SAND2021-2663J

The data-driven future of high energy density physics

Peter W. Hatfield,[1, a)] Jim A. Gaffney,[2, b)] Gemma J. Anderson,[2, c)] Suzanne Ali,[2] Luca Antonelli,[3] Suzan Başeğmez du Pree,[4] Jonathan Citrin,[5] Marta Fajardo,[6] Patrick Knapp,[7] Brendan Kettle,[8] Bogdan Kustowski,[2] Michael J. MacDonald,[2] Derek Mariscal,[2] Madison E. Martin,[2] Taisuke Nagayama,[7] Charlotte A.J. Palmer,[9] J. L. Peterson,[2] Steven Rose,[1, 8] JJ Ruby,[10] Carl Shneider,[11] Matt J.V. Streeter,[8] Will Trickey,[3] and Ben Williams[12]

[1)] *Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK*

[2)] *Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA 94550, USA*

[3)] *York Plasma Institute, Department of Physics, University of York, Heslington, York YO10 5DD, UK*

[4)] *Nikhef, National Institute for Subatomic Physics, Science Park 105, 1098 XG, Amsterdam, Netherlands*

[5)] *DIFFER - Dutch Institute for Fundamental Energy Research, Eindhoven, Netherlands*

[6)] *Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, 1049-001 Lisboa, Portugal*

[7)] *Sandia National Laboratories, 1515 Eubank SE, Albuquerque, New Mexico 87185-1196, USA*

[8)] *Imperial College London, London, SW7 2AZ, UK*

[9)] *School of Mathematics and Physics, Queen's University Belfast, Belfast, BT7 1NN, UK*

[10)] *Laboratory for Laser Energetics, University of Rochester, Rochester, NY, USA*

[11)] *Dutch National Center for Mathematics and Computer Science (CWI), Science Park 123, 1098 XG, Amsterdam, Netherlands*

[12)] *AWE Plc, Aldermaston, Reading, RG4 7PR, United Kingdom*

(Dated: 24 February 2021)

PREFACE

The study of plasma physics under conditions of extreme temperatures, densities and electromagnetic field strengths is significant for our understanding of astrophysics, nuclear fusion and fundamental physics. These extreme physical systems are strongly non-linear and very difficult to understand theoretically or optimize experimentally. Here, we argue that machine learning models and data-driven methods are in the process of reshaping our exploration of these extreme systems that have hitherto proven far too non-linear for human researchers. From a fundamental perspective, our understanding can be helped by the way in which machine learning models can rapidly discover complex interactions in large data sets. From a practical point of view, the newest generation of extreme physics facilities can perform experiments multiple times a second (as opposed to ∼daily) – moving away from human-based control towards automatic control based on real-time interpretation of diagnostic data and updates of the physics model. To make the most of these emerging opportunities, we advance proposals for the community in terms of research design, training, best practices, and support for synthetic diagnostics and data analysis.

'Set the controls for the heart of the Sun' encouraged a 2004 paper[1] (riffing on the 1968 Pink Floyd song), describing the bright future of using Earth based experiments to create conditions similar to inside the Sun in the lab. Seventeen years later substantial advances have been made in this research programme. A question that is presently emerging however is *who* should be at the controls - humans, or artificial intelligences?

In the last few years plasma physics has been steadily beginning to explore the use of modern day data science and artificial intelligence methods to support research goals[2,3]. In this article we will identify data science issues for physics specifically at the extreme[4]; extremely high temperatures, densities or electromagnetic field strengths - which have unique challenges. In particular phenomena at these conditions are highly *non-linear* - small parameter changes can lead to large changes in behaviour. Interpreting extreme physics data typically requires simultaneously comprehending large amounts of complex multi-modal data from multiple different sources. Optimising extreme physics systems requires fine-tuning over large numbers of (often highly correlated) parameters. Artificial Intelligence (AI) methods have proved highly successful at teasing out correlations in large data sets like these and we believe will be crucial for understanding and optimising systems that up to now have been inscrutable. These extreme conditions can be found in astrophysical scenarios, but can also be created using high-energy 'drivers' (often lasers) in the laboratory - millimetre sized plasmas with temperatures and pressures higher than the centre of the Sun. The field has seen an explosion of interest in machine learning techniques because new and future laser facilities have much higher shot (and corresponding data) rates than previous facilities. Data

[a)] Electronic mail: peter.hatfield@physics.ox.ac.uk

[b)] Electronic mail: gaffney3@llnl.gov

[c)] Electronic mail: anderson276@llnl.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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-NA0003525.

from laboratory experiments can help us understand astrophysical plasmas, let us probe new phenomena in particle physics that conventional accelerators can't reach, and might even lead the way to nuclear fusion as a power source. In this paper we will highlight what data science issues are relevant for extreme plasma science, discuss successes in the field, identify what challenges remain, and look towards the future.

One of the most challenging areas of extreme plasma physics is High Energy Density Physics (HEDP), a sub-field dating back to the 1940s that seeks to understand the behaviour of macroscopic matter that is simultaneously at *both* very high temperatures and pressures; typically $>10^7$ K and $>10^6$ bar. At these conditions, several complex areas of physics become relevant and highly coupled, making *ab initio* predictions very challenging. For example, in many circumstances HEDP plasmas can start to inherit properties from both 'classical plasmas' (where the behaviour of the matter can to some degree be thought of as a gas of both ions and free electrons) and condensed matter (matter at solid density where strong interactions between bound electrons are relevant)[5]. Key contemporary problems in HEDP theory include understanding multi-species plasmas, self-consistent emission, absorption, and scattering of radiation, non-equilibrium plasmas, relativistic electron transport, magnetised plasmas, and quantum electrodynamic effects.

Understanding HEDP is of both great theoretical and practical importance. As already discussed, understanding these conditions is key in astrophysics. The field is also a ripe domain for new areas of fundamental physics; it is becoming possible to study particle and nuclear physics through HEDP experiments, as well as novel phenomena that are predicted to only emerge at extreme conditions e.g. the predicted thermal Schwinger process - spontaneous production of electrons and positrons at very high electric field strengths[6]. Extreme physics and high power laser science have given access to exotic forms of matter e.g. new forms of ice[7] and metallic hydrogen[8]. HEDP experiments are also at the forefront of the development of new classes of particle accelerators[9] (e.g. Laser Wakefield acceleration[10], bright gamma ray sources[11], laser-driven ion acceleration[12] and highly efficient neutron generation[13]) with wide ranging application across science including condensed matter physics, material science, and biomedical imaging[14]. Finally, the high temperatures and pressures of HEDP are one way to make nuclear fusion as a clean industrial power source a reality via *inertial confinement fusion*[15] (ICF).

HEDP has a rich heritage of experimentation, currently practiced by thousands of scientists in several large facilities around the world. The National Ignition Facility (NIF) at Lawrence Livermore National Laboratory (LLNL) is the most energetic laser in the world, and is the premier facility working towards ICF[16], as well as operating an exciting Discovery Science programme[17,18]. At the other end of the energy spectrum are high repetition rate lasers (for example Gemini at the Central Laser Facility, CLF[19]), which are much less energetic (and so can typically reach less extreme conditions), but can fire up to many times a second rather than at most a few times a day at NIF. There are many more facilities, each with unique capabilities, and a range of other technologies that are used around the world in HEDP experiments, including gas guns, Z-pinches, proton beams, pulsed power, X-Ray Free-Electron Lasers (XFELs) and ion accelerators (e.g. the Facility for Antiproton and Ion Research[20]). New facilities and upgrades are constantly in planning, and understanding the data we currently have is directly relevant to choices about what facilities we will need in the future. There are also important synergies with closely related areas of physics, for example magnetic confinement fusion (MCF), solar probes, and the detection of high energy cosmic rays. Finally, alongside experimentation, huge computational facilities have traditionally been a key part of HEDP, with the development of many sophisticated simulation codes run on top high performance computing (HPC) facilities. At the time of writing, ~ 6 of the top ten supercomputers in the world are used in some capacity for simulating plasma physics experiments like ICF.

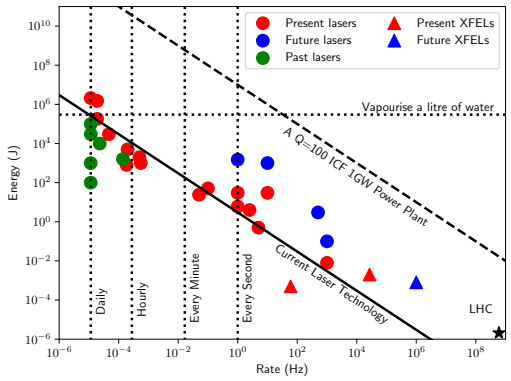

FIG. 1. Shot rates and energy of large high powered laser facilities in different eras. Shot rates and energies plotted are representative rather than definitive; facilities typically can operate in a number of slightly different modes, and in addition other laser properties (long pulse versus short pulse, beam colour etc.) are also important. Note also that higher shot numbers enable more parameter space to be probed, or a higher signal-to-noise ratio to be reached, but do not automatically translate into more data (depending on what diagnostics are used and the experiment performed). Facilities included are: NIF, LMJ, Omega, Gemini, Vulcan, Orion, TA2, Artemis, Vega, Titan, Texas Petawatt, HAPLS, SG-III, CoReLS 4PW, RRCAT 150TW, LULI2000, ALLS, Shanghai Superintense Ultrafast Laser Facility, DRACO (current), Allegra, EPAC DiPOLE, TARANIS-X, Station of Extreme Light (future), Nova, SHIVA, Cyclops, Argus, SG-II, ISKRA-V (past)[21–30], SACLA, the European XFEL (current XFELs) and LCLS-II (a future XFEL)[31–33]. The solid line illustrates the approximate state of current technology. We show for context the collision rate, and energy per collision for the LHC[34], although such figures are not directly comparable. We also indicate with a dashed line the shot rates that would be need to be achieved for an ICF power plant[a], to illustrate the long term aspirations of the field.

[a] For $Q = 100$; an output energy to input energy ratio of 100 per shot

The field is entering the regime where it is necessary to systematically manage large quantities of data, both because the amount of experimental data is set to massively grow, and also because the capacity to simulate huge numbers of exper-

iments is moving beyond the limits of conventional methods. The quantity of experimental data is increasing both because shot rates on facilities are dramatically increasing (see Figure 1), but also because diagnostics are becoming more sophisticated; ∼150GB of data is taken on each NIF shot, and LCLS campaigns have reached ∼70GB per minute[35]. Machine learning[36], Bayesian methods[37], and data-driven science[38] have been common in particle physics and astrophysics for many years, and are having large impacts on other multiscale, highly non-linear areas of physics e.g. climate science and Earth system science[39,40]. Some AI solutions from other fields are likely to be applicable in plasma physics - but HEDP also has its own unique challenges. Specifically we typically want to (sometimes very rapidly) fine tune (either optimising or fitting a model) a large number of parameters for a desired outcome, based on a large number of multi-modal data sets. This is very difficult for humans, as it is very hard to simultaneously comprehend all the different sources and forms of data - but is achievable for AI. In HEDP, the highest performing experiments in the field are now increasingly data driven[41,42]; the vision is to work towards a HEDP science where algorithms are at the centre of design, analysis of experiments, and discovery.

In the following sections we review key challenges and topics in extreme plasma physics, and highlight research areas where data science has dramatically impacted the field. We further lay out what the future might hold for data driven Extreme Physics and HEDP - and how the community must change and adapt how it practices research to make the most of these exciting new approaches.

CHALLENGES

The qualities that make HEDP an exciting area of research also contribute to causing the challenges inherent in any quantitative analysis of data. Table I summarises some key quantitive methods, comparing conventional and emerging approaches, and in the following sub-sections we discuss three primary challenges that data science techniques can help address in HEDP.

Experimental Design and Automation

A key component of HEDP is experimentation. Designing experiments is a hugely complex task, and researchers typically will have a wide range of overlapping goals during this phase. Researchers must consider what specific hypotheses are to be tested, and if the expected data would be sufficient to rule out alternatives. What design to shoot, diagnostic instruments to field, or astronomical observation to make will depend heavily on the specific science goal at hand. In current design approaches there is typically much intuition implicitly at play, and often experiments are built as an extension of what has been done before, limiting the regions of experimental space that are studied. Machine learning techniques offer a possible framework for the intuition of the computational scientists and experimental scientists to be explicitly included in a cohesive picture that considers both measurements that can be made, and which aspects of physics have the most leverage on those measurements. AI-aided design is beginning to be used in the creation of new HEDP experiments[43,44], and we foresee that becoming the norm in the coming years. However, machine learning methods have yet to demonstrate that they can "think outside the box" of pre-defined parameter spaces, therefore for the foreseeable future it will still be necessary to have substantial human input in the design process.

Experiments on state-of-the-art high repetition rate lasers, firing many times a second, cannot be done with a human in the loop - so in this case some algorithmic control is essential. This in principle could lead to huge savings of time, money, and human effort in the near future - allowing them to be redirected to aspects of research where they can be better used. Automating experiments combines control of experimental parameters and real-time analysis of experimental results into a single algorithm. The experimental goals are coded in to the automation, such that choices of how to vary the experimental inputs are made automatically in order to maximise the output.

Automation also allows for active feedback stabilisation of complex processes. This is of particular benefit to HEDP experiments, where many non-linear effects combine to determine the performance of what is a nominally unpredictable process. In this context, AI could be the solution to an otherwise intractable problem; the well-known ability of deep learning models to discover complex interactions[45] could be used to optimize systems that have proven far too non-linear for human researchers.

Data Synthesis

The measurements made in HEDP experiments are often highly integrated; experiments typically don't measure the actual quantity of interest (QoI), and there are usually multiple confounding or nuisance variables that have to be controlled. Isolating a particular aspect of these systems is often not possible, resulting in a measurement of an evolving system subject to different conditions and physical processes. This complexity makes repeatability an issue. In order to isolate individual aspects of the underlying physics, experiments therefore typically require multiple, indirect observations, sometimes spread across several different experimental facilities. At most facilities researchers have developed multiple diagnostics for experiments; for example both x-ray and particle spectra may be measured on a single experiment, along with many other forms of experimental data - all of which might contribute to the determination of a single quantity. The analysis of such increasingly sophisticated interlinked data requires the use of more advanced modelling techniques to make the best use of the available data, and to sensibly quantify the uncertainty on any inferences. Looking forward, large quantities of data will require the development of streamlined automated data analysis tools to avoid read/write bottlenecks[46].

As well as combining data from multiple diagnostics, there

Analytical Task	Scientific Task	Conventional approaches	Limitations of conventional approaches	Emergent or potential approaches
Uncertainty Quantification	Quantifying uncertainty on estimate of some microphysics (e.g. opacity, equation-of-state), characterising uncertainty estimate on laser energy required for ignition	Simulation-based sensitivity studies, basic Poisson/Gaussian uncertainties, not including correlations	Local sensitivities only, relies on estimated uncertainties in underlying parameters, cannot account for simulation bias	MCMC, dropout, bootstrapping, quantification of 'unknown unknowns', invertible neural networks
Regression	Emulating an expensive simulation, building an empirical model of microphysics	Look-up tables, polynomial regression	Struggles in high dimension, incorporating known physics constraints, extrapolation is difficult, often not fast enough	Neural networks, Gaussian processes, autoencoders
Design	Selecting target design parameters, scheduling observation runs	Adjusting parameters by hand, using a combination of code output and Designer judgement	Hard in high dimensions, human intensive, cannot be done quickly, can miss optimal/novel designs	Bayesian optimisation, Genetic/Evolutionary algorithms
Pattern Recognition	Identifying target defects, characterising magnetic perturbations, image segmentation/featurization	Human inspection	Laborious and time consuming, subject to individual biases, uncontrolled approximation to true information content	Convolutional Neural Networks, Deep learning, random forest classifiers, Human-AI hybrid approaches, Application Specific Integrated Circuits incorporated into diagnostics, Generative adversarial networks
Data Synthesis	Combining data from multiple sources (e.g. different instruments)	Researcher guided inference from independently analysed diagnostic data	Difficult to do 'by hand', hard to take advantages of any degeneracies broken	Bayesian inference, data assimilation methods
Classification	Image classification, particle track classification, identify good/bad shots, anomaly detection	Simple analytic criteria, human inspection	Laborious and time consuming, potentially inaccurate	Random forest, decision trees, neural networks, deep learning, Generative adversarial networks
Model Calibration	Update physics models/parameters in the face of experimental data	Trial and error, single-point fitting to data, hand-tuning	Laborious and time consuming, inaccurate or missing treatment of uncertainties, miss multiple solutions, prone to overfitting	Bayesian inference, discrepancy modeling, transfer learning, multi-task learning, physics informed neural networks

TABLE I. Conventional and AI-enabled approaches to tasks in extreme plasma physics

are also challenges in combining data from multiple sources; 'data synthesis', where multiple forms of heterogeneous data are combined in a self-consistent manner in order to construct a more complete picture of the phenomenon of interest. Typically each diagnostic will be analysed separately, and overall conclusions reached heuristically by the researcher. However, combining all available data can help reduce error bars, break degeneracies, and cancel out uncorrelated systematics[47].

The long term vision for the best use of the physics data is to develop systems to combine data from multiple diagnostics on the same shot, multiple shots, shots on different facilities, and finally from different types of facility.

Physics models

The evolution of HEDP experiments is governed by multiple complex, non-linear physics models, each of which has its own range of applicability and uncertainties. Solving these multiphysics computer models requires very expensive simulations, which are often not suited to next-generation HPC platforms. With the increasing desire to explore larger experimental parameter spaces, there is a shifting dynamic between single, incredibly computationally expensive, 'hero' simulations, and large-scale ensembles of simulations which only become meaningful when confronted with experimental data.

Due to numerical approximations, poorly known or unknown model parameters, and missing physics (*model discrepancy*), computer models often do not accurately represent the physical process under study. We can leverage real-world experiments to calibrate our computational models, enabling us to constrain some of the uncertain model parameters; ideally including an Uncertainty Quantification (UQ) analysis[48] and practicing 'data assimilation'[49] that obeys physical laws. A useful approach to this problem is through *Bayesian inversion* (also known as model calibration)[50,51], which allows

prior knowledge to be be included and for which convenient numerical tools exist. In a HEDP context, challenges with this approach include large parameter spaces, expensive models, and very sparse experimental data.

The computer models can often be prohibitively computationally expensive. In this case a *surrogate model (or emulator)* may be useful - running a moderate number of expensive simulations, and training a machine learning algorithm to reproduce what the simulation would have given as an output. Surrogate models are of course themselves only an approximation of the true model, introducing further uncertainty that needs to be accounted for.

Emulation can be done at the macro level (e.g. predicting outputs of a whole experiment), or at the level of individual modules run inline inside of a computer model. The use of emulators opens up an array of new inference methods that would not be practical with the full computational expense of a conventional simulation[52–58].

CASE STUDIES

Here we highlight three key areas where researchers are tackling the challenges described in Section II, and where data science is significantly impacting the practice of extreme plasma science.

Astrophysics

Plasmas are found throughout the Universe: in Solar physics (the centre of the Sun, the solar corona, solar wind); interplanetary, interstellar and intergalactic media; in Earth's and other planetary magnetospheres and ionospheres, and tails of comets; in compact astrophysical objects (white dwarfs, neutron stars, and black holes) and their accretion disks. After direct matter-antimatter annihilation, accretion onto compact objects is the most efficient energy source in the Universe[59]; the Cosmos offers plenty of opportunity to probe extreme physics.

Understanding our closest star is of course of supreme practical importance - alongside curiosity-driven blue sky astrophysical motivation. Space weather refers to conditions on the Sun, in the heliosphere, in the solar wind, and in Earth's magnetosphere, ionosphere, and thermosphere, that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health[60,61]. Changes in the space environment, resulting mainly from changes on the Sun, include modification of the ambient plasma, particulate radiation (electrons, protons and ions), electromagnetic radiation (including radio, visible, UV, X-ray, and gamma radiation), and magnetic and electric fields.

As with many other areas of physics, the amount of data we have on the multi-scale complex physics experiment that is the Sun has massively increased in recent years from Solar space missions e.g. from both spacecraft, such as the Solar and Heliospheric Observatory (SOHO), the Hinode mission, the Solar Dynamics Observatory (SDO), the Parker Solar Probe (PSP) and Solar Orbiter (SolO), and CubeSats, such as the Miniature X-ray Solar Spectrometer (MinXSS). Machine learning has been employed successfully to both forecast and 'now-cast' space weather[62]. It has been used to make predictions and gain insight about: solar wind[63], solar flares[64,65], coronal mass ejections (CMEs)[66,67], Van Allen radiation belts[68], geomagnetically induced currents[69] and the role of auroras as proxy for ionospheric disturbances[70].

Outside our Solar System, large survey telescopes are taking huge amounts of data on astronomical bodies with extreme physics. A full range of data science based techniques are used to both a) identify objects of interest in the large data sets[71], and b) understand their underlying physics. Machine learning and statistical methods have been used to: make a Bayesian constraint on supra-nuclear equations-of-state[72], understand the interiors of exoplanets[73], constrain fundamental stellar parameters from asteroseismic observations[74], classify supernovae[75], identify and infer the properties of white dwarfs[76], classify states of black hole X-ray binaries[77], and emulate radiative transfer during the epoch of cosmological reionization[78].

HEDP experimental facilities can, as discussed, also probe the extreme plasma physics relevant in astrophysical bodies. For example, experiments have measured the equation of state at conditions relevant for the centres of gas giant exoplanets[79,80], tested theories on possible origins of magnetic fields on galactic scales[81] and helped our understanding of white dwarf photosphere spectra[82]. There has recently been great success in applying data science methods to these experiments and making inferences with realistic uncertainties. Measurements of iron opacity at conditions relevant to solar physics on the Z facility for example have modelled a large number of possible sources of uncertainty and systematics, and have combined the data from multiple shots together to get realistic uncertainties. This measurement was statistically inconsistent with conventional opacity predictions and has potentially led to an adjustment of estimates of the metallicity (lithium and higher atomic number elements) of the Sun[83,84]. Similarly, Bayesian estimates from experiment of nuclear reaction rate at conditions relevant for Big Bang nucleosynthesis were found to be 3% different from conventionally assumed; a level of precision needed for cosmological studies[85]. As discussed in the introduction, future facilities will have shot rates that will make multi-shot studies like this the norm. Critically this will make it possible to consistently make realistic uncertainty quantifications of key parameters, and will also give the ability to probe large parts of parameter space (e.g. measure the equation-of-state or opacity at a large number of points in temperature-density space). With these vast quantities of observational and experimental data, the natural next step is to use experimentally calibrated models of microphysics like these in astrophysical models - with the potential to give improved predictions over theory-only models.

There is a huge amount of data on astrophysical plasmas of a wide variety of sources, taken in a huge variety of ways, but unfortunately they are currently held in different forms, by different communities. See Figure 2 for an infogram on

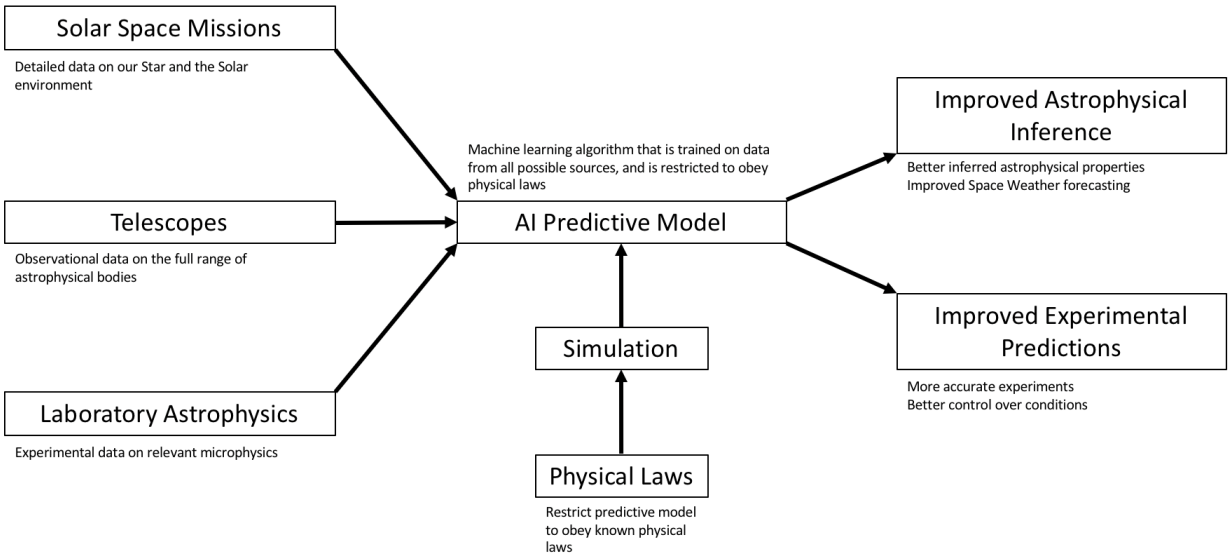

FIG. 2. **Integration of astrophysical information:** combining data from multiple sources. At the centre is a machine learning algorithm that is receiving data from multiple sources, and is able to update its beliefs based on this ('data assimilation'). The resulting data-driven results will take information from both theory, and data, to give more precise predictions, with realistic uncertainties.

how different astrophysical data sets could in future be combined in a data assimilation framework[86,87]. Astronomers have long practiced *'multi-wavelength' astronomy* (astronomy observing at multiple electromagnetic wavelengths), and since $\sim$2015 have practiced *'multi-messenger' astronomy* (astronomy taking data from multiple messengers of information, electromagnetic waves, gravitational waves, cosmic rays and neutrinos). It is now time for *'multi-provenance' astronomy*; integrating data from both observations and experiments into one coherent model of our understanding of astrophysical sources.

Inertial Confinement Fusion

The aim of igniting a self-propagating hydrogen fusion reaction in laboratory-scale plasmas has been a scientific grand challenge for over 50 years. The motivation is clear: a reliable fusion reaction could form the basis for an effectively limitless, clean, and safe energy source[88]. There are good reasons the research has been long-running; generating densities and temperatures high enough to overcome the Coulomb repulsion between nuclei and Bremsstrahlung radiation losses from electrons, over timescales long enough to allow energy break-even, is extremely difficult. In nature, the required conditions are reached in the cores of stars as a result of gravitational collapse and confinement. On Earth, we require even more extreme conditions to account for the much shorter time scales.

In inertial confinement fusion (ICF) studies, millimetre-scale targets are driven to implode using high energy (100s kJ to MJ-class) drivers[89]. For fusion ignition, the implosion and subsequent stagnation needs to generate a high temperature (10^7-10^8 K) hydrogen plasma, surrounded by a compressed ($100\times$ - $1000\times$) fuel layer, confined by its own inward velocity for several hundred picoseconds[90]. Experiments on current ICF facilities like the National Ignition Facility (NIF)[91,92], the Omega Laser Facility[93], and the Z pulsed-power machine[94,95] routinely generate plasmas under solar core conditions and beyond.

ICF experiments present distinct data challenges due to their scale and complexity. Experimental facilities are expensive and are not expected to achieve high repetition-rate operation any time soon. Targets and drivers are very complex resulting in high-dimensional experimental design spaces. Experiments are also highly *integrated*, meaning direct measurement of any figure of merit beyond the raw energy yield is not possible. These factors mean that ICF datasets are always sparse, with multiple confounding factors and uncertain information content; researchers therefore place a very high value on theoretical studies undertaken using multiphysics simulation codes. While cheaper than experiments, the simulations are still expensive, requiring at least CPU-months to complete. They can also have significant bias[96], and therefore require calibration against the available experimental data (from both ICF, and smaller scale experiments focused on relevant phenomena). There is a significant need for new methods that can help with experimental design and optimization, interpretation of experimental data, linking experiments with physics models, as well as making reliable predictions of future experiments. As this work progresses, ICF is becoming a prototypical example of the difficulties associated with science in the data-poor regime.

As with the other examples in this perspective, the funda-

mental data problem is the synthesis of multiple sources of information. Here, the key aim is to efficiently use the sparse information available to update our physics understanding in order to make simulations more predictive of future experiments. Figure 3 shows a potential workflow that fully integrates data-driven and machine-learning methods to achieve this goal. There are three fundamental information sources; experimental data which may comprise ~ 10 'shots' each producing multiple diagnostics with diverse data types; traditional high fidelity simulations which produce single best-physics predictions for each shot; and high-volume ensemble studies which use large numbers of (necessarily lower-fidelity) simulations to investigate competing physics hypothesis and provide protection against overfitting. The optimization and control of these information sources, combination of the resulting data, and updating of models to become more predictive present numerous opportunities for modern data science and machine learning approaches and each stage has seen recent active research.

Work to optimize the three information sources in Figure 3 include the acceleration of multiphysics simulations using deep learning[55,97], infrastructure for intelligent control of large-scale simulations[98], intelligent and data-informed design of experiments[42,43,99], as well as optimization of experimental facilities[100]. These developments have enabled simulation studies of unprecedented size[101] and the generation of open-source ICF datasets[102] that motivate novel deep learning research[54,57,103,104]. AI tools have been applied to the automatic analysis and featurization of complex data types like spectra, images[54,105], and line-of-sight dependent quantities. There has been significant interest in using Bayesian inference to improve diagnostics[106], and to synthesize observations in both focused HEDP experiments[51] and full-scale ICF experiments[50,107–109]. The ultimate aims of using these methods to improve physics understanding, and the reliability of simulations in extrapolating to new designs[48] or facilities, have been addressed though machine learning[42,110], Bayesian model calibration[50], and transfer learning[58,111].

Integrating the recent work we have described into a fully developed workflow similar to Figure 3 is still a significant challenge. Once achieved, however, we expect to gain an unprecedented view of the ICF design space, significantly better understanding of the conditions in current ICF experiments, and an improved understanding of the path to high yield and ignition.

Automation for High Repetition Rates

At the other end of the spectrum to NIF, which can only fire roughly once a day, are high-repetition rate lasers that can fire up to multiple times a second. This means huge amounts of data towards given science goals, but also means large aspects of the experimental process must be automated. To succeed, the automation of experiments requires both control of experimental parameters *and* real-time analysis of experimental results in one single algorithmic process; see Figure 4.

The first step in achieving this is for experimental goals to be coded in to the automation algorithm, such that choices of how to vary the experimental inputs are made automatically in order to maximise the desired output. This approach enables huge increases in efficiency in optimisation and model learning experiments, especially important when these experiments are resource limited. In addition, automation also allows for active feedback stabilisation of complex processes; this is of particular benefit to high energy density and plasma physics experiments, where many non-linear (and also often non-equilibrium) effects combine to determine the performance of what is a nominally unpredictable process. Thought must also be given to other factors that will determine optimal data return; diagnostics used, signal-to-noise ratio achieved, and experimental parameter space covered. Many diagnostics have already been adapted for fast electronic readout[112], however, challenges still remain in adapting x-ray or charged particle imagers and similar detectors for both maximum flexibility, as well as robustness to the harsh radiation and debris environments of HEDP experiments - especially at multi-Hz repetition rates.

This approach may be used to perform the following tasks:

- *Optimisation*: A function of experimental diagnostics is used to calculate the 'fitness' which expresses how closely measurements reflect the desired performance. An iterative procedure is then performed to optimise this fitness value by controlling experiment input parameters. This can be done with any optimisation procedure e.g. an evolutionary approach (i.e. genetic algorithms[41,113,114]), a numerical minimisation method (e.g. Nelder-Mead[114]) or by Bayesian optimisation using a machine learned surrogate model[115116]. This approach can also be used to limit the parameter search to satisfy some safety constraint, e.g. beam loss in a particle accelerator[115]. In general applying these approaches allows for rapid optimisation of experiments in a far more efficient manner than human controlled experiments - and produces much better results.

- *Stabilisation*: Active feedback can improve the stability of experimental performance by rapidly controlling input parameters to counteract oscillation or drifts in the apparatus[117]. This is routinely performed to stabilise component systems, such as alignment of laser beam transport, but can also be applied to highly complex and non-linear experimental phenomena, such as density limit disruptions in tokamaks using the predictions of neural network[118–120]. A stable output source then allows for much better experimental or source application outcomes.

- *Model Inference*: A Bayesian approach to statistical inference and model validation requires incorporating experiment uncertainties from diagnostic data in a rigorous manner, accounting for correlations across all parameter spaces[50]. This can lead not only to better estimates of the uncertainty, but the results of the inference can also dramatically change. Including this approach in real time in the data taking process can be

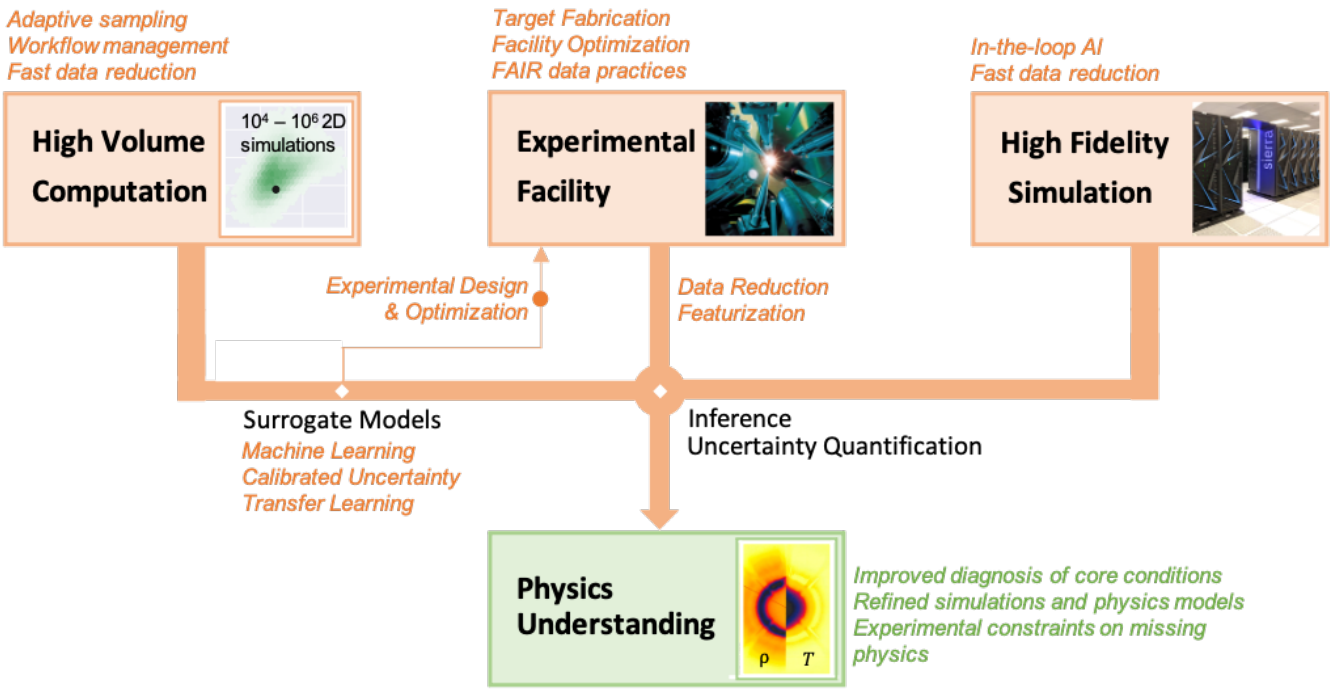

FIG. 3. **Integrating Information Sources in ICF Studies:** Our understanding of ICF implosion physics is based on a combination of high-volume, lower-fidelity simulation ensembles; sparse, difficult-to-diagnose experiments; and best-physics simulations that push the limits of high-performance computing technology. Creating and synthesizing these data into an improved understanding of the physics will require multiple new, complementary techniques from data science, uncertainty quantification (UQ) and artificial intelligence.

used to ensure that the experiment optimally constrains the physical models under examination - getting the most *information* per shot (as opposed to simply the most neutrons/x-rays etc.).

Encoding the scientific goals into the algorithm making the shot choice is just one of many challenges in automating high repetition rate facilities. Not only the physics of interest, but also the laser system and the target setup typically have inherent non-linearities that can make automating knowledge extraction extremely challenging. Small changes in system parameters (laser pulse width, shape, energy, focal spot conditions, target thickness etc.) can lead to large changes in experimental outcomes - requiring very fine control of the entire system. Thus the laser itself and the target must be modelled alongside the physics of interest. This complex multi-modal data must also be analysed as fast as the shot-rate to prevent another bottleneck in the experimental loop. In addition performance of diagnostics themselves might be impaired over time (for example if exposed to large radiation fluxes), requiring further modelling. The goal is for the AI to understand the effect of these diagnosed (and potentially undiagnosed) fluctuations in the system, rather than be confused by it. Human intuition risks misinterpreting evidence when many parameters are changing simultaneously. Finally, data archiving from experiments will rapidly become a challenge. Significant challenges in developing pipelines that can prevent data bottlenecks will become important, i.e. the operating algorithms may have to decide whether to record or destroy data based on quality of inputs and outputs to avoid large amounts of spurious and insignificant data occupying many TB of storage systems. In summary there are two separate challenges that should not be conflated. There is both a) the technical challenge of delivering online feedback and real time data curation and b) the modelling problem of automating knowledge extraction from the complex HEDP data. Researchers should take care to identify what aspects of their specific scientific problem fit into these two categories, and seek appropriate solutions.

The computerised control of experimental parameters increases the convenience for the experimental operators, allows for automatic parameter scans and reduces the likelihood of experimental errors. Enabling experiment automation requires considerable additional investment and effort in the preparation of experimental apparatus and facilities. However, time and resources spent on this endeavour yield large returns once the experiment is fully operational due to the increase in efficiency and productivity.

POLICY PROPOSITIONS

To help the field take advantage of the aforementioned new ways of using data, we make a series of suggestions for how educational and research practices might be beneficially changed.

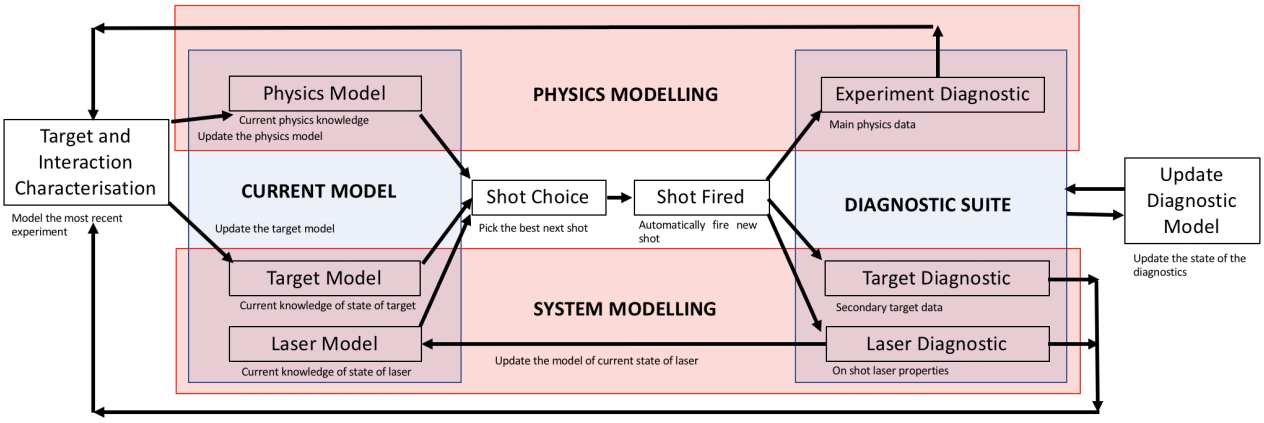

FIG. 4. **High repetition workflow:** the different components of a series of high repetition rate high powered laser. The AI system must have i) a model of its best estimate, with uncertainties, of the physics being probed, ii) a model of the current state of the laser (both so it can better achieve science goals, but also to avoid damaging itself), iii) a model of the target, iv) an algorithm to rapidly select what the next shot must be (depending on science goal), v) a system to actually fire the shot with no human intervention, vi) rapid automated data collection, vii) rapid physics, laser and target modelling, and viii) capacity to update a model of diagnostic performance (if needed).

Education

With the changes in HEDP data acquisition rates and analysis that have been described in this article, it is important to consider whether researchers who are new to the subject have sufficient training in the topics we have discussed.

There are courses in most universities that attempt to teach aspects of the subject; however they are not typically included in the HEDP curriculum. At the time of the writing, only 3 of the 40 Plasma Physics PhD Projects, Programs and Scholarships listed on `www.findaphd.com` contain a data science component.

While the subject is easily within reach for plasma physics graduate students, the unfamiliarity with the Bayesian perspective, and the necessary jargon, render it more difficult than necessary for a self-directed approach. We recommend that, *a minima*, a brief introduction[36–38] be included in one of the Advanced Courses, with the opportunity to take an elective course on Bayesian analysis and Uncertainty Quantification. Ideally, a Bayesian analysis module should be included in future HEDP doctoral programmes, built from the cases presented in this Perspective. In addition to including data science in the general curriculum, existing certificate programs focused on both fundamental ML techniques and multi-disciplinary applications of data science could be leveraged by students and professionals alike. Conferences and workshops like the International Conference on Data Driven Plasma Science (which had its second meeting in 2019) and the 2018 APS Mini-Conference on Machine Learning, Data Science, and Artificial Intelligence in Plasma Research[3] have flourished in the last few years, and these opportunities should be extended to those studying/training.

For all other graduate students and researchers, a complementary approach should be taken, through targeted workshops and schools. We believe that because the field is developing very quickly, it is preferable to consider teaching these topics as part of the existing HEDP workshops that have been established over the last few years for those new to the area. Topics should include Bayesian and frequentist statistics comparisons, UQ, Markov chain Monte Carlo (MCMC), surrogate building, deep neural nets, and optimization techniques for a range of dimensional spaces. Such workshops and meetings may also provide opportunities to engage the general ML community. The HEDP community could also entice ML researchers to collaborate in our field by presenting or organizing focused sessions at ML conferences, releasing datasets, exploring a ML challenge call focused on an application in our field, and pursuing direct research collaborations. Finally we would note that data science skill sets have broad application both in other research areas, as well as in industry. The proposed training is highly transferable and promises to be valuable to students regardless of specific career goals.

Research Practices

The changing nature of the field means practices within the field will also have to change. Researchers will need to become familiar with methods needed to run large numbers of simulations, and tools for storing larger amounts of data[46]. In particular, the field will have to develop data standards so that data is easily compatible between different facilities. Adopting open data practices[102] (like F.A.I.R., findability, accessibility, interoperability, and reusability[122]) wherever possible is also likely to greatly aid collaboration and comparison of data sets, although this will not be possible for all researchers in this area. The importance of these approaches is already well understood by researchers in other fields, for example high en-

ergy physics[123] and astronomy[124]; we foresee a similar level of data curation will soon be necessary in HEDP.

Many of the methods described in this paper require significant computational resources and good synthetic diagnostics, i.e., good simulations of what the data should look like through the actual pipeline the real data goes through. While most research efforts include a computational component, these are often disjoint from the analysis of experimental data. In future it is advisable for experimental time to also have associated funding for computation and analysis, and the development of synthetic diagnostics/data analysis. Diagnostics and experiments should be designed so that collected data can easily be used in conjunction with other shots (e.g. consistent pixel sizes). It may even be necessary that the commissioning of instruments comes with a corresponding budget to develop tools to simulate data as seen by the device. Data analysis is a core component of project commissioning and planning in many other areas of physics (e.g. the Euclid telescope[125]).

High data rates are permitting the probing of low signal-to-noise phenomena e.g. possible beyond Standard Model physics like axions. If laser based accelerators are to play a role in the future of probing new physics in this way, then statistical analysis must be brought up to the same standards as are practiced in High Energy Physics (HEP). In particular we may wish to adopt stringent statistical significance requirements e.g. 5σ for any discovery of beyond Standard Model physics[126]. Similarly using blinding methods are likely to become more necessary, where researchers deliberately hide some aspect of the labelling of the data from themselves to prevent subconscious bias or p-hacking[127]. Using a blinding protocol does also present dangers and challenges; the complexity and bespoke properties common in many laser-plasma experiments make blinding difficult to implement. However, high-rep rates will make blinding strategies much more viable, and may help for analyses where particular outcomes have high psychological significance and/or there are lots of different potentially viable approaches to doing the analysis.

BOX 1
Key conclusions:

1. The application of machine learning and modern data science methods to extreme plasma physics and HEDP is rapidly growing and is aiding in producing realistic uncertainties on predictions
2. Higher repetition rate facilities open up a range of new ways of working; data driven discovery, blinding methods, greater reproducibility, automated data taking
3. Integrating machine learning based approaches into working practices can greatly save money, time, and human effort
4. AI based tools are now often more successful at optimising non-linear extreme physics systems and comprehending multi-modal data than humans

Key recommendations:

1. Researchers should think carefully about how to best use their data: what methods and diagnostics can they use to take the best data, get sensible uncertainties, and coherently combine with other data sets
2. Awards of experimental time and instrument construction should also include greater support for uncertainty quantification, building synthetic diagnostics and data analysis
3. Plasma physics graduate education and national lab training programmes begin to include basic data science courses
4. Researchers should try where possible to practice open science best practice; making code and data available publicly, using shared data standards between different facilities

CONCLUSION

The volume of data on plasma physics at the extremes is rapidly growing and offers the potential for dramatically increased speed of scientific advance. These data are typically multi-modal and describe very non-linear systems, making interpretation challenging for humans, but tractable for AI algorithms. Plasma physics is unlikely to reach the extremely high data rates of HEP in the immediate future, and it is fully the case that AI will not solve all problems in the field. Nonetheless, this new approach in the field offers novel ways of working, and new ways to gain insight - we hope practitioners in the field will be able to find applicability for these methods in their research, see Box 1.

In conclusion, modern data science has a lot to offer extreme plasma physics and HEDP science; the community must act now to identify which areas it will make the biggest impact in and make resources and training available to make the most of these novel approaches.

ACKNOWLEDGMENTS

This Perspective was the result of a meeting at the Lorentz Center, University of Leiden, 13th-17th January 2020. The Lorentz Centre is funded by the Dutch Research Council (NWO) and the University of Leiden. The meeting also had support from the John Fell Oxford University Press (OUP) Research Fund. The organisers are grateful to Tanja Uitbei-

jerse (Lorentz Center) for facilitating a fruitful and productive meeting.

P.W.H. acknowledges funding from the Engineering and Physical Sciences Research Council. A portion of this work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. J.A.G and G.J.A were supported by LLNL Laboratory Directed Research and Development project 18-SI-002. The paper has LLNL tracking number LLNL-JRNL-811857. This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

AUTHOR CONTRIBUTIONS

P.W.H., J.A.G and G.J.A conceived the work and led the writing of the manuscript. All authors contributed to the manuscript and the ideas discussed at the Lorentz Center Meeting.

[1] Rose, S. Set the controls for the heart of the Sun. *Contemporary Physics* **45**, 109–121 (2004).

[2] Spears, B. K. *et al.* Deep learning: A guide for practitioners in the physical sciences. *Physics of Plasmas* **25**, 080901 (2018).

This tutorial paper gives an introduction to scientific machine learning, with examples taken from ICF research

[3] Wang, Z., Peterson, J. L., Rea, C. & Humphreys, D. Special Issue on Machine Learning, Data Science, and Artificial Intelligence in Plasma Research. *IEEE Transactions on Plasma Science* **48**, 1–2 (2020).

[4] Colvin, J. & Larsen, J. *Extreme Physics* (Cambridge University Press, Cambridge, 2013).

[5] Graziani, F., Desjarlais, M. P., Redmer, R. & Trickey, S. B. (eds.) *Frontiers and Challenges in Warm Dense Matter*, vol. 96 of *Lecture Notes in Computational Science and Engineering* (Springer International Publishing, Cham, 2014).

[6] Gould, O., Mangles, S., Rajantie, A., Rose, S. & Xie, C. Observing thermal Schwinger pair production. *Physical Review A* **99**, 052120 (2019).

[7] Millot, M. *et al.* Experimental evidence for superionic water ice using shock compression. *Nature Physics* **14**, 297–302 (2018).

[8] Celliers, P. M. *et al.* Insulator-metal transition in dense fluid deuterium. *Science* **361**, 677–682 (2018).

[9] Joshi, C. & Malka, V. Focus on Laser- and Beam-Driven Plasma Accelerators. *New Journal of Physics* **12**, 045003 (2010).

[10] Hidding, B., Foster, B., Hogan, M. J., Muggli, P. & Rosenzweig, J. B. Directions in plasma wakefield acceleration. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **377**, 20190215 (2019).

[11] Wang, W.-M. *et al.* Collimated ultrabright gamma rays from electron wiggling along a petawatt laser-irradiated wire in the QED regime. *Proceedings of the National Academy of Sciences* **115**, 9911–9916 (2018).

[12] Badziak, J. Laser-driven ion acceleration: methods, challenges and prospects. *Journal of Physics: Conference Series* **959**, 012001 (2018).

[13] Feng, J. *et al.* High-efficiency neutron source generation from photonuclear reactions driven by laser plasma accelerator. *High Energy Density Physics* **36**, 100753 (2020).

[14] Hidding, B. *et al.* Plasma wakefield accelerator research 2019 - 2040: A community-driven uk roadmap compiled by the plasma wakefield accelerator steering committee (pwasc). WorkingPaper (2019).

[15] Nuckolls, J., Wood, L., Thiessen, A. & Zimmerman, G. Laser compression of matter to super-high densities: Thermonuclear applications. *Nature* **239**, 139–142 (1972).

[16] Hurricane, O. A. & Program, I. Overview of Progress and Future Prospects in Indirect Drive Implosions on the National Ignition Facility. *Journal of Physics: Conference Series* **717**, 012005 (2016).

[17] Fournier, K. & K. NIF Monthly Highlights for July 2019. Tech. Rep., Lawrence Livermore National Laboratory (LLNL), Livermore, CA (United States) (2019). URL `http://www.osti.gov/servlets/purl/1548376/`.

[18] Remington, B. A., Rudd, R. E. & Wark, J. S. From microjoules to megajoules and kilobars to gigabars: Probing matter at extreme states of deformation. *Physics of Plasmas* **22**, 090501 (2015).

[19] Trines, R. CLF Annual Report 2017-2018. Tech. Rep., Science and Technology Facilities Council (2019). URL `https://www.clf.stfc.ac.uk/Pages/Annual-Report-2017-18.aspx`.

[20] Sturm, C. & Stöcker, H. The facility for antiproton and ion research FAIR. *Physics of Particles and Nuclei Letters* **8**, 865–868 (2011).

[21] Roso, L. High repetition rate Petawatt lasers. *EPJ Web of Conferences* **167**, 01001 (2018).

[22] Zheng, W. *et al.* Laser performance of the SG-III laser facility. *High Power Laser Science and Engineering* **4**, e21 (2016).

[23] Danson, C. N. *et al.* Petawatt and exawatt class lasers worldwide. *High Power Laser Science and Engineering* **7**, e54 (2019).

This paper reviews some of the many high-powered lasers in use around the world.

[24] *Opportunities in Intense Ultrafast Lasers* (National Academies Press, Washington, D.C., 2018).

[25] Lin, Z. *et al.* SG-II laser elementary research and precision SG-II program. *Fusion Engineering and Design* **44**, 61–66 (1999).

[26] LULI. LULI2000 User Guide. Tech. Rep., Intenses, Laboratoire pour l'Utilisation des Lasers (2019).

[27] Kirillov, G. A., Murugov, V. M., Punin, V. T. & Shemyakin, V. I. High power laser system ISKRA V. *Laser and Particle Beams* **8**, 827–831 (1990).

[28] Zhao, Z. & Wang, D. XFEL Projects in China (2019). Presented at LINAC2018 in Beijing, China, unpublished.

[29] Zhang, Z. *et al.* The laser beamline in SULF facility. *High Power Laser Science and Engineering* **8**, e4 (2020).

[30] Schramm, U. *et al.* First results with the novel petawatt laser acceleration facility in Dresden. In *Journal of Physics: Conference Series*, vol. 874, 12028 (Institute of Physics Publishing, 2017).

[31]EXFEL. European XFEL Facts and Figures. URL `https://www.xfel.eu/facility/overview/facts_amp_figures/index_eng.html`.
[32]LCLS-II. LCLS-II Design & Performance. URL `https://lcls.slac.stanford.edu/lcls-ii/design-and-performance`.
[33]Yabashi, M., Tanaka, H., Tono, K. & Ishikawa, T. Status of the SACLA Facility. *Applied Sciences* **7**, 604 (2017).
[34]CERN FAQ – LHC the guide. Tech. Rep., CERN-Brochure-2017-002-Eng, Education, Communications and Outreach Group, CERN (2017).
[35]MacDonald, M. J. *et al.* Measurement of high-dynamic range x-ray Thomson scattering spectra for the characterization of nano-plasmas at LCLS. *Review of Scientific Instruments* **87**, 11E709 (2016).
[36]Mitchell, T. *Machine Learning* (McGraw-Hill, New York, 1997).
[37]Sivia, D. & Skilling, J. *Data analysis: a Bayesian tutorial* (Oxford University Press).
[38]Brunton, S. L. & Kutz, J. N. *Data-Driven Science and Engineering* (Cambridge University Press, 2019).
[39]Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
[40]Fleming, S. W. & Gupta, H. V. The physics of river prediction. *Physics Today* **73**, 46–52 (2020).
[41]Streeter, M. J. V. *et al.* Temporal feedback control of high-intensity laser pulses to optimize ultrafast heating of atomic clusters. *Applied Physics Letters* **112**, 244101 (2018).
[42]Gopalaswamy, V. *et al.* Tripled yield in direct-drive laser fusion through statistical modelling. *Nature* **565**, 581–586 (2019).

This paper is the first to use data-driven approaches to motivate and actually carry out new ICF experiments.

[43]Hatfield, P. W., Rose, S. J. & Scott, R. H. H. The blind implosion-maker: Automated inertial confinement fusion experiment design. *Physics of Plasmas* **26**, 062706 (2019).
[44]Martin, M., London, R., Goluoglu, S. & Whitley, H. An automated design process for short pulse laser driven opacity experiments. *High Energy Density Physics* **26**, 26–37 (2018).
[45]Raghu, M. & Schmidt, E. A Survey of Deep Learning for Scientific Discovery (2020). 2003.11755.
[46]Thayer, J. *et al.* Data systems for the Linac coherent light source. *Advanced Structural and Chemical Imaging* **3**, 3 (2017).
[47]Bernal, J. L., Peacock, J. A., Bernal, J. L. & Peacock, J. A. Conservative cosmology: combining data with allowance for unknown systematics. *JCAP* **2018**, 002 (2018).

This paper considers how to get realistic uncertainty estimates in the presence of unknown systematics.

[48]Osthus, D., Vander Wiel, S. A., Hoffman, N. M. & Wysocki, F. J. Prediction Uncertainties beyond the Range of Experience: A Case Study in Inertial Confinement Fusion Implosion Experiments. *SIAM/ASA Journal on Uncertainty Quantification* **7**, 604–633 (2019).
[49]Lang, M. & Owens, M. J. A Variational Approach to Data Assimilation in the Solar Wind. *Space Weather* **17**, 59–83 (2019).
[50]Gaffney, J. A. *et al.* Making inertial confinement fusion models more predictive. *Physics of Plasmas* **26**, 082704 (2019).
[51]Kasim, M. F., Galligan, T. P., Topp-Mugglestone, J., Gregori, G. & Vinko, S. M. Inverse problem instabilities in large-scale modeling of matter in extreme conditions. *Physics of Plasmas* **26**, 112706 (2019).
[52]van de Plassche, K. L. *et al.* Fast modeling of turbulent transport in fusion plasmas using neural networks. *Physics of Plasmas* **27**, 022310 (2020).
[53]Meneghini, O. *et al.* Self-consistent core-pedestal transport simulations with neural network accelerated models. *Nuclear Fusion* **57**, 086034 (2017).
[54]Anirudh, R., Thiagarajan, J. J., Bremer, P.-T. & Spears, B. K. Improved surrogates in inertial confinement fusion with manifold and cycle consistencies. *Proceedings of the National Academy of Sciences* **117**, 9741–9746 (2020).
[55]Kluth, G. *et al.* Deep learning for NLTE spectral opacities. *Physics of Plasmas* **27**, 052707 (2020).

This paper uses machine learning to emulate opacity calculations, in order to speed up ICF simulations.

[56]Humbird, K. D., Peterson, J. L. & McClarren, R. G. Deep neural network initialization with decision trees. *IEEE transactions on neural networks and learning systems* **30**, 1286–1295 (2018).

This paper presents a data-driven method for selecting deep neural network architecture and hyperparameters

[57]Kasim, M. F. *et al.* Building high accuracy emulators for scientific simulations with deep neural architecture search (2020). arXiv:2001.08055.
[58]Kustowski, B. *et al.* Transfer Learning as a Tool for Reducing Simulation Bias: Application to Inertial Confinement Fusion. *IEEE Transactions on Plasma Science* **48**, 46–53 (2020).

Transfer learning updates a subset of a pre-trained deep neural network using experimental data. This paper explores its application to sparse ICF datasets

[59]Bambi, C. Astrophysical Black Holes: A Compact Pedagogical Review (2018). 1711.10256.
[60]The National Space Weather Program Strategic Plan. Tech. Rep., CM-P30-1995, Washington D.C. (1995).
[61]Eastwood, J. P. *et al.* The Economic Impact of Space Weather: Where Do We Stand? *Risk Analysis* **37**, 206–218 (2017).
[62]Camporeale, E. The Challenge of Machine Learning in Space Weather: Nowcasting and Forecasting. *Space Weather* **17**, 1166–1207 (2019).
[63]Camporeale, E., Carè, A. & Borovsky, J. E. Classification of Solar Wind With Machine Learning. *Journal of Geophysical Research: Space Physics* **122**, 10,910–10,920 (2017).
[64]Chen, Y. *et al.* Identifying Solar Flare Precursors Using Time Series of SDO/HMI Images and SHARP Parameters. *Space Weather* **17**, 1404–1426 (2019).
[65]Campi, C. *et al.* Feature Ranking of Active Region Source Properties in Solar Flare Forecasting and the Uncompromised Stochasticity of Flare Occurrence. *The Astrophysical Journal* **883**, 150 (2019).
[66]Inceoglu, F. *et al.* Using Machine Learning Methods to Forecast if Solar Flares Will Be Associated with CMEs and SEPs. *The Astrophysical Journal* **861**, 128 (2018).
[67]Bobra, M. G. & Ilonidis, S. Predicting coronal mass ejections using machine learning methods. *The Astrophysical Journal* **821**, 127 (2016).
[68]Sarma, R. *et al.* Bayesian Inference of Quasi-Linear Radial Diffusion Parameters using Van Allen Probes. *Journal of Geophysical Research: Space Physics* **125** (2020).
[69]Camporeale, E. *et al.* A Gray-Box Model for a Probabilistic Estimate of Regional Ground Magnetic Perturbations: Enhancing the NOAA Operational Geospace Model With Machine Learning. *Journal of Geophysical Research: Space Physics* **125**, e27684 (2020). 1912.01038.
[70]Lamb, K. *et al.* Correlation of auroral dynamics and gnss scintillation with an autoencoder. In *Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada* (2019).
[71]Rowlinson, A. *et al.* Identifying transient and variable sources in radio images. *Astronomy and Computing* **27**, 111–129 (2019).
[72]Coughlin, M. W., Dietrich, T., Margalit, B. & Metzger, B. D. Multi-messenger Bayesian parameter inference of a binary neutron-star merger. *Monthly Notices of the Royal Astronomical Society: Letters* **489**, L91–L96 (2019).

This paper uses both gravitational wave, and multi-wavelength electromagnetic data to simultaneously constrain stellar masses, orbital parameters, and supra-nuclear density equations-of-state.

[73]Dorn, C. *et al.* A generalized Bayesian inference method for constraining the interiors of super Earths and sub-Neptunes. *Astronomy & Astrophysics* **597**, A37 (2017).
[74]Bellinger, E. P. *et al.* Fundamental parameters of main-sequence stars in an instant with machine learning. *The Astrophysical Journal* **830**, 31 (2016).
[75]Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O. & Winter, M. K. Photometric supernova classification with machine learning. *The Astrophysical Journal Supplement Series* **225**, 31 (2016).

[76]Kong, X. *et al.* Spectral feature extraction for DB white dwarfs through machine learning applied to new discoveries in the SDSS DR12 and DR14. *PASP* **130**, 084203 (2018).

[77]Huppenkothen, D., Heil, L. M., Hogg, D. W. & Mueller, A. Using machine learning to explore the long-term evolution of GRS 1915+105. *Monthly Notices of the Royal Astronomical Society* **466**, 2364–2377 (2017).

[78]Chardin, J. *et al.* A deep learning model to emulate simulations of cosmic reionization. *MNRAS* **490**, 1055–1065 (2019).

[79]Saumon, D. & Guillot, T. Shock Compression of Deuterium and the Interiors of Jupiter and Saturn. *The Astrophysical Journal* **609**, 1170–1180 (2004).

[80]Smith, R. F. *et al.* Ramp compression of diamond to five terapascals. *Nature* **511**, 330–333 (2014).

[81]Tzeferacos, P. *et al.* Laboratory evidence of dynamo amplification of magnetic fields in a turbulent plasma. *Nature Communications* **9**, 591 (2018).

[82]Falcon, R. E. *et al.* Laboratory measurements of white dwarf photospheric spectral lines: Hβ. *The Astrophysical Journal* **806**, 214 (2015).

[83]Bailey, J. E. *et al.* A higher-than-predicted measurement of iron opacity at solar interior temperatures. *Nature* **517**, 56–59 (2015).

[84]Nagayama, T. *et al.* Systematic Study of L -Shell Opacity at Stellar Interior Temperatures. *Physical Review Letters* **122**, 235001 (2019).

[85]de Souza, R. S., Boston, S. R., Coc, A. & Iliadis, C. Thermonuclear fusion rates for tritium + deuterium using Bayesian methods. *Physical Review C* **99**, 014619 (2019).

[86]Zhelavskaya, I. S., Shprits, Y. Y. & Spasojević, M. Empirical Modeling of the Plasmasphere Dynamics Using Neural Networks. *Journal of Geophysical Research: Space Physics* **122**, 11,227–11,244 (2017).

[87]Zhelavskaya, I. S., Shprits, Y. Y. & Spasojevic, M. Reconstruction of Plasma Electron Density From Satellite Measurements Via Artificial Neural Networks. *Machine Learning Techniques for Space Weather* 301–327 (2018).

[88]Freidberg, J. *Plasma Physics and Fusion Energy* (Cambridge University Press, 2008).

[89]Atzeni, S., Meyer-ter Vehn, J. & Meyer-ter Vehn, J. *The Physics of Inertial Fusion: BeamPlasma Interaction, Hydrodynamics, Hot Dense Matter*. International Series of Monographs on Physics (Clarendon Press, 2004).

[90]Lindl, J. *Inertial Confinement Fusion: The Quest for Ignition and Energy Gain Using Indirect Drive*. Aip-Press Series (AIP Press, 1998).

[91]Campbell, E. M. & Hogan, W. J. The National Ignition Facility - applications for inertial fusion energy and high-energy-density science. *Plasma Physics and Controlled Fusion* **41**, B39–B56 (1999).

[92]Moses, E. I. Ignition on the national ignition facility. *Journal of Physics: Conference Series* **112**, 012003 (2008).

[93]Boehly, T. *et al.* Initial performance results of the omega laser system. *Optics Communications* **133**, 495 – 506 (1997).

[94]Slutz, S. A. High-gain magnetized inertial fusion. *Physical Review Letters* **108** (2012).

[95]Gomez, M. R. *et al.* Experimental demonstration of fusion-relevant conditions in magnetized liner inertial fusion. *Phys. Rev. Lett.* **113**, 155003 (2014).

[96]Dimonte, G. Quantitative metrics for evaluating thermonuclear design codes and physics models applied to the National Ignition Campaign. *Physics of Plasmas* **27**, 052709 (2020).

[97]Yang, C. *et al.* Preparing dense net for automated hydra mesh management via reinforcement learning. Tech. Rep. LLNL-TR-799958, Lawrence Livermore National Laboratory (2019). URL `https://www.osti.gov/servlets/purl/1580017`.

This report details the use of deep neural networks to automatically perform adaptive mesh refinement in radiation-hydrodynamics simulations

[98]Peterson, J. L. *et al.* Merlin: Enabling Machine Learning-Ready HPC Ensembles (2019). arXiv:1912.02892.

[99]Peterson, J. L. *et al.* Zonal flow generation in inertial confinement fusion implosions. *Physics of Plasmas* **24**, 032702 (2017).

This paper presents the first novel ICF design reached using machine learning.

[100]Amorin, C., Kegelmeyer, L. M. & Kegelmeyer, W. P. A hybrid deep learning architecture for classification of microscopic damage on National Ignition Facility laser optics. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **12**, 505–513 (2019).

Here, deep neural network-based image classification is used to detect and classify damage in the NIF laser system, allowing more reliable operation at high energy

[101]Nora, R., Peterson, J. L., Spears, B. K., Field, J. E. & Brandon, S. Ensemble simulations of inertial confinement fusion implosions. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **10**, 230–237 (2017).

[102]Gaffney, J. A. *et al.* The JAG inertial confinement fusion simulation dataset for multi-modal scientific deep learning (2020). URL `https://library.ucsd.edu/dc/object/bb5534097t`.

This dataset is an example of open data practices in ICF, and is one of the first multi-modal scientific datasets to be released by the HEDP community

[103]Thiagarajan, J. J. *et al.* Designing accurate emulators for scientific processes using calibration-driven deep models. *Nature Communications* **11**, 1–10 (2020). 2005.02328.

[104]Hatfield, P. *et al.* Using sparse gaussian processes for predicting robust inertial confinement fusion implosion yields. *IEEE Transactions on Plasma Science* **48**, 14–21 (2020).

[105]Glinsky, M. E. *et al.* Quantification of MagLIF Morphology using the Mallat Scattering Transformation. *Office of Scientific and Technical Information (OSTI)* (2019). URL `http://dx.doi.org/10.2172/1569785`.

This paper uses a novel, physics-motivated deep neural network architecture to featurize images of ICF implosions driven on the Z pulsed-power machine

[106]Palaniyappan, S. *et al.* Hydro-scaling of direct-drive cylindrical implosions at the omega and the national ignition facility. *Physics of Plasmas* **27**, 042708 (2020).

[107]Gaffney, J., Clark, D., Sonnad, V. & Libby, S. Development of a Bayesian method for the analysis of inertial confinement fusion experiments on the NIF. *Nuclear Fusion* **53**, 073032 (2013).

[108]Gaffney, J., Clark, D., Sonnad, V. & Libby, S. Bayesian inference of inaccuracies in radiation transport physics from inertial confinement fusion experiments. *High Energy Density Physics* **9**, 457 – 461 (2013).

[109]Knapp, P. A Bayesian Parameter Estimation Framework for Understanding Fusion Experiments on Z. Tech. Rep. (2018). URL `https://www.osti.gov/biblio/1525597`.

[110]Hsu, A., Cheng, B. & Bradley, P. A. Analysis of NIF scaling using physics informed machine learning. *Physics of Plasmas* **27**, 012703 (2020).

[111]Humbird, K. D., Peterson, J. L., Spears, B. K. & McClarren, R. G. Transfer Learning to Model Inertial Confinement Fusion Experiments. *IEEE Transactions on Plasma Science* **48**, 61–70 (2020).

[112]Götzfried, J. *et al.* Research towards high-repetition rate laser-driven X-ray sources for imaging applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **909**, 286–289 (2018). 1803.05415.

[113]He, Z.-H. *et al.* Coherent control of plasma dynamics. *Nature communications* **6**, 7156 (2015).

[114]Dann, S. J. *et al.* Laser wakefield acceleration with active feedback at 5 Hz. *Physical Review Accelerators and Beams* **22**, 041303 (2019).

[115]Kirschner, J., Mutny, M. M., Hiller, N., Ischebeck, R. & Krause, A. Adaptive and Safe Bayesian Optimization in High Dimensions via One-Dimensional Subspaces. *Proceedings of the 36th International Conference on Machine Learning* 3429–3438 (2019).

This paper describes a Bayesian Optimisation method for a Free Electron Laser that prevents the laser from violating any safety constraints.

[116]Shalloo, R. J. *et al.* Automation and control of laser wakefield accelerators using Bayesian optimization. *Nature Communications* **11**, 1–8 (2020). 2007.14340.

[117]Maier, A. R. *et al.* Decoding Sources of Energy Variability in a Laser-Plasma Accelerator. *Physical Review X* **10**, 031039 (2020).

[118]Team, J. T. *et al.* Hybrid neural network for density limit disruption prediction and avoidance on J-TEXT tokamak. *Nuclear Fusion* (2018).

[119]Fu, Y. *et al.* Machine learning control for disruption and tearing mode avoidance. *Physics of Plasmas* **27**, 022501 (2020).

[120]Kates-Harbeck, J., Svyatkovskiy, A. & Tang, W. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568**, 526–531 (2019).

[121]Blanken, T., Felici, F., Rapson, C., de Baar, M. & Heemels, W. Control-oriented modeling of the plasma particle density in tokamaks and application to real-time density profile reconstruction. *Fusion Engineering and Design* **126**, 87–103 (2018).

[122]Wilkinson, M. D. *et al.* Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 1–9 (2016).

[123]Albertsson, K. *et al.* Machine Learning in High Energy Physics Community White Paper. *Journal of Physics: Conference Series* **1085**, 022008 (2018).

[124]Borne, K. D. *et al.* The Revolution in Astronomy Education: Data Science for the Masses (2009). arXiv:0909.3895.

[125]Pasian, F. *et al.* Science ground segment for the ESA Euclid Mission. In Radziwill, N. M. & Chiozzi, G. (eds.) *Software and Cyberinfrastructure for Astronomy II*, vol. 8451, 845104 (SPIE, 2012).

[126]Lyons, L. Discovering the Significance of 5 sigma (2013). arXiv:1310.1284.

This paper discusses the background, rationale, and advantages and disadvantages of the 5-σ criterion commonly used in particle physics.

[127]Roodman, A. Blind Analysis in Particle Physics. In *Proceedings of the Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology, Stanford Linear Accelerator Center, Stanford, California, September 8th – 11th, 2003* (2003).