complex surface chemical dynamics need to be captured accurately (e.g., plasma-enhanced catalysis or atomic layer deposition/etching) [566], [612]. Rigorous treatment of PSI is, moreover, required if emission from the surfaces may significantly influence the plasma discharge itself. Data-driven PSI models may capture these dynamics at a nonprohibitive computational effort.

The procedure of establishing corresponding data-driven PSI models may differ in detail, but a rather generic scheme is outlined in Fig. 34 as follows: 1) data retrieval from measurements or simulations; 2) feature selection through identification of reliable physical descriptors; and 3) establishing of a regression relating descriptors (model inputs) to targets (model outputs), possibly with uncertainties (systematic or statistical). Each step is indispensable and could require several iterations, depending on the utilized procedure.

While a manifold of surface interaction phenomena may be considered, data-driven approaches to PSI have focused on the analysis of sputtering due to energetic particle impingement (e.g., ions, fast neutrals, and photons). While its fundamental nature may seem simple, it poses a nontrivial problem due to the nonlinear dynamics of the collision cascade in the solid subsequent to interaction. In the absence of a widely applicable analytical description from first principles, data-driven approaches have been suggested to establish generalized relations inferred from the data [54], [613], [614], [615].

1) The amount of data accessible for data-driven PSI modeling of sputtering varies significantly. For instance, a well-defined dataset of experimental sputtering yields for different ion–solid combinations is publicly available [616] and has been successfully used [613]. These are limited to integral information, however, eliminating the details of the flux and energy distributions emanating from the surface. Energy and angle-resolved data from MCSs (with BCA) provide a compromise between computational costs and physical fidelity [54], [615]. Accurate physical simulation data at the atomic level (e.g., MD) are typically sparse and may require data augmentation because the computational cost to obtain large datasets imposes a significant challenge.
2) The process of defining independent features depends on the requirement of physical interpretability. Given a set of possibly correlated physical variables, a subset of descriptive physical parameters has been devised by hierarchical clustering and corresponding descriptor analysis for sputtering yield regression [613]. In contrast, the concept of variational AE ANNs has been applied to provide a descriptive set of latent parameters at the cost of complicated physical interpretability [615]. UQ of physical descriptors using Bayesian analysis has devised confidence bounds in the inference of the sputtering yield, suggesting a more accurate surface binding energy [614].
3) The ultimate goal is the design of a PSI regression task. While kernel ridge regression was successfully applied for the inference of sputtering yields as a function of the incident particle properties [613], GPR has proven capable of simultaneously providing sputtering yields and corresponding uncertainty bounds [617]. Finally, the capability to capture the complex nonlinear relation between incoming ion energy distributions and outgoing energy and angular distributions of sputtered particles using ANNs has been demonstrated. It facilitates detailed PSI evaluation during plasma simulation runtime (cf. Fig. 35) at tremendously reduced computational cost [54], [615].

The outlined steps focus on reported approaches to data-driven PSI modeling of sputtering. An extension of similar procedures to other PSI mechanisms, such as plasma-induced electron emission or surface chemical reactions, is due. For instance, the complex transient interplay between reactive plasma and surface dynamics inherent to plasma catalysis or atmospheric pressure plasma in contact with surfaces/liquids may only be resolved with data-driven PSI models. In this context, data-driven chemical reaction

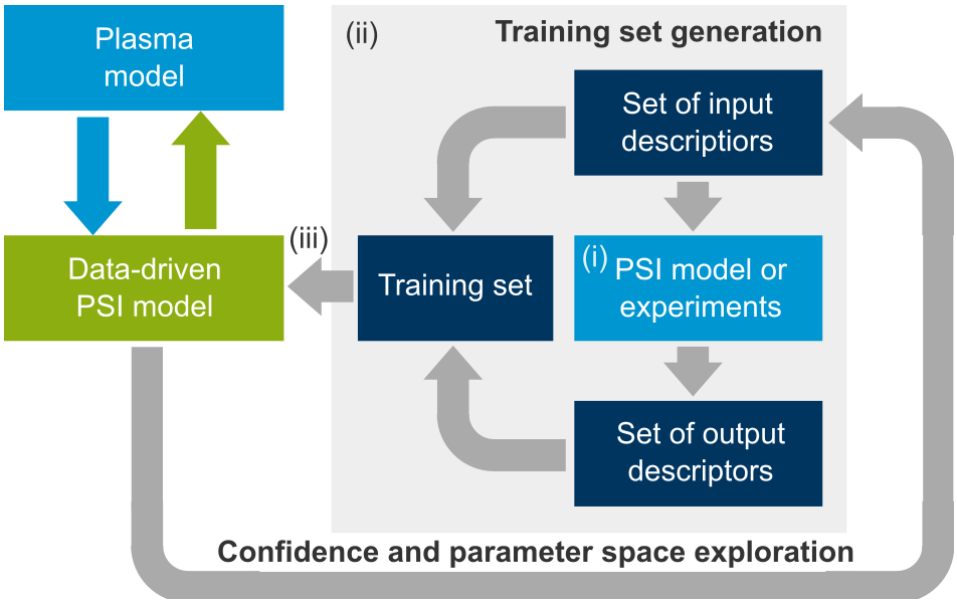

Fig. 34. Schematic of a generic data-driven PSI model.

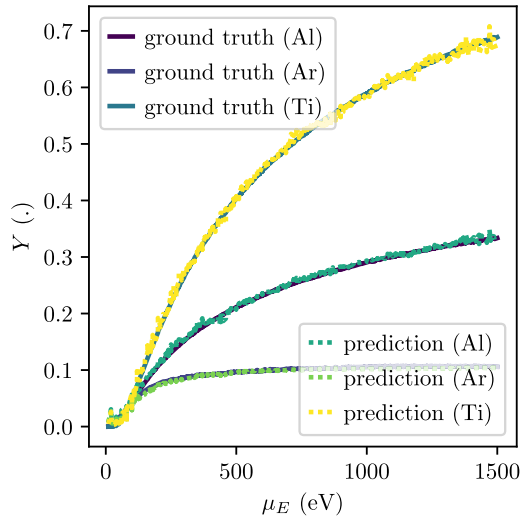

Fig. 35. Yield per sputtered species (Al, Ar, and Ti) as a function of the mean ion energy for an Al_xTi_{1-x} surface with initial stoichiometry $x = 0.3$. Ground truth compared to ANN predictions. Reproduced with the permission of the American Vacuum Society, from [615].

pathway analysis (PWA) [618] and active/transfer learning strategies for computationally costly atomic scale simulations [619] should be considered. Data-driven PSI models may ultimately permit a continuous and high-fidelity physical description of technological plasmas, providing guidelines for future research and exploration.
[Jan Trieschmann]

F. Neural-Network Potentials for the Analysis of Plasma–Surface Interactions With Molecular Dynamics Simulations

The surfaces of a fusion reactor will inevitably be exposed to harsh environmental conditions. Besides neutron fluxes, material erosion and fuel retention will limit their lifetime, especially in the divertor region. Experimental investigations under these conditions are difficult to impossible. Therefore, theoretical materials science is increasingly playing a role to quantify PSIs. On the atomic level, two techniques play a major role in MD simulations [620], where the many-body system is studied in detail by modeling its time evolution and the BCA [621] theory where the path of a projectile ion or atom is determined by a sequence of binary collisions. In BCA, scattering integrals are normally calculated by Monte Carlo methods to average over angular and energetic distributions, and the collision cascades are then derived. The assumption of binary collisions works best at projectile energies from keV and up but not at lower energies where many-body effects are important. In MD, the total PES is the key ingredient. It contains all the information about the system, and the trajectories of all atoms under consideration are derived from it. MD simulations have only recently been applied to systems where bond breaking and bond formation happen since, in this case, analytical potential energy expressions are difficult to derive. Such events are, however, happening all the time in sputtering processes. From humble beginnings, such as the Sutton Chen potential [622], quite successful interaction models, such as the bond-order potentials [623], were devised. They are analytical expressions that can be evaluated quickly on the computer, and especially, the latter was used in the investigation of several plasma-facing materials [624], [625]. However, their construction is demanding in terms of human effort, and their mathematical form is sometimes not flexible enough. About 15 years ago, with the increased employment of ML, techniques were developed to construct the PES nonparametrically with NNs [626] or GAPs [627]. Both methods allow for the necessary flexibility and, being parametrized via quantum chemical calculations, can model PSIs accurately for subsequent use in MD simulations. In the next paragraph, we give an example of typical NN-potential-based MD modeling.

Finding a suitable NN-based PES can be divided into three independent subproblems: 1) converting the cartesian

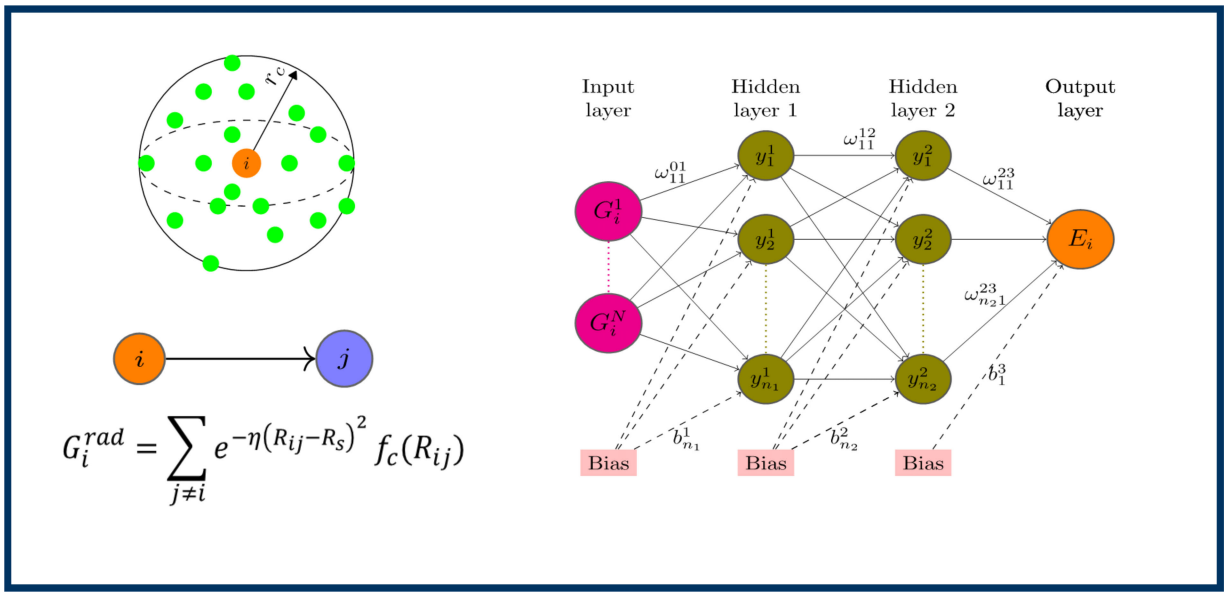

Fig. 36. Left from top to bottom: an atom i and its surrounding atoms, one of the corresponding radial vectors, and one radial symmetry function G^{rad}. Right: two G^{rad} functions are input, and the energy of atom i is an output of an NN with two hidden layers.

coordinates of the atoms into descriptors that can be input to the NN; 2) finding an optimal NN architecture; and 3) training the NN. Subsequently, the MD simulations produce statistically meaningful sample directories that are analyzed with respect to sputtering yields and many other material properties.

The conversion of cartesian coordinates into symmetry-adapted descriptors (1) is necessary because the energy of an atom stems from its environment and must be independent of translation, rotation, and the permutation of like atoms. In the Behler method, the descriptors are radial and angular basis functions, and their coefficients are calculated by projecting the atomic environment onto them. The invariant coefficients are input to the NN. Optimal descriptors are at least as important as having an optimal NN (or GAP) architecture and are an area of ongoing research [628]. The left-hand side of Fig. 36 shows as an example how the weight of a radial symmetry function G^{rad} is derived from all neighbors j of atom i. This is done flexibly for several Gaussian functions with varying exponents and midpoints to construct the radial density.

The NN (2) itself can have various architectures but is often a simple feedforward NN with as many input neurons as basis functions, two hidden layers of the same size, and one output layer delivering the energy for one atom. The atomic energies are then summed up.

Training (2) is the process of finding the best NN weights and offsets, and is similar to other applications of NNs. From simple backpropagation over Marquardt–Levenberg fitting to Kalman filters, many techniques are used. The training data are symmetry-adapted atomic coordinates, and associated energies and forces from trajectories derived by direct quantum chemical MD simulations. Sometimes, the potential energy and the forces are divided into one part that is treated with simple analytical expressions, and the NN takes care of the rest. This is advisable for charged systems where simple electrostatic interactions make up the largest part.

The process of network training is normally iterative. A trained NN is used in MD runs. Some MD configurations are checked by quantum chemical methods if the NN energies and forces are accurate up to a threshold. If they are not, such configurations are used in retraining. After a few cycles, an NN-based PES is obtained that is accurate within the limits of the parameter space.

Then, production runs can be performed like in conventional MD simulations. For calculating sputtering yields where energy and angle of the incoming particle are variable, for each energy/angle combination, about 5000 trajectories are necessary to achieve a good statistic. Fig. 37 shows results from sputtering simulations of a Be_2W surface. The trajectories of MD runs with different angles of the incoming deuterium atoms are analyzed to obtain density distributions (histograms) of the angles with which Be atoms are sputtered away [629]. Similar studies have been performed also for other surfaces as well [630].

In reality, more than two environmental parameters are important, such as the surface temperature and atomistic surface details. Then, unfortunately, the limits of MD are quickly reached due to finite computational resources. MD is also not practical in the MeV range since the integration of the equations of motion would require too small a timestep. MD with ML-based PES is, however, by now an established technique that is increasingly used to study PSI-relevant processes, such as sputtering, retention, diffusion, bubble formation, and diffusion.

Computational materials science is now becoming a useful tool, and the modeling of PSIs by means of MD simulations is evolving rapidly. It is recognized now that the optimal descriptors are of utmost importance, even more so than the mathematical shape of the potential energy function. At the

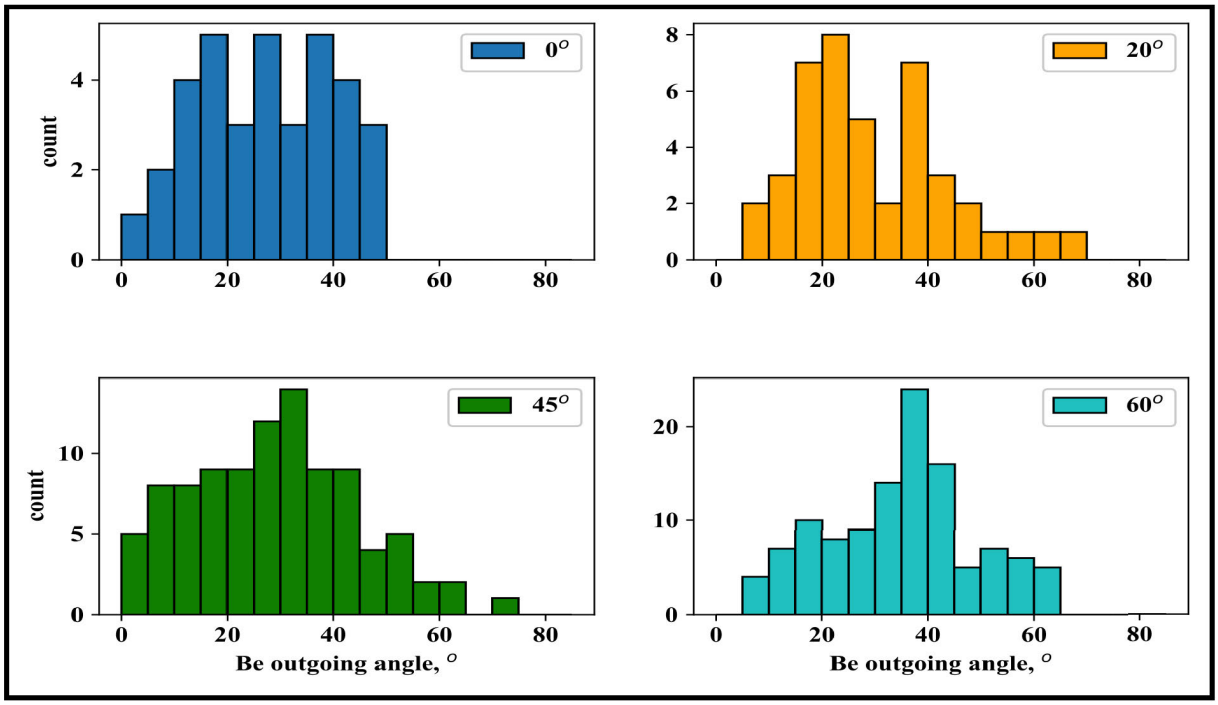

Fig. 37. Effect of the angles (0°, 20°, 45°, and 60° to the surface normal) of deuterium atoms incoming with 100 eV on the angular distributions with which Be atoms are sputtered [629].

same time, one becomes aware that the automatization of the training/simulation/improvement cycle is necessary. This is not trivial, and "active learning" [631] methods can be used to achieve this goal. It is quite possible that the methods described here will be soon available in computer codes in a more standardized fashion and will be used even more. At the same time, improvements and new algorithms are published in short intervals, indicating that ML is far from mature, even as a tool for representing complex potential energy surfaces. [Lei Chen and Michael Probst]

G. ML-Based Numerical Simulation of Low-Temperature Plasmas

Charged-particle transport plays a key role in generating and maintaining low-temperature plasmas. The BE provides the basis for elucidating charged-particle transport in plasmas. However, since the BE is the transport equation in phase space, it has still been limited to simulate the spatiotemporal development of the charged-particle transport by solving the BE numerically. This is due to the curse of dimensionality and exponential growth of computational cost with respect to the dimension. Such difficulty clearly appears in 3-D and higher dimensional simulations using mesh-based methods, such as finite difference methods. PINN has attracted attention to solving the PDE. In the PINN approach, the latent solution of the PDE is represented by ANN, and the ANN is trained to respect both the PDE, often describing the law of physics, and boundary conditions. When a function is represented by ANN, partial derivatives of the function with respect to its variables can be calculated analytically by taking advantage of automatic differentiation; therefore, the PINN approach enables us to solve the PDE without generating grids and meshes, and would allow us to tackle high-dimensional problems. The PINN approach was proposed by Raissi et al. [48]. They demonstrated this approach to solve 1-D Burgers' equation and Shrödinger equation with Dirichlet boundary conditions. The PINN approach has been applied for solving a wide range of problems with various boundary conditions and constraints. Kawaguchi et al. [632] employed this approach for solving the BE for 2-D equilibrium EVDF in Reid's ramp model gas and Ar under dc uniform electric fields with normalization constraint of the EVDF. Rao et al. [633] simulated incompressible laminar flows with Dirichlet and Neumann boundary conditions. Zobeiry and Humfeld [634] applied the PINN approach to solving 1-D and 2-D heat transfer equations with convection boundary conditions. A comprehensive review of the PINN is available in [387]. In this section, the procedure for solving the PDE through the PINN approach is presented. The BE for 3-D equilibrium EVDF under crossed dc uniform electric and magnetic fields in a boundary-free space is chosen as an example of the PDE [635]. Such EVDF would be important to analyze the electron transport properties in magnetized plasmas, which are employed in material processing.

The equilibrium EVDF $f(\boldsymbol{v})$ under dc uniform electric and magnetic fields is governed by

$$\frac{e}{m}(\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B})\cdot\frac{\partial f(\boldsymbol{v})}{\partial \boldsymbol{v}}+\nu_{\text{eff}}f(\boldsymbol{v})-J_c f(\boldsymbol{v})=0 \qquad (10)$$

where $\boldsymbol{E}=(0,0,-E)$ is the electric field, $\boldsymbol{B}=(0,-B,0)$ is the magnetic field, e is the electron charge, m is the electron mass, $\boldsymbol{v}=(v_x,v_y,v_z)$ is the electron velocity, ν_{eff} is the effective ionization collision frequency, and $J_c f(\boldsymbol{v})$ is a collision term. Here, SF_6 is chosen as ambient gas, and collisions between an electron and a gas molecule for elastic, excitation, electron attachment, and ionization are considered in the collision term. Fig. 38 shows the schematic for solving (10) using the PINN approach. The latent solution of (10) is

represented by ANN. How well the ANN respects the PDE and boundary conditions is measured by a loss function

$$\mathcal{L} = \mathcal{L}_{\mathrm{PDE}} + \lambda \mathcal{L}_C \tag{11}$$

where $\mathcal{L}_{\mathrm{PDE}}$ and $\mathcal{L}_C$ represent the residual of the PDE and boundary conditions, respectively, and λ is the parameter controlling $\mathcal{L}_C$. The term $\mathcal{L}_{\mathrm{PDE}}$ is given by

$$\mathcal{L}_{\mathrm{PDE}} = \frac{1}{N_{\mathrm{PDE}}} \sum_{i=1}^{N_{\mathrm{PDE}}} |R(\boldsymbol{v}_i)|^2 \tag{12}$$

where R is the residual of the PDE, namely, the left-hand side of (10), and $\boldsymbol{v}_i = (v_{x,i}, v_{y,i}, v_{z,i})$ denotes a point sampled on the domain of the solution. The partial derivatives of $f(\boldsymbol{v})$ with respect to v_x, v_y, and v_z are calculated by using automatic differentiation. If the Dirichlet boundary conditions were applied, $\mathcal{L}_C$ could be described by

$$\mathcal{L}_C = \frac{1}{N_C} \sum_{j=1}^{N_C} \left| f(\boldsymbol{v}_j) - \hat{f}(\boldsymbol{v}_j) \right|^2 \tag{13}$$

where $\boldsymbol{v}_j$ is sampled on the boundary of the domain and $\hat{f}(\boldsymbol{v}_j)$ is a given value at $\boldsymbol{v}_j$. In the present calculation, the normalization constraint $\int_{-\infty}^{\infty} f(\boldsymbol{v}) d\boldsymbol{v} = 1$ is applied, and the term $\lambda \mathcal{L}_C$ is truncated. Instead, the collision term on $\mathcal{L}_{\mathrm{PDE}}$ is calculated by using normalized EVDF. The ANN has weight and bias parameters, and they are optimized to minimize the value of $\mathcal{L}$ by the gradient descent-based method, such as Adam [636], until the value of $\mathcal{L}$ reaches a minimum. There is flexibility in how to sample points. We can simply sample points by using pseudorandom numbers. The Latin hypercube sampling and quasi-random numbers are used to sample points uniformly. Adaptive sampling method in which the distribution of the sampling points is improved by considering that of $\mathcal{L}$ is proposed [637]. Scaling the ANN input is important, and they should be distributed on $[-1, 1]$. The appropriate architecture of the ANN would vary with the problem to be solved and is tuned empirically by users at present. Designing an effective ANN architecture for solving the PDE accurately has been investigated [638]. Fig. 39 shows the EVDF projected into a v_x–v_z plane and the EEDF calculated from the EVDF. The EVDF and EEDF calculated from the MCS are also shown as reference data. The PINN can successfully reproduce the MCS results. In this calculation, the EVDF is represented by feedforward ANN having 41 700 parameters. The EVDF in the same condition was also calculated using the mesh-based method [639] and was stored in a 3-D array, the size of which is $10\,000 \times 45 \times 750$. Given that the precision of floating points employed in the calculations is the same, the PINN allows us to represent the 3-D EVDF properly with approximately 0.01% of the memory capacity required in the mesh-based method.

A PINN provides a novel mesh-free approach to solve the PDEs, allowing us to deal with high-dimensional problems. For the electron BE, it is confirmed that the PINN approach can significantly reduce the memory capacity required for representing the EVDF properly compared to the mesh-based method. The PINN approach has been applied to various problems regarding fluid dynamics, heat transfer, EMs, and so forth. Combining PINNs for various scientific disciplines would enable us to represent multiphysics systems and would contribute to advances in plasma simulation. In this case, constituent NNs would be trained not so much to minimize their loss functions as to minimize the loss function for the system, for example, the sum of the loss functions for each NN. [Satoru Kawaguchi]

H. Reduction of Chemical Reaction Models

1) Introduction: The number of species that can be formed in plasmas can be considerable. Dozens of electronically excited states may need to be considered to correctly predict the rates of ionization, recombination, and radiative processes, even when the plasma is created in an atomic gas, such as argon [640] or mercury vapor [641]. When the plasma is created in a mixture of molecular gases, the complexity further increases, especially when a rise of the gas temperature results in the onset of a multitude of nonelectronic reactions. Among the many contemporary technologically relevant examples are plasmas in methane (36 species and 367 reactions) [642], air (84 species and 1880 reactions) [643], and carbon dioxide (72 species and 5732 reactions) [644]. Incorporating such chemistries in full into a space- and time-resolved computer simulation may be tempting but is at present hardly feasible. Therefore, an analysis and, when possible, a reduction of such plasma chemistries are called for, and those tasks have been accomplished even for rather complicated chemistries (see [642] and [645]). Although computers have gotten exponentially faster over the past decades, Gustafson's law suggests that the problems that we try to solve with them continuously get bigger as well [646]. Therefore, the need for more systematic and automated methods grows, and it is no surprise that *plasma chemistry reduction* continues to be a subject of great interest.

Like any modeling effort, an attempt to achieve a chemistry reduction should start with a precise statement of the scope of the model and the observables that the model aims to reproduce. If these observables are not influenced by a particular minority species, that species may be removed from the species list. However, in another study, that minority species may be among the key observables, for example, because, in spite of its small abundance, it is responsible for the degradation of the plasma device. Also, the relevant time scales must be part of a model specification. A plasma reactor model may target the plasma behavior on a timescale of milliseconds, and in such a case, it may be desirable to eliminate the nanosecond timescales from the model. However, in a model of a laser-induced fluorescence experiment, these smallest time scales are the relevant ones, and the long-term dynamics of the plasma can be disregarded [647].

This section discusses a number of methods that have been considered for plasma-chemical reduction in the past. Furthermore, recent works that are related to the subject will be summarized. Special attention will be paid to the suitability of methods that originate from adjacent fields of science, such as combustion engineering for plasma-chemical reduction.

2) Timescale-Based Reduction Schemes: The chemical composition of a plasma can be characterized by the particle

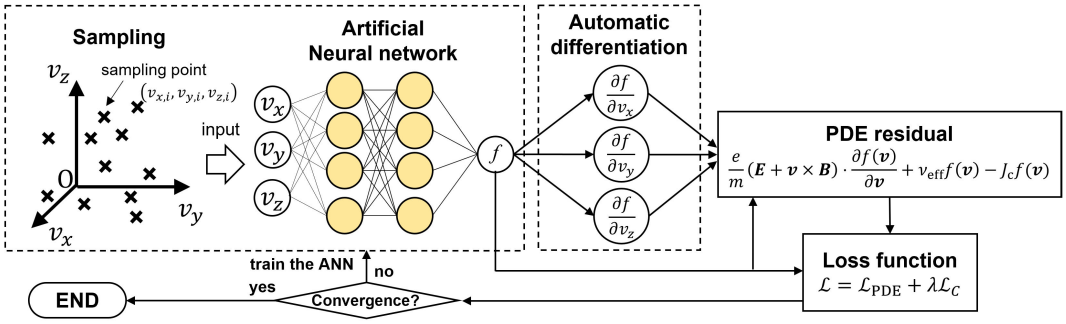

Fig. 38. Schematic of the PINN approach for solving the PDE.

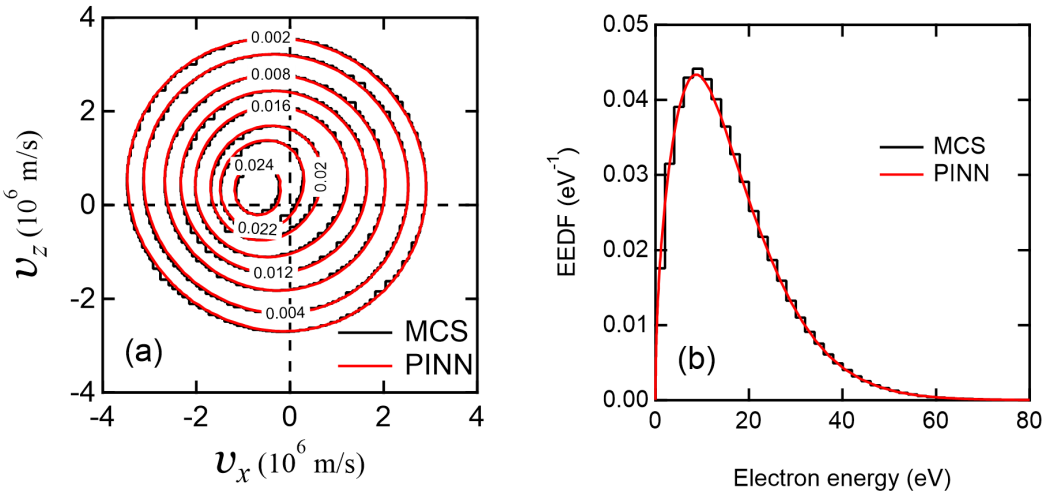

Fig. 39. (a) Contour plot of the EVDF projected into a v_x–v_z plane. (b) EEDF as a function of the electron energy. The strength of the reduced electric field E/N and that of the reduced magnetic flux density B/N are set to 2000 Td (1 Td = 10^{-21} Vm^2) and 2000 Hx (1 Hx = 10^{-27} Tm^3), respectively. Here, $N = 3.535 \times 10^{22}$ m^{-3} denotes the number density of gas molecules.

densities n_i of the components i. The temporal and spatial variations of these components can be calculated from a set of balance equations that are given by

$$\frac{\partial n_i}{\partial t} + \nabla \cdot \vec{\Gamma}_i = S_i \tag{14}$$

where $\vec{\Gamma}_i$ and S_i are the particle flux density and the volumetric production rate of particles of type i. Depending on the transport coefficients, the electric field, and the reaction scheme that underlies the sources and sinks that end up in S_i, the density n_i may be affected by transport or may follow from *chemical equilibrium*, which is to say that $S_i \approx 0$.

The QSSA, which amounts to setting $S_i = 0$ for (near-)equilibrium species, has been around since the early 1900s [648], [649]. In the 1960s, Bates et al. [650] used the QSSA for the excited states in atomic plasma. If the source terms for these states are only due to radiative and electron-impact processes, these are *linear* in the densities of those species, and the authors demonstrated that this allows the elimination of the excited state densities from the system of transport equations, in combination with a correction of the rate coefficients for ionization and recombination for the remaining atom and ion ground state. These corrections account for indirect or *ladder-like* processes. The result is an important tool for chemistry reduction since the number of atomic states that are considered in the transport model is reduced from dozens to only two, without sacrificing the physical validity of the model.

A generalization and a more explicit algebraic perspective on this procedure were provided in [641]. When we bundle the sources and densities of the atomic and ion states in column vectors $\boldsymbol{S} = [\cdots S_i \cdots]^T$ and $\boldsymbol{n} = [\cdots n_i \cdots]^T$, one can write $\boldsymbol{S} = \boldsymbol{M}\boldsymbol{n}$, where the matrix $\boldsymbol{M}$ depends on the electron temperature (through the rate coefficients), the electron density, and the opacities of the plasma for resonant radiation. When the "nonlocal" densities are placed at the top of these vectors, the reduced system can be partitioned in transport-sensitive (t) and local (l) blocks

$$\begin{bmatrix} \boldsymbol{S}_t \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}_{tt} & \boldsymbol{M}_{lt} \\ \boldsymbol{M}_{tl} & \boldsymbol{M}_{ll} \end{bmatrix} \begin{bmatrix} \boldsymbol{n}_t \\ \boldsymbol{n}_l \end{bmatrix}. \tag{15}$$

Solving the second block of equations for $\boldsymbol{n}_l$ and substituting the result in the first block yield

$$\boldsymbol{n}_l = -\boldsymbol{M}_{ll}^{-1}\boldsymbol{M}_{tl}\boldsymbol{n}_t, \quad \boldsymbol{S}_t = \left(\boldsymbol{M}_{tt} - \boldsymbol{M}_{lt}\boldsymbol{M}_{ll}^{-1}\boldsymbol{M}_{tl}\right)\boldsymbol{n}_t. \tag{16}$$

The first equation expresses the densities of the local states in terms of those that are affected by transport. The second equation expresses the sources of the transport-sensitive levels in terms of their densities. The *effective* coefficient matrix contains the direct processes ($\boldsymbol{M}_{tt}$) and a correction for the indirect or ladder-like processes that involve the states that no longer need to be modeled explicitly.

This elaboration demonstrates the technique that underlies many chemical reduction schemes. It shows that the *locality* of species densities can be used to replace differential equations with algebraic ones. It also shows that these species may still

influence the kinetics of the remaining species via indirect processes.

A drawback of the QSSA method is that accurate error estimates can only be obtained by running the solution both with and without QSSA, and comparing the results [649], [651]. A detailed overview of more recent methods for the analysis of chemistries that do not suffer from this problem can be found in [649] and [652]. For the reduction of chemistries based on timescales, a few classes of techniques are available, many of which find a root in combustion engineering. One of the earliest numerical approaches is the CSP, first described in 1985 by Lam [653], [654], [655], [656]. The goal of this family of methods is to automate the process of simplifying systems of differential equations like the ones encountered in chemical reaction systems, a task that up until then was executed manually. Variations include linear CSP, nonlinear CSP, and CSP without eigenvalue decomposition [657].

In 1992, the ILDM family of methods was pioneered by Maas and Pope [651], [658]. This family of methods recognizes that the time scales involved in the chemical reactions in a mixture often span multiple orders of magnitude. The fastest equilibration processes attract the systems toward a low-dimensional subspace in phase space, the so-called low-dimensional manifold. This effect is demonstrated in Fig. 40 for the imaginary chemistry from [652], consisting of species A, B, and C. The reaction space of this chemistry is confined to a 2-D manifold, described by A+B+C = 1. It can be observed that any random initial composition on this surface quickly converges onto a 1-D manifold, before eventually settling at the equilibrium composition, a 0-D manifold. Various methods of finding such manifolds for arbitrarily complex chemistries exist, including FGMs [659], trajectory-generated manifolds (TGMs) [660], and ILDM assisted by in situ adaptive tabulation (ISAT) [661], [662].

Applications of CSP or ILDM to plasma chemistry are still scarce, and an example can be found in [663]. The reason may be a lack of awareness within the community of such reduction methods or the fact that it simply takes more time for techniques to transfer to a different field of science. Another reason is that, in plasmas, more parameters come into play (electron energy and opacities), and often, their gradients are not coaligned, frustrating methods that rely on quasi-1-D behavior, such as FGM [659].

3) Recent Developments, Outlook: Various innovative strategies have been proposed in the past five years. As an example, PCA has been applied to plasmas for the first time [664], [665]. Also, the method of PWA [666] has seen renewed interest [666], [667] and has been applied to large plasma chemistries (see [668] and [669]). More recently, graph theory and ML are being used to extract information from complex chemistries [670], [671]. While an ultimate solution to the problem of plasma-chemical reduction is not yet in sight, these developments bear great promise for the future of the field.
[Jan van Dijk and Rick H. S. Budé]

I. Biological Data and Plasma Medicine

In 2003, when Stoffels et al. [672] first reported on the nonlethal manipulation of mammalian cells by a nonequilibrium plasma ("plasma needle"), a new chapter of plasma physics began. Besides the widely accepted technical application of plasma processes and the inactivation of prokaryotic bacteria reported since the mid-1990s [673], the report highlighted a new facet of plasma and sparked a surge of research projects all around the globe. For the last almost 20 years, a number of breakthroughs have been made, and nonequilibrium atmospheric pressure plasmas, which are, for the sake of biomedical and clinical researchers, often simply called "cold plasmas" or "gas plasmas," have found their way into the clinics and ambulant care with a number of certified medical devices in the market. The number of publications on plasma medicine rose from less than five in 2003/2004 to more than 800 per year (2020, Google Scholar). In the beginning, the new interdisciplinary studies were published in journals with a traditional engineering or physical scope. While these journals still publish data on biomedical plasma research, journals with a broader scope and readership beyond the plasma research community are increasingly targeted. Among these, numerous medical or interdisciplinary journals dominate. With the increasing impact of the research on foreign communities, clinicians, funding agencies, and the public, increasing awareness of the validity, interchangeability, and reproducibility of results can be felt in the community. Adherence to the FAIR data use policies (see Section VIII-E) [674], international approaches to define a universal plasma dose, or actions on standardization are representative for this "coming-of-age" time of the research field. Naturally, this affects all aspects of the topic, but the larger variance of biomedical experiments and the resulting data, and medical safety aspects accelerated the correspondent efforts. When surveying current publications on biomedical aspects of cold plasma, the use of bioinformatics tools has become the normal case [675], [676], [677]. Currently, when proteomics (proteins) or lipidomics (lipids) data are a central piece of the paper, most journals desire the upload of these data into public repositories to ensure their long-term persistence and preservation. A number of dedicated databases have evolved, e.g., the members of the ProteomXchange consortium http://www.proteomexchange.org/, such as MassIVE, PeptideAtlas, or PRIDE for proteomics data; Metabolomics Workbench https://www.metabolomicsworkbench.org/ for small molecules, including lipids; or the Genome Sequence Archive https://ngdc.cncb.ac.cn/gsa/ for genomic information. The major benefit for any research community is the long-term conservation of the data independent of individual working groups, the possibility to share the data with colleagues to allow additional data analysis approaches, and the increase in reliability and reproducibility as defined by the FAIR Guiding Principles for scientific data management and stewardship that were introduced in 2016 https://www.go-fair.org/fair-principles/ [674]. In the plasma science community with a special focus on plasma medicine, a dedicated repository INPTDAT has been established https://www.inptdat.de/, adhering to the FAIR principles, as discussed in Section VIII-E.

To understand the impact of cold physical plasma in biological systems, Wende et al. have deployed methods such

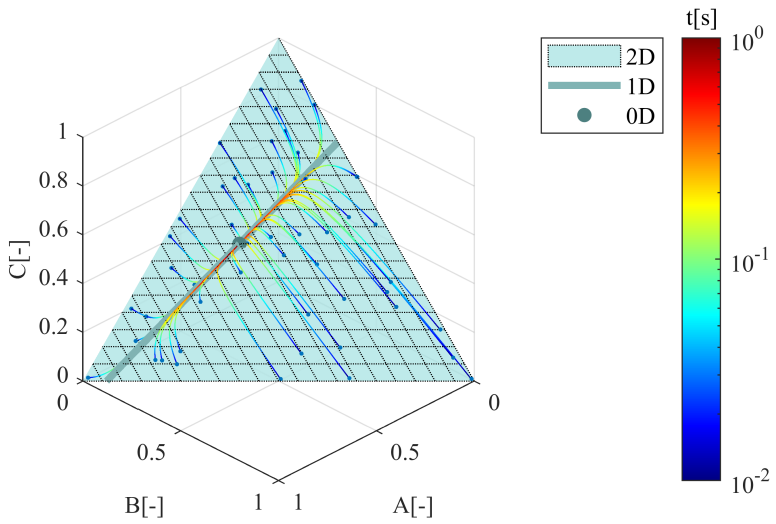

Fig. 40. 0-D, 1-D, and 2-D manifolds for the chemistry described in [652]. The time evolution of random initial compositions on the 2-D manifold is shown, showing that the compositions first converge onto the 1-D manifold, before settling onto the 0-D manifold. Image recreated from [652].

as high-content imaging, flow cytometry, transcriptomics, and proteomics in a number of in vitro and in vivo models, e.g., [675], [678], [679], [680], [681], and [682]. To analyze the significant amount of raw data, softwares such as GeneSpring (Agilent), Kaluza (BeckmanCoulter), Tibco Spotfire (Tibco Software), Byonic (ProteinMetrics), or Proteome Discoverer (Thermo) are used. The bottom line of all studies presented here is the major role that reactive oxygen and nitrogen species occupy to trigger the observed events. Since cellular signaling of both prokaryotic and eukaryotic cells uses the same reactive species, there is a "common language" between the gas-phase phenomena of plasmas and biological systems. However, due to the long distances between the generation and the assumed place of action, a direct contribution by short-lived species, such as singlet and atomic oxygen, or peroxynitrite is questionable. For this reason, we pursue the hypothesis that the short-lived species chemically modify biomolecules in close vicinity to the point of impact. Subsequently, either the chemical energy of the reactive species is preserved, e.g., as a radical or peroxide, or the modified molecule is perceived as a signal molecule; or its functionality is changed significantly. In the first steps of validity testing, it was observed that an MHz-driven dielectric barrier argon jet (kINPen, neoplasm, Germany) has a significant impact on cysteine and tyrosine. The reaction products reflected the gas phase composition and the reactive species formed, permitting its use to compare plasma sources and conditions, and infer plasma liquid chemistry and gas–liquid interphase chemistry [683], [684], [685]. The concept was extended using artificial peptides, providing a more complex chemical environment and a greater variety of chemical structures to be attacked by the plasma-generated species [677], [686]. Again, this approach involved high-resolution mass spectrometry and the use of an advanced software solution to filter the raw data for relevant information on oxidative posttranslational protein modifications (oxPTMs, Byonic, ProteinMetrics, Palo Alto, CA, USA). As a result, the introduction of 17 different oxPTMs was determined along with four main targets: cysteine, methionine, tryptophane, and tyrosine. For example, in the two decapeptides, Ala-Asp-Gln-Gly-His-Leu-Lys-Ser-Trp-Tyr and Ala-Cys-Glu-Gly-Lyl-Ile-leu-Lys-Tyr-Val, the modification nitration (+44.98 m/z, +N + 2O −H) is introduced in dependence on gas-phase composition (Ar ≫ Ar/O_2), plasma source (kINPen ≫ COST jet), and solvent system [H_2O ≫ phosphate-buffered saline (PBS)]. Since an aromatic structure and an acidic pH promote nitration, it is most prominent in tyrosine and water as a solvent. In Fig. 41, the role of the investigated conditions on the extent of amino acid modifications [see Fig. 41(a)] or on the type of observed modification [see Fig. 41(b)] is visualized after statistical analysis by the software package R. The data allow insight into the likelihood that a certain amino acid is modified by a plasma treatment when a specific condition is met and how a certain modification can be triggered by the choice of condition (model) or can be expected in an in vivo setting.

A prominent example is the occurrence of dioxidations (+31.98 m/z, + 2O) that is strictly linked to a direct plasma treatment plus suitable gas phase composition (oxygen admix), setting the stage for singlet oxygen as the underlying reactive species. For further analysis and details, see [677].

The impact of plasma-driven oxPTMs on protein function was shown for a number of proteins. One example is the enzyme phospholipase A2 that is a relevant player in inflammatory processes by supplying unsaturated fatty acids as precursors for signaling molecules. A necessary step in the cleavage of membrane lipids (phosphatidylcholines) is the docking of the proteins C-terminus to the membranes' polar head groups. A plasma treatment by the kINPen disrupts the docking and subsequently enzyme activity, strongly suggesting that the biomedical application of cold plasma may utilize the (in)activation of proteins to achieve effectivity. Via high-resolution mass spectrometry/bioinformatics and MD simulation (GROMACS [56] program package (version 5.0) OPLS-AA/L all-atom force field), the amino acid residue tryptophan 128 was identified to be the target of plasma-derived singlet oxygen dioxidation, yielding a ring-open kynurenine

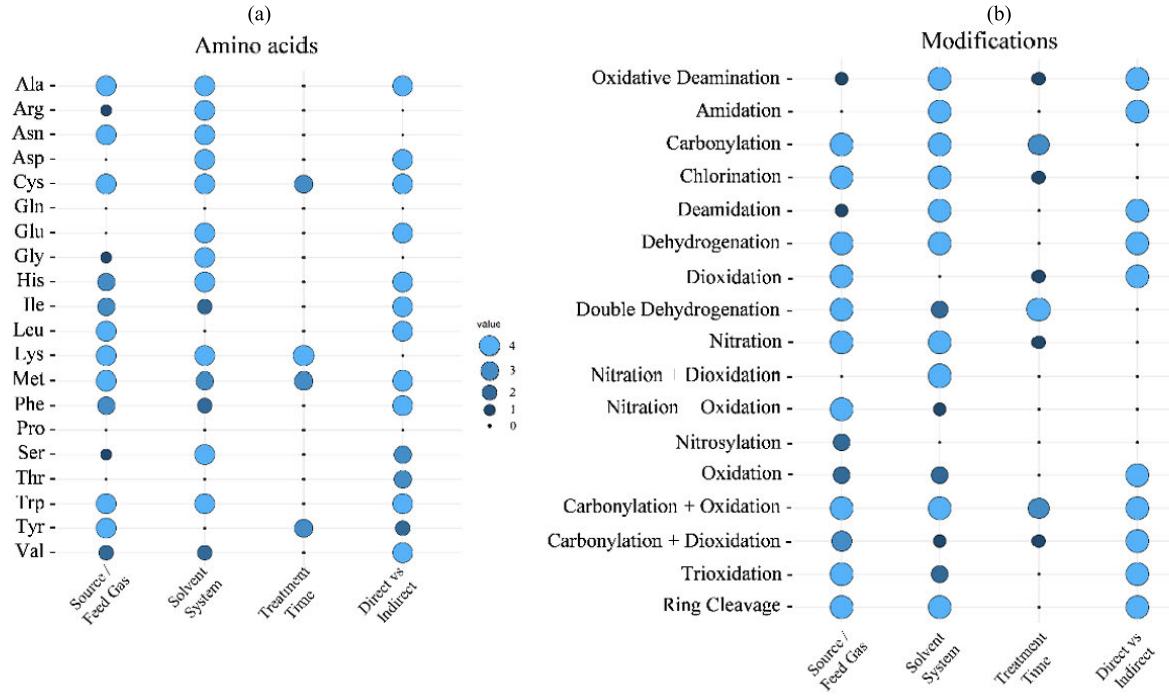

Fig. 41. Impact of plasma source/gas phase composition, solvent treatment time, and treatment mode on the extent of (a) amino acid modification or type of modification introduced by (b) plasma treatment. A large circle indicates a strong correlation. For example, the modification oxidative deamination (replacement of nitrogen by oxygen) is influenced by the solvent type and the treatment mode (direct) but, to a minor extent, only to (b) plasma source or the treatment time. Reprinted from [677]. Copyright 2021 Author(s), licensed under a Creative Commons Attribution (CC BY) License.

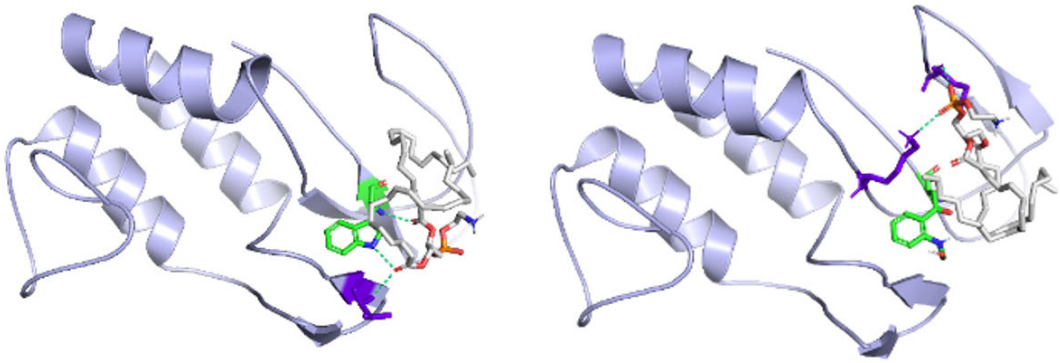

Fig. 42. Impact of argon plasma jet (kINPen) on phospholipase A2 secondary structure. Left: control. Right: after direct plasma treatment. The residue tryptophan 128 is dioxidized, yielding a structural change and inhibiting enzyme function. Reprinted from [676]. Copyright 2021 Author(s), licensed under a CC BY License.

derivative that subsequently distorted the secondary structure of the C-terminal β-sheets of PLA_2 (see Fig. 42) [676]. In a similar manner, it was shown by Clemen et al. [675] in an animal model that protein oxidation triggers a more strict response of the immune system, opening the avenue to plasma-driven cancer vaccination.

In conclusion, the hypothesis that plasma-derived reactive species modify biomolecules that subsequently modulate physiological processes has to be accepted: oxPTMs are introduced not only in model peptides but also in also full proteins, changing their perception and role.
[Kristian Wende]

J. Challenges and Outlook

Plasma processing involves complex physical and chemical systems in nonthermal equilibrium conditions. In addition, spatial and time scales involved in those systems vary widely from the atomic scales to the manufacturing tool scales. For example, in a typical plasma processing tool, macroscopic parameters, such as gas compositions, gas pressure, and applied power to the plasma source, are used as control nobs to form nanometer-scale complex device structures on a wafer surface. The conventional first-principle-based approaches to analyzing plasma processing systems, i.e., numerical solutions to the fundamental physics equations describing the systems, are, in general, not free from input parameters; they typically require fundamental data, such as reaction rates in the gas phase and on surfaces. Furthermore, such approaches are, even if available, typically time-consuming and often accumulate errors arising from inaccurate input parameters in their analyses. Therefore, although such analyses are undoubtedly important for a better understanding of the nature of plasma

processing, more quantitatively reliable and timely analyses are also required for practical applications, such as plasma system control and new process development.

Data-driven approaches may offer solutions to such requirements. For example, a large amount of numerical simulation data and/or measurement data of experimental/manufacturing systems may be used to create machine-learned regression models or surrogate models to predict system characteristics, such as etch rates, sputtering yields, and interatomic forces, as discussed in Sections VII-C, VII-E, and VII-F. Reduction of the dimensions of extremely large datasets to make the data more tractable by computation is also another challenge, as discussed in Section VII-H for chemical reactions in plasmas.

Although a large amount of data may be obtained from individual plasma processing tools and their processed material surfaces, what remains a challenge in plasma technologies is the shortage (or sometimes lack) of fundamental data on elementary processes that can be applied to any processing tools, such as chemical reaction rates of specific surface materials with specific incident gaseous species that characterize the plasma surface interaction. Of course, it is unrealistic to expect to obtain such data for all possible combinations of surfaces and gaseous species exhaustively. However, it is desirable to establish new techniques for HTS to obtain fundamental chemical reaction data associated with desired plasma processing efficiently. In general, the physics of plasmas is better understood than their chemistry, so such chemical data combined with the conventional first-principle-based approaches and the latest data-driven approaches would allow far more accurate analyses of plasma processing and drive faster and more cost-effective development of new processes and better plasma control techniques.
[Satoshi Hamaguchi]

VIII. Plasma and Related Databases

A. Introduction

In the study of any of the different plasmas discussed in this review, a common challenge is to obtain a thorough understanding of the physical and chemical properties of plasmas. In order to determine such properties, it is essential to assemble authoritative databases that allow the design, diagnostics, and monitoring of the plasma. The plasma community has been active in assembling such databases that include the following:

1) AM databases detailing both spectroscopic data (commonly used as plasma diagnostics to identify key plasma species) and collisional data characterizing electron, ion, and photon interactions with those AM species within the plasma and knowledge of both the cross sections and reaction rates for such collisions, both in the gas phase and on the surfaces of the plasma reactor;
2) material databases that provide data on the properties used in the design and operation of plasma systems with databases for fusion reactors being among the most extensive;
3) plasma chemistry databases that provide access to complete and validated data for plasma modeling with preassembled and validated chemistry sets;
4) low-temperature plasma databases that have been among the most common databases since these have been constructed to support specific industrial plasmas, such as those used in the semiconductor, lighting, and medical industries.

However, the compilation of such databases remains a major challenge, and the necessary coordinated infrastructure and funding to build and sustain them have often been lacking. This, in turn, challenges the broader scientific community to recognize that their fields also rely upon the compilation and access to relevant databases, and a united research community must then confront the funders of research (government and industrial), specifying that scientific and technological progress is based upon a strong fundamental bedrock; if this is neglected, then the scientific and technological advances that they require will not occur, and their investment will not be rewarded.

This section reviews the current status of the different databases and gives indications as to present data deficits. Core to all databases is the criteria for data selection: whether the database then recommends datasets or leaves the user to select data is an important parameter. In particular, recommended datasets allow individual models to be cross-correlated. Methods and community practice in establishing recommended datasets will also be presented.
[Nigel J. Mason]

B. Atomic and Molecular Databases

AM processes are elementary processes in plasmas and important to understand the microscopic behavior of plasmas and radiative processes in plasmas. Radiative and collisional processes of atoms and molecules govern the energy balance of plasmas. It is also useful to use emissions from atoms and molecules for spectroscopic diagnostics, for example, to know impurity behavior in fusion plasmas and plasma properties, such as electron temperature and density. AM data, such as wavelengths and transition probabilities of emission lines or collision cross sections, are important fundamental data to describe AM processes. AM databases compile and store such important data since the 1970s and provide them for users in various research fields [687]. In recent years, many databases are available through the internet, and there have been some attempts to provide such data more conveniently for users. As a new attempt, databases are used to train ML methods, for example, to estimate a set of electron-impact cross sections from swarm transport data [688].

There are two kinds of AM databases available: one has evaluated data and the other has original data.

The former databases contain evaluated one value (or one dataset) for one process, e.g., one wavelength for one specific transition and one set of ionization cross sections as a function of collision energy for a specific atom. Data evaluation is done by the organizers of the database in

various ways. The accuracy of data is carefully examined experimentally, by checking the method of the study, or by comparing it with other data, and one value or one dataset is selected and stored in the database. The NIST Atomic Spectra Database [689] is this type of database for atomic wavelengths, transition probabilities, energy levels, and ionization potentials. The atomic database in CHIANTI [690] for spectroscopic diagnostics for solar physics is also this type. The IAEA ALADDIN database contains evaluated data of cross sections and rate coefficients for electron collisions, photon collisions, and heavy particle collisions [691], but several datasets evaluated by different research groups are stored for one process.

The second type of database contains many data for one process obtained by various theoretical or experimental studies. All data or dataset have their references on their origins, and users can track the data source. Users can compare several datasets for one process, such as ionization cross sections for a specific atom, and can evaluate and select data by themselves. The NIFS AM numerical database is this type of database for collision cross sections and rate coefficients for ionization, excitation, recombination, and charge exchange processes of atoms and molecules [692]. Users can compare experimental and theoretical data for one process with a graphic output of the database. Open ADAS [693] is also this type of database for datasets relevant to spectroscopic diagnostics of fusion and solar plasmas. Various theoretical datasets are stored for fundamental data, such as a set of energy levels and electron impact excitation effective collision strengths. Derived data calculated with the ADAS software package are also available, such as photo emissivity coefficients for emission line intensities of an atomic ion. Databases that provide one set of calculated data for one process are also categorized as this second type, such as opacity databases [694], [695].

There are some attempts to access various AM databases from one website. LXCat, the Plasma Data Exchange Project [696], is the project to collect AM data from various databases for low-temperature plasmas and provide them to users from one website. Databases on electron scattering cross sections, differential scattering cross sections, swarm transport data, and online BE solvers are available. VAMDC [697] is also the project to access various databases from one website and provide data with the same XML format. Currently, 46 databases on spectral lines, opacities, and collision cross sections of atoms and molecules are connected to VAMDC, including NIST ASD, CHIANTI, and NIFS databases. The XML schema, XSAMS, was developed under the collaboration coordinated by the IAEA Atomic and Molecular Datat Unit (see Fig. 43).

Current AM numerical databases have been developed and maintained to be available for communities with big efforts by researchers on atomic physics and various plasma physics for many years. Databases on such fundamental data are useful for various applications. Data need from communities give motivation to studies for atomic physicists, and the help and efforts of data providers are largely appreciated. Continuous efforts to maintain these databases must be supported by communities.

[Izumi Murakami]

C. Materials Database

In most industrial plasmas, as well as in fusion plasmas, the plasma is "contained," and therefore, plasma surface interactions are important in determining the operation and characterization of the plasma. Many plasmas are specifically designed to interact with surfaces; for example, atmospheric plasmas are being used to sterilize surfaces in medicine [698], [699], which requires understanding both of the "sterilizing agents" in the plasma (ions, UV photons, and radicals) and the properties of the surfaces. Indeed, medical applications are a good example of the myriad of materials with which plasma may interact—metals, plastics, ceramics, and glass. Plasma treatment is recognized as a valuable method for treating surfaces and may be scaled up for large-scale manufacturing, for example, introducing hydrophobic properties in materials [700]. Plasmas may "activate" processes on surfaces or even activate drugs [701], [702]. Plasma waste remediation and waste treatment [703] require detailed knowledge of plasma surface interactions including with (and in) liquids and may be used even for radioactive waste [704]. However, to date, there are no databases that focus on plasma interactions with such materials, and there have been few studies to explore in detail the physicochemical changes induced by plasmas across such a range of materials. Rather, publications are scattered and often present a limited dataset for one plasma and one material, making cross-comparison difficult.

In contrast, the fusion community has developed a detailed materials database since the materials used in plasma confinement chambers and the plasma–wall interactions are pivotal to the operation and sustaining of a fusion plasma. Accordingly, the fusion plasma community has developed and maintained databases that detail and analyze the properties of relevant materials and their critical parameters for fusion environments. In Europe, this work has been performed under the EUROfusion program with the data recorded in the EUROfusion database and handbook [705]. The database has established protocols to obtain the raw data and introduce screening procedures and data storage to ensure quality and thence acceptance (and adoption) by the international community. Similarly, the IAEA has compiled data and published reviews for many years often resulting from IAEA CRPs, for example, the recent CRP on plasma–wall interaction for irradiated tungsten and tungsten alloys in fusion devices [706]. Such reviews are commonly published in the IAEA's journal series *Atomic and Plasma–Material Interaction Data for Fusion* (APID) with 18 volumes from 1991 to 2019 [707]. Unfortunately, not all these data are yet available online, but IAEA has a large repository of databases: https://amdis.iaea.org/databases/.

Newer resources for nuclear fusion energy research hosted by the IAEA focus on atomistic modeling of candidate materials for fusion reactors: MD simulations of collision cascades (CascadesDB [708]) and DFT simulations of radiation-induced defect structures (DefectDB [709]). These

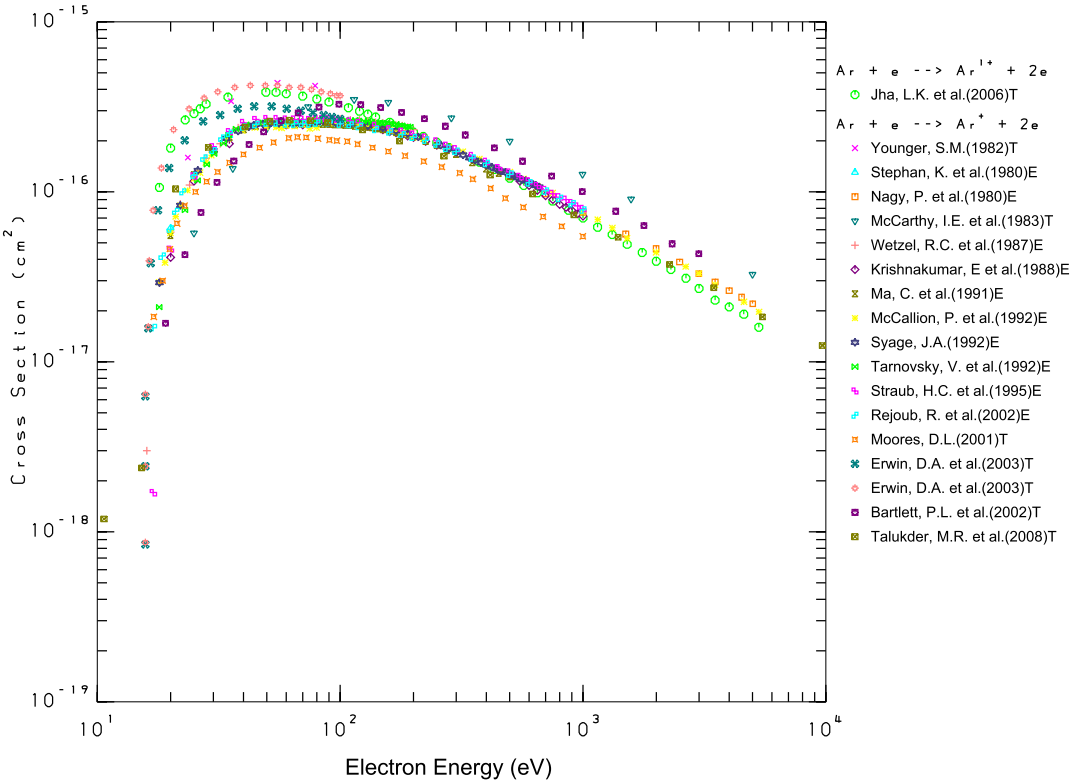

Fig. 43. Example of electron-impact ionization cross section of Ar atom, taken from NIFS database. T or E at the end of each legend indicates theoretical or experimental data.

have been developed and are maintained with the active support of the fusion materials modeling community and, in the case of CascadesDB, provide powerful visualization and data exploration tools [710] and allow downloads in multiple data formats (XML, JSON, and plain text).

A further database, under development, HCDB [711], hosts a heterogeneous collection of data in a hierarchical format, combining the structure of a relational database while providing some of the schemaless flexibility of NoSQL database technologies. For example, experimental results from a round-robin comparison exercise on deuterium retention in standardized steel samples may be stored in the same database as literature values for hydrogen diffusion coefficients in different materials without the need to construct new databases for each of these applications.

The IAEA's Atomic and Molecular Bibliographic Data System, AMBDAS, [712] includes data on surface processes including chemical reactions, desorption, reflection, secondary electron emission, sputtering, trapping (and detrapping), and AM processes on the surface, such as neutralization, ionization, and dissociation. The ALADDIN database [713] has both AM and particle-surface data, and together, these two online databases provide the most detailed and accessible materials data albeit with a focus on the fusion community and the materials used in fusion reactors.

[Nigel J. Mason and Christian Hill]

D. Plasma Chemistry Databases

Plasmas are strong sources of chemistry both in their treatment of surfaces and the (often complex) chemistry within the plasma leading to the creation of reactive species that, in turn, provide the main resource for the action of the plasma. It is, therefore, important that the chemistry of the plasma is understood if the plasma properties are to be characterized and, through this, natural plasma phenomena, such as aurorae unraveled. In the development of industrial plasmas, such chemistry should be both derived and modeled if the plasma's functionality is to be tuned and optimized for plasma usage. Thus, the assembly of plasma chemical databases is an important part of future plasma development.

Tennyson et al. [714], [715] defined three criteria for developing a chemistry-inclusive plasma model: 1) the chemistry should be *complete*, that is, it contains all the important reactions for the given plasma; 2) it should be *consistent*, that is, the reactions should not be unbalanced, thus resulting in the plasma composition being driven away from the true composition; and 3) the plasma chemistry should be *correct*. This last criterion is difficult to demonstrate on purely theoretical grounds alone and, therefore, requires validation by experimental measurements made in plasmas.

Addressing the first criterion, for a given plasma composition, there are sets of species that are present in the plasma and a set of processes, generally called reactions, which will link the species or different states of the species. This reaction set is described as the "chemistry" for that plasma. However, assembling plasma chemistries is far from straightforward since, even for relatively simple systems, such as a microwave molecular nitrogen plasma, some 15 species are necessary to characterize the plasma, including the seven lowest vibrationally excited states of the nitrogen molecule in the ground state $N_2(X^1\Sigma_g^+)v = 0$ to 6, the metastable molecule $N_2(A^3\Sigma_u^+)$, the ground state atom $N(^4S)$, two metastable atoms $N(^2D)$ and

TABLE III
ACTIVELY MAINTAINED DATABASES CONTAINING ELECTRON–MOLECULE COLLISION CROSS SECTIONS AND OTHER DATA TO IMPORTANCE FOR PLASMA MODELING APPLICATIONS

Database	Electron-collision data	Target field	Other data
LXCat [698]	Excitation processes	Plasma physics	Atomic cross sections
QDB [716] [717]	Excitation processes	Technological plasmas	Chemical reaction rates
NIFS [694]	Excitation processes	Fusion	Chemical reaction rates
NFRI [720]	Excitation processes	Fusion	Chemical reaction rates
ALADDIN [715] [721]	Excitation processes	Fusion	Chemical reaction rates
Phys4Entry [722]	Vibrational excitation	Atmospheric re-entry	Heavy particle inelastic cross sections.
BASECOL [723]	Rotational excitation	Astrophysics	Heavy particle inelastic cross sections.
KIDA [719]	Dissociative recombination	Astrophysics	Chemical reaction rates
UfDA [724]	Dissociative recombination	Astrophysics	Chemical reaction rates
IDEADB	Dissociative electron attachement		

$N(^2P)$, and five ionic species N, N^+, N_2^+, N_3^+, and N_4^+ [716]. With these 15 species, more than 100 "reactions" may be necessary to define the inherent plasma chemistry, most of which have never been measured. For even the simplest industrial plasmas, such as those used in etching, the number of reactions taking place may be more than a thousand making it unfeasible to make a "complete" model. It is, therefore, necessary to determine the "critical" or most important reactions to characterize and describe the physical and chemical properties of the plasma. However, since several important reactions remain completely uncharacterized (e.g., those involving molecular radicals), it is possible that models will neglect key processes due to the unavailability of such data.

This lack of data is, therefore, a challenge in meeting the second criterion that the dataset should be "consistent" since some reaction pathways may be indeterminate or even unknown such that production and destruction routes for important reactants may not be complete. For example, in atmospheric pressure plasmas, the role of water (humidity) may be an important criterion and explain differences in day-to-day operations. In atmospheric pressure plasmas, many ions are "solvated," and thus, their chemical properties are altered by their attachment to one or more water molecules, while, during the plasma operation, such clusters may be fragmented releasing reactive ions into the plasma once again. If such cluster chemistry is not accounted for, the true composition and density of reactive species (e.g., OH radicals) will not be accurate resulting in the modeled plasma composition being different from measurements.

The final criterion that the modeled plasma chemistry should be shown to be correct requires some modeled parameters to be measured. Selection of such parameters is not trivial; for example, the number density of some species may rely upon spectroscopic measurements. While spectroscopy may be used to identify species, deriving number densities by spectroscopic measurements is difficult since excited species are populated both by direct excitation and by "cascade" as higher excited atomic/molecular states decay into the lower state, and such cascade cross sections are largely unknown. Such cascade processes are responsible for more than 80% of the formation of metastable species in many plasmas.

Despite these challenges and limitations, plasma chemistry databases have been assembled for different research fields. One of the most complete is the KIDA database [717], a database for astrochemical (interstellar medium and planetary atmospheres) studies that contain over 700 species and up to 10 000 reactions tuned to the low-temperature environment of space. The data have been assembled into several "networks" for specific conditions (e.g., distinctive planetary atmospheres): https://kida.astrochem-tools.org/networks.html. This database provides references to all included reactions while commenting on their validity (making corrections where necessary) and where there are several alternative values that may make recommendations as to the values to use.

The QDB [714] is a commercial database that contains chemistry data for industrial plasma modeling from preassembled and validated chemistry sets, allowing users to assemble their own unique database for their specific plasma. It has about 50 preassembled datasets used in common plasma etching processes incorporating electron, heavy particle, photon collision cross sections, and AM species reaction rates. It also hosts some data for surface processes split into two categories: data for plasma simulations, such as sticking coefficients for atomic oxygen, atomic fluorine, fluorocarbons, and silane radicals; and data for surface mechanisms, such as specific etches, where it provides a set of individual reactions with their associated probabilities.

Such chemical databases are expected to increase in coming years as the chemistry induced by plasmas is utilized in more applications, including medical processes [643], [723] and waste treatment [724], [725].
[Jonathan Tennyson]

E. Low-Temperature Plasma Database

In the field of low-temperature plasma science, central databases providing fundamental data for the analysis, and interpretation of measurement results, theoretical modeling and simulations have been used and maintained for many years. These include, for example, the NIST atomic spectra database [726]; LXCat [727] for electron and ion scattering cross sections, swarm parameters, reaction rates, energy distribution functions, and so on; and Quantemol-DB [714] for plasma species, reactions, and chemistries. However, the results of application-oriented research in the area of low-temperature plasmas are mainly published in traditional journal publications and poorly structured, and often not accessible in digital form for direct reuse. This not only suspends the continuous life cycle of research data but also

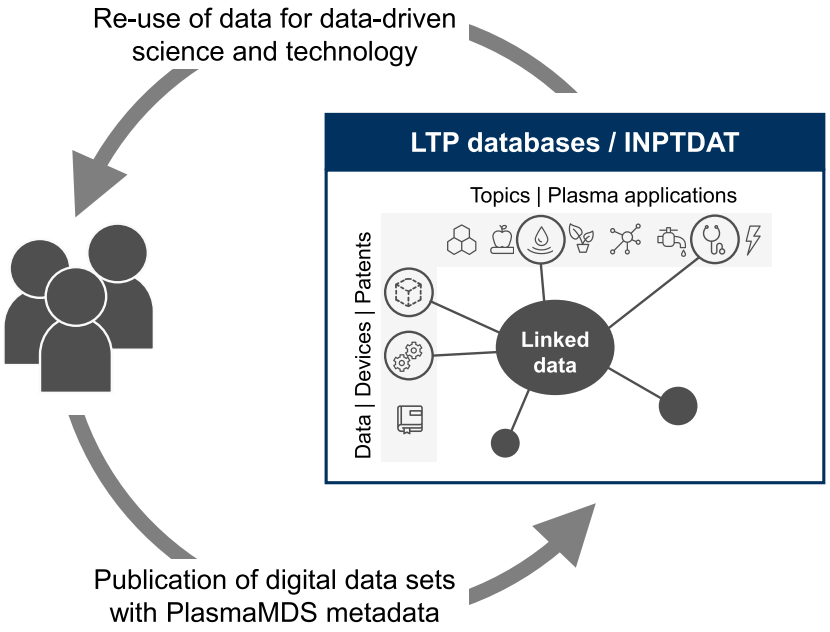

Fig. 44. Concept of a data life cycle supporting data-driven science and technology in low-temperature plasma science by means of INPTDAT and the PlasmaMDS.

inhibits technology transfer since comprehensive datasets for comparison and validation studies are often lacking. In particular, the application of AI/ML methods to data-driven science and technology requires large datasets in well-defined formats. Data must be shared with machine-readable metadata containing information on how the data can be accessed, how they can interoperate with applications or work flows for analysis, storage, and processing, and in which context they can be reused. Initiatives in many research fields are underway to develop or advance systems and standards for documentation and sharing of research data to meet these requirements and to make it easier to find such data and make it interoperable and reusable in accordance with the FAIR data principles [674], [728], [729], [730]. Furthermore, funding agencies and publishers are starting to issue policies requiring researchers to preserve and share the research data collected during the course of a research grant or presented in a paper. Both the practical needs and formal requirements have motivated work on providing a central database for research data in low-temperature plasma science.

In general, three options are available for publishing research data in digital form: first, institutional repositories, which are operated by universities or individual research institutions and accommodate data from all disciplines; second, subject-specific repositories for collecting research data from a specific research area; and third, generic repositories that are open to all types of data from any source, such as Figshare or Zenodo. Each option has its own advantages and disadvantages. Institutional solutions, for example, can be linked easily to local data management and quality assurance processes. Generic repositories generally impose no restrictions or quality criteria on the data, making them particularly easy for individual researchers to use. Subject-specific databases have the advantage over the former that the data can be documented and stored according to appropriate metadata standards and data models. This aspect is particularly important in the context of data-driven research where data should be findable and reusable by automated processes. Many research communities with large-scale experiments, and mostly, homogeneous data already have established solutions, e.g., high-energy physics and astrophysics [375], [731]. Research in low-temperature plasma science, however, is often characterized by small-scale table-top experiments involving diverse methods and devices. Furthermore, application-oriented research in the field of plasma science often involves researchers from other disciplines, such as electrical engineering, biology, and medicine. As a result, research data are extremely heterogeneous, and convenient infrastructures are needed to manage and link these data in the sense of making them available for data-driven research. The data platform INPTDAT and plasma MDS (PlasmaMDS), an MDS for the uniform description of data in the field of applied plasma science, have recently been developed to address this challenge [732]. As illustrated in Fig. 44, the concept underlying these developments is that data obtained in the course of research in a specific subject area by means of a specific experiment and involving specific devices are assigned by the data producers to the respective topic, to a concrete application if applicable, as well as to the experiment, devices, and substrates used. In this way, a graph of linked data and further information, e.g., from patents and device descriptions, is created, and research data available for specific applications, devices, and/or substrates can be found and reused immediately. This is particularly beneficial if similar experiments or devices are used in different subject areas and for various applications. An example from the field of plasma technology is a plasma source being used both in plasma surface technology for the functionalization of materials and in plasma medicine for biomedical applications. Up until now, the data and knowledge gained in the respective fields (plasma surface technology and plasma medicine) have only rarely been brought together and reused in an interdisciplinary manner. The concept implemented by INPTDAT and

PlasmaMDS supports cross-domain reuse of research data by making the data directly accessible for machines and scientists from different fields via linking with topics, applications, methods, and devices. If this approach is further developed and established in the following years according to the needs, and a community consensus on sharing and documenting research data is reached, a basis for the broad application of data-driven research, development, and technology transfer can be achieved. In this endeavor, data do not necessarily have to be collected in a central location but can remain with the data providers and will be linked via uniform metadata descriptions and a common metadata catalog. The research department Plasmas With Complex Interactions, Ruhr University Bochum, has already adapted this approach and, following the example of INPTDAT, set up its own data repository implementing PlasmaMDS [733]. With publicly shared and collaboratively developed software and standards, a basis for further dissemination has been provided [734].

In conclusion, the widespread reuse of data for data-driven research and technology transfer in low-temperature plasma science requires that more data are provided and described in appropriate formats. The open data platform INPTDAT and the PlasmaMDS are only the first steps in this direction. Further work is currently being carried out on semantic cross-linking of data by means of knowledge graphs [735], whereby the participation of the community in developing common terminologies, schemas, and ontologies for the extremely diverse requirements in different applications of low-temperature plasma science and technology will be important in the future. [Markus M. Becker]

F. Selections of Recommended Data

The compilation of data in itself is valuable but when confronted by multiple datasets for the same cross section or reaction how is the user to select one set over another? This is a major challenge for user community members who often do not have a detailed knowledge of the methods by which such data are collected and, thus, cannot easily distinguish between the myriad of data presented to them. When should they use experimental data, and when should they use theoretical data? Is the data collected or calculated by one methodology more reliable than that of another? Is newer data necessarily more reliable than older data? These questions are often asked by the user and modeling communities, and some data providers (such as QDB) offer a service to provide recommended datasets having the expertise to analyze the data and determine recommended and self-consistent datasets. However, more broadly, how are recommended datasets derived and is it necessary?

To answer the question of whether there is a need for recommended datasets, it is only necessary to consider the use of spectroscopy to determine the number density of excited species in a plasma. The cross sections used for a specific spectral emission may be used to determine the number density of the emitting species; if different cross-sectional data are used to calibrate different instruments viewing the plasma, then the same observational data will be "translated" into different number densities. Accordingly, for projects such as JET plasma and in future ITER, it is recognized that agreed cross sections for key diagnostics should be agreed [736].

Similarly, many of the discrepancies between different models may be due to the use of different cross sections and reaction rates rather than different physical and chemical processes included in each model. Unraveling the data used in different models and the influence of the choice of that data has attracted the attention of both data compilers and users in recent years with discussions of the methodology to provide "recommended data" being held in several meetings, for example, those chaired by the IAEA and data centers, such as VAMDC (https://vamdc.org) and VESPA (http://www.europlanet-vespa.eu). Some broad guidelines for recommending datasets have emerged from such meetings.

1) All recommended data should have been previously published and, therefore, have been subject to peer review.
2) Estimates of uncertainties in the data should be provided. This is standard for experimental data but has been less common in theoretical data. However, recently, publishers have required a discussion of uncertainties in theoretical/computational data [737].
3) It is preferable for recommended data to be in datasets rather than individual processes. For example, consider electron scattering cross section data: individual cross sections may be recommended from different sources, but the summation of these individual cross sections should be consistent with the recommended total cross section. Similarly, integrated differential cross sections should be consistent with the recommended integral cross section, summed partial ionization cross sections should be consistent with the recommended total ionization cross section, and momentum transfer cross sections should be consistent with recommended elastic and inelastic cross sections.

These guidelines demonstrate that providing recommended datasets is a challenging exercise and requires a wide knowledge of the methods by which such data is generated and often the researchers involved. Experimental data are often prone to systematic effects that are known to the community; for example, the community may know the energy and angular ranges over which data have been demonstrated to be reliable and ranges in which systematic effects may lead to larger uncertainties. Extrapolation of data over angular ranges to obtain an integral cross section may be known to be problematic in some systems (e.g., electron scattering from targets with dipole moments may show strong forward scattering in regions where experimental errors are large). Some theoretical methods may also be known to be more accurate over some particular energy range. These limitations are not always clear to the general user looking at published data but are known by the community. Therefore, it is the community with its expertise that is best suited to provide recommended datasets. However, with a few exceptions (e.g., Section I and the atmospheric community with its HITRAN database), there are few institutional structures to compile and recommend datasets in part due to a lack of funding for such activities.

International organizations, such as the IAEA, are able to provide a stable and long-term platform for database resources serving particular communities. As computing infrastructure, including cloud computing facilities, become cheaper and more available, this has enabled such institutions to collect and serve a wider variety of data. For example, the well-established ALADDIN database of evaluated plasma collisional data at the IAEA is now supplemented by a larger database of unevaluated data, CollisionDB [738], which accepts (with provenance) data from all published sources and provides a searchable interface enabling such data to be compared, aggregated, and assessed.

Recent European Union activities, leading to the creation of VAMDC and VESPA, are encouraging, but ensuring the sustainability of such efforts remains a challenge. Hence, most data compilations are due to the efforts and enthusiasm of individuals, such as the KIDA and LXCat astrochemistry, and plasma databases with data for individual targets are published by small academic consortia and often result from a specific need that they have identified for other research; Table III gives a summary of active databases in this area. Several initiatives have tried to provide a longer term approach to multiple targets, for example, initiatives to develop recommended datasets for electron scattering from molecular targets used in the semiconductor industry by Christophorou and Olthoff [739] at NIST and, more recently, teams led by Song et al. [740] at the Plasma Technology Research Center, Korea Institute of Fusion Energy, Sout Korea; both are focused on low-temperature plasmas. If such recommended datasets are to be updated and new ones compiled in the future, much greater emphasis and funding support must be given to such activities and the next generation of researchers convinced of the need to participate and lead such initiatives.

G. Challenges and Outlook

The variety of plasma-based systems has, and will continue, to expand from the study of astrochemistry and planetary atmospheres to the use of atmospheric plasmas for waste treatment and medicine. The need to redesign basic industrial plasmas for semiconductor processing using feedstock gases that comply with environmental protection (e.g., low global warming and ozone depletion potentials) has been recognized since the Kyoto protocol (designed in 1997 and entered into force in 2005), but, at the recent COP26 meeting, it was recognized that the targets set for 2020 had not been met and with current global uncertainties, and few are optimistic of new targets being met. The design of new plasma treatment systems and their optimization both in energy and net emissions are likely to become even more important while the need to accelerate the development of commercial nuclear fusion as an alternative to fossil fuels will place new emphasis on knowledge of AM collisions, spectroscopy, and, crucially, surface interactions in such plasmas.

The collection, compilation, and preparation of recommended datasets for plasma studies, therefore, remain one of the most important, yet also the most challenging aspects of modern plasma research. The increasing development of "virtual factories" and the concept of a "digital twin" [741], [742] in which a plasma processing plant and the procedure are modeled prior to construction places increasing emphasis on the quality and quantity of the input data used in such models. However, despite the recognition of the need to collect and compile such data, the community is small, and in many cases, such as AM data, this community is steadily declining in numbers as other areas of science and technology attract more funding. This is a dangerous trend since the production of such data underpins all aspects of plasma technology, from the provision of diagnostics for characterizing the plasma to the design of plasma itself for specific applications.

The amount of data required is already far in excess of that practical to assemble by experiment, with many targets being unsuitable for experimental research (short-lived radicals, highly reactive species, and species that are obtained only from highly toxic precursors); hence, the majority of the data must be evaluated by theoretical calculations with the limited experimental data being used to benchmark such calculations. While semiempirical methods may be attractive to users, and commercial packages, such as Quantemol, are available, they should be used with caution, and the user is advised to cooperate with a more experienced user and take advantage of expert advice where offered (as in the case of Quantemol [743]).

Recently, there have been some attempts to use ML [368], [744] to derive datasets with an ML-based method being to construct a model for predicting total ionization cross sections Q_{ion} of large molecules without the high cost of ab initio calculations. The model is learned from the data composed of the calculated Q_{ion} of the small molecules with fewer constituent atoms and the electron numbers of the corresponding molecules in a training set by an SVM [743]. Initial results are in broad agreement with experimental and semiclassical calculations, so they may be valid for higher energies, but whether they are robust enough for lower energies where the structural properties of the target are important and "resonances" are formed is an open question, and it is such low energy interactions that are most relevant in the myriad of low-temperature industrial plasmas.

In conclusion, the need for the collection and compilation of fundamental data underpinning the operation of plasmas is widely recognized by the community, and there have been several attempts to address the challenge of providing such data to user communities with the creation of several international databases. However, this work remains poorly supported and too often relies on the efforts of a few active individuals, which is not sustainable. A long-term strategy for the maintenance and review of databases is required and should be instilled in the training of the next generation of researchers.
[Nigel J. Mason]

IX. Conclusion

In this review article, the latest studies and their results in DDPS are summarized for various applications ranging from basic plasma physics to nuclear fusion, space and astronomical plasmas, and industrial plasmas. In addition, we presented a review of fundamental data science that serves as the basis for all analytical techniques used in different plasma applications and databases that serve as vital resources for the wide

scientific community. It is seen that many common techniques and ideas are used for different applications. From a large amount of observational or computational data, some important features are extracted by regression or classification techniques, and such features are used to control plasma dynamics or predict certain properties of the system involving plasmas. A large amount of data is also used to construct surrogate models for the systems of interest, and such models can be used as alternatives to the corresponding first-principle-based computation of the system equations. While the first-principle computation of a model system continues to be important for a better understanding of the underlying mechanisms of the system, there are many other important uses of such computation. One of such important uses is the prediction of system behavior. A surrogate model that requires only short or instantaneous computation time can be used to predict the system behaviors in real time. The reduction of large-scale computation is one of the goals that data-driven plasmas science attempts to achieve.

Shortage or lack of experimental data is one of the most important challenges in this field. Probably, this problem is more application-specific, and what kind of data should be collected and in what way depend strongly on the system of interest. Fast and systematic ways of obtaining useful data, such as HTS, will continue to be sought after in this field with specially designed experimental systems. The design of experiments with the Bayesian inference, for example, is also widely used for such purposes.

Although we attempted to cover an extensive range of examples of data-driven analyses in plasma science, what is presented in this review article is by no means exhaustive. Unfortunately, many important studies are still missing in this article. Furthermore, the field is rapidly developing, and within several years, some of the results written here may become obsolete. This is why we named this review article "2022 Review" with the year of publication. We hope to update this review article with more extensive examples of the latest important developments as the field progresses.

[Satoshi Hamaguchi]

ACKNOWLEDGMENT

Zhehui Wang wishes to thank Christoph Räth (Institut für Materialphysik im Weltraum, Deutsches Zentrum für Luft- und Raumfahrt, Germany), Chengkun Huang (Los Alamos National Laboratory, Los Alamos, NM, USA), Ghanshyam Pilania (Los Alamos National Laboratory), and Platon Karpov (Department of Astronomy and Astrophysics, University of California at Santa Cruz, Santa Cruz, CA, USA) for stimulating discussions. All examples described in this article are available and fully documented in the open-source AidaPy software at https://gitlab.com/aidaspace/aidapy. For Section VII-B, the authors would like to thank Kaihan Ashtiani, Michal Danek, Paul Franzen, Sassan Roham, and other colleagues and mentors at Lam Research Corporation for valuable discussions over the years to help them better understand the impact that digital transformation will have on the semiconductor industry. The work in Section VII-C was technically supported by the Executive Vice-President Jinwoo Park, Jaehyung Lee, Jeonggen Yoo, and Insso Cho of Samsung Display Company Ltd. In Section VII-D, they would like to thank Alok Ranjan and Hiromasa Mochiki for their support and valuable feedback. For Section VII-E, Jan Trieschmann acknowledges valuable input from Tobias Gergs, in particular, Fig. 35, helpful discussions with Borislav Borislavov, and continuous support from Thomas Mussenbrock. The work in Section VII-F was partly carried out within the framework of the EUROfusion Consortium. Views and opinions expressed here do not necessarily reflect those of the European Union or the European Commission. For Section VIII-B, Izumi Murakami acknowledges many atomic physicists supporting the AM database activities and users from the communities. The Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy under Contract 89233218CNA000001. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-NA-0003525.

Rushil Anirudh, Peer-Timo Bremer, Jim A. Gaffney, Kelli Humbird, Bogdan Kustowski, Tammy Ma, J. Luc Peterson, Brian K. Spears, Jayaraman Thiagarajan, Brian Van Essen, and Xueqiao Xu are with the Lawrence Livermore National Laboratory, Livermore, CA 94550 USA (e-mail: anirudh1@llnl.gov; bremer5@llnl.gov; gaffney3@llnl.gov; humbird1@llnl.gov; kustowski1@llnl.gov; ma8@llnl.gov; peterson76@llnl.gov; spears9@llnl.gov; jayaramanthi1@llnl.gov; vanessen1@llnl.gov; xu2@llnl.gov).

Rick Archibald, Ana Gainaru, Scott Klasky, and David Pugmire are with the Oak Ridge National Laboratory and the Department of Electrical and Computer Engineering, Oak Ridge, TN 37830 USA (e-mail: archibaldrk@ornl.gov; gainarua@ornl.gov; klasky@ornl.gov; pugmire@ornl.gov).

M. Salman Asif is with the University of California at Riverside, Riverside, CA 92521 USA (e-mail: sasif@ece.ucr.edu).

Markus M. Becker and Kristian Wende are with the Leibniz Institute for Plasma Science and Technology (INP), 17489 Greifswald, Germany (e-mail: markus.becker@inp-greifswald.de; kristian.wende@inp-greifswald.de).

Sadruddin Benkadda is with CNRS, PIIM UMR 7345, Aix-Marseille University, 13007 Marseille, France (e-mail: sadruddin.benkadda@univ-amu.fr).

Rick H. S. Budé and Jan van Dijk are with the Department of Applied Physics, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: r.h.s.bude@tue.nl; J.v.Dijk@tue.nl).

C. S. Chang and R. M. Churchill are with the Princeton Plasma Physics Laboratory, Princeton, NJ 08540 USA (e-mail: cschang@pppl.gov; rchurchi@pppl.gov).

Lei Chen is with the Institute of Ion Physics and Applied Physics, University of Innsbruck, 6020 Innsbruck, Austria (e-mail: Lei.Chen@iter.org).

Jonathan Citrin is with the Dutch Institute for Fundamental Energy Research (DIFFER), 5600 HH Eindhoven, The Netherlands, and also with the Science and Technology of Nuclear Fusion Group, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: J.Citrin@differ.nl).

Walter Gekelman is with the Department of Physics and Astronomy, University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: gekelman@physics.ucla.edu).

Tom Gibbs is with NVIDIA, Santa Clara, CA 95051 USA (e-mail: tgibbs@nvidia.com).

Satoshi Hamaguchi is with the Center for Atomic and Molecular Technologies, Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan (e-mail: hamaguch@ppl.eng.osaka-u.ac.jp).

Christian Hill is with the Department of Nuclear Sciences and Applications, International Atomic Energy Agency, 1400 Vienna, Austria (e-mail: Ch.Hill@iaea.org).

Sören Jalas is with the Center for Free-Electron Laser Science and the Department of Physics, Universität Hamburg, 22761 Hamburg, Germany (e-mail: soeren.jalas@desy.de).

Satoru Kawaguchi is with the Division of Information and Electronic Engineering, Graduate School of Engineering, Muroran Institute of Technology, Muroran 050-8585, Japan (e-mail: skawaguchi@mmm.muroran-it.ac.jp).

Gon-Ho Kim is with the Department of Nuclear Engineering, Seoul National University, Seoul 151-741, Republic of Korea (e-mail: ghkim@snu.ac.kr).

Manuel Kirchen and Rob J. Shalloo are with the Deutsches Elektronen-Synchrotron (DESY), 22607 Hamburg, Germany (e-mail: manuel.kirchen@desy.de; rob.shalloo@desy.de).

John L. Kline, Wenting Li, Alexander Scheinker, Zhehui Wang, and Xinhua Zhang are with the Los Alamos National Laboratory, Los Alamos, NM 87545 USA (e-mail: jkline@lanl.gov; wenting@lanl.gov; ascheink@lanl.gov; zwang@lanl.gov; sean.xinhua.zhang@gmail.com).

Karl Krushelnick is with the Center for Ultrafast Optical Science, University of Michigan, Ann Arbor, MI 48109, USA (e-mail: kmkr@umich.edu).

Giovanni Lapenta is with the Department of Mathematics, University of Leuven (KU Leuven), 3000 Leuven, Belgium (e-mail: giovanni.lapenta@kuleuven.be).

Nigel J. Mason is with the Department of Physical Sciences, University of Kent, CT2 7NH Canterbury, U.K. (e-mail: n.j.mason@kent.ac.uk).

Ali Mesbah is with the Department of Chemical and Biomolecular Engineering, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: mesbah@berkeley.edu).

Craig Michoski is with the Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: michoski@gmail.com).

Todd Munson is with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439 USA (e-mail: tmunson@mcs.anl.gov).

Izumi Murakami is with the National Institute for Fusion Science, National Institutes of Natural Sciences, Gifu 509-5292, Japan, and also with Graduate Institute for Advanced Studies, SOKENDAI, Gifu 509-5292, Japan (e-mail: murakami.izumi@nifs.ac.jp).

Habib N. Najm is with Sandia National Laboratories, Livermore, CA 94551 USA (e-mail: hnnajm@sandia.gov).

K. Erik J. Olofsson, Brian Sammuli, and David P. Schissel are with General Atomics, San Diego, CA 92186 USA (e-mail: olofsson@fusion.gat.com; sammuli@fusion.gat.com; schissel@fusion.gat.com).

Seolhye Park and Jaegu Seong are with Samsung Display Company Ltd., Asan-si 31454, Republic of Korea (e-mail: druyvesteyndf@gmail.com; seongjaegu@gmail.com).

Michael Probst is with the Institute of Ion Physics and Applied Physics, University of Innsbruck, 6020 Innsbruck, Austria, and also with the School of Molecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology, Rayong 21210, Thailand (e-mail: Michael.Probst@uibk.ac.at).

Kapil Sawlani is with Lam Research Corporation, Fremont, CA 94538 USA (e-mail: Kapil.Sawlani@lamresearch.com).

Jun Shinagawa and Peter Ventzek are with Tokyo Electron America, Inc., Austin, TX 78741 USA (e-mail: Jun.Shinagawa@us.tel.com; peter.ventzek@us.tel.com).

Jonathan Tennyson is with the Department of Physics and Astronomy, University College London, WC1E 6BT London, U.K. (e-mail: j.tennyson@ucl.ac.uk).

Catalin M. Ticoş is with the National Institute for Laser, Plasma and Radiation Physics, 077125 Măgurele, Romania (e-mail: catalin.ticos@elinp.ro).

Jan Trieschmann is with the Theoretical Electrical Engineering, Faculty of Engineering, Kiel University, 24143 Kiel, Germany (e-mail: jt@tf.uni-kiel.de).

Haimin Wang and Jason T. L. Wang are with the Institute for Space Weather Sciences, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: haimin.wang@njit.edu; jason.t.wang@njit.edu).

Hiroshi Yamada and Tatsuya Yokoyama are with the Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8568, Japan (e-mail: yamada.hiroshi@edu.k.u-tokyo.ac.jp; yokoyama.tatsuya@qst.go.jp).

Authors' Contributions

Sadruddin Benkadda, C. S. Chang, Satoshi Hamaguchi, Nigel J. Mason, Brian K. Spears, and Zhehui Wang acted as the section conveners of this review article. Especially, C. S. Chang hosted the manuscript website, and organized and edited Sections II and IV, Satoshi Hamaguchi for Section VII, Nigel J. Mason for Section VIII, Brian K. Spears for Section V, and Zhehui Wang for Sections III and VI. Satoshi Hamaguchi and Zhehui Wang also checked and edited the final version of this entire article. The author(s) of each subsection is(are) listed at the end of the subsection.

References

[1] M. L. Green et al., "Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies," *Appl. Phys. Rev.*, vol. 4, no. 1, Mar. 2017, Art. no. 011105.

[2] T. Hey, S. Tansley, and K. Tolle, Eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA, USA: Microsoft Corporation, 2009.

[3] F. Wilczek, "Physics in 100 years," 2015, *arXiv:1503.07735*.

[4] J. You, "DARPA sets out to automate research," *Science*, vol. 347, no. 6221, p. 465, Jan. 2015.

[5] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

[6] D. Liu, M. R. Maxey, and G. E. Karniadakis, "Simulations of dynamic self-assembly of paramagnetic microspheres in confined microgeometries," *J. Micromech. Microeng.*, vol. 15, no. 12, pp. 2298–2308, Dec. 2005.

[7] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 2. San Francisco, CA, USA: Morgan Kaufmann, 1995, pp. 1137–1143.

[8] I. Guyon, M. Nikravesh, and S. G. A. L. A. Zadeh, *Feature Extraction*. Springer, 2006.

[9] S. Li, N. Marsaglia, C. Garth, J. Woodring, J. Clyne, and H. Childs, "Data reduction techniques for simulation, visualization and data analysis," *Comput. Graph. Forum*, vol. 37, no. 6, pp. 422–447, Sep. 2018.

[10] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.

[11] J. Seward. (1996). *Bzip2 and Libbzip2*. [Online]. Available: http://www.bzip.org

[12] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 520–540, Jun. 1987.

[13] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 337–343, May 1977.

[14] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 5, pp. 530–536, Sep. 1978.

[15] P. Deutsch, "DEFLATE compressed data format specification version 1.3," ALADDIN Enterprises, Tech. Rep., 1996.

[16] GNU. (2020). *The GZIP Home Page*. [Online]. Available: http://www.gzip.org/

[17] J. L. Gailly and M. Adler. (2017). *Zlib: A Massively Spiffy Yet Delicately Unobtrusive Compression Library*. [Online]. Available: https://zlib.net/

[18] P. Lindstrom and M. Isenburg, "Fast and efficient compression of floating-point data," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 1245–1250, Oct. 2006.

[19] Z. Gong et al., "MLOC: Multi-level layout optimization framework for compressed scientific data exploration with heterogeneous access patterns," in *Proc. 41st Int. Conf. Parallel Process. (ICPP)*, Sep. 2012, pp. 239–248.

[20] J. Diffenderfer, A. L. Fox, J. A. Hittinger, G. Sanders, and P. G. Lindstrom, "Error analysis of ZFP compression for floating-point data," *SIAM J. Sci. Comput.*, vol. 41, no. 3, pp. A1867–A1898, Jan. 2019.

[21] S. Lakshminarasimhan et al., "ISABELA for effective in situ compression of scientific data," *Concurrency Comput., Pract. Exper.*, vol. 25, no. 4, pp. 524–540, Feb. 2013.

[22] S. Di and F. Cappello, "Fast error-bounded lossy HPC data compression with SZ," in *Proc. IEEE Int. Parallel Distrib. Process. Symp. (IPDPS)*, Chicago, IL, USA, Jun. 2016, pp. 730–739.

[23] S. Li, S. Jaroszynski, S. Pearse, L. Orf, and J. Clyne, "VAPOR: A visualization package tailored to analyze simulation data in earth system science," *Atmosphere*, vol. 10, no. 9, p. 488, Aug. 2019.

[24] A. Skodras, C. Christopoulis, and T. Ebrahimi, "The JPEG2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36–58, Sep. 2001.

[25] P. Lindstrom, "Fixed-rate compressed floating-point arrays," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2674–2683, Dec. 2014.

[26] K. Zhao et al., "SDRBench: Scientific data reduction benchmark for lossy compressors," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 2716–2724.

[27] J. Zhang, X. Zhuo, A. Moon, H. Liu, and S. W. Son, "Efficient encoding and reconstruction of HPC datasets for checkpoint/restart," in *Proc. 35th Symp. Mass Storage Syst. Technol. (MSST)*, May 2019, pp. 79–91.

[28] M. Ainsworth, O. Tugluk, B. Whitney, and S. Klasky, "Multilevel techniques for compression and reduction of scientific data—The multivariate case," *SIAM J. Sci. Comput.*, vol. 41, no. 2, pp. A1278–A1303, Jan. 2019.

[29] R. Ballester-Ripoll, P. Lindstrom, and R. Pajarola, "TTHRESH: Tensor compression for multidimensional visual data," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 9, pp. 2891–2903, Sep. 2020.

[30] K. Weiss and L. De Floriani, "Simplex and diamond hierarchies: Models and applications," *Comput. Graph. Forum*, vol. 30, no. 8, pp. 2127–2155, Dec. 2011.

[31] K. Wu et al., "FastBit: Interactively searching massive data," *J. Phys., Conf. Ser.*, vol. 180, Jul. 2009, Art. no. 012053.

[32] M. Larsen et al., "A flexible system for in situ triggers," in *Proc. Workshop Situ Infrastructures Enabling Extreme-Scale Anal. Vis. (ISAV)*, Dallas, TX, USA, Nov. 2018, pp. 1–6.

[33] J. Ling et al., "Using feature importance metrics to detect events of interest in scientific computing applications," in *Proc. IEEE 7th Symp. Large Data Anal. Vis. (LDAV)*, Oct. 2017, pp. 55–63.

[34] M. Salloum, J. C. Bennett, A. Pinar, A. Bhagatwala, and J. H. Chen, "Enabling adaptive scientific workflows via trigger detection," in *Proc. 1st Workshop Situ Infrastruct. Enabling Extreme-Scale Anal. Vis. (ISAV)*, New York, NY, USA, Nov. 2015, pp. 41–45.

[35] C. Heine et al., "A survey of topology-based methods in visualization," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 643–667, Jun. 2016.

[36] J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D. H. Rogers, and M. Petersen, "An image-based approach to extreme scale in situ visualization and analysis," in *Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal.*, Nov. 2014, pp. 424–434.

[37] J. Choi et al., "Neural data compression for physics plasma simulation," in *Proc. Neural Compress., From Inf. Theory Appl.-Workshop ICLR*, 2021, pp. 1–5.

[38] A. Glaws, R. King, and M. Sprague, "Deep learning for in situ data compression of large turbulent flow simulations," *Phys. Rev. Fluids*, vol. 5, no. 11, Nov. 2020, Art. no. 114602.

[39] T. Liu, J. Wang, Q. Liu, S. Alibhai, T. Lu, and X. He, "High-ratio lossy compression: Exploring the autoencoder to compress scientific data," *IEEE Trans. Big Data*, vol. 9, no. 1, pp. 22–36, Feb. 2023.

[40] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.

[41] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[42] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[43] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[44] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems," *J. Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 113–127, Jul. 2006.

[45] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. NIPS*, vol. 14, 2001, pp. 585–591.

[46] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 10, pp. 5591–5596, May 2003.

[47] P. Ray, S. S. Reddy, and T. Banerjee, "Various dimension reduction techniques for high dimensional data analysis: A review," *Artif. Intell. Rev.*, vol. 54, pp. 1–43, Jan. 2021.

[48] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.

[49] G. Pang and G. E. Karniadakis, "Physics-informed learning machines for partial differential equations: Gaussian processes versus neural networks," in *Emerging Frontiers in Nonlinear Science*. Cham, Switzerland: Springer, 2020, pp. 323–343.

[50] M. A. Miller, R. M. Churchill, A. Dener, C. S. Chang, T. Munson, and R. Hager, "Encoder–decoder neural network for solving the nonlinear Fokker–Planck–Landau collision operator in XGC," *J. Plasma Phys.*, vol. 87, no. 2, Apr. 2021.

[51] C. Ma, B. Zhu, X.-Q. Xu, and W. Wang, "Machine learning surrogate models for Landau fluid closure," *Phys. Plasmas*, vol. 27, no. 4, Apr. 2020, Art. no. 042502.

[52] R. Maulik, N. A. Garland, J. W. Burby, X.-Z. Tang, and P. Balaprakash, "Neural network representability of fully ionized plasma fluid model closures," *Phys. Plasmas*, vol. 27, no. 7, Jul. 2020, Art. no. 072106.

[53] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, "Machine learning–accelerated computational fluid dynamics," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 21, 2021, Art. no. e2101784118.

[54] F. Krüger, T. Gergs, and J. Trieschmann, "Machine learning plasma-surface interface for coupling sputtering and gas-phase transport simulations," *Plasma Sources Sci. Technol.*, vol. 28, no. 3, Mar. 2019, Art. no. 035002.

[55] B. Enders et al., "Cross-facility science with the superfacility project at LBNL," in *Proc. IEEE/ACM 2nd Annu. Workshop Extreme-Scale Experiment-Loop Comput. (XLOOP)*, Nov. 2020, pp. 1–7.

[56] E. Deelman et al., "Pegasus, a workflow management system for science automation," *Future Gener. Comput. Syst.*, vol. 46, pp. 17–35, May 2015.

[57] B. Ludäscher et al., "Scientific workflow management and the Kepler system," *Concurr. Comput., Pract. Exper.*, vol. 18, Aug. 2006, Art. no. 10391065.

[58] A. Merzky, M. Santcroos, M. Turilli, and S. Jha, "Radical-pilot: Scalable execution of heterogeneous and dynamic workloads on supercomputers," 2015, *arXiv:1512.08194*.

[59] J. Cummings et al., "EFFIS: An end-to-end framework for fusion integrated simulation," in *Proc. 18th Euromicro Conf. Parallel, Distrib. Netw.-Based Process.*, Feb. 2010, pp. 428–434.

[60] J. F. Lofstead, S. Klasky, K. Schwan, N. Podhorszki, and C. Jin, "Flexible IO and integration for scientific codes through the adaptable IO system (ADIOS)," in *Proc. 6th Int. Workshop Challenges Large Appl. Distrib. Environ. (CLADE)*. New York, NY, USA: Association for Computing Machinery, Jun. 2008, pp. 15–24.

[61] R. Barreto et al., "Managing and monitoring scientific workflows through dashboards," in *Proc. Microsoft eSci. Workshop*, 2007, p. 108.

[62] K. Mehta et al., "A codesign framework for online data analysis and reduction," in *Proc. IEEE/ACM Workflows Support Large-Scale Sci. (WORKS)*, Nov. 2019, pp. 11–20.

[63] E. Suchyta et al., "The exascale framework for high fidelity coupled simulations (EFFIS): Enabling whole device modeling in fusion science," *Int. J. High Perform. Comput. Appl.*, vol. 36, no. 1, pp. 106–128, Jan. 2022.

[64] E. D'Azevedo et al., "The fusion code XGC: Enabling kinetic study of multi-scale edge turbulent transport in ITER," Simmetrix, Clifton Park, NY, USA, Tech. Rep., Nov. 2017.

[65] S. Nickolay, E.-S. Jung, R. Kettimuthu, and I. Foster, "Towards accommodating real-time jobs on HPC platforms," 2021*arXiv:2103.13130*.

[66] D. Wang, E.-S. Jung, R. Kettimuthu, I. Foster, D. J. Foran, and M. Parashar, "Supporting real-time jobs on the IBM blue gene/Q: Simulation-based study," in *Job Scheduling Strategies for Parallel Processing* (Lecture Notes in Computer Science), vol. 10773, 2018.

[67] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*. New York, NY, USA: Springer-Verlag, 1991.

[68] H. N. Najm, "Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics," *Annu. Rev. Fluid Mech.*, vol. 41, no. 1, pp. 35–52, Jan. 2009.

[69] O. L. Maître and O. Knio, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*, 1st ed. Springer, Apr. 2010.

[70] R. Smith, *Uncertainty Quantification: Theory, Implementation and Applications*. Philadelphia, PA, USA: SIAM, 2013.

[71] C. Soize, *Uncertainty Quantification: An Accelerated Course With Advanced Applications in Computational Engineering*. Springer, 2017.

[72] L. Tenorio, *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems*. Philadelphia, PA, USA: SIAM, 2017.

[73] D. Sivia, *Data Analysis: A Bayesian Tutorial*. Oxford Science, 1996.

[74] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models (with discussion)," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 63, no. 3, pp. 425–464, 2001.

[75] B. P. Carlin and T. A. Louis, *Bayesian Methods for Data Analysis*. Boca Raton, FL, USA: Chapman & Hall, 2011.

[76] R. Ghanem, "Ingredients for a general purpose stochastic finite elements implementation," *Comput. Methods Appl. Mech. Eng.*, vol. 168, pp. 19–34, Jan. 1999.

[77] O. Le Maître, O. Knio, H. Najm, and R. Ghanem, "A stochastic projection method for fluid flow: I. Basic formulation," *J. Comput. Phys.*, vol. 173, pp. 481–511, Nov. 2001.

[78] B. J. Debusschere, H. N. Najm, P. P. Pébay, O. M. Knio, R. G. Ghanem, and O. P. Le Maître, "Numerical challenges in the use of polynomial chaos representations for stochastic processes," *SIAM J. Sci. Comput.*, vol. 26, no. 2, pp. 698–719, Jan. 2004.

[79] O. Le Maître, M. Reagan, H. Najm, R. Ghanem, and O. Knio, "A stochastic projection method for fluid flow: II. Random process," *J. Comput. Phys.*, vol. 181, pp. 9–44, Sep. 2002.

[80] D. Xiu and G. E. Karniadakis, "Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos," *Comput. Methods Appl. Mech. Eng.*, vol. 191, no. 43, pp. 4927–4948, Sep. 2002.

[81] D. Xiu and G. E. Karniadakis, "The Wiener–Askey polynomial chaos for stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 24, no. 2, pp. 619–644, Jan. 2002.

[82] D. Ghosh and R. Ghanem, "Stochastic convergence acceleration through basis enrichment of polynomial chaos expansions," *Int. J. Numer. Meth. Eng.*, vol. 73, no. 2, pp. 162–174, 2008.

[83] Y. M. Marzouk, H. N. Najm, and L. A. Rahn, "Stochastic spectral methods for efficient Bayesian solution of inverse problems," *J. Comput. Phys.*, vol. 224, no. 2, pp. 560–586, Jun. 2007.

[84] Y. M. Marzouk and H. N. Najm, "Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems," *J. Comput. Phys.*, vol. 228, no. 6, pp. 1862–1902, Apr. 2009.

[85] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[86] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007.

[87] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[88] B. Moore and B. Natarajan, "A general framework for robust compressive sensing based nonlinear regression," in *Proc. IEEE 7th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2012, pp. 225–228.

[89] V. Barthelmann, E. Novak, and K. Ritter, "High dimensional polynomial interpolation on sparse grids," *Adv. Comput. Math.*, vol. 12, pp. 273–288, Mar. 2000.

[90] D. Xiu and J. S. Hesthaven, "High-order collocation methods for differential equations with random inputs," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 1118–1139, Jan. 2005.

[91] I. Babuška, F. Nobile, and R. Tempone, "A stochastic collocation method for elliptic partial differential equations with random input data," *SIAM Rev.*, vol. 52, no. 2, pp. 317–355, Jan. 2010.

[92] A. Narayan and J. D. Jakeman, "Adaptive leja sparse grid constructions for stochastic collocation and high-dimensional approximation," *SIAM J. Sci. Comput.*, vol. 36, no. 6, pp. A2952–A2983, Jan. 2014.

[93] M. K. Stoyanov and C. G. Webster, "A dynamically adaptive sparse grids method for quasi-optimal interpolation of multidimensional functions," *Comput. Math. Appl.*, vol. 71, no. 11, pp. 2449–2465, Jun. 2016.

[94] L. Grasedyck, D. Kressner, and C. Tobler, "A literature survey of low-rank tensor approximation techniques," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 53–78, Aug. 2013.

[95] I. V. Oseledets, "Constructive representation of functions in low-rank tensor formats," *Constructive Approximation*, vol. 37, no. 1, pp. 1–18, Feb. 2013.

[96] A. A. Gorodetsky and J. D. Jakeman, "Gradient-based optimization for regression in the functional tensor-train format," *J. Comput. Phys.*, vol. 374, pp. 1219–1238, Dec. 2018.

[97] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems*, vol. 8, D. S. Teuretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA, USA: MIT Press, 1996.

[98] A. G. Wilson, D. A. Knowles, and Z. Ghahramani, "Gaussian process regression networks," 2011, pp. 1–17, *arXiv:1110.4411*.

[99] I. Bilionis, N. Zabaras, B. A. Konomi, and G. Lin, "Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification," *J. Comput. Phys.*, vol. 241, pp. 212–239, May 2013.

[100] R. K. Tripathy and I. Bilionis, "Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification," *J. Comput. Phys.*, vol. 375, pp. 565–588, Dec. 2018.

[101] C. Fiorina, A. Scolaro, D. Siefman, M. Hursin, and A. Pautz, "Artificial neural networks as surrogate models for uncertainty quantification and data assimilation in 2-D/3-D fuel performance studies," *J. Nucl. Eng.*, vol. 1, no. 1, pp. 54–62, Nov. 2020.

[102] M. S. Eldred, L. W. T. Ng, M. F. Barone, and S. P. Domino, "Multifidelity uncertainty quantification using spectral stochastic discrepancy models," in *Handbook of Uncertainty Quantification*. Cham, Switzerland: Springer, 2017, pp. 991–1036.

[103] B. Peherstorfer, K. Willcox, and M. Gunzburger, "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *SIAM Rev.*, vol. 60, no. 3, pp. 550–591, Jan. 2018.

[104] C. M. Fleeter, G. Geraci, D. E. Schiavazzi, A. M. Kahn, and A. L. Marsden, "Multilevel and multifidelity uncertainty quantification for cardiovascular hemodynamics," *Comput. Methods Appl. Mech. Eng.*, vol. 365, Jun. 2020, Art. no. 113030.

[105] T. K. West and B. D. Phillips, "Multifidelity uncertainty quantification of a commercial supersonic transport," *J. Aircr.*, vol. 57, no. 3, pp. 491–500, May 2020.

[106] A. Gelman, B. Carlin, and J. B. Stern, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1995.

[107] C. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer Texts in Statistics). Springer, 2004.

[108] J. M. Marin and C. P. Robert, *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York, NY, USA: Springer, 2007.

[109] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Boca Raton, FL, USA: Chapman & Hall, 2006.

[110] H. Haario, E. Saksman, and J. Tamminen, "An adaptive Metropolis algorithm," *Bernoulli*, vol. 7, no. 2, pp. 223–242, 2001.

[111] H. Haario, M. Laine, A. Mira, and E. Saksman, "DRAM: Efficient adaptive MCMC," *Statist. Comput.*, vol. 16, no. 4, pp. 339–354, Dec. 2006.

[112] A. Solonen, P. Ollinaho, M. Laine, H. Haario, J. Tamminen, and H. Järvinen, "Efficient MCMC for climate model parameter estimation: Parallel adaptive chains and early rejection," *Bayesian Anal.*, vol. 7, no. 3, pp. 715–736, Sep. 2012.

[113] S. Byrne and M. Girolami, "Geodesic Monte Carlo on embedded manifolds," *Scandin. J. Statist.*, vol. 40, no. 4, pp. 825–845, Dec. 2013.

[114] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *J. Roy. Soc. Interface*, vol. 6, no. 31, pp. 187–202, Feb. 2009.

[115] G. W. Peters, Y. Fan, and S. A. Sisson, "On sequential Monte Carlo, partial rejection control and approximate Bayesian computation," *Statist. Comput.*, vol. 22, no. 6, pp. 1209–1222, Nov. 2012.

[116] M. A. Beaumont, "Approximate Bayesian computation," *Annu. Rev. Statist. Appl.*, vol. 6, no. 1, pp. 379–403, 2019.

[117] H. Haario, L. Kalachev, and J. Hakkarainen, "Generalized correlation integral vectors: A distance concept for chaotic dynamical systems," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 25, no. 6, Jun. 2015, Art. no. 063102.

[118] T. Craciunescu and A. Murari, "Geodesic distance on Gaussian manifolds for the robust identification of chaotic systems," *Nonlinear Dyn.*, vol. 86, no. 1, pp. 677–693, Oct. 2016.

[119] A. Beskos, M. Girolami, S. Lan, P. E. Farrell, and A. M. Stuart, "Geometric MCMC for infinite-dimensional inverse problems," *J. Comput. Phys.*, vol. 335, pp. 327–351, Apr. 2017.

[120] T. Cui, K. J. H. Law, and Y. M. Marzouk, "Dimension-independent likelihood-informed MCMC," *J. Comput. Phys.*, vol. 304, pp. 109–137, Jan. 2016.

[121] T. Cui and O. Zahm, "Data-free likelihood-informed dimension reduction of Bayesian inverse problems," *Inverse Problems*, vol. 37, no. 4, Mar. 2021, Art. no. 045009.

[122] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton, NJ, USA: Princeton Univ. Press, 2010.

[123] R. Ghanem, D. Higdon, and H. Owhadi, Eds., *Handbook of Uncertainty Quantification*. Springer, 2017.

[124] *The DAKOTA Project: Large-Scale Engineering Optimization and Uncertainty Analysis*. Accessed: Jan. 15, 2015. [Online]. Available: https://dakota.sandia.gov/

[125] B. Debusschere et al. *UQ Toolkit*. [Online]. Available: http://www.sandia.gov/UQToolkit

[126] U. Villa, N. Petra, and O. Ghattas, "HIPPYlib: An extensible software framework for large-scale inverse problems," *J. Open Source Softw.*, vol. 3, no. 30, p. 940, Oct. 2018.

[127] R. Ghanem, "Probabilistic characterization of transport in heterogeneous media," *Comput. Methods Appl. Mech. Eng.*, vol. 158, nos. 3–4, pp. 199–220, Jun. 1998.

[128] R. Ghanem and S. Dham, "Stochastic finite element analysis for multiphase flow in heterogeneous porous media," *Transp. Porous Media*, vol. 32, pp. 239–262, Sep. 1998.

[129] B. Crestel, G. Stadler, and O. Ghattas, "A comparative study of structural similarity and regularization for joint inverse problems governed by PDEs," *Inverse Problems*, vol. 35, no. 2, Dec. 2018, Art. no. 024003.

[130] D. Lucor and G. E. Karniadakis, "Noisy inflows cause a shedding-mode switching in flow past an oscillating cylinder," *Phys. Rev. Lett.*, vol. 92, no. 15, Apr. 2004, Art. no. 154501.

[131] D. Lucor and M. S. Triantafyllou, "Parametric study of a two degree-of-freedom cylinder subject to vortex-induced vibrations," *J. Fluids Struct.*, vol. 24, no. 8, pp. 1284–1293, Nov. 2008.

[132] R. Preuss and U. von Toussaint, "Investigations on Bayesian uncertainty quantification with two examples," *AIP Conf.*, vol. 1757, no. 1, 2016, Art. no. 060001.

[133] B. Acharya, B. D. Carlsson, A. Ekström, C. Forssén, and L. Platter, "Uncertainty quantification for proton–proton fusion in chiral effective field theory," *Phys. Lett. B*, vol. 760, pp. 584–589, Sep. 2016.

[134] O. Cekmer, K. Sargsyan, S. Blondel, H. N. Najm, D. E. Bernholdt, and B. D. Wirth, "Uncertainty quantification for incident helium flux in plasma-exposed tungsten," *Int. J. Uncertainty Quantification*, vol. 8, no. 5, pp. 429–446, 2018.

[135] J. Lakhlili, O. Hoenen, O. O. Luk, and D. P. Coster, "Uncertainty quantification for multiscale fusion plasma simulations with VECMA toolkit," in *Computational Science—ICCS 2020*, V. V. Krzhizhanovskaya et al., Eds. Cham, Switzerland: Springer, 2020, pp. 719–730.

[136] T. Xiao and M. Frank, "A stochastic kinetic scheme for multi-scale plasma transport with uncertainty quantification," *J. Comput. Phys.*, vol. 432, May 2021, Art. no. 110139.

[137] D. P. Coster, "Quantification of the uncertainty arising from atomic physics in edge plasmas," *Nucl. Mater. Energy*, vol. 33, Oct. 2022, Art. no. 101282.

[138] H. N. Najm, R. D. Berry, C. Safta, K. Sargsyan, and B. J. Debusschere, "Data free inference of uncertain parameters in chemical models," *Int. J. Uncertainty Quantification*, vol. 4, no. 2, pp. 111–132, 2014.

[139] M. Khalil, K. Chowdhary, C. Safta, K. Sargsyan, and H. N. Najm, "Inference of reaction rate parameters based on summary statistics from experiments," in *Proc. 36th Symp. Int. Combustion*, 2016.

[140] L. Hakim, G. Lacaze, M. Khalil, K. Sargsyan, H. Najm, and J. Oefelein, "Probabilistic parameter estimation in a 2-step chemical kinetics model for n-dodecane jet autoignition," *Combustion Theory Model.*, vol. 22, no. 3, pp. 446–466, May 2018.

[141] B. Ganapathysubramanian and N. Zabaras, "A non-linear dimension reduction methodology for generating data-driven stochastic input models," *J. Comput. Phys.*, vol. 227, no. 13, pp. 6612–6637, Jun. 2008.

[142] P. Chen and N. Zabaras, "Uncertainty quantification for multiscale disk forging of polycrystal materials using probabilistic graphical model techniques," *Comput. Mater. Sci.*, vol. 84, pp. 278–292, Mar. 2014.

[143] S. Hagen, "The mind's eye," *Rochester Rev.*, vol. 32, Mar./Apr. 2012.

[144] H. Childs et al., "Visit: An end-user tool for visualizing and analyzing very large data," in *High Performance Visualization: Enabling Extreme-Scale Scientific Insight*, Oct. 2012, pp. 357–372.

[145] J. P. Ahrens, B. Geveci, and C. C. Law, "Paraview: An end-user tool for large-data visualization," in *The Visualization Handbook*, 2005.

[146] B. Whitlock, J. M. Favre, and J. S. Meredith, "Parallel in situ coupling of simulation with a fully featured visualization system," in *Proc. Eurographics Symp. Parallel Graph. Vis.*, T. Kuhlen et al., Eds. The Eurographics Association, 2011.

[147] U. Ayachit et al., "ParaView catalyst: Enabling in situ data analysis and visualization," in *Proc. 1st Workshop Situ Infrastruct. Enabling Extreme-Scale Anal. Vis.*, Nov. 2015, pp. 25–29.

[148] D. Pugmire et al., "Visualization as a service for scientific data," in *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI*, J. Nichols, B. Verastegui, A. B. Maccabe, O. Hernandez, S. Parete-Koon, and T. Ahearn, Eds. Cham, Switzerland: Springer, 2020, pp. 157–174.

[149] K. Moreland et al., "VTK-m: Accelerating the visualization toolkit for massively threaded architectures," *IEEE Comput. Graph. Appl.*, vol. 36, no. 3, pp. 48–58, May 2016.

[150] D. Pugmire et al., "Fides: A general purpose data model library for streaming data," in *High Performance Computing*, H. Jagode, H. Anzt, H. Ltaief, and P. Luszczek, Eds. Cham, Switzerland: Springer, 2021, pp. 495–507.

[151] R. Tchoua et al., "Collaborative monitoring and analysis for simulation scientists," in *Proc. Int. Symp. Collaborative Technol. Syst.*, Jun. 2010, pp. 235–244.

[152] W. Gekelman et al., "Nonlocal ohms law, plasma resistivity, and reconnection during collisions of magnetic flux ropes," *Astrophys. J.*, vol. 853, no. 1, p. 33, Jan. 2018.

[153] W. Gekelman et al., "The upgraded large plasma device, a machine for studying frontier basic plasma physics," *Rev. Sci. Instrum.*, vol. 87, no. 2, Feb. 2016, Art. no. 025105.

[154] W. Gekelman, T. DeHaas, R. Prior, and A. Yates, "Using topology to locate where fully 3D reconnection occurs," *Social Netw. Appl. Sci.*, vol. 2, p. 2187, 2020.

[155] J. Kress et al., "Comparing time-to-solution for in situ visualization paradigms at scale," in *Proc. IEEE 10th Symp. Large Data Anal. Vis. (LDAV)*, Oct. 2020, pp. 22–26.

[156] J. Kress et al., "Opportunities for cost savings with in-transit visualization," in *High Performance Computing*, P. Sadayappan, B. L. Chamberlain, G. Juckeland, and H. Ltaief, Eds. Cham, Switzerland: Springer, 2020, pp. 146–165.

[157] J. Kress et al., "Comparing the efficiency of in situ visualization paradigms at scale," in *High Performance Computing*, M. Weiland, G. Juckeland, C. Trinitis, and P. Sadayappan, Eds. Cham, Switzerland: Springer, 2019, pp. 99–117.

[158] M. Raji, A. Hota, and J. Huang, "Scalable web-embedded volume rendering," in *Proc. IEEE 7th Symp. Large Data Anal. Vis. (LDAV)*, Oct. 2017, pp. 45–54.

[159] W. Gekelman, "The many faces of the shear alfvén wave," American Physical Society (APS)-DPP Plenary Talk Chicago, Tech. Rep., 2020.

[160] W. Gekelman, "The laser," *Nat. Geographic*, vol. 165, no. 3, 1984.

[161] W. Gibson, *The Peripheral*. New York, NY, USA: G.P. Putnam and Sons, 2014.

[162] J. B. Rawlings, D. Q. Mayne, and M. Diehl, *Model Predictive Control: Theory, Computation, and Design*, vol. 2. Madison, WI, USA: Nob Hill Publishing, 2017.

[163] E. Maljaars et al., "Control of the tokamak safety factor profile with time-varying constraints using MPC," *Nucl. Fusion*, vol. 55, no. 2, Feb. 2015, Art. no. 023001.

[164] D. Gidon, D. B. Graves, and A. Mesbah, "Effective dose delivery in atmospheric pressure plasma jets for plasma medicine: A model predictive control approach," *Plasma Sources Sci. Technol.*, vol. 26, no. 8, Jul. 2017, Art. no. 085005.

[165] C. Wan, Z. Yu, F. Wang, X. Liu, and J. Li, "Experiment data-driven modeling of tokamak discharge in EAST," *Nucl. Fusion*, vol. 61, no. 6, Jun. 2021, Art. no. 066015.

[166] D. Gidon, H. S. Abbas, A. D. Bonzanini, D. B. Graves, J. M. Velni, and A. Mesbah, "Data-driven LPV model predictive control of a cold atmospheric plasma jet for biomaterials processing," *Control Eng. Pract.*, vol. 109, Apr. 2021, Art. no. 104725.

[167] G. Dong, X. Wei, J. Bao, G. Brochard, Z. Lin, and W. Tang, "Deep learning based surrogate model for first-principles global simulations of fusion plasmas," 2021, *arXiv:2106.10849*.

[168] A. A. Kaptanoglu, K. D. Morgan, C. J. Hansen, and S. L. Brunton, "Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 104, no. 1, Jul. 2021, Art. no. 015206.

[169] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, "Learning-based model predictive control: Toward safe learning in control," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 3, no. 1, pp. 269–296, May 2020.

[170] L. Lin et al., "Introducing adaptive cold atmospheric plasma: The perspective of adaptive cold plasma cancer treatments based on real-time electrochemical impedance spectroscopy," *Phys. Plasmas*, vol. 27, no. 6, Jun. 2020, Art. no. 063501.

[171] A. D. Bonzanini, D. B. Graves, and A. Mesbah, "Learning-based SMPC for reference tracking under state-dependent uncertainty: An application to atmospheric pressure plasma jets for plasma medicine," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 2, pp. 611–624, Mar. 2022.

[172] A. D. Bonzanini, J. A. Paulson, G. Makrygiorgos, and A. Mesbah, "Fast approximate learning-based multistage nonlinear model predictive control using Gaussian processes and deep neural networks," *Comput. Chem. Eng.*, vol. 145, Feb. 2021, Art. no. 107174.

[173] A. D. Bonzanini, J. A. Paulson, D. B. Graves, and A. Mesbah, "Toward safe dose delivery in plasma medicine using projected neural network-based fast approximate NMPC," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 5279–5285, 2020.

[174] S. Lucia, M. Kögel, P. Zometa, D. E. Quevedo, and R. Findeisen, "Predictive control, embedded cyberphysical systems and systems of systems—A perspective," *Annu. Rev. Control*, vol. 41, pp. 193–207, 2016.

[175] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[176] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.

[177] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 2, no. 1, pp. 253–279, May 2019.

[178] G. W. Lindsay, "Convolutional neural networks as a model of the visual system: Past, present, and future," *J. Cognit. Neurosci.*, vol. 33, no. 10, pp. 2017–2031, Sep. 2021.

[179] C. M. Bishop and C. M. Roach, "Fast curve fitting using neural networks," *Rev. Sci. Instrum.*, vol. 63, no. 10, pp. 4450–4456, Oct. 1992.

[180] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Art. Int. Statist. (PMLR)*, vol. 9, 2010, pp. 249–256.

[181] J. T. Kwok and I. W. Tsang, "Linear dependency between ϵ and the input noise in ϵ-support vector regression," *IEEE Trans. Neural Netw.*, vol. 14, pp. 544–553, May 2003, doi: 10.1109/TNN.2003.810604.

[182] S. Sundararajan and S. Keerthi, "Predictive approaches for choosing hyperparameters in Gaussian processes," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 12, 1999.

[183] Z. Wang et al., "Large-amplitude electron density and H_α fluctuations in the sustained spheromak physics experiment," *Nucl. Fusion*, vol. 42, no. 6, p. 643, 2002.

[184] J. Chalaturnyk and R. Marchand, "A first assessment of a regression-based interpretation of Langmuir probe measurements," *Frontiers Phys.*, vol. 7, p. 63, May 2019.

[185] D. J. Clayton et al., "Electron temperature profile reconstructions from multi-energy SXR measurements using neural networks," *Plasma Phys. Controlled Fusion*, vol. 55, no. 9, Sep. 2013, Art. no. 095015.

[186] F. A. Matos, D. R. Ferreira, and P. J. Carvalho, "Deep learning for plasma tomography using the bolometer system at JET," *Fusion Eng. Des.*, vol. 114, pp. 18–25, Jan. 2017.

[187] C. M. Samuell, A. G. Mclean, C. A. Johnson, F. Glass, and A. E. Jaervinen, "Measuring the electron temperature and identifying plasma detachment using machine learning and spectroscopy," *Rev. Sci. Instrum.*, vol. 92, no. 4, Apr. 2021, Art. no. 043520.

[188] S. H. Lee, J. H. Lee, I. Yamada, and J. S. Park, "Development of a neural network technique for KSTAR Thomson scattering diagnostics," *Rev. Sci. Instrum.*, vol. 87, no. 11, Nov. 2016, Art. no. 11E533.

[189] J. van den Berg, I. Abramovic, N. J. L. Cardozo, and D. Moseev, "Fast analysis of collective Thomson scattering spectra on Wendelstein 7-X," *Rev. Sci. Instrum.*, vol. 89, no. 8, Aug. 2018, Art. no. 083507.

[190] S. B. Korsholm et al., "Collective Thomson scattering capabilities to diagnose fusion plasmas," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 623, no. 2, pp. 677–680, Nov. 2010.

[191] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, Jul. 2011.

[192] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[193] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[194] A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.

[195] S. Mallat, *A Wavelet Tour of Signal Processing*. Amsterdam, The Netherlands: Elsevier, 1999.

[196] E. Kalapanidas, N. Avouris, M. Craciun, and D. Neagu, "Machine learning algorithms: A study on noise sensitivity," in *Proc. 1st Balkan Conf. Inform. (BCI)*, Nov. 2003.

[197] B. T. Wolfe et al., "Neural network for 3D inertial confinement fusion shell reconstruction from single radiographs," *Rev. Sci. Instrum.*, vol. 92, no. 3, Mar. 2021, Art. no. 033547.

[198] K. Fujii, I. Yamada, and M. Hasuo, "Machine learning of noise in LHD Thomson scattering system," *Fusion Sci. Technol.*, vol. 74, nos. 1–2, pp. 57–64, Aug. 2018.

[199] A.-E. Ilesanmi and T.-O. Ilesanmi, "Methods for image denoising using convolutional neural network: A review," *Complex Intell. Syst.*, vol. 7, no. 5, 2021, Art. no. 21792198.

[200] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[201] J. D. Usala, A. Maag, T. Nelis, and G. Gomez, "Compressed sensing spectral imaging for plasma optical emission spectroscopy," *J. Anal. At. Spectrometry*, vol. 31, no. 11, p. 2198, 2016.

[202] M. C. M. Cheung et al., "Multi-component decomposition of astronomical spectra by compressed sensing," *Astrophys. J.*, vol. 882, no. 1, p. 13, 2019.

[203] I. Bright, G. Lin, and J. N. Kutz, "Compressive sensing based machine learning strategy for characterizing the flow around a cylinder with limited pressure measurements," *Phys. Fluids*, vol. 25, no. 12, Dec. 2013, Art. no. 127102.

[204] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, 2017.

[205] M. F. Duarte et al., "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[206] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, Apr. 2007, pp. 128–133.

[207] F. Yang, H. Jiang, Z. Shen, W. Deng, and D. Metaxas, "Adaptive low rank and sparse decomposition of video using compressive sensing," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 1016–1020.

[208] J. V. Shi, A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk, "Video compressive sensing for dynamic MRI," *BMC Neurosci.*, vol. 13, no. S1, p. P183, Jul. 2012.

[209] C. Zhao, S. Ma, J. Zhang, R. Xiong, and W. Gao, "Video compressive sensing reconstruction via reweighted residual sparsity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1182–1195, Jun. 2017.

[210] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: An overview," *IEEE Trans. Consum. Electron.*, vol. 46, no. 4, pp. 1103–1127, Nov. 2000.

[211] A. Puri and A. Eleftheriadis, "MPEG-4: An object-based multimedia coding standard supporting mobile applications," *Mobile Netw. Appl.*, vol. 3, no. 1, pp. 5–32, 1998.

[212] G. J. Sullivan, P. N. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," *Proc. SPIE*, vol. 5558, pp. 454–475, Nov. 2004.

[213] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consum. Electron.*, vol. 38, no. 1, pp. 18–34, Feb. 1992.

[214] M. S. Asif, F. Fernandes, and J. Romberg, "Low-complexity video compression and compressive sensing," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2013, pp. 579–583.

[215] Y. Li, W. Dai, J. Zou, H. Xiong, and Y. F. Zheng, "Structured sparse representation with union of data-driven linear and multilinear subspaces model for compressive video sampling," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5062–5077, Oct. 2017.

[216] W. Dai, Y. Li, J. Zou, H. Xiong, and Y. F. Zheng, "Fully decomposable compressive sampling with joint optimization for multidimensional sparse representation," *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 603–616, Feb. 2018.

[217] J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. D. Rodrigues, "Adaptive-rate reconstruction of time-varying signals with application in compressive foreground extraction," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3651–3666, Jul. 2016.

[218] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "FSRNet: End-to-end learning face super-resolution with facial priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2492–2501.

[219] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.

[220] A. Bora, A. Jalal, E. Price, and A. Dimakis, "Compressed sensing using generative models," in *Proc. ICML*, 2017.

[221] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[222] R. Heckel and P. Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," in *Proc. ICLR*, 2018.

[223] M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, "Deep fully-connected networks for video compressive sensing," *Digit. Signal Process.*, vol. 72, pp. 9–18, Jan. 2018.

[224] R. Lin, Y. Zhang, H. Wang, X. Wang, and Q. Dai, "Deep convolutional neural network for decompressed video enhancement," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2016, p. 617.

[225] P. Pan, Z. Xu, Y. Yang, F. Wu, and Y. Zhuang, "Hierarchical recurrent neural encoder for video representation with application to captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1029–1038.

[226] S. Santurkar, D. Budden, and N. Shavit, "Generative compression," in *Proc. Picture Coding Symp. (PCS)*, Jun. 2018, pp. 258–262.

[227] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using lstms," in *Proc. ICML*, 2015, pp. 843–852.

[228] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2242–2251.

[229] R. Hyder and M. S. Asif, "Generative models for low-dimensional video representation and reconstruction," *IEEE Trans. Signal Process.*, vol. 68, pp. 1688–1701, 2020.

[230] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[231] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[232] P. Hand and V. Voroninski, "Compressed sensing from phaseless Gaussian measurements via linear programming in the natural parameter space," 2016, *arXiv:1611.05985*.

[233] D. Van Veen, A. Jalal, M. Soltanolkotabi, E. Price, S. Vishwanath, and A. G. Dimakis, "Compressed sensing with deep image prior and learned regularization," 2018, *arXiv:1806.06438*.

[234] A. Creswell and A. A. Bharath, "Inverting the generator of a generative adversarial network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 1967–1974, Jul. 2019.

[235] P. Bojanowski, A. Joulin, D. Lopez-Pas, and A. Szlam, "Optimizing the latent space of generative networks," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 599–608.

[236] Z. C. Lipton and S. Tripathi, "Precise recovery of latent vectors from generative adversarial networks," 2017, *arXiv:1702.04782*.

[237] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *Proc. ECCV*, 2016, pp. 597–613.

[238] P. Hand, O. Leong, and V. Voroninski, "Phase retrieval under a generative prior," in *Proc. NIPS*, 2018, pp. 9154–9164.

[239] V. Shah and C. Hegde, "Solving linear inverse problems using GAN priors: An algorithm with provable guarantees," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 4609–4613.

[240] R. Hyder, V. Shah, C. Hegde, and M. S. Asif, "Alternating phase projected gradient descent with generative priors for solving compressive phase retrieval," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7705–7709.

[241] G. Jagatap and C. Hegde, "Algorithmic guarantees for inverse imaging with untrained network priors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 14832–14842.

[242] G. Mataev, P. Milanfar, and M. Elad, "DeepRED: Deep image prior powered by RED," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Oct. 2019.

[243] R. Heckel and M. Soltanolkotabi, "Compressive sensing with untrained neural networks: Gradient descent finds the smoothest approximation," in *Proc. Int. Conf. Mach. Learn.*, 2020.

[244] M. Z. Darestani and R. Heckel, "Can un-trained neural networks compete with trained neural networks at image reconstruction?" 2020, *arXiv:2007.02471*.

[245] B. I. Cohen et al., "The numerical tokamak project: Simulation of turbulent transport," *Comput. Phys. Commun.*, vol. 87, nos. 1–2, pp. 1–15, May 1995.

[246] C. Holland et al., "Implementation and application of two synthetic diagnostics for validating simulations of core tokamak turbulence," *Phys. Plasmas*, vol. 16, no. 5, May 2009, Art. no. 052301.

[247] L. Shi et al., "Synthetic diagnostics for fusion plasmas," *Rev. Sci. Instrum.*, vol. 87, no. 11, 2016, Art. no. 11D303.

[248] A. B. Kukushkin, V. S. Neverov, A. G. Alekseev, S. W. Lisgo, and A. S. Kukushkin, "Synthetic H-Alpha diagnostics for ITER: Inverse problems and error estimations for strong non-maxwellian effects and intense divertor stray light," *Fusion Sci. Technol.*, vol. 69, no. 3, pp. 628–642, May 2016.

[249] M. Hoppe, O. Embréus, R. A. Tinguely, R. S. Granetz, A. Stahl, and T. Fülöp, "SOFT: A synthetic synchrotron diagnostic for runaway electrons," *Nucl. Fusion*, vol. 58, no. 2, Feb. 2018, Art. no. 026032.

[250] Z. J. Yang et al., "Development of intelligent control module for the J-TEXT electron cyclotron emission imaging system," *Fusion Eng. Des.*, vol. 153, Apr. 2020, Art. no. 111494.

[251] G. J. Kramer, R. Nazikian, and E. J. Valeo, "Simulation of optical and synthetic imaging using microwave reflectometry," *Plasma Phys. Controlled Fusion*, vol. 46, no. 4, pp. 695–710, Apr. 2004.

[252] J. Duarte et al., "Computed stereo lensless X-ray imaging," *Nature Photon.*, vol. 13, no. 7, pp. 449–453, Jul. 2019.

[253] V. F. Shevchenko, R. G. L. Vann, S. J. Freethy, and B. K. Huang, "Synthetic aperture microwave imaging with active probing for fusion plasma diagnostics," *J. Instrum.*, vol. 7, no. 10, Oct. 2012, Art. no. P10016.

[254] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[255] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[256] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[257] A. Krizhevsky, "Learning multiple layers of features from tiny images," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2009.

[258] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[259] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[260] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[261] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, p. 4681.

[262] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[263] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4569–4579.

[264] D. Nie et al., "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018, doi: 10.1109/TBME.2018.2814538.

[265] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[266] T. M. Quan, T. Nguyen-Duc, and W. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.

[267] L. Yang, D. Zhang, and G. E. Karniadakis, "Physics-informed generative adversarial networks for stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 42, no. 1, pp. A292–A317, Jan. 2020.

[268] W. M. Tang and V. S. Chan, "Advances and challenges in computational plasma science," *Plasma Phys. Controlled Fusion*, vol. 47, no. 2, pp. R1–R34, Feb. 2005.

[269] K. D. Humbird, J. L. Peterson, J. Salmonson, and B. K. Spears, "Cognitive simulation models for inertial confinement fusion: Combining simulation and experimental data," *Phys. Plasmas*, vol. 28, Apr. 2021, Art. no. 042709.

[270] T. Tajima and J. M. Dawson, "Laser-electron accelerator," *Phys. Rev. Lett.*, vol. 43, no. 4, p. 267, 1979.

[271] S. P. D. Mangles et al., "Mono-energetic relativistic electron beams from intense laser plasma interactions," *Nature*, vol. 431, p. 535, 2004.

[272] E. L. Clark et al., "Energetic heavy-ion and proton generation from ultraintense laser-plasma interactions with solids," *Phys. Rev. Lett.*, vol. 85, no. 8, pp. 1654–1657, Aug. 2000.

[273] G. Sarri et al., "A table-top laser-based source of short, collimated, ultra-relativistic positron beams," *Phys. Rev. Lett.*, vol. 25, May 2013, Art. no. 255002.

[274] Z.-H. He, B. Hou, V. Lebailly, J. A. Nees, K. Krushelnick, and A. G. R. Thomas, "Coherent control of plasma dynamics," *Nature Commun.*, vol. 6, no. 1, p. 7156, May 2015.

[275] R. J. Shalloo et al., "Automation and control of laser wakefield accelerators using Bayesian optimization," *Nature Commun.*, vol. 11, no. 1, p. 6355, Dec. 2020.

[276] M. Litos et al., "High-efficiency acceleration of an electron beam in a plasma wakefield accelerator," *Nature*, vol. 515, no. 7525, pp. 92–95, 2014.

[277] G. Loisch et al., "Observation of high transformer ratio plasma wakefield acceleration," *Phys. Rev. Lett.*, vol. 121, no. 6, Aug. 2018, Art. no. 064801.

[278] V. Yakimenko et al., "Facet-II accelerator research with beams of extreme intensities," in *Proc. TUOBB*, 2016.

[279] C. Joshi et al., "Plasma wakefield acceleration experiments at FACET II," *Plasma Phys. Controlled Fusion*, vol. 60, no. 3, Mar. 2018, Art. no. 034001.

[280] E. Adli et al., "Acceleration of electrons in the plasma wakefield of a proton bunch," *Nature*, vol. 561, no. 7723, pp. 363–367, 2018.

[281] Y. Li, W. Cheng, L. H. Yu, and R. Rainer, "Genetic algorithm enhanced by machine learning in dynamic aperture optimization," *Phys. Rev. Accel. Beams*, vol. 21, no. 5, May 2018, Art. no. 054601.

[282] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelmann, "Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems," *Phys. Rev. Accel. Beams*, vol. 23, no. 4, Apr. 2020, Art. no. 044601.

[283] M. Kranjčević, B. Riemann, A. Adelmann, and A. Streun, "Multiobjective optimization of the dynamic aperture using surrogate models based on artificial neural networks," *Phys. Rev. Accel. Beams*, vol. 24, no. 1, Jan. 2021, Art. no. 014601.

[284] C. Emma, A. Edelen, M. J. Hogan, B. O'Shea, G. White, and V. Yakimenko, "Machine learning-based longitudinal phase space prediction of particle accelerators," *Phys. Rev. Accel. Beams*, vol. 21, no. 11, Nov. 2018, Art. no. 112802.

[285] A. Hanuka et al., "Accurate and confident prediction of electron beam longitudinal properties using spectral virtual diagnostics," *Sci. Rep.*, vol. 11, no. 1, p. 2945, Feb. 2021.

[286] J. Zhu, Y. Chen, F. Brinker, W. Decking, S. Tomin, and H. Schlarb, "Deep learning-based autoencoder for data-driven modeling of an RF photoinjector," 2021, *arXiv:2101.10437*.

[287] A. Scheinker, S. Gessner, C. Emma, and A. L. Edelen, "Adaptive model tuning studies for non-invasive diagnostics and feedback control of plasma wakefield acceleration at FACET-II," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 967, Jul. 2020, Art. no. 163902.

[288] Y. Li, Y. Hao, W. Cheng, and R. Rainer, "Analysis of beam position monitor requirements with Bayesian Gaussian regression," 2019, *arXiv:1904.05683*.

[289] J. Duris et al., "Bayesian optimization of a free-electron laser," *Phys. Rev. Lett.*, vol. 124, no. 12, Mar. 2020, Art. no. 124801.

[290] Y. Hao, Y. Li, M. Balcewicz, L. Neufcourt, and W. Cheng, "Reconstruction of storage ring 's linear optics with Bayesian inference," 2019, *arXiv:1902.11157*.

[291] Y. Li, R. Rainer, and W. Cheng, "Bayesian approach for linear optics correction," *Phys. Rev. Accel. Beams*, vol. 22, no. 1, Jan. 2019, Art. no. 012804.

[292] M. McIntire et al., "Bayesian optimization of FEL performance at LCLS," in *Proc. 7th Int. Part. Accel. Conf.*, 2016.

[293] E. Fol, R. Tomás, and G. Franchetti, "Supervised learning-based reconstruction of magnet errors in circular accelerators," *Eur. Phys. J. Plus*, vol. 136, no. 4, pp. 1–19, Apr. 2021.

[294] E. Fol et al., "Unsupervised machine learning for detection of faulty beam position monitors," in *Proc. 10th Int. Part. Accel. Conf. (IPACb)*, Melbourne, VIC, Australia, vol. 2668, 2019.

[295] P. Arpaia et al., "Machine learning for beam dynamics studies at the CERN large hadron collider," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 985, Jan. 2021, Art. no. 164652.

[296] E. Fol, J. C. de Portugal, G. Franchetti, and R. Tomás, "Optics corrections using machine learning in the LHC," in *Proc. 10th Int. Part. Accel. Conf. (IPACb)*, 2019, pp. 3990–3993.

[297] A. Adelmann, "On nonintrusive uncertainty quantification and surrogate model construction in particle accelerator modeling," *SIAM/ASA J. Uncertainty Quantification*, vol. 7, no. 2, pp. 383–416, Jan. 2019.

[298] F. H. O'Shea, N. Bruchon, and G. Gaio, "Policy gradient methods for free-electron laser and terahertz source optimization and stabilization at the Fermi free-electron laser at elettra," *Phys. Rev. Accel. Beams*, vol. 23, no. 12, Dec. 2020, Art. no. 122802.

[299] N. Bruchon et al., "Basic reinforcement learning techniques to control the intensity of a seeded free-electron laser," *Electronics*, vol. 9, no. 5, p. 781, May 2020.

[300] V. Kain et al., "Sample-efficient reinforcement learning for CERN accelerator control," *Phys. Rev. Accel. Beams*, vol. 23, no. 12, Dec. 2020, Art. no. 124801.

[301] S. Hirlaender and N. Bruchon, "Model-free and Bayesian ensembling model-based deep reinforcement learning for particle accelerator control demonstrated on the Fermi FEL," 2020, *arXiv:2012.09737*.

[302] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[303] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2528–2535.

[304] L. Castrejón, Y. Aytar, C. Vondrick, H. Pirsiavash, and A. Torralba, "Learning aligned cross-modal representations from weakly aligned data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2940–2949.

[305] Y.-F. Shen, R. Pokharel, T. J. Nizolek, A. Kumar, and T. Lookman, "Convolutional neural network-based method for real-time orientation indexing of measured electron backscatter diffraction patterns," *Acta Mater.*, vol. 170, pp. 118–131, May 2019.

[306] A. Scheinker and M. Krstić, *Model-Free Stabilization by Extremum Seeking*. Springer, 2017.

[307] A. Scheinker, "Adaptive machine learning for time-varying systems: Low dimensional latent space tuning," *J. Instrum.*, vol. 16, no. 10, Oct. 2021, Art. no. P10008.

[308] A. Scheinker et al., "Model-independent tuning for maximizing free electron laser pulse energy," *Phys. Rev. Accel. Beams*, vol. 22, no. 8, Aug. 2019, Art. no. 082802.

[309] A. Scheinker et al., "Online multi-objective particle accelerator optimization of the AWAKE electron beam line for simultaneous emittance and orbit control," *AIP Adv.*, vol. 10, no. 5, May 2020, Art. no. 055320.

[310] A. Scheinker and S. Gessner, "Adaptive method for electron bunch profile prediction," *Phys. Rev. Special Topics-Accel. Beams*, vol. 18, no. 10, Oct. 2015, Art. no. 102801.

[311] A. Scheinker and D. Scheinker, "Extremum seeking for optimal control problems with unknown time-varying systems and unknown objective functions," *Int. J. Adapt. Control Signal Process.*, vol. 35, no. 7, pp. 1143–1161, Jul. 2021.

[312] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and A. Lutman, "Demonstration of model-independent control of the longitudinal phase space of electron beams in the linac-coherent light source with femtosecond resolution," *Phys. Rev. Lett.*, vol. 121, no. 4, Jul. 2018, Art. no. 044801.

[313] A. Scheinker and R. Pokharel, "Adaptive 3D convolutional neural network-based reconstruction method for 3D coherent diffraction imaging," *J. Appl. Phys.*, vol. 128, no. 18, Nov. 2020, Art. no. 184901.

[314] A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, "Adaptive deep learning for time-varying systems with hidden parameters: Predicting changing input beam distributions of compact particle accelerators," *Sci. Rep.*, vol. 11, pp. 1–11, 2021.

[315] A. Scheinker, "Adaptive machine learning for robust diagnostics and control of time-varying particle accelerator components and beams," *Information*, vol. 12, no. 4, p. 161, Apr. 2021.

[316] T. Tajima and J. M. Dawson, "Laser electron accelerator," *Phys. Rev. Lett.*, vol. 43, p. 267, Jul. 1979.

[317] P. Chen, J. M. Dawson, R. W. Huff, and T. Katsouleas, "Acceleration of electrons by the interaction of a bunched electron beam with a plasma," *Phys. Rev. Lett.*, vol. 54, no. 7, pp. 693–696, Feb. 1985.

[318] E. L. Clark et al., "Measurements of energetic proton transport through magnetized plasma from intense laser interactions with solids," *Phys. Rev. Lett.*, vol. 84, no. 4, pp. 670–673, Jan. 2000.

[319] S. Kneip et al., "Bright spatially coherent synchrotron X-rays from a table-top source," *Nature Phys.*, vol. 6, no. 12, pp. 980–983, Dec. 2010.

[320] W. Wang et al., "Free-electron lasing at 27 nanometres based on a laser wakefield accelerator," *Nature*, vol. 595, no. 7868, pp. 516–520, Jul. 2021.

[321] F. Albert et al., "2020 roadmap on plasma accelerators," *New J. Phys.*, vol. 23, no. 3, 2021, Art. no. 031101.

[322] R. Lehe, M. Kirchen, I. A. Andriyash, B. B. Godfrey, and J.-L. Vay, "A spectral, quasi-cylindrical and dispersion-free particle-in-cell algorithm," *Comput. Phys. Commun.*, vol. 203, pp. 66–82, Jun. 2016.

[323] M. J. V. Streeter et al., "Temporal feedback control of high-intensity laser pulses to optimize ultrafast heating of atomic clusters," *Appl. Phys. Lett.*, vol. 112, no. 24, Jun. 2018, Art. no. 244101.

[324] S. J. D. Dann et al., "Laser wakefield acceleration with active feedback at 5 Hz," *Phys. Rev. Accel. Beams*, vol. 22, Apr. 2019, Art. no. 041303.

[325] J. Lin et al., "Adaptive control of laser-wakefield accelerators driven by mid-IR laser pulses," *Opt. Exp.*, vol. 27, no. 8, p. 10912, 2019.

[326] T. Nayuki et al., "Production of a MeV proton with 30 mJ laser energy by optimizing the focusing spot using a deformable mirror," *Rev. Sci. Instrum.*, vol. 76, no. 7, Jul. 2005, Art. no. 073305.

[327] M. Noaman-ul-Haq et al., "Controlling laser driven protons acceleration using a deformable mirror at a high repetition rate," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 883, pp. 191–195, Mar. 2018.

[328] S. Jalas et al., "Bayesian optimization of a laser-plasma accelerator," *Phys. Rev. Lett.*, vol. 126, no. 10, Mar. 2021, Art. no. 104801.

[329] M. Kirchen et al., "Optimal beam loading in a laser-plasma accelerator," *Phys. Rev. Lett.*, vol. 126, no. 17, Apr. 2021, Art. no. 174801.

[330] J. Lin et al., "Beyond optimization—Supervised learning applications in relativistic laser-plasma experiments," *Phys. Plasmas*, vol. 28, no. 8, Aug. 2021, Art. no. 083102.

[331] B. Z. Djordjević et al., "Modeling laser-driven ion acceleration with deep learning," *Phys. Plasmas*, vol. 28, no. 4, Apr. 2021, Art. no. 043105.

[332] B. Z. Djordjević et al., "Characterizing the acceleration time of laser-driven ion acceleration with data-informed neural networks," *Plasma Phys. Controlled Fusion*, vol. 63, no. 9, Sep. 2021, Art. no. 094005.

[333] A. R. Maier et al., "Decoding sources of energy variability in a laser-plasma accelerator," *Phys. Rev. X*, vol. 10, no. 3, Aug. 2020, Art. no. 031039.

[334] L. Rovige et al., "Demonstration of stable long-term operation of a kilohertz laser-plasma accelerator," *Phys. Rev. Accel. Beams*, vol. 23, no. 9, Sep. 2020, Art. no. 093401.

[335] D. Guénot et al., "Relativistic electron beams driven by kHz single-cycle light pulses," *Nature Photon.*, vol. 11, no. 5, pp. 293–296, May 2017.

[336] F. Salehi, M. Le, L. Railing, M. Kolesik, and H. M. Milchberg, "Laser-accelerated, low-divergence 15-MeV quasimonoenergetic electron bunches at 1 kHz," *Phys. Rev. X*, vol. 11, no. 2, Jun. 2021, Art. no. 021055.

[337] S. C. Leemann et al., "Demonstration of machine learning-based model-independent stabilization of source properties in synchrotron light sources," *Phys. Rev. Lett.*, vol. 123, no. 19, Nov. 2019, Art. no. 194801.

[338] A. Sanchez-Gonzalez et al., "Accurate prediction of X-ray pulse properties from a free-electron laser using machine learning," *Nature Commun.*, vol. 8, Jun. 2017, Art. no. 15461.

[339] O. Convery, L. Smith, Y. Gal, and A. Hanuka, "Uncertainty quantification for virtual diagnostic of particle accelerators," *Phys. Rev. Accel. Beams*, vol. 24, no. 7, Jul. 2021, Art. no. 074602.

[340] A. Hanuka et al., "Physics model-informed Gaussian process for online optimization of particle accelerators," *Phys. Rev. Accel. Beams*, vol. 24, no. 7, Jul. 2021, Art. no. 072802.

[341] R. Roussel, A. Hanuka, and A. Edelen, "Multiobjective Bayesian optimization for online accelerator tuning," *Phys. Rev. Accel. Beams*, vol. 24, no. 6, Jun. 2021, Art. no. 062801.

[342] A. Mamun and P. K. Shukla, *Introduction to Dusty Plasma Physics*. Bristol, U.K.: IoP, 2001.

[343] D. A. Mendis and M. Rosenberg, "Cosmic dusty plasma," *Annu. Rev. Astron. Astrophys.*, vol. 32, no. 1, pp. 419–463, Sep. 1994.

[344] R. L. Merlino and J. A. Goree, "Dusty plasma in the laboratory, industry, and space," *Phys. Today*, vol. 57, pp. 32–38, 2004.

[345] Z. Wang et al., "Physics of dust in magnetic fusion devices," in *New Aspects of Plasma Physics: Proceedings of the 2007 ITCP Summer College on Plasma Physics*. Singapore: World Scientific Publishing, 2008, pp. 394–475.

[346] (2006). *Supernova Dust Factory in M74*. Accessed: Feb. 27, 2022. [Online]. Available: https://www.spitzer.caltech.edu/image/sig06-018-supernova-dust-factory-in-m74

[347] A. Piel and A. Melzer, "Dynamical processes in complex plasmas," *Plasma Phys. Controlled Fusion*, vol. 44, no. 1, pp. R1–R26, Jan. 2002.

[348] V. Fortov, A. Ivlev, S. Khrapak, A. Khrapak, and G. Morfill, "Complex (dusty) plasmas: Current status, open issues, perspectives," *Phys. Rep.*, vol. 421, nos. 1–2, pp. 1–103, Dec. 2005.

[349] H. M. Thomas et al., "Complex plasma laboratory PK-3 plus on the international space station," *New J. Phys.*, vol. 10, Mar. 2008, Art. no. 033036.

[350] E. Thomas Jr. U. Konopka, R. L. Merlino, and M. Rosenberg, "Initial measurements of two-and three-dimensional ordering, waves, and plasma filamentation in the magnetized dusty plasma experiment," *Phys. Plasmas*, vol. 23, Mar. 2016, Art. no. 055701.

[351] A. V. Ivlev et al., "First observation of electrorheological plasmas," *Phys. Rev. Lett.*, vol. 100, Mar. 2008, Art. no. 095003.

[352] N. N. Rao, P. K. Shukla, and M. Y. Yu, "Dust-acoustic waves in dusty plasmas," *Planet. Space Sci.*, vol. 38, p. 543, Apr. 1990.

[353] A. Barkan, R. L. Merlino, and N. D'Angelo, "Laboratory observation of the dust-acoustic wave mode," *Phys. Plasmas*, vol. 2, no. 10, pp. 3563–3565, Oct. 1995.

[354] J. Pramanik, B. M. Veeresha, G. Prasad, A. Sen, and P. K. Kaw, "Experimental observation of dust-acoustic wave turbulence," *Phys. Lett. A*, vol. 312, nos. 1–2, pp. 84–90, Jun. 2003.

[355] P.-C. Lin and L. I, "Interacting multiscale acoustic vortices as coherent excitations in dust acoustic wave turbulence," *Phys. Rev. Lett.*, vol. 120, no. 13, Mar. 2018, Art. no. 135004.

[356] Z. Wang et al., "Microparticle cloud imaging and tracking for data-driven plasma science," *Phys. Plasmas*, vol. 27, no. 3, Mar. 2020, Art. no. 033703.

[357] C. M. Ticoş, D. Ticoş, and J. D. Williams, "Kinetic effects in a plasma crystal induced by an external electron beam," *Phys. Plasmas*, vol. 26, no. 4, Apr. 2019, Art. no. 043702.

[358] C. M. Ticoş, D. Ticoş, and J. D. Williams, "Pushing microscopic matter in plasma with an electron beam," *Plasma Phys. Controlled Fusion*, vol. 62, no. 2, Feb. 2020, Art. no. 025003.

[359] H. Wang, Y. Liu, and S. Wang, "Dense velocity reconstruction from particle image velocimetry/particle tracking velocimetry using a physics-informed neural network," *Phys. Fluids*, vol. 34, no. 1, Jan. 2022, Art. no. 017116.

[360] C. Dietz, J. Budak, T. Kamprich, M. Kretschmer, and M. H. Thoma, "Phase transition in electrorheological plasmas," *Contrib. Plasma Phys.*, vol. 61, no. 10, Nov. 2021, Art. no. e202100079.

[361] Z. Ding et al., "Machine learning combined with Langmuir probe measurements for diagnosis of dusty plasma of a positive column," *Plasma Sci. Technol.*, vol. 23, no. 9, Sep. 2021, Art. no. 095403.

[362] H. Huang, M. Schwabe, and C.-R. Du, "Identification of the interface in a binary complex plasma using machine learning," *J. Imag.*, vol. 5, no. 3, p. 36, Mar. 2019.

[363] Z. Ding, L. S. Matthews, and T. W. Hyde, "A machine learning based Bayesian optimization solution to non-linear responses in dusty plasmas," *Mach. Learn., Sci. Technol.*, vol. 2, no. 3, Sep. 2021, Art. no. 035017.

[364] L. Wang, "Discovering phase transitions with unsupervised learning," *Phys. Rev. B, Condens. Matter*, vol. 94, no. 19, Nov. 2016, Art. no. 195105.

[365] J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nature Phys.*, vol. 13, no. 5, pp. 431–434, May 2017.

[366] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, "Machine learning in materials informatics: Recent applications and prospects," *npj Comput. Mater.*, vol. 3, no. 1, Dec. 2017, Art. no. 54.

[367] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost," *Chem. Sci.*, vol. 8, no. 4, p. 3192, 2017.

[368] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, pp. 547–555, Jul. 2018.

[369] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman, "Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities," *Inf. Fusion*, vol. 50, pp. 71–91, Oct. 2019.

[370] X. Yan, J. Zhu, M. Kuang, and X. Wang, "Aerodynamic shape optimization using a novel optimizer based on machine learning techniques," *Aerosp. Sci. Technol.*, vol. 86, pp. 826–835, Mar. 2019.

[371] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," *Annu. Rev. Fluid Mech.*, vol. 52, no. 1, pp. 477–508, Jan. 2020.

[372] W. Li and D. Deka, "PPGN: Physics-preserved graph networks for real-time fault location in distribution systems with limited observation and labels," 2021, *arXiv:2107.02275*.

[373] E. Mjolsness and D. DeCoste, "Machine learning for science: State of the art and future prospects," *Science*, vol. 293, no. 5537, pp. 2051–2055, Sep. 2001.

[374] R. A. Heinonen and P. H. Diamond, "Turbulence model reduction by deep learning," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 101, Jun. 2020, Art. no. 061201R.

[375] P. W. Hatfield et al., "The data-driven future of high-energy-density physics," *Nature*, vol. 593, pp. 351–361, May 2021.

[376] J. Kates-Harbeck, A. Svyatkovskiy, and W. Tang, "Predicting disruptive instabilities in controlled fusion plasmas through deep learning," *Nature*, vol. 568, no. 7753, pp. 526–531, Apr. 2019.

[377] S. Joung et al., "Deep neural network Grad–Shafranov solver constrained with measured magnetic signals," *Nucl. Fusion*, vol. 60, no. 1, Jan. 2020, Art. no. 016034.

[378] A. Gonoskov, E. Wallin, A. Polovinkin, and I. Meyerov, "Employing machine learning for theory validation and identification of experimental conditions in laser-plasma physics," *Sci. Rep.*, vol. 9, p. 7043, May 2019.

[379] M. T. Hagan, H. B. Bemuth, M. H. Beale, and O. D. Jesus, *Neural Network Design*, 2nd ed., 2014.

[380] P. A. M. Dirac, "Quantum mechanics of many-electron systems," *Proc. Roy. Soc. London A, Containing Papers Math. Phys. Character*, vol. 123, no. 792, p. 714, 1929.

[381] E. Agliari, A. Barra, P. Sollich, and L. Zdeborová, "Machine learning and statistical physics: Preface," *J. Phys. A, Math. Theor.*, vol. 53, no. 50, Nov. 2020, Art. no. 500401.

[382] M. S. Parsons, "Interpretation of machine-learning-based disruption models for plasma control," *Plasma Phys. Controlled Fusion*, vol. 59, no. 8, Aug. 2017, Art. no. 085001.

[383] A. Radovic et al., "Machine learning at the energy and intensity frontiers of particle physics," *Nature*, vol. 560, no. 7716, pp. 41–48, Aug. 2018.

[384] M. Raissi and G. E. Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," *J. Comput. Phys.*, vol. 357, pp. 125–141, Mar. 2018.

[385] C. Rackauckas et al., "Universal differential equations for scientific machine learning," 2020, *arXiv:2001.04385*.

[386] G. Carleo et al., "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, Dec. 2019, Art. no. 054002.

[387] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021.

[388] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," 2016, *arXiv:1611.09347*.

[389] E. M. Stoudenmire and D. J. Schwab, "Supervised learning with quantum-inspired tensor networks," 2016, *arXiv:1605.05775*.

[390] K. Duraisamy, G. Iaccarino, and H. Xiao, "Turbulence modeling in the age of data," *Annu. Rev. Fluid Mech.*, vol. 51, no. 1, pp. 357–377, Jan. 2019.

[391] D. J. Gross, "The role of symmetry in fundamental physics," *Proc. Nat. Acad. Sci. USA*, vol. 93, no. 25, pp. 14256–14259, Dec. 1996.

[392] E. Kauderer-Abrams, "Quantifying translation-invariance in convolutional neural networks," 2017, *arXiv:1801.01450*.

[393] R. Gens and P. Domingos, "Deep symmetry networks," in *Proc. NIPS*, 2014.

[394] T. S. Cohen, M. Gieger, J. Köhler, and M. Welling, "Spherical CNNs," in *Proc. ICRL*, 2018.

[395] L. Sarra, A. Aiello, and F. Marquardt, "Renormalized mutual information for artificial scientific discovery," *Phys. Rev. Lett.*, vol. 126, no. 20, May 2021, Art. no. 200601.

[396] E. Narita, M. Honda, M. Nakata, M. Yoshida, and N. Hayashi, "Quasilinear turbulent particle and heat transport modelling with a neural-network-based approach founded on gyrokinetic calculations and experimental data," *Nucl. Fusion*, vol. 61, no. 11, Nov. 2021, Art. no. 116041.

[397] T. Cartier-Michaud, D. Galassi, P. Ghendrih, P. Tamain, F. Schwander, and E. Serre, "A posteriori error estimate in fluid simulations of turbulent edge plasmas for magnetic fusion in tokamak using the data mining iPoPe method," *Phys. Plasmas*, vol. 27, no. 5, May 2020, Art. no. 052507.

[398] E. C. Howell and J. D. Hanson, "Development of a non-parametric Gaussian process model in the three-dimensional equilibrium reconstruction code V3FIT," *J. Plasma Phys.*, vol. 86, no. 1, Feb. 2020, Art. no. 905860102.

[399] D. Moreau et al., "Combined magnetic and kinetic control of advanced tokamak steady state scenarios based on semi-empirical modelling," *Nucl. Fusion*, vol. 55, no. 6, Jun. 2015, Art. no. 063011.

[400] K. E. J. Olofsson, A. Soppelsa, T. Bolzonella, and G. Marchiori, "Subspace identification analysis of RFX and T2R reversed-field pinches," *Control Eng. Pract.*, vol. 21, no. 7, pp. 917–929, Jul. 2013.

[401] I. R. Goumiri et al., "Modeling and control of plasma rotation for NSTX using neoclassical toroidal viscosity and neutral beam injection," *Nucl. Fusion*, vol. 56, no. 3, Mar. 2016, Art. no. 036023.

[402] W. Treutterer et al., "Management of complex data flows in the ASDEX upgrade plasma control system," *Fusion Eng. Des.*, vol. 87, no. 12, pp. 2039–2044, Dec. 2012.

[403] M. A. Chilenski et al., "Improved profile fitting and quantification of uncertainty in experimental measurements of impurity transport coefficients using Gaussian process regression," *Nucl. Fusion*, vol. 55, no. 2, Feb. 2015, Art. no. 023012.

[404] M. A. Chilenski et al., "Experimentally testing the dependence of momentum transport on second derivatives using Gaussian process regression," *Nucl. Fusion*, vol. 57, no. 12, Dec. 2017, Art. no. 126013.

[405] A. Mathews, M. Francisquez, J. W. Hughes, D. R. Hatch, B. Zhu, and B. N. Rogers, "Uncovering turbulent plasma dynamics via deep learning from partial observations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 104, no. 2, Aug. 2021, Art. no. 025205.

[406] W. Treutterer et al., "Integrated operation of diagnostic and control systems," *Fusion Eng. Des.*, vol. 86, nos. 6–8, pp. 465–470, Oct. 2011.

[407] C. Michoski, M. Milosavljević, T. Oliver, and D. R. Hatch, "Solving differential equations using deep neural networks," *Neurocomputing*, vol. 399, pp. 193–212, Jul. 2020.

[408] Z. Long, Y. Lu, X. Ma, and B. Dong, "PDE-Net: Learning PDEs from data," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 3208–3216.

[409] B. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. Kutz, and S. Brunton, "PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data," 2020, *arXiv:2004.08424*.

[410] C. B. Delahunt and J. N. Kutz, "A toolkit for data-driven discovery of governing equations in high-noise regimes," 2021, *arXiv:2111.04870*.

[411] Z. Long, Y. Lu, and B. Dong, "PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network," *J. Comput. Phys.*, vol. 399, Dec. 2019, Art. no. 108925.

[412] C. S. Chang et al., "Constructing a new predictive scaling formula for ITER's divertor heat-load width informed by a simulation-anchored machine learning," *Phys. Plasmas*, vol. 28, no. 2, Feb. 2021, Art. no. 022501.

[413] E. A. Baltz et al., "Achievement of sustained net plasma heating in a fusion experiment with the optometrist algorithm," *Sci. Rep.*, vol. 7, no. 1, p. 6425, Jul. 2017.

[414] J. A. Gaffney et al., "Making inertial confinement fusion models more predictive," *Phys. Plasmas*, vol. 26, no. 8, Aug. 2019, Art. no. 082704.

[415] R. Myers, D. Montgomery, and C. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2016.
[416] D. Montgomery, *Design and Analysis of Experiments*. Hoboken, NJ, USA: Wiley, 2017.
[417] N. Nisan and A. Ronen, "Algorithmic mechanism design," *Games Econ. Behav.*, vol. 35, nos. 1–2, pp. 166–196, 2001.
[418] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, *Algorithmic Game Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[419] D. Kress, S. Meiswinkel, and E. Pesch, "Mechanism design for machine scheduling problems: Classification and literature overview," *OR Spectr.*, vol. 40, no. 3, pp. 583–611, Jul. 2018.
[420] A. Dinklage, H. Dreier, R. Fischer, S. Gori, R. Preuss, and U. V. Toussaint, "Integrated data analysis for fusion: A Bayesian tutorial for fusion diagnosticians," *AIP Conf.*, vol. 988, pp. 471–480, Mar. 2008.
[421] A. Pavone, J. Svensson, S. Kwak, M. Brix, and R. C. Wolf, "Neural network approximated Bayesian inference of edge electron density profiles at JET," *Plasma Phys. Controlled Fusion*, vol. 62, no. 4, Apr. 2020, Art. no. 045019.
[422] A. Tejero-Cantero et al., "Sbi: A toolkit for simulation-based inference," *J. Open Source Softw.*, vol. 5, no. 52, p. 2505, Aug. 2020.
[423] C. Furia and R. M. Churchill, "Normalizing flows for likelihood-free inference with fusion simulations," *Plasma Phys. Control Fusion*, vol. 64, p. 13, Aug. 2022, Art. no. 104003, doi: 10.1088/1361-6587/ac828d.
[424] A. Jalalvand et al., "Alfvén eigenmode classification based on ECE diagnostics at DIII-D using deep recurrent neural networks," *Nucl. Fusion*, vol. 62, no. 2, Dec. 2021, Art. no. 026007.
[425] J. E. Lee, P. H. Seo, J. G. Bak, and G. S. Yun, "A machine learning approach to identify the universality of solitary perturbations accompanying boundary bursts in magnetized toroidal plasmas," *Sci. Rep.*, vol. 11, no. 1, p. 3662, Feb. 2021.
[426] R. M. Churchill, B. Tobias, Y. Zhu, and D.-D. team, "Deep convolutional neural networks for multi-scale time-series classification and application to tokamak disruption prediction using raw, high temporal resolution diagnostic data," *Phys. Plasmas*, vol. 27, no. 6, Jun. 2020, Art. no. 062510.
[427] R. Churchill et al., "A framework for international collaboration on ITER using large-scale data transfer to enable near real-time analysis," in *Proc. IAEA Fusion Data Process., Validation, Anal. Tech. Meeting*, 2019.
[428] R. Kube et al., "Near real-time streaming analysis of big fusion data," *Plasma Phys. Controlled Fusion*, vol. 64, no. 3, Feb. 2022, Art. no. 035015.
[429] J. Wesson, *Tokamaks*, 4th ed. Oxford, U.K.: Oxford Univ. Press, 2011, ch. 7.7, pp. 378–380.
[430] A. H. Boozer, "Theory of tokamak disruption," *Phys. Plasmas*, vol. 19, no. 5, 2012, Art. no. 058101.
[431] M. Lehnen et al., "Disruptions in ITER and strategies for their control and mitigation," *J. Nucl. Mater.*, vol. 463, pp. 39–48, Aug. 2015.
[432] I. Bandyopadhyay et al., "Summary of the IAEA technical meeting on plasma disruptions and their mitigation," *Nucl. Fusion*, vol. 61, no. 7, Jul. 2021, Art. no. 077001.
[433] P. C. de Vries et al., "Survey of disruption causes at JET," *Nucl. Fusion*, vol. 51, no. 5, May 2011, Art. no. 053018.
[434] J. X. Zhu, C. Rea, K. Montes, R. S. Granetz, R. Sweeney, and R. A. Tinguely, "Hybrid deep-learning architecture for general disruption prediction across multiple tokamaks," *Nucl. Fusion*, vol. 61, no. 2, Feb. 2021, Art. no. 026007.
[435] A. Pau et al., "A machine learning approach based on generative topographic mapping for disruption prevention and avoidance at JET," *Nucl. Fusion*, vol. 59, no. 10, Oct. 2019, Art. no. 106017.
[436] C. Rea, K. J. Montes, K. G. Erickson, R. S. Granetz, and R. A. Tinguely, "A real-time machine learning-based disruption predictor in DIII-D," *Nucl. Fusion*, vol. 59, no. 9, Sep. 2019, Art. no. 096016.
[437] C. Rea, K. J. Montes, A. Pau, R. S. Granetz, and O. Sauter, "Progress toward interpretable machine learning–based disruption predictors across tokamaks," *Fusion Sci. Technol.*, vol. 76, no. 8, pp. 912–924, Nov. 2020.
[438] R. A. Tinguely, K. J. Montes, C. Rea, R. Sweeney, and R. S. Granetz, "An application of survival analysis to disruption prediction via random forests," *Plasma Phys. Controlled Fusion*, vol. 61, no. 9, Sep. 2019, Art. no. 095009.
[439] W. H. Hu et al., "Real-time prediction of high-density EAST disruptions using random forest," *Nucl. Fusion*, vol. 61, no. 6, Jun. 2021, Art. no. 066034.
[440] T. Yokoyama et al., "Likelihood identification of high-beta disruption in JT-60U," *Plasma Fusion Res.*, vol. 16, p. 1402073, May 2021.
[441] F. M. Poli, "Integrated tokamak modeling: When physics informs engineering and research planning," *Phys. Plasmas*, vol. 25, no. 5, May 2018, Art. no. 055602.
[442] J. Degrave et al., "Magnetic control of tokamak plasmas through deep reinforcement learning," *Nature*, vol. 602, no. 7897, pp. 414–419, 2022.
[443] M. D. Boyer, S. Kaye, and K. Erickson, "Real-time capable modeling of neutral beam injection on NSTX-U using neural networks," *Nucl. Fusion*, vol. 59, no. 5, May 2019, Art. no. 056008.
[444] S. M. Morosohk, M. D. Boyer, and E. Schuster, "Accelerated version of NUBEAM capabilities in DIII-D using neural networks," *Fusion Eng. Des.*, vol. 163, Feb. 2021, Art. no. 112125.
[445] K. L. van de Plassche et al., "Fast modeling of turbulent transport in fusion plasmas using neural networks," *Phys. Plasmas*, vol. 27, no. 2, Feb. 2020, Art. no. 022310.
[446] A. Ho et al., "Neural network surrogate of QuaLiKiz using JET experimental data to populate training space," *Phys. Plasmas*, vol. 28, p. 22, Mar. 2021, Art. no. 032305, doi: 10.1063/5.0038290.
[447] S. Van Mulders et al., "Rapid optimization of stationary tokamak plasmas in RAPTOR: Demonstration for the ITER hybrid scenario with neural network surrogate transport model QLKNN," *Nucl. Fusion*, vol. 61, no. 8, Aug. 2021, Art. no. 086019.
[448] O. Meneghini et al., "Neural-network accelerated coupled core-pedestal simulations with self-consistent transport of impurities and compatible with ITER IMAS," *Nucl. Fusion*, vol. 61, no. 2, Feb. 2021, Art. no. 026006.
[449] S. Morosohk, T. Rafiq, E. Schuster, O. Meneghini, and M. Boyer, "Control-oriented current-profile response modeling using neural network accelerated versions of TGLF and NUBEAM for DIII-D," in *Proc. APS Division Plasma Phys. Meeting Abstr.*, 2020, Paper GP19-029.
[450] S. M. Morosohk, A. Pajares, T. Rafiq, and E. Schuster, "Neural network model of the multi-mode anomalous transport module for accelerated transport simulations," *Nucl. Fusion*, vol. 61, no. 10, Oct. 2021, Art. no. 106040.
[451] A. Merlo et al., "Proof of concept of a fast surrogate model of the VMEC code via neural networks in Wendelstein 7-X scenarios," *Nucl. Fusion*, vol. 61, no. 9, Sep. 2021, Art. no. 096039.
[452] A. Piccione, J. W. Berkery, S. A. Sabbagh, and Y. Andreopoulos, "Physics-guided machine learning approaches to predict the ideal stability properties of fusion plasmas," *Nucl. Fusion*, vol. 60, no. 4, Apr. 2020, Art. no. 046033.
[453] D. Humphreys and A. Kupresanin, "Report of the workshop on advancing fusion with machine learning research," U.S. Dept. Energy Office Sci., Office Fusion Energy Sci. Office Adv. Sci. Comput. Res., Tech. Rep., Sep. 2019. Accessed: Jul. 10, 2023. [Online]. Available: https://www.osti.gov/biblio/1615240
[454] B. B. McHarg, "The DIII-D computing environment: Characteristics and recent changes," *Fusion Eng. Des.*, vol. 48, nos. 1–2, pp. 77–82, Aug. 2000.
[455] (2020). *MDSplus*. [Online]. Available: http://mdsplus.org/
[456] (2020). *BeeGFS Parallel File System*. [Online]. Available: https://www.beegfs.io
[457] B. S. Sammuli et al., "TokSearch: A search engine for fusion experimental data," *Fusion Eng. Design*, vol. 129, pp. 12–15, Apr. 2018.
[458] (2020). *Microsoft SQL Server*. [Online]. Available: https://www.microsoft.com/en-us/sql-server
[459] Wikipedia. *Federated Database System*. Accessed: Jul. 10, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Federated_database_system
[460] R. Nora, J. L. Peterson, B. K. Spears, J. E. Field, and S. Brandon, "Ensemble simulations of inertial confinement fusion implosions," *Stat. Anal. Data Mining, ASA Data Sci. J.*, vol. 10, no. 4, pp. 230–237, Aug. 2017.
[461] R. Anirudh, J. J. Thiagarajan, P.-T. Bremer, and B. K. Spears, "Improved surrogates in inertial confinement fusion with manifold and cycle consistencies," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 18, pp. 9741–9746, Apr. 2020.
[462] R. Anirudh, J. J. Thiagarajan, S. Liu, P.-T. Bremer, and B. K. Spears, "Exploring generative physics models with scientific priors in inertial confinement fusion," in *Proc. Mach. Learn. Phys. Sci. Workshop NeurIPS*, 2019.
[463] B. Kustowski et al., "Suppressing simulation bias in multi-modal data using transfer learning," *Mach. Learn., Sci. Technol.*, vol. 3, no. 1, Mar. 2022, Art. no. 015035.

[464] B. Kustowski, J. A. Gaffney, B. K. Spears, G. J. Anderson, J. J. Thiagarajan, and R. Anirudh, "Transfer learning as a tool for reducing simulation bias: Application to inertial confinement fusion," *IEEE Trans. Plasma Sci.*, vol. 48, no. 1, pp. 46–53, Jan. 2020.

[465] P. T. Springer et al., "A 3D dynamic model to assess the impacts of low-mode asymmetry, aneurysms and mix-induced radiative loss on capsule performance across inertial confinement fusion platforms," *Nucl. Fusion*, vol. 59, no. 3, Dec. 2018, Art. no. 032009.

[466] A. Hsu, B. Cheng, and P. A. Bradley, "Analysis of NIF scaling using physics informed machine learning," *Phys. Plasmas*, vol. 27, no. 1, Jan. 2020, Art. no. 012703.

[467] J. J. Ruby et al., "Energy flow in thin shell implosions and explosions," *Phys. Rev. Lett.*, vol. 125, no. 21, Nov. 2020, Art. no. 215001.

[468] J. J. Ruby, J. A. Gaffney, J. R. Rygg, Y. Ping, and G. W. Collins, "High-energy-density-physics measurements in implosions using Bayesian inference," *Phys. Plasmas*, vol. 28, no. 3, Mar. 2021, Art. no. 032703.

[469] V. Gopalaswamy et al., "Tripled yield in direct-drive laser fusion through statistical modelling," *Nature*, vol. 565, no. 7741, pp. 581–586, Jan. 2019.

[470] V. Gopalaswamy et al., "Using statistical modeling to predict and understand fusion experiments," *Phys. Plasmas*, vol. 28, no. 12, Dec. 2021, Art. no. 122705.

[471] J. B. Nakhleh, M. G. Fernández-Godino, M. J. Grosskopf, B. M. Wilson, J. Kline, and G. Srinivasan, "Exploring sensitivity of ICF outputs to design parameters in experiments using machine learning," *IEEE Trans. Plasma Sci.*, vol. 49, no. 7, pp. 2238–2246, Jul. 2021.

[472] W. E. Lewis et al., "Deep-learning-enabled Bayesian inference of fuel magnetization in magnetized liner inertial fusion," *Phys. Plasmas*, vol. 28, no. 9, Sep. 2021, Art. no. 092701.

[473] K. D. Humbird, J. L. Peterson, and R. G. Mcclarren, "Deep neural network initialization with decision trees," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1286–1295, May 2019.

[474] M. F. Kasim et al., "Building high accuracy emulators for scientific simulations with deep neural architecture search," *Mach. Learn., Sci. Technol.*, vol. 3, no. 1, Dec. 2021, Art. no. 015013.

[475] B. K. Spears et al., "Deep learning: A guide for practitioners in the physical sciences," *Phys. Plasmas*, vol. 25, no. 8, Aug. 2018, Art. no. 080901.

[476] J. A. Gaffney et al., "The JAG inertial confinement fusion simulation dataset for multi-modal scientific deep learning," UC San Diego Library Digit. Collections, Tech. Rep., Dec. 2019. Accessed:Jul. 10, 2023, doi: 10.6075/J0RV0M27.

[477] K. L. Baker et al., "High-performance indirect-drive cryogenic implosions at high adiabat on the national ignition facility," *Phys. Rev. Lett.*, vol. 121, no. 13, Sep. 2018, Art. no. 135001.

[478] D. T. Casey et al., "The high velocity, high adiabat, 'Bigfoot' campaign and tests of indirect-drive implosion scaling," *Phys. Plasmas*, vol. 25, May 2018, Art. no. 056308.

[479] M. M. Marinak et al., "Three-dimensional HYDRA simulations of national ignition facility targets," *Phys. Plasmas*, vol. 8, no. 5, pp. 2275–2280, May 2001.

[480] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn.*, in Proceedings of Machine Learning Research, vol. 48, M. F. Balcan and K. Q. Weinberger, Eds. New York, NY, USA, Jun. 2016, pp. 1050–1059.

[481] G. J. Anderson, J. A. Gaffney, B. K. Spears, P.-T. Bremer, R. Anirudh, and J. J. Thiagarajan, "Meaningful uncertainties from deep neural network surrogates of large-scale numerical simulations," 2020, *arXiv:2010.13749*.

[482] S. A. Jacobs et al., "Parallelizing training of deep generative models on massive scientific datasets," in *Proc. IEEE Int. Conf. Cluster Comput. (CLUSTER)*, Sep. 2019, pp. 1–10.

[483] S. A. Jacobs, N. Dryden, R. Pearce, and B. Van Essen, "Towards scalable parallel training of deep neural networks," in *Proc. Mach. Learn. HPC Environ.*, Nov. 2017, pp. 1–9.

[484] S. A. Jacobs et al., "Enabling rapid COVID-19 small molecule drug design through scalable deep learning of generative models," *Int. J. High Perform. Comput. Appl.*, vol. 35, no. 5, pp. 469–482, Sep. 2021.

[485] J. L. Peterson et al., "Zonal flow generation in inertial confinement fusion implosions," *Phys. Plasmas*, vol. 24, no. 3, Mar. 2017, Art. no. 032702.

[486] P. W. Hatfield, S. J. Rose, and R. H. H. Scott, "The blind implosion-maker: Automated inertial confinement fusion experiment design," *Phys. Plasmas*, vol. 26, no. 6, Jun. 2019, Art. no. 062706.

[487] K. D. Humbird, J. L. Peterson, B. K. Spears, and R. G. McClarren, "Transfer learning to model inertial confinement fusion experiments," *IEEE Trans. Plasma Sci.*, vol. 48, no. 1, pp. 61–70, Jan. 2020.

[488] N. N. Vazirani et al., "Coupling 1D xRAGE simulations with machine learning for graded inner shell design optimization in double shell capsules," *Phys. Plasmas*, vol. 28, no. 12, Dec. 2021, Art. no. 122709.

[489] R. F. da Silva et al., "Workflows community summit: Bringing the scientific workflows community together," 2021, *arXiv:2103.09181*.

[490] H. Alfvén, "The plasma universe," *Phys. Today*, vol. 39, pp. 22–27, 1986.

[491] J. L. Ballester et al., "Partially ionized plasmas in astrophysics," *Space Sci. Rev.*, vol. 214, p. 58, Mar. 2018.

[492] A. G. Lyne et al., "A double-pulsar system: A rare laboratory for relativistic gravity and plasma physics," *Science*, vol. 303, no. 5661, pp. 1153–1157, Feb. 2004.

[493] M. Bonitz, Z. A. Moldabekov, and T. S. Ramazanov, "Quantum hydrodynamics for plasmas—Quo vadis?" *Phys. Plasmas*, vol. 26, no. 9, Sep. 2019, Art. no. 090601.

[494] Y. Zhang and Y. Zhao, "Astronomy in the big data era," *Data Sci. J.*, vol. 14, p. 11, May 2015.

[495] M. F. Skrutskie et al., "The two micron all sky survey (2MASS)," *Astron. J.*, vol. 133, no. 2, pp. 1163–1183, 2006.

[496] A. R. Azari, J. W. Lockhart, M. W. Liemohn, and X. Jia, "Incorporating physical knowledge into machine learning for planetary space physics," 2020, *arXiv:2006.01927*.

[497] G. Longo, E. Merényi, and P. Tiňo, "Foreword to the focus issue on machine intelligence in astronomy and astrophysics," *Publications Astronomical Soc. Pacific*, vol. 131, no. 1004, Nov. 2019, Art. no. 100101.

[498] M. Z. Alom et al., "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*.

[499] A. Koul, S. Ganju, M. Kasam, and J. Parr, "SpaceML: Distributed open-source research with citizen scientists for the advancement of space technology for NASA," 2020, *arXiv:2012.10610*.

[500] Z. Wang et al., "Billion-pixel X-ray camera (BiPC-X)," *Rev. Sci. Instrum.*, vol. 92, no. 4, Apr. 2021, Art. no. 043708.

[501] M. Krainak et al., "Integrated photonics for NASA applications," *Proc. SPIE*, vol. 10899, Mar. 2019, Art. no. 108990F.

[502] A. N. Dinkelaker and A. Rahman, "Astrophotonics: Processing starlight," *Europhys. News*, vol. 52, no. 1, pp. 22–24, 2021.

[503] D. R. Thompson et al., "Onboard machine learning classification of images by a cubesat in earth orbit," *AI Matters*, vol. 1, no. 4, pp. 38–40, Jun. 2015.

[504] Z. Zhang, G. Xu, and J. Song, "Cubesat could detection based on JPEG2000 compression and deep learning," *Adv. Mech. Eng.*, vol. 10, no. 10, pp. 1–10, 2018.

[505] A. Maskey and M. Cho, "CubeSatNet: Ultralight convolutional neural network designed for on-orbit binary image classification on a 1U CubeSat," *Eng. Appl. Artif. Intell.*, vol. 96, Nov. 2020, Art. no. 103952.

[506] J. Manning et al., "Machine-learning space applications on small-sat platforms with tensorflow," in *Proc. AIAA/USU Conf. Small Satell.*, vol. 32, 2018.

[507] S. Aigrain, H. Parviainen, S. Roberts, S. Reece, and T. Evans, "Robust, open-source removal of systematics in Kepler data," *Monthly Notices Roy. Astronomical Soc.*, vol. 471, no. 1, p. 759, 2017.

[508] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *Proc. ACM SIGGRAPH Papers*, 2006, p. 787.

[509] A. D. Cobb et al., "An ensemble of Bayesian neural networks for exoplanetary atmospheric retrieval," 2019, *arXiv:1905.10659*.

[510] T. Zingales and I. P. Waldmann, "ExoGAN: Retrieving exoplanetary atmospheres using deep convolutional generative adversarial networks," 2018, *arXiv:1806.02906*.

[511] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," 2016, *arXiv:1607.07539*.

[512] I. Pesenson, S. J. Carey, and B. Mccollum, "High dimensional data reduction, image inpainting and their astronomical applications," in *Proc. ASP Conf. D*, vol. 411, 2009.

[513] A. Szenicer et al., "A deep learning virtual instrument for monitoring extreme UV solar spectral irradiance," *Sci. Adv.*, vol. 5, no. 10, 2019, Art. no. eaaw6548.

[514] H. Liu et al., "Inferring vector magnetic fields from Stokes profiles of GST/NIRIS using a convolutional neural network," *Astrophys. J.*, vol. 894, no. 1, p. 70, May 2020.

[515] H. Jiang et al., "Identifying and tracking solar magnetic flux elements with deep learning," *Astrophys. J. Suppl. Ser.*, vol. 250, no. 1, p. 5, Aug. 2020.

[516] G. Fu, F. Y. Shih, and H. Wang, "Automatic detection of magnetic flux emergings in the solar atmosphere from full-disk magnetogram sequences," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2174–2185, Nov. 2008.

[517] H. Song, C. Tan, J. Jing, H. Wang, V. Yurchyshyn, and V. Abramenko, "Statistical assessment of photospheric magnetic features in imminent solar flare predictions," *Sol. Phys.*, vol. 254, no. 1, pp. 101–125, Jan. 2009.

[518] Y. Yuan, F. Y. Shih, J. Jing, and H.-M. Wang, "Automated flare forecasting using a statistical learning technique," *Res. Astron. Astrophys.*, vol. 10, no. 8, pp. 785–796, Aug. 2010.

[519] Y. Yuan, F. Y. Shih, J. Jing, and H. Wang, "Solar flare forecasting using sunspot-groups classification and photospheric magnetic parameters," in *Physics of Sun and Star Spots*, vol. 273, D. P. Choudhary and K. G. Strassmeier, Eds. International Astronomical Union, 2011, pp. 446–450.

[520] C. Liu, N. Deng, J. T. L. Wang, and H. Wang, "Predicting solar flares using SDO/HMI vector magnetic data products and the random forest algorithm," *Astrophys. J.*, vol. 843, no. 2, p. 104, Jul. 2017.

[521] Y. Abduallah, J. T. L. Wang, Y. Nie, C. Liu, and H. Wang, "DeepSun: Machine-learning-as-a-service for solar flare prediction," *Res. Astron. Astrophys.*, vol. 21, no. 7, p. 160, Aug. 2021.

[522] H. Liu, C. Liu, J. T. L. Wang, and H. Wang, "Predicting solar flares using a long short-term memory network," *Astrophys. J.*, vol. 877, no. 2, p. 121, Jun. 2019.

[523] X. Wang et al., "Predicting solar flares with machine learning: Investigating solar cycle dependence," *Astrophys. J.*, vol. 895, no. 1, p. 3, May 2020.

[524] H. Liu, C. Liu, J. T. L. Wang, and H. Wang, "Predicting coronal mass ejections using SDO/HMI vector magnetic data products and recurrent neural networks," *Astrophys. J.*, vol. 890, no. 1, p. 12, Feb. 2020.

[525] A. Malanushenko, N. Flyer, and S. Gibson, "Convolutional neural networks for predicting the strength of the near-earth magnetic field caused by interplanetary coronal mass ejections," *Frontiers Astron. Space Sci.*, vol. 7, p. 62, Sep. 2020.

[526] T. Kim et al., "Solar farside magnetograms from deep learning analysis of STEREO/EUVI data," *Nature Astron.*, vol. 3, no. 5, pp. 397–400, Mar. 2019.

[527] C. J. D. Baso and A. A. Ramos, "Enhancing SDO/HMI images using deep learning," *Astron. Astrophys.*, vol. 614, p. A5, Jun. 2018.

[528] L. Spitzer and R. Härm, "Transport phenomena in a completely ionized gas," *Phys. Rev.*, vol. 89, no. 5, pp. 977–981, Mar. 1953.

[529] S. Braginskii, "Transport processes in a plasma," *Rev. Plasma Phys.*, vol. 1, p. 205, 1965.

[530] G. W. Hammett and F. W. Perkins, "Fluid moment models for Landau damping with application to the ion-temperature-gradient instability," *Phys. Rev. Lett.*, vol. 64, no. 25, pp. 3019–3022, Jun. 1990.

[531] Z. Chang and J. D. Callen, "Unified fluid/kinetic description of plasma microinstabilities. Part I: Basic equations in a sheared slab geometry," *Phys. Fluids B, Plasma Phys.*, vol. 4, no. 5, pp. 1167–1181, May 1992.

[532] P. B. Snyder, G. W. Hammett, and W. Dorland, "Landau fluid models of collisionless magnetohydrodynamics," *Phys. Plasmas*, vol. 4, no. 11, 1997, Art. no. 39743985.

[533] M. V. Umansky, A. M. Dimits, I. Joseph, J. T. Omotani, and T. D. Rognlien, "Modeling of tokamak divertor plasma for weakly collisional parallel electron transport," *J. Nucl. Mater.*, vol. 463, pp. 506–509, Aug. 2015.

[534] Z. Guo and X.-Z. Tang, "Parallel heat flux from low to high parallel temperature along a magnetic field line," *Phys. Rev. Lett.*, vol. 108, no. 16, Apr. 2012, Art. no. 165005.

[535] P. Hunana et al., "New closures for more precise modeling of Landau damping in the fluid framework," *Phys. Rev. Lett.*, vol. 121, no. 13, Sep. 2018, Art. no. 135101.

[536] L. Wang, B. Zhu, X.-Q. Xu, and B. Li, "A landau-fluid closure for arbitrary frequency response," *AIP Adv.*, vol. 9, no. 1, Jan. 2019, Art. no. 015217.

[537] L. Wang, X. Q. Xu, B. Zhu, C. Ma, and Y.-A. Lei, "Deep learning surrogate model for kinetic landau-fluid closure with collision," *AIP Adv.*, vol. 10, no. 7, Jul. 2020, Art. no. 075108.

[538] A. Shukla, D. R. Hatch, W. Dorland, and C. Michoski, "A learned closure method applied to phase mixing in a turbulent gradient-driven gyrokinetic system in simple geometry," *J. Plasma Phys.*, vol. 88, no. 1, Feb. 2022, Art. no. 905880115.

[539] T. Xiao and M. Frank, "Using neural networks to accelerate the solution of the Boltzmann equation," *J. Comput. Phys.*, vol. 443, Oct. 2021, Art. no. 110521.

[540] J. Han, C. Ma, Z. Ma, and W. E, "Uniformly accurate machine learning-based hydrodynamic models for kinetic equations," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 44, pp. 21983–21991, Oct. 2019.

[541] J. B. Scoggins, J. Han, and M. Massot, "Machine learning moment closures for accurate and efficient simulation of polydisperse evaporating sprays," in *Proc. AIAA Scitech Forum*, Jan. 2021, p. 1786.

[542] J. Huang, Y. Cheng, A. J. Christlieb, and L. F. Roberts, "Machine learning moment closure models for the radiative transfer equation I: Directly learning a gradient based closure," *J. Comput. Phys.*, vol. 453, Mar. 2022, Art. no. 110941.

[543] J. Huang, Y. Cheng, A. J. Christlieb, L. F. Roberts, and W.-A. Yong, "Machine learning moment closure models for the radiative transfer equation II: Enforcing global hyperbolicity in gradient based closures," 2021, *arXiv:2105.14410*.

[544] S. Schotthöfer, T. Xiao, M. Frank, and C. D. Hauck, "A structure-preserving surrogate model for the closure of the moment system of the Boltzmann equation using convex deep neural networks," 2021, *arXiv:2106.09445*.

[545] O. Meneghini et al., "Self-consistent core-pedestal transport simulations with neural network accelerated models," *Nucl. Fusion*, vol. 57, no. 8, Aug. 2017, Art. no. 086034.

[546] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[547] D. Biskamp, "Magnetic reconnection in plasmas," *Astrophys. Space Sci.*, vol. 242, no. 1, pp. 165–207, 1996.

[548] S. Servidio, W. H. Matthaeus, M. A. Shay, P. A. Cassak, and P. Dmitruk, "Magnetic reconnection in two-dimensional magnetohydrodynamic turbulence," *Phys. Rev. Lett.*, vol. 102, no. 11, Mar. 2009, Art. no. 115003.

[549] M. V. Goldman, D. L. Newman, and G. Lapenta, "What can we learn about magnetotail reconnection from 2D PIC harris-sheet simulations?" *Space Sci. Rev.*, vol. 199, nos. 1–4, pp. 651–688, Mar. 2016.

[550] J. Burch et al., "Electron-scale measurements of magnetic reconnection in space," *Science*, vol. 352, no. 6290, 2016, Art. no. aaf2939.

[551] G. Lapenta, "Detecting reconnection sites using the Lorentz transformations for electromagnetic fields," *Astrophys. J.*, vol. 911, no. 2, p. 147, Apr. 2021.

[552] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annu. Rev. Statist. Appl.*, vol. 6, pp. 355–378, Jan. 2019.

[553] R. Dupuis, M. V. Goldman, D. L. Newman, J. Amaya, and G. Lapenta, "Characterizing magnetic reconnection regions using Gaussian mixture models on particle velocity distributions," *Astrophys. J.*, vol. 889, no. 1, p. 22, Jan. 2020.

[554] M. V. Goldman, D. L. Newman, J. P. Eastwood, and G. Lapenta, "Multibeam energy moments of multibeam particle velocity distributions," *J. Geophys. Res., Space Phys.*, vol. 125, no. 12, Dec. 2020, Art. no. e2020JA028340.

[555] J. MacQueen et al., "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Oakland, CA, USA, vol. 1, 1967, pp. 281–297.

[556] M. V. Goldman, D. L. Newman, J. P. Eastwood, G. Lapenta, J. L. Burch, and B. Giles, "Multi-beam energy moments of compound measured ion velocity distributions," 2021, *arXiv:2101.10470*.

[557] C. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Springer, 2006.

[558] A. Hu et al., "Identifying magnetic reconnection in 2D hybrid vlasov Maxwell simulations with convolutional neural networks," *Astrophys. J.*, vol. 900, no. 1, p. 86, Sep. 2020.

[559] M. Sisti, F. Finelli, G. Pedrazzi, M. Faganello, F. Califano, and F. D. Ponti, "Detecting reconnection events in kinetic vlasov hybrid simulations using clustering techniques," *Astrophys. J.*, vol. 908, no. 1, p. 107, Feb. 2021.

[560] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, vol. 96, 1996, pp. 226–231.

[561] M. Lieberman and A. Lichtenberg, "Principles of plasma discharges and materials processing: Second edition," in *Principles of Plasma Discharges and Materials Processing*, vol. 30, M. A. Lieberman and A. J. Lichtenberg, Eds., 2nd ed. Hoboken, NJ, USA: Wiley, Sep. 2005, p. 800.

[562] T. Makabe and Z. L. Petrovic, *Plasma Electronics*. Boca Raton, FL, USA: Routledge, 2016.

[563] G. S. Oehrlein and S. Hamaguchi, "Foundations of low-temperature plasma enhanced materials synthesis and etching," *Plasma Sources Sci. Technol.*, vol. 27, no. 2, Feb. 2018, Art. no. 023001.

[564] I. Adamovich et al., "The 2017 plasma roadmap: Low temperature plasma science and technology," *J. Phys. D: Appl. Phys.*, vol. 50, no. 32, p. 323001, 2017.

[565] K.-D. Weltmann et al., "The future for plasma science and technology," *Plasma Process. Polym.*, vol. 16, no. 1, 2019, Art. no. 1800118.

[566] A. Bogaerts et al., "The 2020 plasma catalysis roadmap," *J. Phys. D: Appl. Phys.*, vol. 53, no. 44, Oct. 2020, Art. no. 443001.

[567] H. Bauer, O. Burkakcy, P. Kenevan, S. Lingemann, K. Pototzky, and B. Wiseman, "Semiconductor design and manufacturing: Achieving leading-edge capabilities," McKinsey & Company, New York, NY, USA, Tech. Rep., Aug. 2020. Accessed: Jul. 10, 2023. [Online]. Available: https://www.mckinsey.com/industries/industrials-and-electronics/our-insights/semiconductor-design-and-manufacturing-achieving-leading-edge-capabilities

[568] R. E. Fontana Jr., and G. M. Decad, "Moore's law realities for recording systems and memory storage components: HDD, tape, NAND, and optical," *AIP Adv.*, vol. 8, no. 5, May 2018, Art. no. 056506.

[569] A. Mesbah and D. B. Graves, "Machine learning for modeling, diagnostics, and control of non-equilibrium plasmas," *J. Phys. D: Appl. Phys.*, vol. 52, no. 30, Jul. 2019, Art. no. 30LT02.

[570] B. R. Barricelli, E. Casiraghi, and D. Fogli, "A survey on digital twin: Definitions, characteristics, applications, and design implications," *IEEE Access*, vol. 7, pp. 167653–167671, 2019.

[571] S. Oh, Y. Jung, S. Kim, I. Lee, and N. Kang, "Deep generative design: Integration of topology optimization and generative models," *J. Mech. Des.*, vol. 141, no. 11, Nov. 2019.

[572] B. Darbon, M. Grould, N. Clément, J. Baderot, S. Martinez, and J. Foucher, "Parametric modeling of patterned object with variations and expert tuning of edge placement," *Proc. SPIE*, vol. 11615, Feb. 2021, Art. no. 116150K.

[573] J. Na, G. Kim, S.-H. Kang, S.-J. Kim, and S. Lee, "Deep learning-based discriminative refocusing of scanning electron microscopy images for materials science," *Acta Mater.*, vol. 214, Aug. 2021, Art. no. 116987.

[574] Y. Suzuki, S. Iwashita, T. Sato, H. Yonemichi, H. Moki, and T. Moriya, "Machine learning approaches for process optimization," in *Proc. Int. Symp. Semiconductor Manuf. (ISSM)*, Dec. 2018, pp. 1–4.

[575] F. Tanaka, H. Sato, N. Yoshii, and H. Matsui, "Materials informatics for process and material co-optimization," in *Proc. Int. Symp. Semiconductor Manuf. (ISSM)*, Dec. 2018, pp. 1–3.

[576] I. Chakroun, T. J. Ashby, S. Das, S. Halder, R. Wuyts, and W. Verachtert, "Using unsupervised machine learning for plasma etching endpoint detection," in *Proc. ICPRAM*, 2020, pp. 273–279.

[577] J. O'Leary, K. Sawlani, and A. Mesbah, "Deep learning for classification of the chemical composition of particle defects on semiconductor wafers," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 1, pp. 72–85, Feb. 2020.

[578] U. Batool, M. I. Shapiai, M. Tahir, Z. H. Ismail, N. J. Zakaria, and A. Elfakharany, "A systematic review of deep learning for silicon wafer defect recognition," *IEEE Access*, vol. 9, pp. 116572–116593, 2021.

[579] C. Jeong et al., "Bridging TCAD and AI: Its application to semiconductor design," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5364–5371, Nov. 2021.

[580] J. Moyne and J. Iskandar, "Big data analytics for smart manufacturing: Case studies in semiconductor manufacturing," *Processes*, vol. 5, no. 3, p. 39, Jul. 2017.

[581] C. Morariu, O. Morariu, S. Răileanu, and T. Borangiu, "Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems," *Comput. Ind.*, vol. 120, Sep. 2020, Art. no. 103244.

[582] W.-K. Loh and J.-Y. Yun, "A parallel algorithm for robust fault detection in semiconductor manufacturing processes," *Cluster Comput.*, vol. 17, no. 3, pp. 643–651, Sep. 2014.

[583] F. Cheng, H. Huang, and C. Kao, "Developing an automatic virtual metrology system," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 1, pp. 181–188, Jan. 2012.

[584] H. Du Lee, M. H. Lee, H. W. Cho, C. H. Han, and K. S. Chang, "Online quality monitoring using multivariate statistical methods in continuous—Stirred MMA-VA copolymerization process," *Korean J. Chem. Eng.*, vol. 35, p. 605, 1997.

[585] S. J. van Albada and P. A. Robinson, "Transformation of arbitrary distributions to the normal distribution with application to EEG test–retest reliability," *J. Neurosci. Methods*, vol. 161, no. 2, pp. 205–211, Apr. 2007.

[586] S. Park, S. Jeong, Y. Jang, S. Ryu, H. Roh, and G. Kim, "Enhancement of the virtual metrology performance for plasma-assisted oxide etching processes by using plasma information (PI) parameters," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 3, pp. 241–246, Aug. 2015.

[587] S. Park et al., "Application of PI-VM for management of the metal target plasma etching processes in OLED display manufacturing," *Plasma Phys. Controlled Fusion*, vol. 61, no. 1, Nov. 2018, Art. no. 014032.

[588] S. Park et al., "Cause analysis of the faults in HARC etching processes by using the PI-VM model for OLED display manufacturing," *Plasma Processes Polym.*, vol. 16, no. 9, Sep. 2019, Art. no. 1900030.

[589] S. Park et al., "Micro-range uniformity control of the etching profile in the OLED display mass production referring to the PI-VM model," *Phys. Plasmas*, vol. 28, no. 10, Oct. 2021, Art. no. 103505.

[590] S. Park et al., "Predictive control of the plasma processes in the OLED display mass production referring to the discontinuity qualifying PI-VM," *Phys. Plasmas*, vol. 27, no. 8, Aug. 2020, Art. no. 083507.

[591] J.-W. Kwon et al., "Development of virtual metrology using plasma information variables to predict Si etch profile processed by $SF_6/O_2/Ar$ capacitively coupled plasma," *Materials*, vol. 14, no. 11, p. 3005, Jun. 2021.

[592] Y. Jang et al., "Characteristics of a plasma information variable in phenomenology-based, statistically-tuned virtual metrology to predict silicon dioxide etching depth," *Current Appl. Phys.*, vol. 19, no. 10, pp. 1068–1075, Oct. 2019.

[593] S. Park, J.-M. Choe, H.-J. Roh, and G.-H. Kim, "Characteristics of a non-Maxwellian electron energy distribution in a low-pressure argon plasma," *J. Korean Phys. Soc.*, vol. 64, no. 12, pp. 1819–1827, Jun. 2014.

[594] J. T. Gudmundsson, "On the effect of the electron energy distribution on the plasma parameters of an argon discharge: A global (volume-averaged) model study," *Plasma Sources Sci. Technol.*, vol. 10, no. 1, pp. 76–81, Jan. 2001.

[595] SEMI. *What is Smart Manufacturing?* Accessed: Nov. 30, 2021. [Online]. Available: https://www.semi.org/en/industry-groups/smart-manufacturing/what-is-smart-manufacturing

[596] J. Moyne, E. Castillo, and A. M. Hurwitz, *Advanced Process Control in the Semiconductor Industry*. Boca Raton, FL, USA: CRC Press, 2001, pp. 33–62.

[597] C. D. Himmel and G. S. May, "Advantages of plasma etch modeling using neural networks over statistical techniques," *IEEE Trans. Semicond. Manuf.*, vol. 6, no. 2, pp. 103–111, May 1993.

[598] M.-H. Hung, T.-H. Lin, F.-T. Cheng, and R.-C. Lin, "A novel virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing," *IEEE/ASME Trans. Mechatronics*, vol. 12, no. 3, pp. 308–316, Jun. 2007.

[599] S. Lynn, J. Ringwood, and N. Macgearailt, "Gaussian process regression for virtual metrology of plasma etch," in *Proc. IET Irish Signals Syst. Conf.*, Jul. 2010, pp. 42–47.

[600] J. Shinagawa, "Methods and systems for chamber matching and monitoring," U.S. Patent 10 622 219, Dec. 6, 2016.

[601] S. Lynn, "Virtual metrology for plasma etch process," Ph.D. thesis, Dept. Electron. Eng., Nat. Univ. Ireland, Maynooth, Ireland, 2011.

[602] D. H. Kim and S. J. Hong, "Use of plasma information in machine-learning-based fault detection and classification for advanced equipment control," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 3, pp. 408–419, Aug. 2021.

[603] Y. Osano and K. Ono, "Atomic-scale cellular model and profile simulation of poly-Si gate etching in high-density chlorine-based plasmas: Effects of passivation layer formation on evolution of feature profiles," *J. Vac. Sci. Technol. B, Microelectron. Nanometer Struct. Process., Meas., Phenomena*, vol. 26, no. 4, pp. 1425–1439, 2008.

[604] D. J. Cooperberg, V. Vahedi, and R. A. Gottscho, "Semiempirical profile simulation of aluminum etching in a Cl_2/BCl_3 plasma," *J. Vac. Sci. Technol. A, Vac., Surf., Films*, vol. 20, no. 5, pp. 1536–1556, Sep. 2002.

[605] R. A. Gottscho, C. W. Jurgensen, and D. J. Vitkavage, "Microscopic uniformity in plasma etching," *J. Vac. Sci. Technol. B, Microelectron. Nanometer Struct. Process., Meas., Phenomena*, vol. 10, no. 5, pp. 2133–2147, 1992.

[606] J. Shinagawa, "Sensor-to-sensor matching methods for chamber matching," U.S. Patent 10 916 411, Aug. 13, 2018.

[607] T. Tsutsui and T. Matsuzawa, "Virtual metrology model robustness against chamber condition variation using deep learning," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 428–433, Nov. 2019.

[608] D. Gidon, D. B. Graves, and A. Mesbah, "Predictive control of 2D spatial thermal dose delivery in atmospheric pressure plasma jets," *Plasma Sources Sci. Technol.*, vol. 28, no. 8, Aug. 2019, Art. no. 085001.

[609] C. Woelfel et al., "Control-oriented plasma modeling and controller design for reactive sputtering," *IFAC J. Syst. Control*, vol. 16, Jun. 2021, Art. no. 100142.

[610] A. M. Ito, A. Takayama, and H. Nakamura, "Triple hybrid simulation method for tungsten fuzzy nanostructure formation," *Plasma Fusion Res.*, vol. 13, Jun. 2018, Art. no. 3403061.

[611] K. Denpoh, P. Moroz, T. Kato, and M. Matsukuma, "Multiscale plasma and feature profile simulations of plasma-enhanced chemical vapor deposition and atomic layer deposition processes for titanium thin film fabrication," *Jpn. J. Appl. Phys.*, vol. 59, no. SH, May 2020, Art. no. SHHB02.

[612] *National Academies of Sciences, Engineering, and Medicine, Plasma Science: Enabling Technology, Sustainability, Security, and Exploration*, Nat. Academies Press, Washington, DC, USA, 2021.

[613] H. Kino, K. Ikuse, H.-C. Dam, and S. Hamaguchi, "Characterization of descriptors in machine learning for data-based sputtering yield prediction," *Phys. Plasmas*, vol. 28, no. 1, Jan. 2021, Art. no. 013504.

[614] R. Preuss, R. Arredondo, and U. von Toussaint, "Bayesian determination of parameters for plasma-wall interactions," *Entropy*, vol. 21, no. 12, p. 1175, Nov. 2019.

[615] T. Gergs, B. Borislavov, and J. Trieschmann, "Efficient plasma-surface interaction surrogate model for sputtering processes based on autoencoder neural networks," *J. Vac. Sci. Technol. B, Microelectron.*, vol. 40, no. 1, Jan. 2022, Art. no. 012802.

[616] Y. Yamamura and H. Tawara, "Energy dependence of ion-induced sputtering yields from monatomic solids at normal incidence," *At. Data Nucl. Data Tables*, vol. 62, no. 2, pp. 149–253, Mar. 1996.

[617] K. Ikuse and S. Hamaguchi. (2021). *GPR-based Sputtering Yield Prediction*. [Online]. Available: http://www.camt.eng.osaka-u.ac.jp/hamaguchi/SY/s

[618] Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, "To address surface reaction network complexity using scaling relations machine learning and DFT calculations," *Nature Commun.*, vol. 8, no. 1, p. 14621, Mar. 2017.

[619] A. Diaw et al., "Multiscale simulation of plasma flows using active learning," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 102, no. 2, Aug. 2020, Art. no. 023310.

[620] B. J. Alder and T. E. Wainwright, "Studies in molecular dynamics. I. General method," *J. Chem. Phys.*, vol. 31, no. 2, pp. 459–466, Aug. 1959.

[621] M. T. Robinson, "Computer simulation studies of high-energy collision cascades," *Nucl. Instrum. Methods Phys. Res. B, Beam Interact. Mater. At.*, vol. 67, nos. 1–4, pp. 396–400, Apr. 1992.

[622] A. P. Sutton and J. Chen, "Long-range Finnis–Sinclair potentials," *Phil. Mag. Lett.*, vol. 61, no. 3, pp. 139–146, Mar. 1990.

[623] J. Tersoff, "New empirical approach for the structure and energy of covalent systems," *Phys. Rev. B, Condens. Matter*, vol. 37, no. 12, pp. 6991–7000, Apr. 1988.

[624] A. Lasa, K. Schmid, and K. Nordlund, "Modelling of W–Be mixed material sputtering under D irradiation," *Phys. Scripta*, vol. T159, Apr. 2014, Art. no. 014059.

[625] E. Safi, C. Björkas, A. Lasa, K. Nordlund, I. Sukuba, and M. Probst, "Atomistic simulations of the effect of reactor-relevant parameters on be sputtering," *J. Nucl. Mater.*, vol. 463, pp. 805–809, Aug. 2015.

[626] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.*, vol. 98, no. 14, Apr. 2007, Art. no. 146401.

[627] N. Bernstein, G. Csányi, and V. L. Deringer, "De novo exploration and self-guided learning of potential-energy surfaces," *npj Comput. Mater.*, vol. 5, no. 1, pp. 1–9, Oct. 2019.

[628] A. V. Shapeev, "Moment tensor potentials: A class of systematically improvable interatomic potentials," *Multiscale Model. Simul.*, vol. 14, no. 3, pp. 1153–1173, Jan. 2016.

[629] L. Chen, A. Kaiser, M. Probst, and S. Shermukhamedov, "Sputtering of the beryllium tungsten alloy Be2W by deuterium atoms: Molecular dynamics simulations using machine learned forces," *Nucl. Fusion*, vol. 61, no. 1, Dec. 2021, Art. no. 016031.

[630] L. Chen, I. Sukuba, M. Probst, and A. Kaiser, "Iterative training set refinement enables reactive molecular dynamics via machine learned forces," *RSC Adv.*, vol. 10, no. 8, pp. 4293–4299, Jan. 2020.

[631] *Wikipedia*. Accessed: May 20, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Activelearning(machinelearning)

[632] S. Kawaguchi, K. Takahashi, H. Ohkama, and K. Satoh, "Deep learning for solving the Boltzmann equation of electrons in weakly ionized plasma," *Plasma Sources Sci. Technol.*, vol. 29, no. 2, Feb. 2020, Art. no. 025021.

[633] C. Rao, H. Sun, and Y. Liu, "Physics-informed deep learning for incompressible laminar flows," *Theor. Appl. Mech. Lett.*, vol. 10, no. 3, pp. 207–212, Mar. 2020.

[634] N. Zobeiry and K. D. Humfeld, "A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications," *Eng. Appl. Artif. Intell.*, vol. 101, May 2021, Art. no. 104232.

[635] S. Kawaguchi and T. Murakami, "Physics-informed neural networks for solving the Boltzmann equation of the electron velocity distribution function in weakly ionized plasmas," *Jpn. J. Appl. Phys.*, vol. 61, no. 8, Aug. 2022, Art. no. 086002.

[636] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[637] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, "DeepXDE: A deep learning library for solving differential equations," *SIAM Rev.*, vol. 63, no. 1, pp. 208–228, Jan. 2021.

[638] S. Wang, Y. Teng, and P. Perdikaris, "Understanding and mitigating gradient flow pathologies in physics-informed neural networks," *SIAM J. Sci. Comput.*, vol. 43, no. 5, pp. A3055–A3081, Jan. 2021.

[639] H. Sugawara, "Configuration of propagator method for calculation of electron velocity distribution function in gas under crossed electric and magnetic fields," *Plasma Sci. Technol.*, vol. 21, no. 9, Sep. 2019, Art. no. 094001.

[640] A. Bogaerts, R. Gijbels, and J. Vlček, "Collisional-radiative model for an argon glow discharge," *J. Appl. Phys.*, vol. 84, no. 1, pp. 121–136, Jul. 1998.

[641] J. V. Dijk, B. Hartgers, J. Jonkers, and J. V. D. Mullen, "Collisional radiative models with multiple transport-sensitive levels—Application to high electron density mercury discharges," *J. Phys. D: Appl. Phys.*, vol. 34, no. 10, pp. 1499–1509, May 2001.

[642] C. De Bie, B. Verheyde, T. Martens, J. van Dijk, S. Paulussen, and A. Bogaerts, "Fluid modeling of the conversion of methane into higher hydrocarbons in an atmospheric pressure dielectric barrier discharge," *Plasma Processes Polym.*, vol. 8, no. 11, pp. 1033–1058, Nov. 2011.

[643] W. V. Gaens and A. Bogaerts, "Kinetic modelling for an atmospheric pressure argon plasma jet in humid air," *J. Phys. D: Appl. Phys.*, vol. 46, no. 27, Jun. 2013, Art. no. 275201.

[644] P. M. J. Koelman et al., "A comprehensive chemical model for the splitting of CO_2 in non-equilibrium plasmas," *Plasma Processes Polym.*, vol. 14, pp. 1–20, Apr. 2017.

[645] A. Bogaerts, C. De Bie, R. Snoeckx, and T. Kozák, "Plasma based CO_2 and CH_4 conversion: A modeling perspective," *Plasma Processes Polym.*, vol. 14, no. 6, Jun. 2017, Art. no. 1600070.

[646] J. L. Gustafson, "Reevaluating Amdahl's law," *Commun. ACM*, vol. 31, no. 5, pp. 532–533, May 1988.

[647] H. W. P. van der Heijden, M. G. H. Boogaarts, S. Mazouffre, J. A. M. van der Mullen, and D. C. Schram, "Time-resolved experimental and computational study of two-photon laser-induced fluorescence in a hydrogen plasma," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 61, no. 4, pp. 4402–4409, Apr. 2000.

[648] M. Bodenstein, "Eine theorie der photochemischen reaktionsgeschwindigkeiten," *Zeitschrift Physikalische Chem.*, vol. 85U, no. 1, pp. 329–397, Oct. 1913.

[649] A. S. Tomlin, T. Turanyi, and M. J. Pilling, "Mathematical tools for the construction, investigation and reduction of combustion mechanisms," in *Comprehensive Chemical Kinetics*, vol. 35, M. J. Pilling, Ed. Amsterdam, The Netherlands: Elsevier, 1997, ch. 4, pp. 293–437.

[650] D. R. Bates, A. E. Kingston, and R. W. P. McWhirter, "Recombination between electrons and atomic ions, I. Optically thin plasmas," *Proc. Roy. Soc. London A, Math. Phys. Sci.*, vol. A267, p. 297, May 1962.

[651] U. Maas and S. B. Pope, "Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space," *Combustion Flame*, vol. 88, nos. 3–4, pp. 239–264, Mar. 1992.

[652] U. Maas, "Some aspects of time-reversal in chemical kinetics," *Entropy*, vol. 22, no. 12, p. 1386, Dec. 2020.

[653] S. H. Lam, "Singular perturbation for stiff equations using numerical methods," in *Recent Advances in the Aerospace Sciences: In Honor of Luigi Crocco on His Seventy-Fifth Birthday*, C. Casci and C. Bruno, Eds. Boston, MA, USA: Springer, 1985, pp. 3–19.

[654] S. H. Lam and D. A. Goussis, "Understanding complex chemical kinetics with computational singular perturbation," in *Proc. 22nd Symp. Int. Combustion*, 1989, vol. 22, no. 1, pp. 931–941.

[655] S. H. Lam and D. A. Goussis, "The CSP method for simplifying kinetics," *Int. J. Chem. Kinetics*, vol. 26, no. 4, pp. 461–486, Apr. 1994.
[656] S. H. Lam, "Model reductions with special CSP data," *Combustion Flame*, vol. 160, pp. 270–711, Dec. 2013.
[657] P. Zhao and S. H. Lam, "Toward computational singular perturbation (CSP) without eigen-decomposition," *Combustion Flame*, vol. 209, pp. 63–73, Nov. 2019.
[658] U. Maas and S. B. Pope, "Implementation of simplified chemical kinetics based on intrinsic low-dimensional manifolds," in *Proc. 24th Symp. Combustion*, 1992, vol. 24, no. 1, pp. 103–112s.
[659] J. A. van Oijen, A. Donini, R. J. M. Bastiaans, J. H. M. T. T. Boonkkamp, and L. P. H. de Goey, "State-of-the-art in premixed combustion modeling using flamelet generated manifolds," *Prog. Energy Combustion Sci.*, vol. 57, pp. 30–74, Nov. 2016.
[660] S. B. Pope and U. Maas, "Simplifying chemical kinetics: Trajectory-generated low-dimensional manifolds," Mech. Aerosp. Eng. Rep., Ithaca, NY, USA, Tech. Rep. FDA, 1993, pp. 11–93.
[661] S. B. Pope, "Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation," *Combustion Theory Model.*, vol. 1, no. 1, pp. 41–63, Jan. 1997.
[662] T. Ding, T. Readshaw, S. Rigopoulos, and W. P. Jones, "Machine learning tabulation of thermochemistry in turbulent combustion: An approach based on hybrid flamelet/random data and multiple multilayer perceptrons," *Combustion Flame*, vol. 231, Sep. 2021, Art. no. 111493.
[663] T. Rehman, E. Kemaneci, W. Graef, and J. Van Dijk, "Simplifying plasma chemistry via ILDM," *J. Phys., Conf. Ser.*, vol. 682, pp. 1–7, Feb. 2016.
[664] K. Peerenboom, A. Parente, T. Kozák, A. Bogaerts, and G. Degrez, "Dimension reduction of non-equilibrium plasma kinetic models using principal component analysis," *Plasma Sources Sci. Technol.*, vol. 24, no. 2, Jan. 2015, Art. no. 025004.
[665] A. Bellemans, T. Magin, A. Coussement, and A. Parente, "Reduced-order kinetic plasma models using principal component analysis: Model formulation and manifold sensitivity," *Phys. Rev. Fluids*, vol. 2, no. 7, Jul. 2017, Art. no. 073201.
[666] R. Lehmann, "An algorithm for the determination of all significant pathways in chemical reaction systems," *J. Atmos. Chem.*, vol. 47, no. 1, pp. 45–78, Jan. 2004.
[667] A. H. Markosyan, A. Luque, F. J. Gordillo-Vázquez, and U. Ebert, "PumpKin: A tool to find principal pathways in plasma chemical models," *Comput. Phys. Commun.*, vol. 185, no. 10, pp. 2697–2702, Oct. 2014.
[668] P. Koelman, D. Yordanova, W. Graef, S. T. Mousavi, and J. V. Dijk, "Uncertainty analysis with a reduced set of input uncertainties selected using pathway analysis," *Plasma Sources Sci. Technol.*, vol. 28, no. 7, Jul. 2019, Art. no. 075009.
[669] J. Kruszelnicki, A. M. Lietz, and M. J. Kushner, "Atmospheric pressure plasma activation of water droplets," *J. Phys. D: Appl. Phys.*, vol. 52, no. 35, Jul. 2019, Art. no. 355207.
[670] T. Murakami and O. Sakai, "Rescaling the complex network of low-temperature plasma chemistry through graph-theoretical analysis," *Plasma Sources Sci. Technol.*, vol. 29, no. 11, Nov. 2020, Art. no. 115018.
[671] M. Hanicinec, S. Mohr, and J. Tennyson, "Fast species ranking for iterative species-oriented skeletal reduction of chemistry sets," *Plasma Sources Sci. Technol.*, vol. 29, no. 12, Dec. 2020, Art. no. 125024.
[672] E. Stoffels, I. E. Kieft, and R. E. J. Sladek, "Superficial treatment of mammalian cells using plasma needle," *J. Phys. D: Appl. Phys.*, vol. 36, no. 23, pp. 2908–2913, Dec. 2003.
[673] M. Laroussi, "Sterilization of contaminated matter with an atmospheric pressure plasma," *IEEE Trans. Plasma Sci.*, vol. 24, no. 3, pp. 1188–1191, Jun. 1996.
[674] M. D. Wilkinson et al., "The FAIR guiding principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, 2016, Art. no. 160018.
[675] R. Clemen et al., "Gas plasma technology augments ovalbumin immunogenicity and OT-II T cell activation conferring tumor protection in mice," *Adv. Sci.*, vol. 8, no. 10, May 2021, Art. no. 2003395.
[676] Z. Nasri et al., "Singlet-oxygen-induced phospholipase a 2 inhibition: A major role for interfacial tryptophan dioxidation," *Chem.-A Eur. J.*, vol. 27, no. 59, pp. 14702–14710, Oct. 2021.
[677] S. Wenske, J.-W. Lackmann, L. M. Busch, S. Bekeschus, T. von Woedtke, and K. Wende, "Reactive species driven oxidative modifications of peptides—Tracing physical plasma liquid chemistry," *J. Appl. Phys.*, vol. 129, no. 19, May 2021, Art. no. 193305.
[678] A. Schmidt et al., "Non-thermal plasma activates human keratinocytes by stimulation of antioxidant and phase II pathways," *J. Biol. Chem.*, vol. 290, no. 11, pp. 6731–6750, Mar. 2015.
[679] S. Bekeschus, J. Moritz, A. Schmidt, and K. Wende, "Redox regulation of leukocyte-derived microparticle release and protein content in response to cold physical plasma-derived oxidants," *Clin. Plasma Med.*, vols. 7–8, pp. 24–35, Dec. 2017.
[680] S. Bekeschus et al., "High throughput image cytometry micronucleus assay to investigate the presence or absence of mutagenic effects of cold physical plasma," *Environ. Mol. Mutagenesis*, vol. 59, no. 4, pp. 268–277, May 2018.
[681] K. Zocher et al., "Profiling microalgal protein extraction by microwave burst heating in comparison to spark plasma exposures," *Algal Res.*, vol. 39, May 2019, Art. no. 101416.
[682] A. Schmidt, S. Bekeschus, K. Wende, B. Vollmar, and T. von Woedtke, "A cold plasma jet accelerates wound healing in a murine model of full-thickness skin wounds," *Exp. Dermatol.*, vol. 26, no. 2, pp. 156–162, Feb. 2017.
[683] K. Wende et al., "On a heavy path—Determining cold plasma-derived short-lived species chemistry using isotopic labelling," *RSC Adv.*, vol. 10, no. 20, pp. 11598–11607, Mar. 2020.
[684] G. Bruno, S. Wenske, J.-W. Lackmann, M. Lalk, T. von Woedtke, and K. Wende, "On the liquid chemistry of the reactive nitrogen species peroxynitrite and nitrogen dioxide generated by physical plasmas," *Biomolecules*, vol. 10, p. 1687, Dec. 2020.
[685] G. Bruno, T. Heusler, J.-W. Lackmann, T. von Woedtke, K.-D. Weltmann, and K. Wende, "Cold physical plasma-induced oxidation of cysteine yields reactive sulfur species (RSS)," *Clin. Plasma Med.*, vol. 14, Jun. 2019, Art. no. 100083.
[686] S. Wenske, J.-W. Lackmann, S. Bekeschus, K.-D. Weltmann, T. von Woedtke, and K. Wende, "Nonenzymatic post-translational modifications in peptides by cold plasma-derived reactive oxygen and nitrogen species," *Biointerphases*, vol. 15, no. 6, Nov. 2020, Art. no. 061008.
[687] I. Murakami, M. Osakabe, M. Nishiura, A. Oda, J. Sugawara, and S. Hamaguchi, "Construction and use of atomic and molecular databases," *J. Plasma Fusion Res.*, vol. 88, p. 35, 2012.
[688] P. W. Stokes, D. G. Cocks, M. J. Brunger, and R. D. White, "Determining cross sections from transport coefficients using deep neural networks," *Plasma Sources Sci. Technol.*, vol. 29, no. 5, May 2020, Art. no. 055009.
[689] A. Kramida, Y. Ralchenko, J. Reader, and NIST ASD Team. (2021). *NIST Atomic Spectra Database (Version 5.9)*. [Online]. Available: https://physics.nist.gov/asd
[690] K. P. Dere, E. Landi, H. E. Mason, B. C. M. Fossi, and P. R. Young, "CHIANTI—An atomic database for emission lines: I. Wavelengths greater than 50 Å," *Astron. Astrophys. Suppl. Ser.*, vol. 125, no. 1, pp. 149–173, Oct. 1997.
[691] *ALADDIN, Numerical Database Maintained by the IAEA Nuclear Data Section A+M Data Unit*. Accessed: Jul. 10, 2023. [Online]. Available: https://amdis.iaea.org/databases/
[692] I. Murakami, M. Kato, M. Emoto, D. Kato, H. A. Sakaue, and T. Kawate, "NIFS atomic and molecular numerical database for collision processes," *Atoms*, vol. 8, no. 4, p. 71, Oct. 2020.
[693] *Open-Adas*. Accessed: Jul. 10, 2023. [Online]. Available: https://open.adas.ac.uk
[694] W. Cunto and C. Mendoza, "The opacity project—The topbase atomic datbase," *Revista Mexicana Astronomia Astrofisica*, vol. 23, p. 107, 1992.
[695] K. Olsen et al. *NIST-LANL Lanthanide Opacity Database*. Accessed: Jul. 10, 2023. [Online]. Available: https://nlte.nist.gov/OPAC/opacitysearch.html
[696] E. Carbone et al., "Data needs for modeling low-temperature non-equilibrium plasmas: The LXCat project, history, perspectives and a tutorial," *Atoms*, vol. 9, no. 1, p. 16, Feb. 2021.
[697] D. Albert et al., "A decade with VAMDC: Results and ambitions," *Atoms*, vol. 8, no. 4, p. 76, 2020.
[698] A. Sakudo, Y. Yagyu, and T. Onodera, "Disinfection and sterilization using plasma technology: Fundamentals and future perspectives for biological applications," *Int. J. Mol. Sci.*, vol. 20, no. 20, p. 5216, Oct. 2019.
[699] J. Šimončicová, S. Kryštofová, V. Medvecká, K. Ďurišová, and B. Kaliňáková, "Technical applications of plasma treatments: Current state and perspectives," *Appl. Microbiol. Biotechnol.*, vol. 103, no. 13, pp. 5117–5129, Jul. 2019.

[700] A. Zille, F. R. Oliveira, and A. P. Souto, "Plasma treatment in textile industry," *Plasma Processes Polym.*, vol. 12, no. 2, pp. 98–131, Feb. 2015.

[701] M. Laroussi, "Plasma medicine: A brief introduction," *Plasma*, vol. 1, no. 1, pp. 47–60, Feb. 2018.

[702] L. Gao, X. Shi, and X. Wu, "Applications and challenges of low temperature plasma in pharmaceutical field," *J. Pharmaceutical Anal.*, vol. 11, no. 1, pp. 28–36, May 2021.

[703] R. C. Sanito, S.-J. You, and Y.-F. Wang, "Application of plasma technology for treating e-waste: A review," *J. Environ. Manage.*, vol. 288, Jun. 2021, Art. no. 112380.

[704] E. S. P. Prado, F. S. Miranda, L. G. de Araujo, G. Petraconi, and M. R. Baldan, "Thermal plasma technology for radioactive waste treatment: A review," *J. Radioanalytical Nucl. Chem.*, vol. 325, no. 2, pp. 331–342, Aug. 2020.

[705] M. Gorley et al., "The EUROfusion materials property handbook for DEMO in-vessel components—Status and the challenge to improve confidence level for engineering data," *Fusion Eng. Des.*, vol. 158, Sep. 2020, Art. no. 111668.

[706] *Plasma-Wall Interaction With Irradiated Tungsten and Tungsten Alloys in Fusion Devices*, At. Plasma Interact. Data, IAEA, 2019, vol. 18.

[707] *Atomic and Plasma Interaction Data*, IAEA Publication Series, 1991.

[708] *CascadesDB Database of Molecular Dynamics Simulations of Collisional Cascades*, IAEA. Accessed: Jul. 10, 2023. [Online]. Available: https://cascadesdb.iaea.org/

[709] *DefectDB Database of Density Functional Theory Calculations of Radiation-Induced Defects in Nuclear Materials*, IAEA. Accessed: Jul. 10, 2023. [Online]. Available: https://amdis.iaea.org/databases/

[710] U. Bhardwaj, H. Hemani, M. Warrier, and N. Semwel, "CSaransh: Software suite to study molecular dynamics simulations of collision cascades," *J. Open Source Softw.*, vol. 4, no. 41, 2019. Accessed: Jul. 10, 2023, doi: 10.21105/joss.01461.

[711] HCDB. Accessed: Jul. 10, 2023. [Online]. Available: https://db-amdis.org/hcdb

[712] *AMBDAS: Atomic and Molecular Bibliographic Data System*, IAEA. Accessed: Jul. 10, 2023. [Online]. Available: https://amdis.iaea.org/db/ambdas/

[713] *ALADDIN: A Labelled Atomic Databases*, IAEA. Accessed: Jul. 10, 2023. [Online]. Available: https://amdis.iaea.org/databases/

[714] J. Tennyson et al., "QDB: A new database of plasma chemistries and reactions," *Plasma Sources Sci. Technol.*, vol. 26, no. 5, Apr. 2017, Art. no. 055014.

[715] J. Tennyson et al., "The 2021 release of the quantemol database (QDB) of plasma chemistries and reactions," *Plasma Sources Sci. Technol.*, vol. 31, no. 9, Dec. 2022, Art. no. 095020.

[716] M. Klute et al., "Modelling of a miniature microwave driven nitrogen plasma jet and comparison to measurements," *Plasma Sources Sci. Technol.*, vol. 30, no. 6, Jun. 2021, Art. no. 065014.

[717] V. Wakelam et al., "A kinetic database for astrochemistry (KIDA)," *Astrophysical J. Suppl. Ser.*, vol. 199, no. 1, p. 21, Mar. 2012.

[718] J.-H. Park et al., "A new version of the plasma database for plasma physics in the data center for plasma properties," *Appl. Sci. Converg. Technol.*, vol. 29, no. 1, pp. 5–9, Jan. 2020.

[719] R. A. Hulse, "The ALADDIN atomic physics database system," *AIP Conf.*, vol. 206, no. 1, pp. 63–72, 1990.

[720] R. Celiberto et al., "Atomic and molecular data for spacecraft re-entry plasmas," *Plasma Sources Sci. Technol.*, vol. 25, no. 3, Jun. 2016, Art. no. 033004.

[721] M.-L. Dubernet et al., "BASECOL2012: A collisional database repository and web service within the virtual atomic and molecular data centre (VAMDC)," *Astron. Astrophys.*, vol. 553, p. A50, May 2013.

[722] D. McElroy, C. Walsh, A. J. Markwick, M. A. Cordiner, K. Smith, and T. J. Millar, "The UMIST database for astrochemistry 2012," *Astron. Astrophys.*, vol. 550, p. A36, Feb. 2013.

[723] D. Braný, D. Dvorská, E. Halašová, and H. Škovierová, "Cold atmospheric plasma: A powerful tool for modern medicine," *Int. J. Mol. Sci.*, vol. 21, no. 8, p. 2932, Apr. 2020.

[724] M. Magureanu, C. Bradu, and V. I. Parvulescu, "Plasma processes for the treatment of water contaminated with harmful organic compounds," *J. Phys. D: Appl. Phys.*, vol. 51, no. 31, Jul. 2018, Art. no. 313002.

[725] M. Magureanu, F. Bilea, C. Bradu, and D. Hong, "A review on non-thermal plasma treatment of water contaminated with antibiotics," *J. Hazardous Mater.*, vol. 417, Sep. 2021, Art. no. 125481.

[726] Y. Ralchenko, "NIST atomic spectra database," *Memorie Della Societa Astronomica Italiana Supplementi*, vol. 8, p. 96, 2005.

[727] L. C. Pitchford et al., "LXCat: An open-access, web-based platform for data needed for modeling low temperature plasmas," *Plasma Processes Polym.*, vol. 14, nos. 1–2, Jan. 2017, Art. no. 1600098.

[728] D. Coster et al., "Towards making fusion data FAIR," in *Proc. IEEE 17th Int. Conf. eScience (eScience)*, Sep. 2021, pp. 233–234.

[729] C. Fabry, A. Pittner, V. Hirthammer, and M. Rethmeier, "Recommendations for an open science approach to welding process research data," *Weld. World*, vol. 65, no. 9, pp. 1661–1669, Sep. 2021.

[730] X. Chen et al., "Open is not enough," *Nature Phys.*, vol. 15, no. 2, pp. 113–119, 2019.

[731] M. Garofalo, A. Botta, and G. Ventre, "Astrophysics and big data: Challenges, methods, and tools," *Proc. Int. Astronomical Union*, vol. 12, pp. 345–348, Oct. 2016.

[732] S. Franke, L. Paulet, J. Schäfer, D. O'Connell, and M. M. Becker, "Plasma-MDS, a metadata schema for plasma science with examples from plasma technology," *Sci. Data*, vol. 7, no. 1, p. 439, Dec. 2020.

[733] *Research Data Repository of the RUB Research Department Plasmas With Complex Interactions*. Accessed: Nov. 30, 2021. [Online]. Available: https://rdpcidat.rub.de

[734] *Github Repository for Plasma-MDS*. Accessed: Nov. 30, 2021. [Online]. Available: https://github.com/plasma-mds/

[735] M. M. Becker et al., "Metadata schema and ontologies for FAIR research data in plasma technology," Zenodo, 2020, doi: 10.5281/zenodo.4091401.

[736] H.-K. Chung et al., Eds., *Uncertainty Assessment and Benchmark Experiments for Atomic and Molecular Data for Fusion Applications: Summary Report of an IAEA Technical Meeting*. Vienna, Austria: IAEA, Dec. 2006.

[737] H.-K. Chung et al., "Uncertainty estimates for theoretical atomic and molecular data," *J. Phys. D: Appl. Phys.*, vol. 49, no. 36, Aug. 2016, Art. no. 363002.

[738] *CollisionDB*, IAEA, Vienna, Austria. Accessed: Jul. 10, 2023. [Online]. Available: https://db-amdis.org/collisiondb/

[739] L. G. Christophorou and J. K. Olthoff, "Electron interactions with plasma processing gases: Present status and future needs," *Appl. Surf. Sci.*, vol. 192, nos. 1–4, pp. 309–326, May 2002.

[740] M.-Y. Song and J.-S. Yoon, "Cross sections for electron collisions with methane," *J. Phys. Chem. Reference Data*, vol. 44, no. 2, Jun. 2015, Art. no. 023101.

[741] E. Yildiz, C. Møller, and A. Bilberg, "Virtual factory: Digital twin based integrated factory simulations," *Proc. CIRP*, vol. 93, pp. 216–221, Jan. 2020.

[742] A. Cho et al., "A planning study for virtual DEMO development in Korea," *Fusion Eng. Des.*, vol. 176, Mar. 2022, Art. no. 113026.

[743] B. Cooper et al., "Quantemol electron collisions (QEC): An Enhanced expert system for performing electron molecule collision calculations using the R-matrix method," *Atoms*, vol. 7, no. 4, p. 11, Oct. 2019, doi: 10.3390/atoms7040097.

[744] L. Zhong, "Fast prediction of electron-impact ionization cross sections of large molecules via machine learning," *J. Appl. Phys.*, vol. 125, no. 18, May 2019, Art. no. 183302.