nuclear fusion
IAEA

CONFERENCE REPORT • OPEN ACCESS

Summary of the 5th IAEA technical meeting on fusion data processing, validation and analysis (FDPVA)

To cite this article: M. Xu *et al* 2026 *Nucl. Fusion* **66** 017002

OPEN ACCESS
IOP Publishing | International Atomic Energy Agency
Nuclear Fusion
Nucl. Fusion **66** (2026) 017002 (24pp)
https://doi.org/10.1088/1741-4326/ae048d

Conference Report

Summary of the 5th IAEA technical meeting on fusion data processing, validation and analysis (FDPVA)

M. Xu[1], D. Mazon[2], M. Barbarino[3,*], W. Biel[4], R.M. Churchill[5], R. Fischer[6], K. Fujii[7], P. Jain[3], A. Murari[8,9], S.D. Pinches[10], P. Rodriguez-Fernandez[11], J. Stillerman[11], J. Vega[12], G. Verdoolaege[13], M. Yokoyama[14], P. Abreu[10], S. Ahmed[15], J. Alhage[13], F. Almuhisen[2], M. Bergmann[6], D. Pereira Botelho[16], L. Caputo[13], S. Carli[17], R. Castro[12], T. Craciunescu[18], F. Deeba[19], F. Esquembre[20], K. Giil[21], Y. Gu[22], J. Hall[13], J. Hollocombe[23], X. Huang[24,25], A. Jardin[26], R. Jorge[27], Y. Li[22], Y. Liu[1], S. McIntosh[10], E. Peluso[28], R. Rossi[28], M. Ruiz[29], J. De Rycke[13], M. Schneider[10], M. Sertoli[30], A. Puig Sitjes[6], D. Stieglitz[6], Y. Tan[31], H. Weisen[32], H. Wu[13], I. Wyss[28] and L. Zang[1]

[1] Institute of Modern Physics, Fudan University, Shanghai, China
[2] CEA, IRFM, F-13108, Saint-Paul-lez-Durance, France
[3] Physics Section, Division of Physical and Chemical Sciences, Department of Nuclear Sciences and Applications, International Atomic Energy Agency, Vienna International Centre, PO Box 100, 1400 Vienna, Austria
[4] Forschungszentrum Jülich GmbH, Jülich, Germany
[5] Princeton Plasma Physics Laboratory, Princeton, NJ, United States of America
[6] Max Planck Institut fuer Plasmaphysik, Garching, Germany
[7] Oak Ridge National Laboratory, Oak Ridge, TN, United States of America
[8] Consorzio RFX (CNR, ENEA, INFN, Università di Padova, Acciaierie Venete SpA), Padova, Italy
[9] Istituto per la Scienza e la Tecnologia dei Plasmi, CNR, Padova, Italy
[10] ITER Organization, Saint-Paul-lez-Durance, France
[11] MIT Plasma Science and Fusion Center, Boston, MA, United States of America
[12] CIEMAT, Madrid, Spain
[13] Department of Applied Physics, Ghent University, Gent, Belgium
[14] National Institute for Fusion Science, Gifu, Japan
[15] UiT The Arctic University of Norway, Tromsø, Norway
[16] Renaissance Fusion, Fontaine, France
[17] Department of Mechanical Engineering, KU Leuven, Leuven, Belgium
[18] National Institute for Laser, Plasma and Radiation Physics, Magurele-Bucharest, Romania
[19] Pakistan Tokamak Plasma Research Institute, Islamabad, Pakistan
[20] Department of Mathematics, Universidad de Murcia, Murcia, Spain
[21] Korea Institute of Fusion Energy, Daejeon, Korea, Republic Of
[22] University of Science and Technology of China, Hefei, Anhui, China
[23] UKAEA, Culham, United Kingdom of Great Britain and Northern Ireland

[*] Author to whom any correspondence should be addressed.

[24] Hebei Key Laboratory of Compact Fusion, Langfang, China
[25] ENN Science and Technology Development Co., Ltd, Langfang, China
[26] Institute of Nuclear Physics Polish Academy of Sciences (IFJ PAN), Radzikowskiego 152, 31-342 Krakow, Poland
[27] Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[28] Department of Industrial Engineering, University of Rome Tor Vergata, Rome, Italy
[29] Universidad Politécnica de Madrid, Madrid, Spain
[30] Tokamak Energy Ltd, 173 Brook Drive, Milton Park, Oxfordshire OX14 4SD, United Kingdom of Great Britain and Northern Ireland
[31] Department of Engineering Physics, Tsinghua University, Beijing, China
[32] V.N.Karazin Kharkiv National University, Faculty of Physics and Technology, Kharkiv, Ukraine

E-mail: M.Barbarino@iaea.org

Received 16 April 2025, revised 28 July 2025
Accepted for publication 8 September 2025
Published 27 November 2025

Abstract

The purpose of the 5th International Atomic Energy Agency technical meeting on fusion data processing, validation and analysis (FDPVA) (Ghent University, Ghent, Belgium, 12–15 June 2023) was to provide a platform during which a set of topics relevant to FDPVA were discussed with the view of meeting the needs of next step fusion devices such as ITER. The validation and analysis of experimental data obtained from diagnostics used to characterize fusion plasmas are crucial for a knowledge-based understanding of the physical processes governing the dynamics of these plasmas. This paper presents the recent progress and achievements in the domain of plasma diagnostics data analysis and synthetic diagnostics reported at the meeting, including concept description of new devices; fusion databases; integrated data analysis; inverse problems; uncertainty propagation, verification and validation; probabilistic methods and machine learning. The relevant results underline trends observed in the current major fusion confinement devices.

Keywords: plasma diagnostics, fusion databases, integrated data analysis, machine learning

(Some figures may appear in colour only in the online journal)

1. Introduction

The International Atomic Energy Agency (IAEA) fosters global collaboration by facilitating the exchange of scientific and technical expertise. Its efforts focus on bridging gaps in fusion physics, engineering, and safety, supporting the progress of ITER, and coordinating efforts towards the realization of future fusion plants.

The 5th IAEA technical meeting on fusion data processing, validation and analysis (FDPVA) was held in Ghent University, Ghent, Belgium from 12–15 June 2023. Previous meetings in this series were held in Vienna, Austria [1] (Virtual event, 2021), Vienna, Austria [2] (2019), Boston, United States of America (2017), and Nice, France (2015).

44 experts from 16 Member states (Belgium, China, France, Germany, Italy, Japan, Norway, Pakistan, Poland, Portugal, Romania, South Korea, Spain, Ukraine, United Kingdom and United States of America) and 1 International Organization (ITER Organization) presented the most prominent results in the area of instruments, methods and mathematical solutions for research in the field of nuclear fusion and plasma physics.

The meeting particularly aimed at fostering discussions about research and development results that set out or underline trends observed in the current major fusion confinement devices. Progress was shown on the following topics during the meeting: (i) next step/new fusion device concepts: data challenges and design optimization (NSC); (ii) information retrieval, dimensionality reduction and visualization in fusion databases (DBs); (iii) integrated data analysis (IDA) and synthetic diagnostics; (iv) inverse problems (INV); (v) analysis of time series, images and video: detection, identification and prediction (TIV); (vi) uncertainty propagation, verification and validation in modeling codes and data fusion (UNC); and (vii) advances in data science, probabilistic methods and machine learning (ADV).

2. Topical summary

The scientific programme and paper selection were the responsibility of the Programme Committee composed of the following members: Min Xu (Chair, China), Didier Mazon (Co-chair, France), Geert Verdoolaege (Host, Belgium),

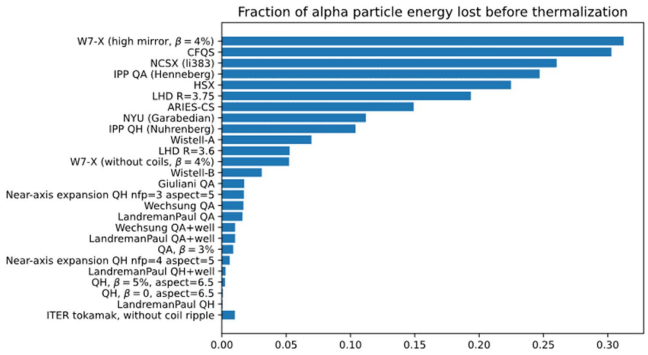

Figure 1. Fraction of alpha particle energy lost before thermalization for different cases.

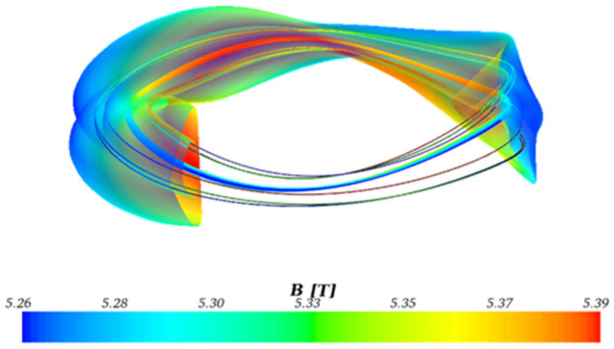

Figure 2. Example of the use of the NEAT tracer code for different magnetic field.

Wolfgang Biel (Germany), Rainer Fischer (Germany), Andrea Murari (Italy), Simon Pinches (ITER Organization), Masayuki Yokoyama (Japan), Jesus Vega (Spain), Michael Churchill (United States of America), Keisuke Fujii (United States of America), Pablo Rodriguez-Fernandez (United States of America) and Josh Stillerman (United States of America).

A brief summary of the seven topics of interest for the FDPVA was presented, focusing on the main highlights and progress, including outcome of the general discussions.

2.1. Next step/new fusion device concepts: data challenges and design optimization (NSC)

This session, chaired by D. Mazon and co-chaired by P. Rodriguez-Fernandez, was devoted to the concept description of new devices focusing on the challenges and design optimization.

The latest results concerning the design of stellarator devices in the field of stellarator optimization were presented [3]. One of the main drivers of the latest developments, the code SIMSOPT, was introduced, which allows designs with increasingly better particle confinement (see figure 1). The development of the near-axis expansion theory is also reported, which is now being explored as a potential method to create initial conditions for more complex optimization scenarios.

The presentation [3] included direct constructions using the near-axis expansion, direct fast particle optimization, direct turbulence optimization, and direct coil optimization. In the near-axis optimization front, it was shown how a recently created DB allowed the training of a neural network to solve the inverse problem: obtain a stellarator configuration given desired geometric quantities such as elongation and rotational transform. In the fast particle front, a new particle tracer code NEAT was developed (see figure 2) and benchmarked using different magnetic field geometries and was publicly available on GitHub [4]. The direct turbulence optimization is based on either a quasi-linear or a non-linear approach that runs a gyrokinetic simulation at each iteration of the optimization routine [5]. Finally, the direct coil optimization showed how a single-stage optimization allowed for higher-performing configurations with better fast particle confining properties and simpler coils [6].

Surrogate-based optimization techniques were presented to accelerate performance predictions [7]. Predicting kinetic profiles and fusion performance of the core of tokamak plasmas has remained a formidable task, as numerical simulations of realistic plasma conditions can be very computationally expensive, particularly for the case of modeling core turbulence dynamics. A newly developed technique, PORTALS [8], leverages advances in surrogate-based and Bayesian optimization to accelerate the convergence of transport solvers. Simulations of macroscopic transport dynamics together with nonlinear, delta-f turbulence can now be done routinely with high physics fidelity, and steady-state solutions are found at a moderate cost. PORTALS is coupled to the CGYRO [9] and NEO [10] codes, and predictions of plasmas in the DIII-D [11], JET, SPARC [8, 12] and ITER [13] tokamaks were presented.

A standard PORTALS workflow [14] is schematically showed in figure 3. The surrogate-based techniques present important advantages, such as the potential re-use of training data to investigate variations in scenario parameters, e.g. input power or boundary conditions. Figure 4 shows an example evolution of the residual (difference between transport and target fluxes) in a standard PORTALS simulation, where the residual for each channel (electron energy, ion energy and electron particle) steadily decreases with iteration number, until convergence is evaluated to occur after 12 iterations. It also shows the result of simulating the SPARC primary reference discharge (PRD) [15, 16] with nominal input power (11 MW) and with full input power (25 MW), all consistent with nonlinear turbulence as simulated with the CGYRO code. This work represents a significant step towards the development of fast, yet accurate predictive frameworks for the core of tokamaks. The combination of high physics fidelity and techniques from machine learning has enabled some of the most accurate predictions of core plasma performance to date.

Renaissance fusion, a nuclear fusion start-up, is pioneering a novel approach in the development of High Temperature Superconductor (HTS) coils for stellarators [17]. The fabrication and assembly of these coils are substantially

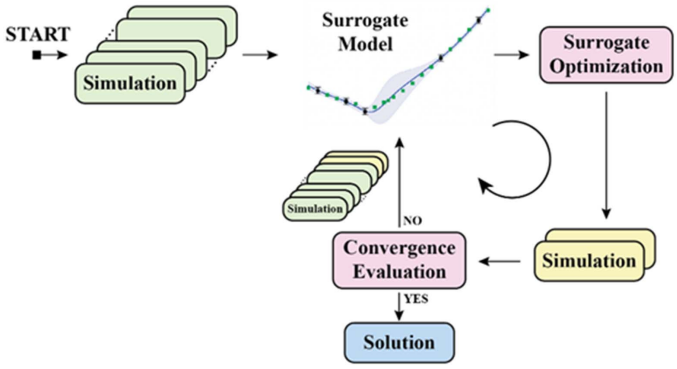

Figure 3. Typical PORTALS workflow to find steady-state solutions to core transport dynamics using surrogate-based optimization techniques.

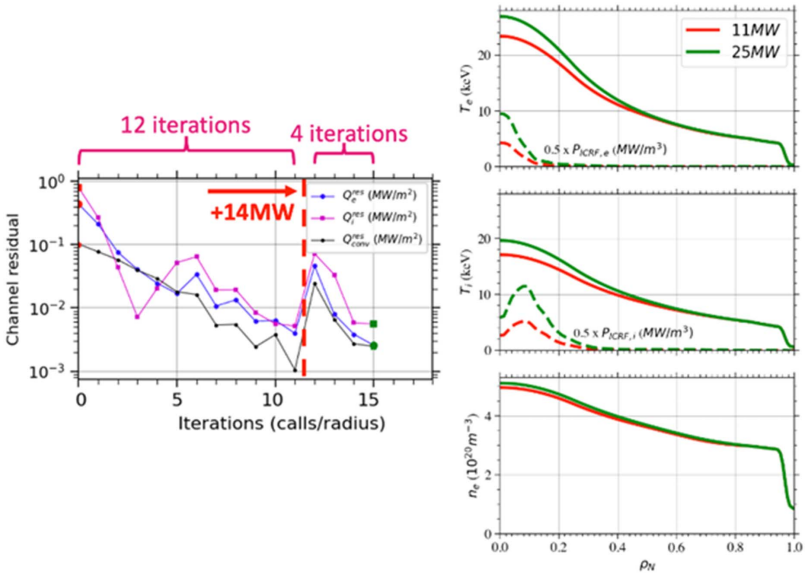

Figure 4. (left) Typical residual vs iteration behavior of PORTALS simulations, with a restart at iteration 12. (right) Simulated SPARC PRD profiles for nominal and full input power.

simplified using streamlined coil-winding surfaces and wide laser-patterned HTS surfaces. Geometric complexity is shifted from the coil winding surface (CWS) to its current distribution. Renaissance fusion adopts simple CWSs, e.g. piecewise cylindrical surfaces, at the cost of more intricate current distributions on said surfaces. These current patterns are obtained through laser-engraved grooves, which precisely direct the currents through wide HTS surfaces. Numerical optimizations of grooved conductive surfaces are presented, accurately generating target magnetic fields. As a first step towards generating stellarator fields, the design tool is applied to two axisymmetric magnets: a gyrotron magnet [18] and a magnetic resonance imaging (MRI) magnet generating highly uniform fields. Given the target field and a CWS, the current pattern generating that field is found, where a Tikhonov-regularized least-squares fitting is employed [19]. Figures 5(*a*) and (*b*) shows the results of such optimization for the gyrotron, and figures 5(*c*) and (*d*) for the MRI. Experimental validation of a gyrotron magnet was also presented. A gyrotron magnet prototype was built by juxtaposing aluminum discs of varying widths to reproduce the corrugated conductor conceived numerically. Good agreement is shown between the measured and target profiles, with an error of 1% or less in the region of interest.

A high-speed data acquisition (DAQ) system of negative ion source breakdown is developed [20]. In long pulse experiments conducted under negative ion source based neutral beam injection (NNBI), the occurrence of breakdown can lead to damage of the ion source device. Low-speed DAQ systems operating in NNBI have low sampling rates, which make it difficult to characterize the changes of each key electrical signal. To solve this problem, a high-speed DAQ system based on random trigger method, pre-trigger method, multi-threading and MDSplus DB is proposed. It collects electrical signals for a short time before and after breakdown. The current sampling rate of the system can reach up to 2 M Sa s^{-1}, which can provide high-precision DAQ of key experimental parameters such as the current and voltage data generated during NNBI experiment, providing data support for researchers to precisely analyze the possible causes of system failures and establish corresponding data samples for fault prediction implemented

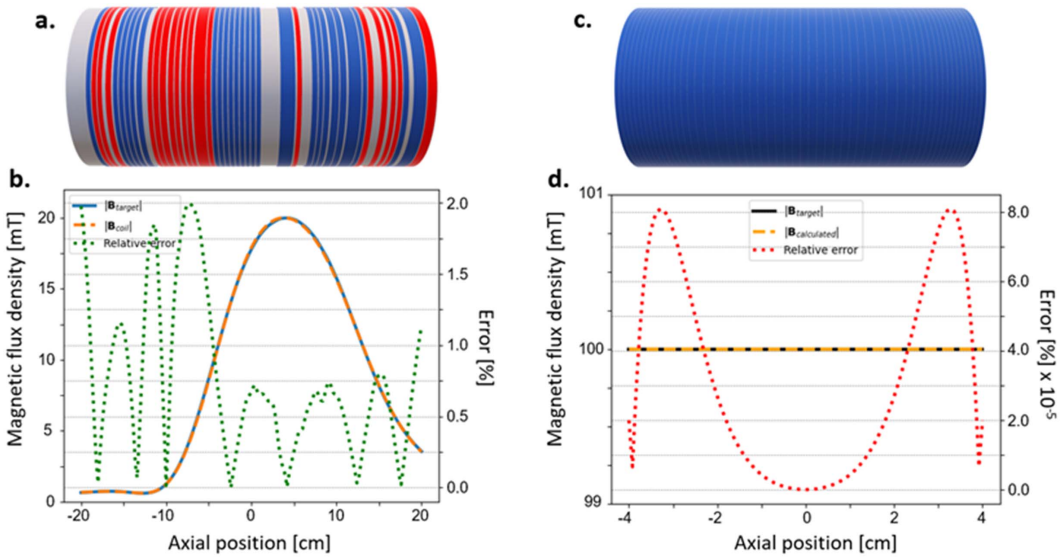

Figure 5. (*a*) Schematic design of a cylindrical conductor, optimally grooved to generate a gyrotron magnetic field profile. The magnet comprises 43 aluminum coils connected in series, where blue and red denote clockwise and counterclockwise currents and grey refers to insulators. (*b*) Desired magnetic field profile in the gyrotron (blue) and actual field (orange) generated by the configuration in a, along with the resulting local field error (green, dotted). (*c*) Like (*a*), but for an MRI magnetic field, requiring 54 coils. (*d*) Like (*b*), but for the MRI configuration in (*c*).

based on artificial intelligence. Experimental results show that the system is stable and reliable, and the acquired waveforms verify that there are subtle changes in electrical signals before and during the breakdown.

New machines, such as the ST40 high-field spherical tokamak [21], typically start with a small subset of diagnostic systems and gradually increase the diagnostic suite to include more complex and comprehensive systems. To make the most of each ST40 operation phase, forward models of various diagnostic systems have been developed [22] to aid diagnostic design, provide consistency-checks of the commissioned instruments, develop analysis methods, and constrain higher-level parameters for interpretation and modeling. The details of the various forward models implemented to date were presented. These include passive spectroscopy diagnostics (e.g. X-ray crystal spectrometer XRCS measuring He-like argon [23, 24] and charge-exchange recombination spectroscopy CXRS [25]), interferometry, filtered visible diodes (e.g. for Bremsstrahlung measurements) [26], Thomson scattering (TS), bolometric and SXR-filtered diode cameras. The different levels of complexity of the models are examined, reviewing their limitations when run stand-alone or integrated in analysis and modeling workflows. Examples of diagnostic design efforts were presented, as well as results from recent high ion temperature ST40 plasma discharges reported in [21, 27–29].

A diagnostic design example is shown in figure 6, where a tangential bolometer array has been modeled to test for resolving power and inversion capabilities. While this system can resolve strong radiation peaking and in-out asymmetries, its coverage limitations close to the LFS boundary will limit the ability to constrain edge gradients (d). Moreover, finite volume of sight (VOS) effects are expected to play a non-negligible role since on ST40 timescales are fast (milliseconds), and gradients are large. Figure 6(*e*) shows preliminary VOS tests (shaded areas) to investigate the brightness variation across the viewing cone, while VOS capabilities are being developed.

The analysis of soft and hard x-ray spectra from EXL-50 was presented [30], which has revealed non-thermal electron energy distributions. The EXL-50 spherical torus is a part of the ENN compact fusion project, which utilizes electron cyclotron wave heating and current drive at both 28 GHz and 50 GHz frequencies, supplemented by a neutral beam injection (NBI) system. X-ray diagnostics installed on the EXL-50 faced challenges such as low bremsstrahlung intensities attributable to low plasma densities and the interference from secondary emissions of high-energy x-rays, see figure 7. Initially, a toroidal silicon drift detector array system was employed for soft x-ray spectral measurements [31]. However, the low intensities of thin-target bremsstrahlung signals, arising from both low plasma temperatures and densities, necessitated a shift in approach. The solution involved the deployment of three individual detectors near the midplane port. This modification not only enhanced signal intensity but also introduced the capability to conduct scans across the plasma radius.

The hard x-ray diagnostic efforts also revealed the presence of thick-target emissions and secondary emissions, necessitating additional measures to mitigate these undesired signals [32]. Further analysis was conducted on high-energy x-ray spectra, which were acquired during time segments with constant plasma current after the switch-off of the electron cyclotron wave. The result revealed a linear dependence on plasma

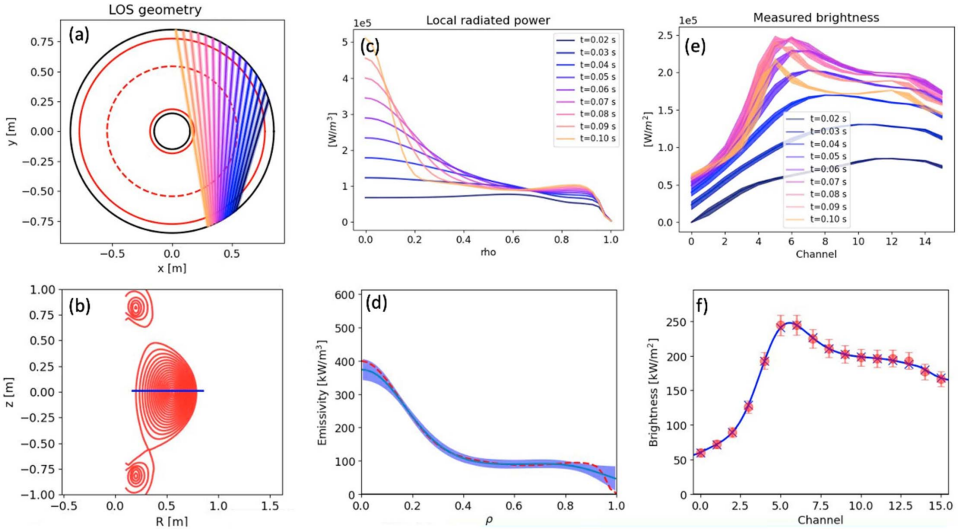

Figure 6. Forward model of a 16-channel tangential bolometer array and phantom test of inversion methodologies: *xy* and RZ line of sight (LOS) projections (*a*), (*b*), phantom local radiated power and measured brightness for a phantom radiation peaking event (*c*), (*e*), test of 1D inversion for one time-point ((*d*), phantom in red, reconstruction in blue) and comparison of back-calculated values vs phantom brightness ((*f*), same color coding).

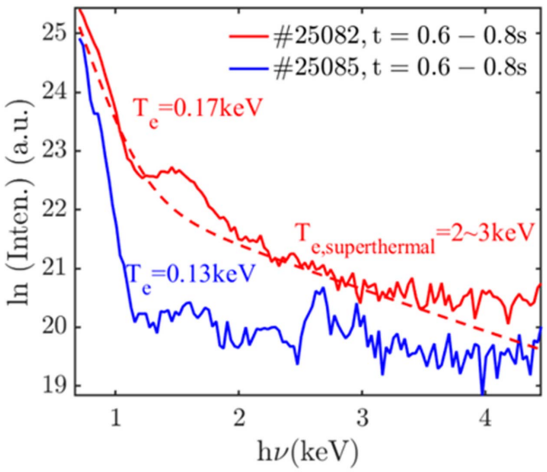

Figure 7. SXR spectra recorded during two discharges, illustrating the presence of superthermal electrons in the low-density discharge #25082.

current. This observation suggested the energy spectra corresponded to thin-target bremsstrahlung. Comparative studies with simulation tools such as CQL3D and GENRAY corroborated the experimental findings, showing that the combined effects of multiple electron cyclotron resonances and multi-pass absorption result in a pronounced asymmetric energetic tail in the electron distribution [33]. This tail was predominantly localized near the second harmonic layer, aligning with the asymmetry observed in the hard x-ray intensity distribution. These results indicate the generation of toroidally asymmetric energetic electrons near the second harmonic layer, emphasizing the need for robust shielding and optimized detection pathways to enhance diagnostic accuracy.

A feed-forward pulse design tool was developed to facilitate the initial design of candidate voltage and current waveforms for a given set of target separatrix shapes [34]. In the context of plasma scenarios in magnetic confinement devices, a pulse design represents the specification or plan for a future pulse. Generation of these plans requires an understanding of the scenario's goals, any constraints that restrict the design space, and a list of assumptions that must be made in order to pose a well-formed problem. The feed-forward pulse design tool was used as an actor within the workflow for the ITER Pulse Design Simulator, as illustrated in figure 8, which could also be used for the waveform design on any tokamak.

The feed-forward pulse design tool presented is part of the NOVA free-boundary equilibrium code. This code includes the effects of passive conducting structures as well as an automatic treatment of non-linear constraints such as coil force and field limits, and plasma-wall gaps. These features free the designer to concentrate on core design aspects, separating themselves from algorithmic details encoded within NOVA's computationally light feed-forward pulse design tool. A verification of this tool is made via comparisons to mature scenario simulations analyzed by the DINA code. The computation of a defeatured plasma scenario with a length of ∼650 s from breakdown to termination requires a wall clock simulation time of ∼5 s run on a laptop computer. A comparison of coil force and maximum field estimates evaluated at the end of this pulse show a good agreement between the DINA and NOVA codes.

The performance prediction of the current NNBI facility was examined [35] based on the physical characteristics of each key parameter of the ion source. NNBI is currently being prioritized for long-pulse experiments, with a particular emphasis on the production of ion sources that can generate beam lines possessing the appropriate energy and density for efficient tokamak NBI. To assess the viability of selected parameters prior to conducting the experiment, a straightforward static performance prediction model was developed. This model enables the estimation of the potential state of

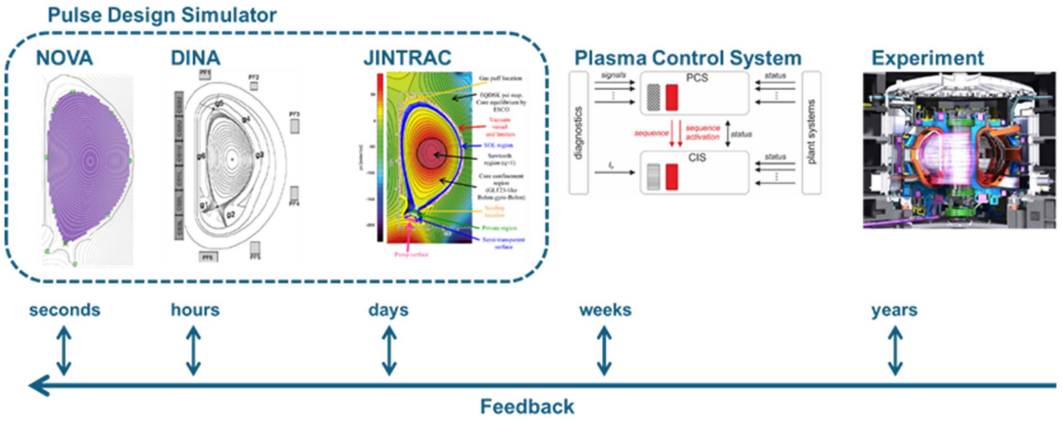

Figure 8. A cartoon illustration of a collection of feedback loops ordered in logarithmic time that provide information to guide future designs. Iteration at the fast end of a Pulse Design Simulator updates on a seconds timescale whilst cycles at the experiment/machine end of this cartoon requires years to decades before new information is obtained.

the experiment based on the input parameters, as well as the identification of any conditions that may lead to parameter breakdown. Additionally, the model provides an approximation of the negative ion output in the event of parameter matching. The experimental data was acquired through data processing and automatic system. Subsequently, a performance prediction DB was created utilizing the acquired experimental data. After performing data preprocessing, the experimental historical data is partitioned into three subsets. To mitigate the potential bias introduced by the data splitting process, a dynamic verification method known as 'K-fold cross-validation' is employed. The prediction of performance using the current parameters is now considered a nonlinear problem. In order to solve this nonlinear problem and establish the corresponding static performance prediction model, the back propagation (BP) neural network is utilized. The construction of the entire model involves two processes: the feedforward process and the back-propagation process. During the feedforward process, the output data of each node is compared with the expected data vector. Subsequently, the gradient of the error with respect to each node is calculated and the weights for each node are adjusted accordingly [36]. The static performance prediction model can avoid the ineffective shot more effectively and improve the performance of the long-pulse NNBI experiment, and provide a good encouragement for the NNBI dynamic performance preview in the next step.

2.2. Information retrieval, dimensionality reduction and visualization in fusion DBs

The session on information retrieval, dimensionality reduction and visualization in fusion DBs was chaired by Stillerman and co-chaired by Xu.

Make Informative and Nice Trends (MINT) is an ITER graphical data visualization and exploration tool designed for plant engineers, operators, and physicists [37]. Its requirements were gathered through interviews with various stakeholders, and its architecture was planned for a long-term project such as ITER. As such, a modular design and clear definition of generic interfaces (abstraction layer) were crucial, providing a robust foundation for future adaptations to new plotting, processing, and GUI libraries. The MINT application relies on an independent plotting library, which acts as a wrapper for the choice of underlying graphical libraries. Data selection and retrieval were also developed as a separate module, with a well-defined data object interface for easy integration of additional data sources. The processing layer is also a separate module, supporting algebraic and user-defined functions.

A framework for describing distributed real-time control systems was presented [38]. It was highlighted that modern real-time plasma control systems (PCS) will be modular and distributed, providing advantages such as component isolation, simplicity, robustness, scalability, and the ability to utilize heterogeneous execution environments. The associated challenges of system complexity and communication-related delays were also discussed. The framework was detailed, explaining its components and the schema for describing the interfaces between components, function blocks, and the deployment of real-time actors on computers running a real-time framework. This solution is implementation-agnostic and can generate artifacts to drive specific real-time systems. As a demonstration, the framework was applied to the control system of a small magnetic levitation device. Deployment stubs for both MDSplus/MARTe2 and SCDDS/MARTe2 were produced, and these were used to implement controllers deployed to operate the device. It was also noted that this metadata framework can be applied to modular simulation environments.

The first steps in designing and developing an intra-pulse data analysis chain for ITER was presented [39]. An infrastructure is set up with on-demand calculations governed by one or more supervisors, which launches different workflows. The supervisor will manage the workflows, the pulse time points to be calculated and the general status of the computational infrastructure. Plasma parameters can thus be calculated during the pulse from diagnostic measurements. A first implementation of a workflow has been shown using synthetic magnetic measurements, and from there inferring the plasma

equilibrium. Initially, a set of magnetic measurements are calculated. This requires knowledge of the poloidal flux from the desired ITER scenarios, together with the corresponding plasma current and the machine description of different components that determine the behavior (like passive structures, wall, and toroidal field coils). A Bayesian inference process is used to add uncertainties and interpolate the data, ensuring a more realist set of signals. An important aspect of this synthetic diagnostic for magnetics data is the introduction of a frequency-dependent noise (lower power at high frequencies) which closely mimics the usual hardware noise. This data is written to self-describing netCDF file(s) and will be used as input information to the real-time processes as implemented by the magnetics plant systems. To save network bandwidth, the data is encoded. Then the data is streamed for archiving and stored as HDF5 files. This part is executed in the Plant Operation Zone network (POZ), with the aim to simulate a complete signal acquisition chain of the magnetics diagnostic. From here, they are handled as real plant signals, being transferred to the external plant network (XPOZ), down sampled and used as the initial data for a short intra-pulse analysis workflow. Here, an equilibrium reconstruction is calculated, which is then displayed in the temporary control room live display.

Most DBs in fusion research are devoted to a single topic, such as energy confinement, H-modes, profiles or disruptions. In order to allow for a wider range of analysis, modeling and validation tasks, a broad-based multi-purpose DB, JETPEAK, has been developed for JET [40]. This DB currently includes 27 065 stationary state ($\partial/\partial t \approx 0$) samples, and near 1000 scalar, 1D (profiles) and 2D (R&Z dependent) variables grouped into topical structures. A similar DB has been created for the TCV tokamak, comprising 65 000 samples reaching back to early TCV operation in the nineties. The breadth and flexibility of these DBs allows them to be used for a wide variety of investigations such as modeling tasks, confinement scaling, testing, validation and benchmarking of algorithms and modeling codes, long term monitoring of device conditions as well as for documentation.

An Integrated Control System (ICS) of the NBI system is designed [41] to coordinate the operation of various subsystems of NBI while ensuring steady-state operation of entire system and the safety of experimenters. The control system adopts a distributed design structure to balance the system load. It is designed based on the idea of Internet of Things (IoT), which includes comprehensive perception, transmission and intelligent processing. IoT identifies 'things' from the perception layer, transmit the perceived data through the network layer, and finally realize the corresponding applications. Similarly, NBI ICS needs the following three elements: the first is Control, which must be able to complete the control tasks accurately, timely and precisely; the second is the need for reliable communication, including the communication between the various nodes of the system and the acquisition and storage of data; the third is its ability to complete and optimize the accuracy and credibility of the prediction model through the massive amount of data. A distributed control system design for NBI based on IoT idea can be shown in figure 9.

As an example, if the EAST NBI ICS is built using the OPC UA protocol (a way to build an IoT network), it can utilize its special address space to connect multiple subsystems of the NBI system to eliminate interoperability barriers between different hardware platforms and automation software. In addition, it enables communication between any node of the system and the accurate exchange of data and control across multiple platforms and protocols. The concept of distributed objects can also be used to transform the digital model of multiple subsystems of NBI so that different software can control the devices as if they were invoking objects. In terms of communication, according to its official introduction, there is no need to make large-scale modifications to the existing system, only the OPC UA server, OPC UA client or OPC UA protocol stack needs to be added to the initialization part of each subsystem in order to make it work automatically. Therefore, IoT allows interoperability between subsystems of different architectures, systems and platforms. Also, if a part of the system needs to be added or removed, this will be easier to accomplish due to the reduced coupling.

Designing smart ICS using IoT is just a concept now, and the actual implementation may encounter various challenges, including privacy, security and economy issues. Therefore, the realization of intelligent control systems may still have a long way to go.

The session also focused on data analysis tools to ease exploitation of complex fusion data sets. In this regard, there were four talks, two on service-based analysis and machine learning platforms, one on integrated modeling & analysis suite (IMAS), and one on the data processing of neutral beam data. An underlying theme of all of them is to tame the complexity of data sets and high-level analysis.

Machine learning methods are used as a service to optimize tokamak operations [42]. Tokamak operations and datasets of fusion plasma research are complex and hard to analyze, and traditional methods usually struggle to handle this complexity. To address these challenges, a local artificial intelligence-as-a-service (L-AI-aaS) platform is developed for WEST tokamak. L-AI-aaS platform aims to enhance and optimize data analysis by integrating data from various sources involved in WEST Tokamak experiments. Once the data is integrated, the platform will utilize AI algorithms to allow rapid insights extraction that are important for both ongoing and future experiments, such as disruption and anomaly detection in pulse. It can also assist in discovering new phenomena and accelerate the researcher outcomes. Moreover, the L-AI-aaS platform will serve as a decision-support tool for researchers. The AI-trained models and services on the platform will provide researchers with increased visibility into the data, thereby helping them in making informed decisions about their experiments. This is an important step forward in operational fusion plasma research, as it helps researchers to focus on designing and interpreting experiments, rather than focusing on data management and preliminary analysis.

The AI operations pipeline consists of data labeling, contextualization, and quality checks (see figure 10). In addition, a proposed workflow for an assisted labeling service is demonstrated. Note that the platform is currently in the prototype

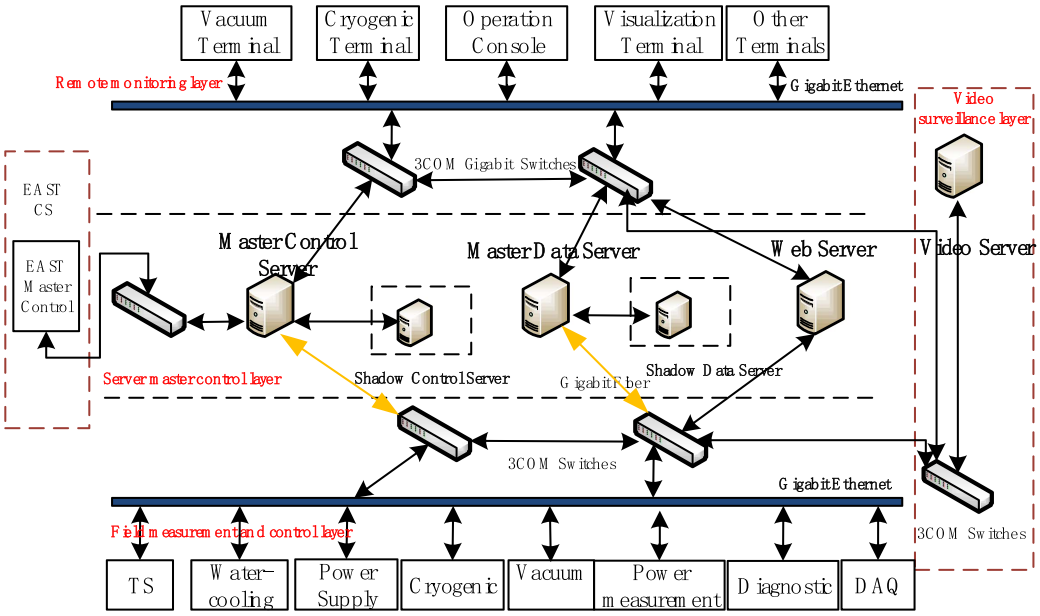

Figure 9. Distributed design of NBI based on IoT.

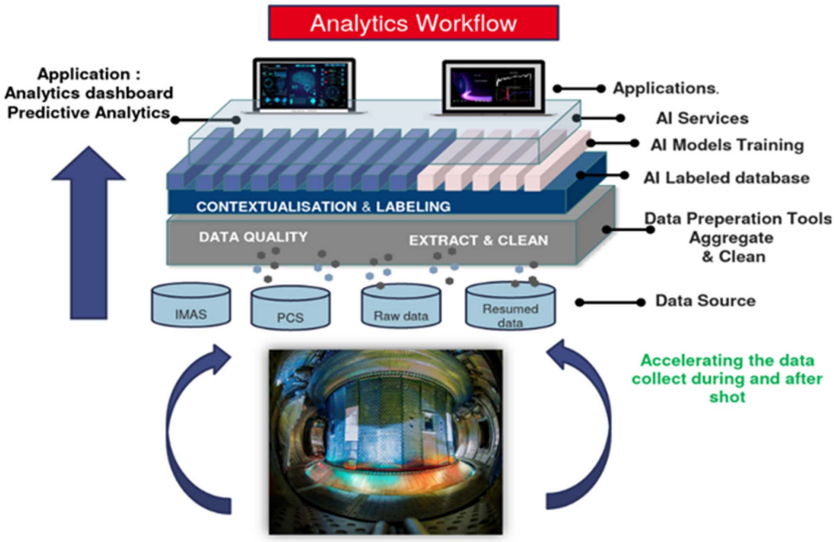

Figure 10. Local artificial intelligence-as-a-service (L-AI-aaS) architecture. Reproduced with permission from [42].

stage and is still under development. For example, algorithms like dynamic time warping (DTW) and K-means clustering can be used to extract shape-based similarity measures in time-series DB. These algorithms are useful in applications where the goal is to find a set of pulses that carry similar physical information for further analysis. Figure 11 shows how clustering techniques group sets of pulses from the DB based on their plasma current values over time. Implementing this approach has its own challenges, including the complexity of managing data from multiple sources, the need for high computing power, and the possibility of algorithmic biases. Despite these challenges, AI applications still hold the potential to advance fusion plasma research.

A new web-based user interface for a distributed data analysis system has been developed [43]. This tool focuses on providing a simple-to-use, federated platform to access to (annotated) data and fine-trained models, share validated algorithms and dedicated hardware, allow for new unanticipated development, and ease communication, reproducibility and replicability. This makes results easier to obtain, communicate and reproduce or replicate. In this system the code is specified by drawing a graph of the operations. The user specified graph is fed to a 'coordinator' which sequences and distributes the tasks, as shown in figure 12. This provides secure, remote access to data and computing power allowing users to operate from any device. It has an intuitive and easy to learn UI. Work is performed in a distributed environment. The system allows integration of code, common and private, different programming and support the building of a community in a federated form.

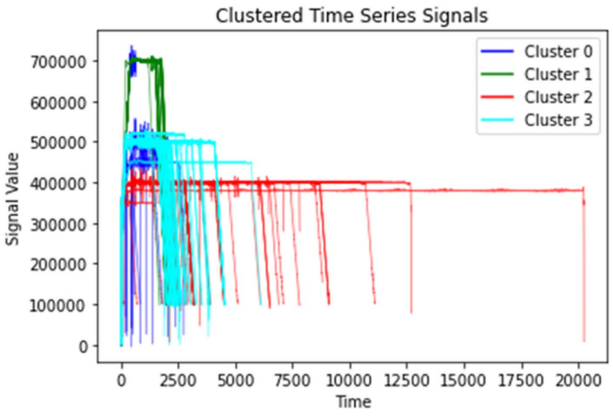

Figure 11. Results of dynamic time warping (DTW) and K-means clustering to identify similarities among different pulses in database, focusing on plasma current signals. Reproduced with permission from [42].

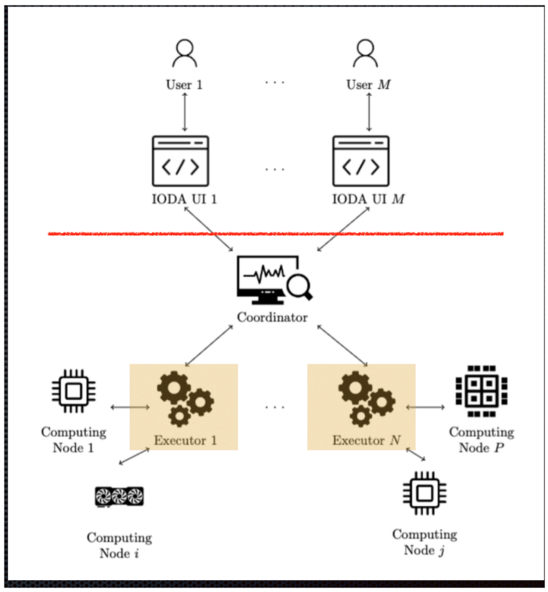

Figure 12. A 'coordinator' which sequences and distributes the tasks. Reproduced with permission from [44]. / 2023 The Author(s). Published by Elsevier B.V. CC BY-NC 4.0.

A new simulation data management tool has been developed [45], which presents new solutions to the ITER simulation management and remote access, and facilitates findable, accessible, interoperable, and reusable (FAIR) data access to the ITER simulation catalogue. While the IMAS [46] is being developed further, the number of simulations available in IMAS is increasing both at ITER and within the Members. As a result, tools are needed to manage, curate and expose the large IMAS simulation DBs to the potential users. Ad-hoc solutions have been developed in some cases, such as the ITER scenario DB (2500+ simulations) and its associated Python scripts that register some meta-data (stored in a yaml file associated with each simulation) and then browse or query this recorded information to identify the simulations of interest. This solution, while simple to set up, is not generic (scripts are tightly bound to ITER while IMAS allows for simulation on any machine), is not meant to scale to a very large number of simulation counts (queries need to go through all yaml files) and lacks important features. In order to address this need for a general purpose tool for managing simulations in IMAS, a simulation data management tool (SimDB), has been developed. Using SimDB simulations are ingested along with meta-data that records the input, output and information about the code that captures the simulation provenance and ensures reproducibility of the simulation. Each simulation is given a globally unique identifier (UUID), and can be pushed to one or more remote archives where the data can be validated and made available to other users. The ingested simulations to be queried via a command-line interface or via a web frontend using a flexible query syntax. A SimDB catalogue has been set-up at ITER with meta-data from the dataset description and summary IDSs being made available and queryable, making the SimDB meta-data directly interoperable with other IMAS-based catalogues. In addition to SimDB, the IMAS access-layer has been extended to allow for remote access of IMAS data, using the simulation URIs returned from the SimDB queries. Based on the UDA client-server solution, this will allow for secured authentication of users and for full or partial access to the IDS data as well as controllable batching of requests to improve the performance depending on the capabilities of the network. The public SimDB servers, the SimDB unique identifier and the remote access URI provides a method to unambiguously refer to IMAS data in publications, improving the FAIRness of IMAS.

The design of a NBI experimental data processing system was presented [47]. NBI requires parameter tuning and optimization before it can be formally put into operation, so its operational data is of critical importance to experimenters. This presentation focuses on the design and implementation of the NBI data processing system based on the Experimental Physics and Industrial Control System (EPICS), which is a three-tier distributed system architecture consisting of 'task processing', 'storage processing' and 'interaction processing'. A data processing scheme between memories is proposed based on the double buffering algorithm and MMAP technology. The principle of double buffering is setting two buffers and realizing uninterrupted data processing process through MMAP. While the experimental data is read into the IOC via 'ai' for visualization, the data files are uploaded asynchronously to the data server for long-term storage. Through testing, the data processing rate is about 5 times of the old. In addition, considering the scalability and compatibility of system, a device model is proposed based on EPICS to unify the device abstraction processing format and standardize the subsequent device development. The asynchronous data processing flow is shown in figure 13. The data update and transmission rate of channel archiver is the same as the scanning rate of EPICS, so a data transmission structure is designed based on TCP/IP protocol to coping with the higher data sampling rate.

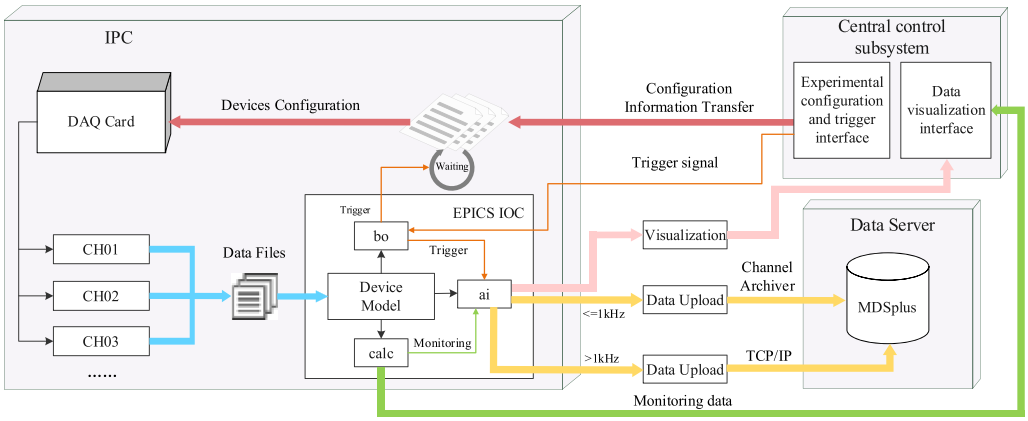

Figure 13. The asynchronous data processing flow. Reprinted from [48], Copyright (2021), with permission from Elsevier.

After testing on NBI testbed, the system showed a significant improvement over the old method in terms of data processing. This system will provide more real-time data processing for NBI experiments and will be used to cope with more complex pulse experiments in the future.

The talks in this session all talked of systems to reduce the difficulty of accessing complex data sets. At least three of them were either explicitly or implicitly talking about FAIR (Findable, Accessible, Interoperable, Repeatable) principles. In our distributed, collaborative, complex field these tools are needed for effective science.

2.3. IDA and synthetic diagnostics

The IDA session, chaired by R. Fischer and co-chaired by G. Verdoolaege, focused on recent advancements on the interpretation of experimental data and its augmentation with physics modeling information.

The session started with an overview on the status and prospects of IDA for present and future fusion devices where a new modular IDA framework was introduced and a first workflow for ITER kinetic profile estimation was presented [49].

For machine control and safety as well as physics studies, present and future fusion devices have to analyze a huge amount of measurements coming from many redundant and complementary diagnostics. IDA in the framework of Bayesian probability theory provides a concept to analyze a coherent combination of measured data from heterogeneous diagnostics including their statistical and systematic uncertainties and to combine them with modeling information. Based on more than 20 years of experience in applying IDA at various fusion devices, on various combinations of diagnostics data and various parameter sets, a generic IDA code package was recently developed to provide a modular and flexible basic python code to be applied to present and next generation fusion devices. The IDA ingredients were summarized and the status of the newly developed IDA platform was presented. The framework allows for modularity with respect to multi-fidelity forward models and prior information from low-fidelity profile smoothness constraints up to high-fidelity physics constraints on profile gradients from transport modeling. Various options for the estimation of physical parameters of interest and their uncertainty determination were shown. The linkage with the ITER:IMAS data base and recent applications are presented in the overview talk.

The second talk addressed an example of density estimation combining reflectometry, TS and lithium beam excitation spectroscopy (LIB) data and how to validate and handle inconsistent diagnostics data employing IDA based on Bayesian probability theory [50]. Reliable forward models for diagnostic, linking parameter space to data space, are essential for predicting measurements accurately enough to be used in a probabilistic IDA approach. This enables the identification of systematic differences in profiles estimated from individual diagnostics. IDA at ASDEX Upgrade determines the density profiles based on interferometry, TS, LIB, thermal helium beam, and swept O-Mode reflectometry (REF). The results from independent diagnostics are not always in agreement, with situational differences beyond the diagnostic uncertainties. These differences can originate from numerous sources like invalid assumptions in the forward models, uncertainties in physical hyper-parameters, insufficient calibration of diagnostics or time-dependent drifts, cross-calibration of diagnostics under critical assumptions. This was exemplified by a combined analysis of the LIB and TS diagnostics which was recently complemented by a swept O-Mode reflectometry diagnostic [51] in the important region around the separatrix. This allows for validation studies of the diagnostics, as the forward models generally rely on individual specifications of the diagnostics—like measurement location and calibration—and, additionally, the forward models might use mathematical simplifications to enhance evaluation speed. The newly developed implementation of the reflectometry forward model and testing procedures are summarized. Common pitfalls in test procedures and how to detect and fix them are illustrated. First results of a validation study on the uncertainties and discrepancies between the three diagnostics are showed based on experimental data from AUG. The joint analysis of LIB and REF solves the initialization problem of stand-alone REF data analysis. While REF agrees with LIB in the scrape of

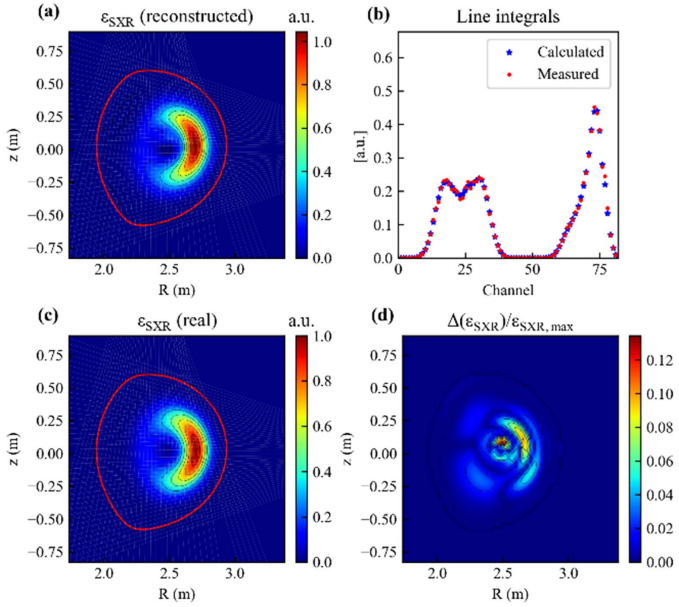

Figure 14. Reconstruction of synthetic SXR emissivity: (*a*) reconstructed emissivity, (*b*) measured and calculated line-integrated emissivity, (*c*) real emissivity (*d*) reconstructed error, normalized by the maximum emissivity. Reproduced from [52], with permission from Springer Nature.

layer and close to the separatrix, there is significant disagreement with respect to the TS data close to the pedestal top. Understanding the limitations of diagnostics and their forward models is essential for the interpretation and evaluation of experiments, especially when profiles are used as input to large modeling codes.

A third example of IDA was shown for soft x-ray tomography and tungsten impurity concentration estimation at WEST [52]. The accumulation of heavy impurities like tungsten (W) in the plasma core can be dangerous as it causes significant radiative power loss, which may eventually lead to a disruption. An accurate estimation of tungsten impurity concentration is crucial for impurity control. The emissivity of plasma radiation from a given species S relies on its density *n*_S, cooling factor *L*_S (dependent primarily on *T*_e) and electron density *n*_e, therefore diagnostics measuring the plasma emission, for instance the soft x-ray (SXR) diagnostic, can be used to infer the tungsten concentration. The local SXR emissivity is reconstructed from the line-integrated emissivity using a Gaussian process tomography method based on Bayesian probability theory [53, 54]. This method has been validated on synthetic data with good performance (figure 14). After applying a new error model containing background noises [55], the tests on real WEST data also yield satisfactory results. The radial tungsten concentration profile is jointly estimated along with the kinetic profiles using measurements from SXR, interferometry (for electron density) and electron cyclotron emission (ECE, for electron temperature) by employing the IDA approach [56]. An example tungsten concentration profile reconstructed from synthetic data matches well in the center of the plasma and the re-integrated SXR emissivity is in good agreement with the synthetic SXR measurements. However, further outside the central plasma the uncertainty of reconstructed c_W grows (figure 15), because at low n_e and T_e the SXR radiation from tungsten is almost negligible. The full Bayesian inference based on MCMC sampling is computationally too expensive. Preliminary investigations have been done on fast inference of density profile by means of a neural network surrogate model trained on synthetic interferometry data.

Neural networks are leveraged in a simulation-based inference (SBI) for a Lyman-alpha diagnostic at the DIII-D tokamak for fast sampling to create a posterior of the neutral density [57]. Inferring physics parameters from experimental data is a key analysis need for physicists. Often it is desired for this inference process to have grounding in detailed models contained in simulation. SBI are techniques which utilize simulations in the forward model and create approximate posteriors which are faithful to the simulation. Recently neural networks have been leveraged in SBI to flexibly represent the underlying Bayesian inference process. The approximate posteriors generated can be sampled from quickly, which is attractive for fast between shot-analysis. Ongoing work was shown in applying SBI to experimental fusion diagnostics, focusing on optical diagnostics. A synthetic diagnostic for the Lyman-alpha diagnostic (LLAMA) at the DIII-D tokamak was created, using the CHERAB code for spectroscopic diagnostics, and the KN1D neutral transport code for relatively fast neutral density transport. By generating many thousands of samples of synthetic plasma input profiles, and obtaining output of the forward model, SBI can be leveraged for creating an approximate posterior. Neural networks representing normalizing flow for accurate replication of the data distribution are used. This creates a neural network which can be sampled

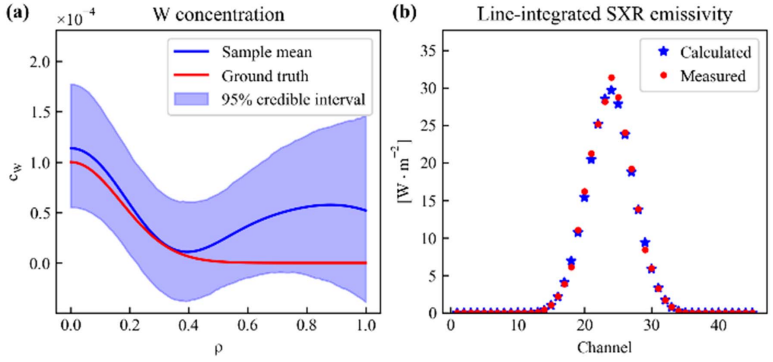

Figure 15. (*a*) W concentration estimated from synthetic data and (*b*) re-integrated SXR emissivity. Reproduced from [52], with permission from Springer Nature.

from quickly to create a posterior of the neutral density given the measured LLAMA line-integrated radiance.

Turbulence plays a critical role in limiting our ability to raise plasma core temperature and density profiles. Kinetic profiles are expected to not only align with experimental data but also adhere to our expectations of turbulence-driven logarithmic gradients. An augmentation of IDA of experimental data with modeling data at ASDEX Upgrade was proposed using ASTRA-TGLF simulations for constraining the parameter space of profile gradients to adhere to our expectations of turbulence-driven logarithmic gradients [58]. Bayesian probability theory is employed to augment the IDA framework using experimental profile data with modeling data from ASTRA-TGLF simulations [59, 60]. This combination was tested by comparing heat and particle balances (calculated by RABBIT [61] and Torbeam [62]) with their respective fluxes (calculated by TGLF-SAT2 and NCLASS [56]) for plasma scenarios in both L-mode and H-mode. For a proof-of-concept work, in which the uncertainties of the simulation are predefined, electron heat flux match was improved by a factor of three, particularly in the mid-radius region where a predetermined 10% uncertainty was assumed. The ion heat flux was found to be less sensitive to the electron profile chosen but largely dependent on the ion profile which is not yet fitted with a kinetic modeling prior. The particle flux match showed promise, especially for cases with higher flux, though there is still room for refinement (especially as particle flux modeling becomes more of a focus for transport codes). To ensure that the kinetic model does not overrule the measurements, the residuals of various diagnostics are validated for different plasma scenarios. It was found that the kinetic modeling prior does not deteriorate the diagnostic match significantly. To further safeguard against a potential negative impact of the kinetic model on the profiles, a Student's *t*-distribution was considered for the modeling data, which accounts for outliers due to its long tails. The Student's *t*-distribution was shown to be especially valuable when the kinetic model failed to predict a large 2/1 neoclassical tearing mode due to missing physics. The theory informed profiles have been tested as a starting point for high-fidelity GENE validation simulations. Using the new profiles for GENE applications could save large computational resources as the input is closer to the final predicted gradients.

In integrated modeling (IM) workflows synthetic diagnostics are coupled as persistent actors in Bayesian inference analysis via the use of the Muscle3 coupling library [63, 64]. An overview of the diagnostic models available in the IMAS [46] to model the ITER instrumentation systems were given. Some use cases were described in which synthetic diagnostics are applied to perform physics data analysis and develop plasma modeling tools. The use of diagnostic models in workflows doing Bayesian inference analysis were presented where the concept of persistent actor framework was employed. Synthetic diagnostics help to development the ITER PCS [65] and its simulation platform (PCSSP) [66] through the design of its support functions and the application of its control algorithms inside co-simulations combining IMAS models with Matlab/Simulink controllers. This type of co-simulation is made possible via the use of the Muscle3 coupling library within the so-called persistent actor framework [67]. This framework facilitates the communication between various actors (models) in an integrated simulation across languages and domains. A closed-loop prototype was presented where the plasma density measurement is simulated by an interferometer model that provides the signals through the real time data network (represented by the real_time_data Interface Data Structure or IDS) to a Matlab/Simulink controller. This controller in return sends a command, still through the real-time-data IDS, to a gas puff model that adjusts the gas injection. As such, an external source of particles is injected into a transport model, which evolves the plasma density accordingly. Methods to extend this prototype to more sophisticated plasma simulators were discussed, since the persistent actor framework can be used for co-simulations between PCSSP controllers and high fidelity or pulse design simulators, e.g. in the context of free-boundary control simulations for the validation and verification of models, workflows and controllers.

Finally, a Bayesian approach for estimating the kinematic viscosity model in reversed-field pinch fusion plasmas was presented [68]. A fundamental feature of reversed-field pinch fusion plasmas is the formation of helical self-organized states. In the past few decades, MHD theory and numerical simulations have played a key role in describing these states. An important parameter is the dimensionless Hartmann number [69], which is determined by the resistivity and the viscosity.

Table 1. Bayes factors between the different viscosity models.

Viscosity models being compared	Log Bayes factor
Perpendicular Braginskii vs. gyro Braginskii	212
Gyro Braginskii vs. ITG	259
ITG vs. Finn	353
Finn vs. parallel Braginskii	653

It can be interpreted as the electromagnetic equivalent of the Reynolds number and it turns out to be the ruling parameter in the 3D nonlinear visco-resistive magnetohydrodynamics activity. However, there is no consensus regarding the theoretical model for the kinematic viscosity coefficient. There are five candidate models according to the various momentum transport theories developed for hot magnetized plasmas: three classical viscosities derived from the closure procedure leading to the Braginskii equations, the ion temperature gradient viscosity, describing a mode that damps the velocity fluctuations, and the Finn anomalous viscosity according to the Rechester–Rosenbluth model. The viscosities and the Hartmann number were calculated using measurements from RFX-mod. A power-law dependence was then sought between the Hartmann number and the amplitude of the $m = 0$, 1 secondary modes. Our approach, using Bayesian statistics, outperforms the previous analysis based on simple least squares fitting. First, by computing the Bayes factor [70], it is inferred that a constant relative error is a better model for the uncertainty in the regression analysis. Second, errors on the plasma parameters and their role in error propagation were taken into consideration. Third, Bayes factors between the different viscosity models were used to infer the optimal viscosity model, in a more robust way compared to the earlier approach based on correlation coefficients and simulations (table 1).

The optimal model, identified through the Bayesian procedure, agrees with physical motivation [71]. More generally, the work has demonstrated the potential of the Bayesian approach in other model selection problems in fusion, using a rigorous and robust statistical methodology.

2.4. Inverse problems (INV)

The section on inverse problems was chaired by G. Verdoolaege and co-chaired by Mazon and Murari. In science, an inverse problem can be defined in full generality as the task of calculating from a set of observations the factors that generated them [72]. Such problems are called inverse because they are meant to derive the causes from their effects. They can therefore be considered the opposite of forward problems, whose objective is calculating the effects of causes. Many data-centric problems in fusion are 'inverse' in nature, i.e. they involve extracting unknown parameters and causes from observations. Indeed many fundamental measurements, being based on the plasma natural emission, require some form of inversion to be interpreted and provide the required physical information: the measurements to obtain the magnetic topology, tomographies, videos and nuclear detectors are just some examples. In general, the interpretation of these measurements needs solving mathematically ill posed inversions and therefore present issues such as: estimating the confidence intervals in the results, dealing with the consequences of noise, minimizing bias effects. Various approaches to address these difficulties have been demonstrated for the case of tomographic inversion of emitted radiation measurements [73].

Accurate measurement and control of radiation emitted by thermonuclear plasmas is crucial for successful operation of fusion reactors. This information is relevant to many aspects ranging from power balances and impurity transport to power exhaust and disruptions. Current tokamaks use bolometers to measure plasma emission, but they only provide line-integrated values and require an inversion technique for obtaining local information. Tomography inversion is a commonly used approach for obtaining high spatial resolution reconstruction of the emissivity, but it is slow and unsuitable for real-time use. The main present day research efforts are therefore focused on two main aspects: (a) the improvement of the quality and reliability of the offline inversion algorithms, (b) the development high time resolution low spatial resolution inversion methods [74].

On the Joint European Torus (JET) first and more recently on ASDEX-Upgrade (AUG), an Expectation Maximization algorithm has been implemented to derive the maximum likelihood (ML) agreement between the line integrated measurements of the radiation, and the reconstructed tomograms representing specific poloidal emissive distributions. The main and most distinctive feature of the method is the capability of estimating the variance related to the reconstructed tomogram and, consequently, of evaluating the uncertainties in the derived quantities.

Since the first implementations on JET, devoted to the gamma and neutron measurements [75, 76], in the last years more attention has been devoted to the bolometric tomography [77, 78]. Dedicated studies have been performed to improve the outputs of the ML tomographies [79] and consequently, to increase the reliability, quality and accuracy of the derived quantities. The algorithm developed can handle missing or unreliable LOSs due to faults that might occur during an experimental campaign, as well as systematic errors and outliers in the measurements [80, 81]. More recently, two upgrades have been developed and implemented to: (a) minimize the risk of producing artifacts, unavoidable and unwanted features that can strongly influence heat transport and turbulence studies (see figure 16) [82]; (b) to handle the asymmetric brightness on LOSs, due strong gas puffing close to one of the bolometer arrays [83]. The developed algorithm is therefore probably one of the most complete and advanced available nowadays in the fusion community. Having proved the portability between devices, efforts have been spent, and are also currently on going, to develop a real-time version compatible with the ITER fast controller platform. Such efforts succeeded at reducing by a factor ten the time interval required for estimating a reconstruction, paving the way at least for an intershot application of the ML code in future versions [84].

Plasma instabilities can be associated with many radiation patterns differing in localization, shape intensity and time

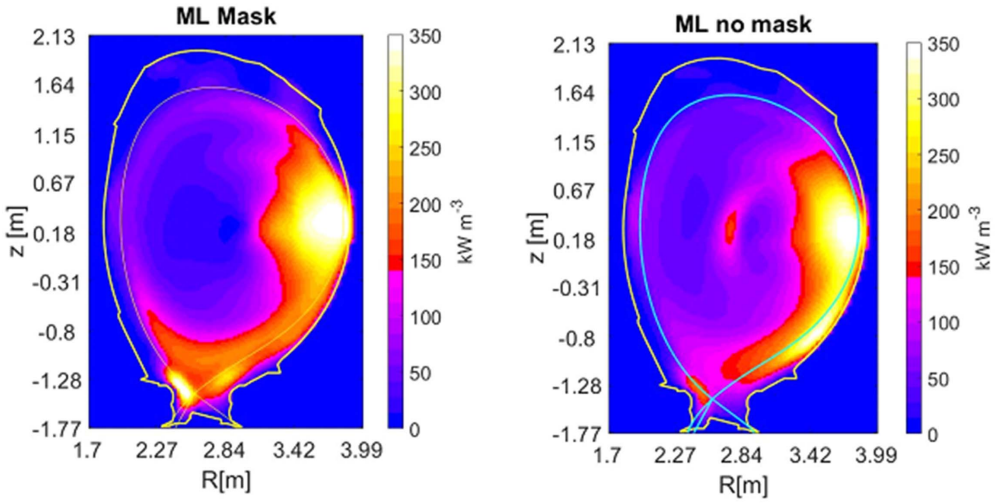

Figure 16. Example of ML tomography showing the effect of correcting of the artifact in the core. Left: corrected tomogram. Right: non corrected tomogram. Reproduced from [82]. © IOP Publishing Ltd. All rights reserved.

evolution. Some of them can rapidly yield to the loss of plasma confinement [85]. A fast inversion technique that provides a fast time reconstruction has been developed [86]. Indeed by decreasing the grid resolution allows a fast inversion of the line integrals, even though a lower spatial accuracy is obtained. Nevertheless, by selecting accurately the regions of interest it is possible to maintain sufficient local information. Different regions of interest can be defined by grouping the lines of sight from the horizontal and vertical cameras and intersecting them in the right combinations (see figure 17). Then the inversion is performed by using a nonnegative linear least-squares algorithm. The reliability of the method has been demonstrated by analyzing numerical generated patterns and the accuracy is evaluated considering different shapes and positions of emitted regions. Further validation of the method is offered by comparison with the established ML tomography reconstruction on different discharges of JET with a metallic wall. The radiation estimated by the fast inversion tomography is typically within 20% of the value calculated with the well-established and validated ML approach. Finally, an analysis of the main radiation patterns has been performed with this developed method in order to understand the mechanism, which can induce the plasma to disrupt. The achieved results suggest that fast inversion techniques, of the type developed for JET and AUG, are promising tools for real-time radiation monitoring and control in fusion reactors.

2.5. Analysis of time series, images and video: detection, identification and prediction (TIV)

This session was chaired by Vega and co-chaired by Murari and was devoted to the automatic recognition of physics events in time series, images and video-movies.

Even if the understanding of the tokamak configuration has progressed significantly in the last years, these devices are all plagued by the collapses of the plasma called disruptions [87]. In the next generation of devices, their consequences can be harmful for the integrity of the devices, in addition to slowing down the scientific exploitation as in present day experiments [88, 89]. Moreover, devices with metallic plasma facing components, similar to those foreseen in future reactors, are also vulnerable in this respect, particularly when operated allow q_95 around 3. Since first principle models based are not mature enough to provide reliable disruption predictions in real time, data driven classifiers based on machine learning technologies are being developed [90]. The present day research in this field is moving into two main directions (a) improving the indicators and the training methodologies to provide tools adequate to perform avoidance and not only mitigation (b) developing techniques and methodologies to perform reliable predictions for mitigation with a limited number of signals, situation typical of the beginning of the operation in new devices.

With regard to the first topic, all JET discharges in campaigns C38 to C41 have been analyzed. They include high power experiments in deuterium, full tritium, and D–T, for a total of almost 1700 pulses (see table 2). Only intentional disruptions and a few discharges missing essential data have been excluded. The results of a systematic analysis of these campaigns have allowed to confirm the main plasma dynamic routes leading to the beginning of the current quench, showed in figure 18 [91]. Almost all disruptions are preceded by anomalies in the radiation patterns, which either cause or reveal the approaching collapse of the configuration. Given the influence of these radiation anomalies on the kinetic profiles and the magnetic instabilities, a series of innovative and specific elaborations of the various measurements, compatible with real time deployment, is required. The data-driven indicators derived from these measurements can be interpreted in terms of physics-based models, which allow determining the sequence of macroscopic events leading to disruptions. The computational and warning times are such that the control systems of future devices are expected to have more than sufficient notice to deploy effective prevention

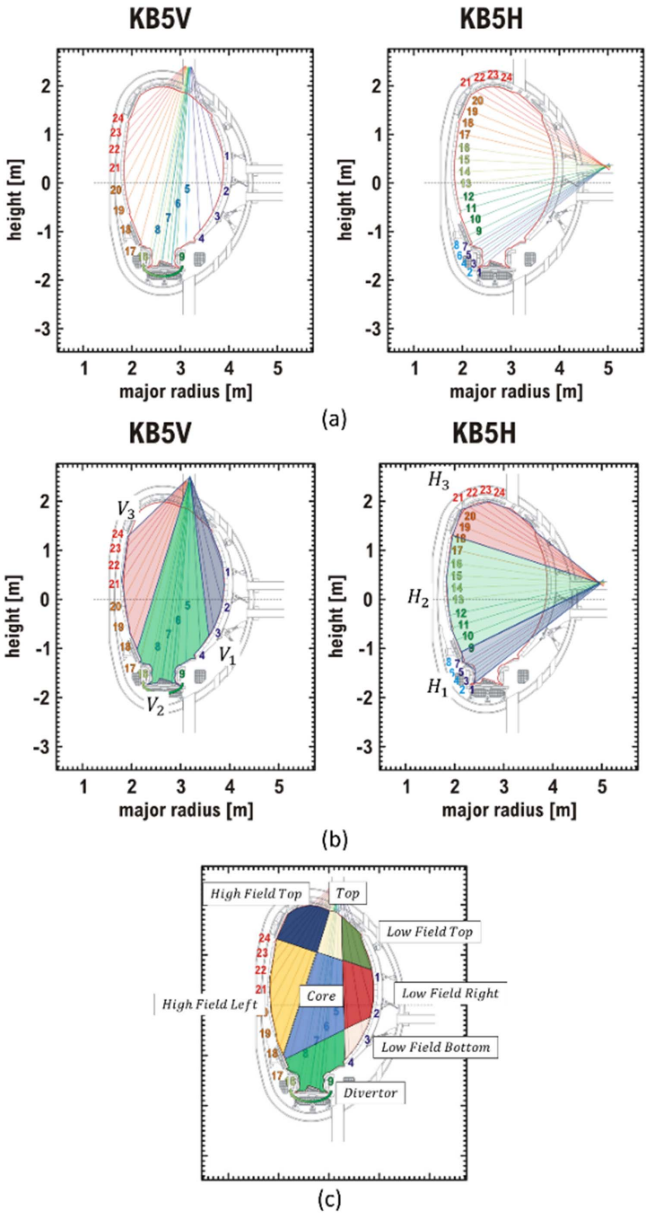

Figure 17. Definition of the macroviews and macropixels for the fats time resolution tomography on JET. Reprinted from [85], with the permission of AIP Publishing.

Table 2. Statistical results for the discharges in JET campaigns from C38 to C41, including experiments from high power deuterium to full tritium and DT. Only seven disruptions do not present a radiation anomaly.

	C38	C39	C40	C41	Total	Hybrid
Total	907	168	310	298	1683	442
Safe	653	133	184	171	1141	336
Disruptive	254	35	126	127	542	106
Flat top disruption	62	23	33	33	151	
Ramp down disruption	192	12	93	94	391	

and avoidance measures [89]. It should also be mentioned that all the developed predictors implement various forms of open world learning [92, 93]. They can therefore start predicting with a minimum number of examples, in principle one disruptive and one safe discharge. Moreover, they adapt autonomously to the evolution of the experimental programme. Transfer from different machine has also been successfully achieved [94].

With regard to the development of predictors using a minimum of diagnostics, an important aspect is to determine the perdition time horizon of each signal. In this direction, a series of methods, based on the time series analysis of the main plasma diagnostic signals, are used to determine when significant changes in the plasma dynamics of the tokamak configuration occur, indicating the onset of drifts towards a disruption. Dynamical indicators, such as embedding dimension, 0–1 chaos test, recurrence plots measures, but also information theoretic criteria, such as information impulse function

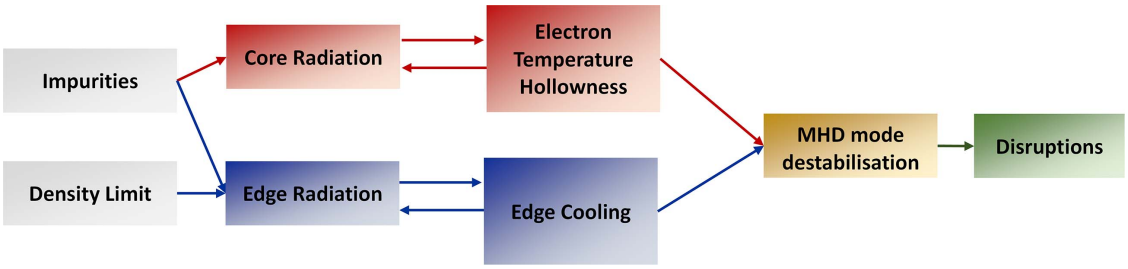

Figure 18. Main types of plasma evolution leading to a disruption.

quantifying information without entropy, have been tested to detect the time intervals when the plasma dynamics drifts towards situations that are likely to lead to disruptions [95]. The refined techniques have been tested with signals, such as the plasma current or the line integrated density (LID), which are expected to be available on the first day of operation of any major future device. The methods allow a good estimate of the intervals, in which the anomalous behaviors manifest themselves, which is very useful for building significantly more appropriate training sets for various kinds of disruption predictors. As they are based on completely different mathematical principles, they are providing robust information about these intervals. On JET the length of these intervals is of the order of 300 ms, more than sufficient to deploy mitigation techniques. Some of the developed methods may also be implemented themselves as stand-alone predictors for real time deployment, since they require computational times of the order of 1 ms.

Concerning the second topic mentioned above and related to the development of techniques to perform reliable disruption predictions for mitigation with a very limited number of signals, it is important to mention its vital importance in the beginning of the operation of new devices. For example, the JT-60SA Tokamak will start the operation with a reduced number of diagnostics [96]. In particular, its list of real-time signals connected to disruptions is very restricted and they show a limited capability to recognize incoming disruptions with enough anticipation time. In this respect, it is important to emphasize the lack of a real-time mode lock signal in JT-60SA, mainly when this signal is a typical reference to identify forthcoming disruptive events. However, a LID signal will be available in real-time and it has been recently shown its real-time capability to predict disruptions [97].

Due to the fact that JT-60SA will start the operation with a C-wall, disruption prediction with a LID signal has been accomplished with the JET DB and C-wall discharges. A total number of 1437 discharges have been analyzed (85 unintentional disruptions and 1352 non-disruptive shots). The LID signal chosen for this has been a vertical chord through the plasma center.

The dataset of 1437 discharges has been processed in chronological way as it would happen in the real production of discharges. The first 42 shots in the DB are non-disruptive and, therefore, a first predictor is obtained with 42 non-disruptive and 1 disruptive discharges. The predictor is based on a single signal: the LID signal. Because the simple amplitude is not

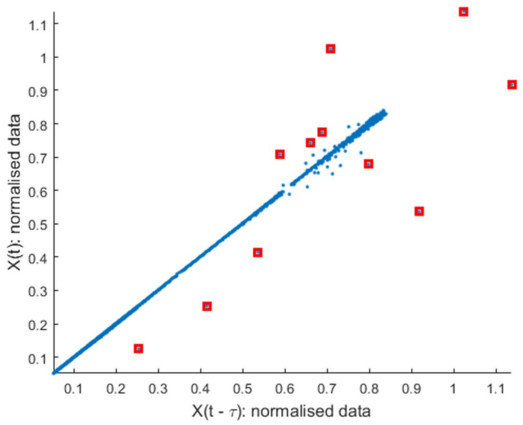

Figure 19. Disruptive and non-disruptive examples corresponding to 42 non-disruptive discharges and 1 disruptive discharge.

enough to distinguish between disruptive and non-disruptive behaviors, the space of consecutive samples is used [98]. This is a two-dimensional space in which the amplitudes of consecutive samples are related. If the temporal evolution of the signals is smooth, the points are concentrated around the diagonal. Abrupt changes in the amplitude reveal the presence of plasma instabilities (plasma anomalies, in general) and the points appear far enough from the diagonal.

Figure 19 shows the disruptive (red squares) and non-disruptive examples (blue points) to create a first predictor. This classifier is determined with Support Vector Machines (SVM) and a Radial Basis Function (RBF) kernel. By applying this predictor to the rest of discharges in the dataset, the results that are obtained appear in table 3 [99].

Table 3 shows that a LID signal can be used to predict disruptions. More than 95% of the disruptions are detected. However, it is important to mention that half of them are tardy alarms (they are detected after the disruption). On the other hand, the rate of false alarms is very high.

By following an adaptive approach, (i.e. the predictor is retrained after missed or tardy alarms), results are improved and are shown in table 4 [97]. In total, 10 retrainings have been carried out (1 after a missed alarm and 9 after tardy alarms). A comparison with table 3 reveals a better success rate and a drastic reduction of both false alarms and average warning time.

Results in table 4 support the fact that a single LID signal can be used to predict disruptions. In fact, the average

Table 3. Results with the training set of figure 19 and an SVM model with RBF kernel. It corresponds to 1394 discharges (84 disruptive shots and 1310 non-disruptive shots).

Success rate	Success rate with positive warning time	Tardy alarms	False alarms	Average warning time
95.24% (80/84)	45.24% (38/84)	50% (42/84)	11.53% (151/1310)	3.3 s

Table 4. Results after 10 retrainings. The successive models are obtained with SVM and RBF kernels.

Success rate	Success rate with positive warning time	Tardy alarms	False alarms	Average warning time
98.81% (83/84)	88.10% (74/84)	10.71% (9/84)	2.14% (28/1310)	0.558 s

Table 5. Results after eight re-trainings with the database of 1394 JET discharges.

Success rate	Success rate with positive warning time	Tardy alarms	False alarms	Average warning time
98.81% (83/84)	90.48% (76/84)	8.33% (7/84)	1.15% (15/1310)	0.112 s

warning time (hundreds of ms) is similar to other predictors, for example the APODIS predictor installed in the JET real-time network [100].

It is important to note the good results in terms of disruption prediction that are shown in table 4. They correspond to the use of a single signal (a LID signal) and without taking into account the mode lock that is the reference signal used for disruption prediction from years ago [101, 102]. A question that arises is: could the mode lock be used to improve the results of table 4? To answer this question, it should be noted again that the case that is being analyzed contemplates the fact that the mode lock signal is not available in real-time. In spite of this, could off-line versions of the mode lock be used to increase the good results of table 4? Vega *et al* have shown in [99] that this is possible by making use of the learning under privileged information (LUPI) paradigm [103]. The LUPI application to disruption prediction consists of using LID and mode lock data at training time to create a predictor but using only the real-time information of the LID signal to make predictions. Again, SVM with RBF kernels has been used for this purpose with the JET DB mentioned previously. Re-trainings have been carried out after tardy and missed alarms. Results are shown in table 5.

From table 5, it can be deduced that the application of the LUPI paradigm with the LID and the mode lock signals at training time and just the LID signal at execution time, improves the prediction results: better success rate with positive warning time, better tardy alarm rate, better false alarm rate and better warning times for mitigation purposes.

In addition to disruption prediction, the session devoted to 'Analysis of time series, images and video', provided cutting edge results in several aspects. Concerning the need of real-time analysis, Ruiz *et al* [104] summarized the implementation of two intelligent data processing applications in real-time heterogeneous platforms. The first one was related to discriminate between gamma and neutron pulses acquired using a scintillator through the use of deep learning techniques implemented with 1D convolutional neural networks (CNNs) and high-sampling rate analog to digital converters (ADCs) [105]. The second application [106] uses the 'connected components labelling' algorithm to detect hot spots in an image acquired with a high speed camera.

Also, there were contributions connected to the subject of event detections. In particular, a comparative study of methods and its application to edge-localized modes [107] was discussed. Moreover, a specific technique to find jumps in waveforms was presented [108]. In addition to these, there was a talk about the detection of thermal events by means of machine learning with application to the feedback control of thermal loads in Wendelstein 7-X [109].

Especially relevant was the presentation about predictive maintenance for both diagnostic and control subsystems [110]. Emphasis was given to failure prediction to prevent unexpected downtime. To this end, data-driven models based on past historical data need to be developed. Particular methods of machine learning techniques can help in the model development, for instance, non-supervised clustering, outlier recognition, change point detection or threshold techniques. Examples were given with the ohmic heating circuit at JET.

As it is well known, nonlinear coupling of modes is important to study instabilities. Bi-spectral analysis is the routine tool for these purposes. Zang presented an alternative method based on the Hilbert transform to track the phase evolution of a coherent mode [111]. The time resolution of this method was shown to be better than the corresponding to wavelet bi-spectral analysis. Results were presented with both HL-2A data and Heliotron J data.

Moreover, intermittent fluctuation studies of far-scrape-off layer time series measurements from gas puff imaging and mirror Langmuir probes at various line-averaged densities in Alcator C-Mod were shown [112]. The contribution reported

the scalings in the intermittency, flux and mean waiting times and mean amplitudes in the framework of stochastic modeling in Alcator C-Mod and how this changes with core plasma density [113].

2.6. Uncertainty propagation, verification and validation in modeling codes and data fusion (UNC)

The session on uncertainty propagation, verification and validation in modeling codes and data fusion was chaired by Pinches and co-chaired by Fujii and Yokoyama.

The presentation of Stefano Carli [114] shows how adjoint sensitivity analysis enables uncertainty quantification (UQ) for plasma edge codes. Plasma edge codes like SOLPS-ITER [115] are currently the main tools for interpreting exhaust scenarios in existing experiments, and for designing next step fusion reactors like ITER and DEMO. These codes typically couple a multi-fluid model for the plasma with a kinetic model for the neutral particles. While the former is implemented in a (deterministic) Finite Volume setting, the latter is solved with (stochastic) Monte Carlo (MC) methods, making simulations computationally expensive. Moreover, several sources of uncertainty are present throughout the complex simulation chain: starting from the magnetic equilibrium reconstruction, which forms the basis for the plasma mesh; going over unresolved physical phenomena requiring closure terms and related parameters, such as anomalous transport coefficients; and finally, a plethora of uncertain input parameters, such as atomic physics rates or boundary conditions. Therefore, UQ for model validation with plasma edge codes appears a challenging task, which is presently precluded by the high computational costs.

The adjoint sensitivity analysis is based on algorithmic differentiation (AD), which provides floating-point accurate sensitivities for complex codes in a semi-automatic way [116]. Such sensitivities are then fed to gradient-based optimization methods, which are employed to solve the backward UQ problem, also known as parameter estimation or model calibration. Casting this estimation into a Bayesian maximum a posterior (MAP) setting, one can consistently account for information and uncertainties from different diagnostics [117]. The main limitation of this framework is that currently only an approximate fluid neutral model can be employed. The more accurate MC model requires dealing with statistical noise in the sensitivity computation. It is shown that this can be accommodated in a discrete adjoint setting [118] and report on first results employing AD [119]. Finally, it is shown how a combination of finite differences and adjoint sensitivities, also known as an in-parts adjoint technique, allows sensitivity propagation throughout the whole simulation chain, including magnetic equilibrium reconstruction [120].

The mapping of experimental data to IMAS IDSs is discussed [121] as a prerequisite for the validation of tools both directly against experimental data and also in comparison with existing tools used on today's research facilities. The suite of tools being developed to support the preparations for ITER operation, including data interpretation and analysis, and the refinement of the ITER research plan, are underpinned by a common data representation that forms the basis around which the IMAS is built and which strives to make fusion data more FAIR. Adopting a common standard for the representation of data allows tools to be interoperable and for them to be tested and validated on present day experimental data, with the aim of accelerating the transition from initial testing to production-ready applications that would otherwise have to wait for the start of ITER operations and the production of ITER data. The data model itself is described by a data dictionary that follows a well-defined life-cycle and evolves in response to community needs, with most changes arising from its application to new use cases while improving data reusability. In this presentation, the mapping of experimental data to IMAS IDSs, both dynamic and static, as well as their accessibility, are discussed. Effort has already started on many devices to begin to map their experimental data into IDSs including AUG, DIII-D, EAST, JET, KSTAR, MAST-U, and TCV, whilst on WEST their plasma reconstruction chain [122] is now wholly based upon the IMAS data representation. In addition to the validation of software tools and workflows, the populated data structures can also be validated using a recently developed tool that forms part of the IDStools package. This uses extensible rules to validate against generic physics and data constraints, as well as use case-specific rules, e.g. for a particular device such as ITER or when developing a specific DB for specific events such as disruptions.

2.7. Advances in data science, probabilistic methods and machine learning (ADV)

The session on advances in data science, probabilistic methods and machine learning was chaired by Churchill and co-chaired by Verdoolaege.

Kwon and Lee [123] presented the Multi-Scale Recurrent Transformer system, a deep learning model for forecasting the temperature of superconducting coils, as it is important to predict the temperature of superconducting magnets to prevent excessive temperature rise of coils. The system recurrently predicts future temperature data of the superconducting coil using previous data obtained from a multi-scale KSTAR PF coil dataset and latent data calculated from previous time step. A multi-scale temperature subsampling approach is applied in the model to learn both the details and the overall structure of the temperature data effectively.

Liu [124] presented the data processing of radially distributed BES measurements related to the two-dimension mapping of the avalanche structure and its related impurity analysis in the HL-2A neutral beam heated H-mode plasmas. To gain deeper understanding of the generation of avalanche, the cross-correlation function (CCF) analysis has been performed to the radially distributed BES channels illuminating a radially elongated structure. In addition, it is demonstrated that the avalanche gets energy from and modulates ambient turbulence via nonlinear interaction from the bi-spectrum analysis of the density and magnetic fluctuation data. The impurity behavior during avalanche is also investigated by numerically simulating the impurity transport process. The impurity

data analysis suggests that the avalanche provide a transport channel for avoidance of heavy impurity accumulation.

Hall [125] investigated the issue of the strongly reduced scaling of the global energy confinement time with machine size in ELMy H-mode tokamak plasmas, when comparing the recent ITPA20 scalings with the IPB98(y,2) scaling law. One of the notably different dependencies lies in a considerably weaker scaling with the device's major radius, with an exponent α_R reduced from quadratic to almost linear. Using optimization and clustering techniques, a subset of the DB has been identified as exhibiting very weak size scaling (α_R $= 0.377$), hence maximally contributing to the reduced size dependence seen in the overall data set. This subset is localized in dimensionless space, governed by normalized gyroradius, collisionality and pressure, as confirmed by random forest classification. Interestingly, in this space, the operational point of future devices like ITER is situated in a region of higher size dependence (α_R $= 1.647$).

3. Round table discussion

A round-table discussion was held on the last day of the meeting, covering several topics that were treated or introduced at the workshop. The aim was to summarize the discussions that took place during the meeting and to stimulate interaction about future directions that the field should explore, driven by the most urgent needs as currently perceived in this domain. The discussion was chaired by Mazon.

It was stressed that the FDPVA community is on the frontier of fusion data interpretation, to help ensure that data are processed correctly, fully validated, and adequately interpreted and used. So far, the FDPVA technical meeting serves as an excellent platform to exchange scientific information within the FDPVA community. Nevertheless, propagation of knowledge, methods and best practices towards the whole fusion community should still be enhanced.

The participation of private companies in the meeting was strongly appreciated by many. It is a trend that is hoped to continue toward future meetings.

Furthermore, it was noted that the FDPVA community can learn from data-centric communities in other scientific fields, like astronomy, climatology, metrology, etc. This applies not only to the methods, but also to practices like data sharing. It was suggested to invite speakers from such communities at the next meeting in 2025.

One of the important messages emerging from the discussion was the need for benchmarking and standardization of methods and tools. This was already stressed during the previous FDPVA meeting in 2021, notably for the case of anomaly detection. However, the same is true for other methods and tools, like equilibrium reconstruction, profile or tomographic reconstruction, synthetic diagnostics, modeling codes, etc.

Strongly linked to the need for benchmarking is the availability of standard (multi-machine) DBs that are open to the community, containing both synthetic and experimental data. This is important for data sharing, reproducibility of experiments and analyses, comparison and benchmarking across devices, validation of theory or modeling efforts, scaling and extrapolation to new devices, etc. Even though various dedicated DBs exist at the level of institutes or working groups, the number of publicly available multi-machine DBs (e.g. ITPA, EUROfusion) remains limited. Nevertheless, it is customary that larger DBs start from relatively modest projects, although it is recommended that, from the start, the DB is designed according to good practices. The IMAS framework and its data dictionary could be a starting point for this, since it can provide an interface for different machines. IMAS can also provide guidance for names, data structures (including uncertainties), standards for data quality, etc. Another example is the FAIR4FUSION project, which has the goal to make fusion data from multiple devices more widely available to the fusion community, or to other communities, funding bodies and the broader public. The next edition of the meeting should devote a good deal of attention to issues of data sharing and interoperability.

Also noted at the 2021 meeting was the potential for application of anomaly detection to a wide range of events in the plasma and various machine systems and components. Although one of the primary applications remains disruption prediction, the presentation at the 2023 conference of a broader variety of applications of event detection can certainly be seen as a positive evolution. Similarly, the need for demonstration of (real-time) applications that go beyond the proof-of-concept stage is increasingly addressed.

A special issue, featuring selected papers from this technical meeting, has been published in Plasma Physics and Controlled Fusion.

4. Conclusion

The 5th IAEA technical meeting on FDPVA (June 12–15, 2023, Ghent University, Ghent, Belgium), was very successful. With the growth of massive measurement systems and data for future fusion reactors, data analysis is moving towards faster, more systematic and smarter directions. The recent highlights and progress in the above topics have been briefly summarized in this report.

The next meeting, i.e. the 6th IAEA technical meeting on FDPVA, will take place in 2025.

Acknowledgments

The views and opinions expressed herein do not necessarily reflect those of the ITER Organization. The authors would like to thank all the topical chairs, co-chairs and Programme Committee members of 5th IAEA Technical Meeting on Fusion Data Processing, Validation and Analysis for their dedicated efforts to organize this meeting, and thank all the colleagues who presented their excellent work in this meeting. In particular, we would like to thank IAEA, and the staff and students at Ghent University who provided excellent services that made this very successful and memorable event possible. Last but not least, we would like to thank Dr Cong Meng at

Southwestern Institute of Physics who helped to prepare this summary.

ORCID iDs

M. Xu 0009-0001-3059-7026
D. Mazon 0000-0001-5560-2277
M. Barbarino 0000-0003-3066-8952
R.M. Churchill 0000-0001-5711-746X
R. Fischer 0009-0000-6205-4731
K. Fujii 0000-0003-0390-9984
P. Jain 0000-0003-3427-2090
A. Murari 0000-0002-3932-3865
S.D. Pinches 0000-0003-0132-945X
P. Rodriguez-Fernandez 0000-0002-7361-1131
J. Stillerman 0000-0003-0901-0806
J. Vega 0000-0002-1622-3984
G. Verdoolaege 0000-0002-2640-4527
M. Yokoyama 0000-0001-8856-1483
P. Abreu 0000-0003-2113-7922
S. Ahmed 0000-0002-7732-5765
J. Alhage 0009-0003-5702-939X
F. Almuhisen 0000-0002-9160-8200
M. Bergmann 0009-0005-6052-6420
D. Pereira Botelho 0009-0000-5889-1924
L. Caputo 0009-0007-7441-5670
S. Carli 0000-0001-6581-2745
R. Castro 0000-0003-4199-9335
T. Craciunescu 0000-0002-0012-4260
F. Deeba 0009-0004-8241-0237
F. Esquembre 0000-0003-4550-0183
K. Giil 0000-0003-1015-5228
Y. Gu 0000-0001-9956-388X
J. Hollocombe 0000-0001-7159-158X
X. Huang 0000-0002-7736-7594
A. Jardin 0000-0003-4910-1470
R. Jorge 0000-0003-2941-6571
Y. Li 0009-0002-4799-8456
E. Peluso 0000-0002-6829-2180
R. Rossi 0000-0003-4414-6119
M. Ruiz 0000-0002-1337-0110
J. De Rycke 0000-0002-4593-1680
M. Schneider 0000-0003-4881-6861
M. Sertoli 0000-0003-1528-6307
D. Stieglitz 0000-0002-4541-320X
Y. Tan 0000-0002-1139-4303
H. Weisen 0000-0001-6211-8096
H. Wu 0009-0004-8249-4477
I. Wyss 0000-0001-7625-8401
L. Zang 0000-0001-9533-6057

References

[1] Gonzalez de Vicente S.M. *et al* 2023 Summary report of the 4th IAEA Technical Meeting on Fusion Data Processing, Validation and Analysis (FDPVA) *Nucl. Fusion* **63** 047001
[2] Mazon D. *et al* 2020 Summary report of the 3rd IAEA technical meeting on fusion data processing validation and analysis (FDPVA) *Nucl. Fusion* **60** 097002
[3] Jorge R. The direct optimization framework in stellarator design: transport and turbulence optimization (available at: https://conferences.iaea.org/event/346/contributions/27298/)
[4] NEAT, publicly available on GitHub (available at: https://github.com/rogeriojorge/NEAT)
[5] Jorge R. *et al* 2023 (arXiv:2301.09356 [physics.plasm-ph])
[6] Jorge R., Goodman A., Landreman M., Rodrigues J. and Wechsung F. 2023 Single-stage stellarator optimization: combining coils with fixed boundary equilibria *Plasma Phys. Control. Fusion* **65** 074003
[7] Rodriguez-Fernandez P. Bayesian optimization techniques to accelerate burning-plasma and reactor simulations *5th Technical Meeting on Fusion Data Processing, Validation and Analysis,* (*Ghent, Belgium*) (available at: https://conferences.iaea.org/event/346/contributions/27297/)
[8] Rodriguez-Fernandez P., Howard N.T. and Candy J. 2022 Nonlinear gyrokinetic predictions of SPARC burning plasma profiles enabled by surrogate modeling *Nucl. Fusion* **62** 076036
[9] Candy J., Belli E.A. and Bravenec R.V. 2016 A high-accuracy Eulerian gyrokinetic solver for collisional plasmas *J. Comput. Phys.* **324** 73–93
[10] Belli E.A. and Candy J. 2008 Kinetic calculation of neoclassical transport including self-consistent electron and impurity dynamics *Plasma Phys. Control. Fusion* **50** 095010
[11] Howard N.T. *et al* 2024 Simultaneous reproduction of experimental profiles, fluxes, transport coefficients, and turbulence characteristics via nonlinear gyrokinetic profile predictions in a DIII-D ITER similar shape plasma *Phys. Plasmas* **31** 032501
[12] Rodriguez-Fernandez P., Howard N.T., Saltzman A., Shoji L., Body T., Battaglia D.J., Hughes J.W., Candy J., Staebler G.M. and Creely A.J. 2024 Core performance predictions in projected SPARC first-campaign plasmas with nonlinear CGYRO *Phys. Plasmas* **31** 062501
[13] Howard N.T., Rodriguez-Fernandez P., Holland C. and Candy J. 2025 Prediction of performance and turbulence in ITER burning plasmas via nonlinear gyrokinetic profile prediction *Nucl. Fusion* **65** 016002
[14] Rodriguez-Fernandez P., Howard N.T., Saltzman A., Kantamneni S., Candy J., Holland C., Balandat M., Ament S. and White A.E. 2024 Enhancing predictive capabilities in fusion burning plasmas through surrogate-based optimization in core transport solvers *Nucl. Fusion* **64** 076034
[15] Creely A.J. *et al* 2020 Overview of the SPARC tokamak *J. Plasma Phys.* **86** 865860502
[16] Rodriguez-Fernandez P. *et al* 2022 Overview of the SPARC physics basis towards the exploration of burning-plasma regimes in high-field, compact tokamaks *Nucl. Fusion* **62** 042003
[17] Pereira Botelho D. *et al* 2024 arXiv:2409.20143 Simplified magnet design and manufacture based on patterning of wide conductors
[18] Jelonnek J. *et al* 2017 Design considerations for future DEMO gyrotrons: a review on related gyrotron activities within EUROfusion *Fusion Eng. Des.* **123** 241–6
[19] Landreman M. 2017 An improved current potential method for fast computation of stellarator coil shapes *Nucl. Fusion* **57** 046003
[20] Zhao Y. Development of high-speed data acquisition system of negative ion source breakdown (available at: https://conferences.iaea.org/event/346/contributions/27300/)

[21] McNamara S.A.M. *et al* 2023 Achievement of ion temperatures in excess of 100 million degrees Kelvin in the compact high-field spherical tokamak ST40 *Nucl. Fusion* **63** 054002

[22] Sertoli M. From MVPs to full models: a stepwise development of diagnostic forward models in constant support of diagnostic design, data analysis, instrument consistency and discharge modelling on the ST40 tokamak (available at: https://conferences.iaea.org/event/346/contributions/27304/)

[23] OPEN-ADAS, Atomic Data and Analysis Structure (available at: https://open.adas.ac.uk/)

[24] Marchuk O. *et al* 2006 Comparison of impurity transport model with measurements of He-like spectra of argon at the tokamak TEXTOR *Plasma Phys. Control. Fusion* **48** 1633

[25] Wood J., Lomanowski B., Delabie E., Willett H.V., Sertoli M. and Varje J. 2023 Characterisation of ion temperature and toroidal rotation on the ST40 tokamak *J. Instrum.* **18** C03019

[26] Morita S. 1994 IPP III/1999 (available at: https://core.ac.uk/download/pdf/210800245.pdf)

[27] Kaye S.M. *et al* 2022 *APS Conf.* CP11.00016

[28] Thomas P.R. *et al* 2022 *APS Conf.* YI02.00005

[29] Sertoli M. *et al* 2022 *APS Conf.* CP11.00014

[30] Huang X. X-ray data validation and analysis on the EXL-50 spherical torus (available at: https://conferences.iaea.org/event/346/contributions/27312/)

[31] Huang X.L. *et al* 2021 Toroidal soft x-ray array on the EXL-50 spherical tokamak *Rev. Sci. Instrum.* **92** 053501

[32] Cheng S.K., Zhu Y.B., Chen Z.Y., Li Y.X., Bai R.H., Chen B., Huang X.L., Dai L.L. and Liu M.S. 2021 Tangential hard x-ray diagnostic array on the EXL-50 spherical tokamak *Rev. Sci. Instrum.* **92** 043513

[33] Banerjee D. *et al* 2022 Investigation of the effectiveness of 'multi-harmonic' electron cyclotron current drive in the non inductive EXL-50 ST *J. Phys.: Conf. Ser.* **2397** 012011

[34] McIntosh S. Constrained feed-forward waveforms for tokamak plasma pulse design (available at: https://conferences.iaea.org/event/346/contributions/27305/)

[35] Yang L. Static performance prediction of long-pulse negative ion based neutral beam injection experiment (available at: https://conferences.iaea.org/event/346/contributions/27305/)

[36] Zhang L., Wang F., Xu B., Chi W., Wang Q. and Sun T. 2018 Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI *Neural Comput. Appl.* **30** 1425–44

[37] Castro R. MINT, ITER interactive data visualization tool (available at: https://conferences.iaea.org/event/346/contributions/27308/)

[38] Stillerman J. Metadata framework for distributed real-time control systems (available at: https://conferences.iaea.org/event/346/contributions/27307/)

[39] Paulo Abreu End-to-end intra-pulse data analysis at ITER: first steps from magnetics to live display (available at: https://conferences.iaea.org/event/346/contributions/27309/)

[40] Weisen H., Sirén P., Varje J. and Ghani Z. 2024 *Fusion Sci. Technol.* **81** 244–58

[41] Gu Y., Hu C., Li Y., Zhao Y., Cui Q. and Xie Y. Design concept of intelligent integrated control system for neutral beam injection (available at: https://conferences.iaea.org/event/346/contributions/27311/)

[42] Almuhisen F. Optimizing tokamak operations using machine learning methods as a service (available at: https://cea.hal.science/cea-04273595/document)

[43] Esquembre F. IODA: a new federated web platform for collaboration and sharing of data analysis resources in Fusion Data Research (available at: https://conferences.iaea.org/event/346/contributions/27330/)

[44] Esquembre F., Chacón J., Saenz J., Vega J. and Dormido-Canto S. 2023 *Fusion Eng. Des.* **197** 114049

[45] Hollocombe J. IMAS simulation management and remote data access for ITER (available at: https://conferences.iaea.org/event/346/contributions/27329/)

[46] Imbeaux F. *et al* 2015 Design and first applications of the ITER integrated modelling & analysis suite *Nucl. Fusion* **55** 123006

[47] Gu Y. The design of NBI experimental data processing system (available at: https://conferences.iaea.org/event/346/contributions/27331/)

[48] Gu Y., Fan M., Zhao Y. and Zhang X. 2021 The upgrade of data processing and storage system for EAST NBI *Fusion Eng. Des.* **173** 112849

[49] Fischer R. Status and prospects of integrated data analysis for present and future fusion devices (available at: https://conferences.iaea.org/event/346/contributions/27301/)

[50] Stieglitz D. Validation of diagnostics for kinetic profiles at ASDEX Upgrade using integrated data analysis (IDA) (available at: https://conferences.iaea.org/event/346/contributions/27319/)

[51] Stieglitz D.J., Santos J. and Fischer R. 2023 Implementation and validation of swept density reflectometry for integrated data analysis at ASDEX Upgrade *Rev. Sci. Instrum.* **94** 043503

[52] Wu H., Jardin A., Mazon D. and Verdoolaege G. 2024 Estimation of the radial tungsten concentration profiles from soft x-ray measurements at WEST with Bayesian integrated data analysis *J. Fusion Energy* **43**

[53] Li D., Svensson J., Thomsen H., Medina F., Werner A. and Wolf R. 2013 Bayesian soft X-ray tomography using non-stationary Gaussian Processes *Rev. Sci. Instrum.* **84** 083506

[54] Wang T., Mazon D., Svensson J., Li D., Jardin A. and Verdoolaege G. 2018 Incorporating magnetic equilibrium information in Gaussian process tomography for soft X-ray spectroscopy at WEST *Rev. Sci. Instrum.* **89** 10F103

[55] Moser K., Bock A., David P., Bernert M. and Fischer R. (ASDEX Upgrade Team) 2022 Gaussian process tomography at ASDEX Upgrade with magnetic equilibrium information and nonstationary kernels *Fusion Sci. Technol.* **78** 607–16

[56] Fischer R., Fuchs C.J., Kurzan B., Suttrop W. and Wolfrum E. 2010 Integrated data analysis of profile diagnostics at ASDEX Upgrade *Fusion Sci. Technol.* **58** 675–84

[57] Churchill R. Simulation-based inference with optical diagnostics (available at: https://conferences.iaea.org/event/346/contributions/27302/)

[58] Bergmann M. Integrated data analysis augmented by kinetic modelling (available at: https://conferences.iaea.org/event/346/contributions/27317/)

[59] Pereverzev G. and Yushmanov P.N. 2002 *ASTRA Automated System for TRansport Analysis in a Tokamak* (Max-Planck-Institut für Plasmaphysik, Garching, IPP 5/98)

[60] Staebler G.M., Kinsey J.E. and Waltz R.E. 2007 A theory-based transport model with comprehensive physics *Phys. Plasmas* **14** 055909

[61] Weiland M., Bilato R., Dux R., Geiger B., Lebschy A., Felici F., Fischer R., Rittich D. and van Zeeland M. 2018 Rabbit: real-time simulation of the NBI fast-ion distribution *Nucl. Fusion* **58** 082032

[62] Poli E. *et al* 2018 TORBEAM 2.0, a paraxial beam tracing code for electron-cyclotron beams in fusion plasmas for extended physics applications *Comput. Phys. Commun.* **225** 36–46

[63] Houlberg W.A., Shaing K.C., Hirshman S.P. and Zarnstorff M.C. 1997 Bootstrap current and neoclassical transport in tokamaks of arbitrary collisionality and aspect ratio *Phys. Plasmas* **4** 3230–42

[64] Schneider M. On the use of synthetic diagnostics as persistent actors in integrated modelling workflows (available at: https://conferences.iaea.org/event/346/contributions/27318/)

[65] Snipes J.A. *et al* 2021 ITER plasma control system final design and preparation for first plasma *Nucl. Fusion* **61** 106036

[66] Walker M. *et al* 2014 A simulation environment for ITER PCS development *Fusion Eng. Des.* **96** 716

[67] Veen L.E. and Hoekstra A.G. Easing multiscale model design and coupling with MUSCLE 3 *Comp. Science—ICCS 2020* vol 12142 (Springer) pp 425–38

[68] De Rycke J. A Bayesian approach for estimating the kinematic viscosity model in reversed-field pinch fusion plasmas (available at: https://conferences.iaea.org/event/346/contributions/27316/)

[69] Montgomery D. 1992 Resistivity profile and instability of the plane sheet pinch *Plasma Phys. Control* **34** 41157

[70] Morey R.D., Romeijn J.-W. and Rouder J.N. 2016 The philosophy of Bayes factors and the quantification of statistical evidence *J. Math. Psychol.* **72** 6

[71] Vivenzi N. *et al* 2022 Kinematic viscosity estimates in reversed-field pinch fusion plasmas *J. Phys.: Conf. Ser.* **2397** 012010

[72] Bunge M. 2019 Inverse Problems *Found. Sci.* **24** 483–525

[73] Odstrcil M., Mlynar J., Odstrcil T., Alper B. and Murari A. 2012 Modern numerical methods for plasma tomography optimisation *Nucl. Instrum. Methods Res. A* **686** 156–61

[74] Jardin A., Mazon D., Bielecki J., Dworak D., Guibert D., Król K., Savoye-Peysson Y., Scholz M. and Walkowiak J. 2024 Validating and speeding up x-ray tomographic inversions in tokamak plasmas *Plasma Phys. Control. Fusion* **66** 085010

[75] Craciunescu T., Bonheure G., Kiptily V., Murari A., Soare S., Tiseanu I. and Zoita V. 2008 The maximum likelihood reconstruction method for JET neutron tomography *Nucl. Instrum. Methods Phys. Res. A* **595** 623–30

[76] Craciunescu T., Bonheure G., Kiptily V., Murari A., Tiseanu I. and Zoita V. 2009 A comparison of four reconstruction methods for JET neutron and gamma tomography *Nucl. Instrum. Methods Phys. Res. A* **605** 374–83

[77] Craciunescu T., Peluso E., Murari A. and Gelfusa M. 2018 Maximum likelihood bolometric tomography for the determination of the uncertainties in the radiation emission on JET TOKAMAK *Rev. Sci. Instrum.* **89** 053504

[78] Murari A., Peluso E., Craciunescu T., Lowry C., Aleiferis S., Carvalho P. and Gelfusa M. 2020 Investigating the thermal stability of highly radiative discharges on JET with a new tomographic method *Nucl. Fusion* **60** 046030

[79] Peluso E., Craciunescu T., Murari A., Carvalho P. and Gelfusa M. 2019 A comprehensive study of the uncertainties in bolometric tomography on JET using the maximum likelihood method *Rev. Sci. Instrum.* **90** 123502

[80] Peluso E. *et al* 2019 On the effects of missing chords and systematic errors on a new tomographic method for JET bolometry *Fusion Eng. Des.* **146** 2124–9

[81] Gelfusa M., Craciunescu T., Peluso E., Giacomelli L., Kiptily V., Reux C., Szepesi G. and Murari A. (JET Contributors) 2021 A maximum likelihood tomographic method applied to JET gamma ray emission during the current quench *Fusion Eng. Des.* **168** 112637

[82] Peluso E. *et al* 2022 Dealing with artefacts in JET iterative bolometric tomography using masks *Plasma Phys. Control. Fusion* **64** 045013

[83] Peluso E., Murari A., Craciunescu T., Carvalho P., Gelfusa M., Gaudio P. and Wyss I. Correction of JET bolometric maximum likelihood tomography for local gas puffing 2023 *Plasma Phys. Control. Fusion* **65** 075003

[84] Ruiz M., Nieto J., Costa V., Craciunescu T., Peluso E., Vega J. and Murari A. 2022 Acceleration of an algorithm based on the maximum likelihood bolometric tomography for the determination of uncertainties in the radiation emission on JET using heterogeneous platforms *Appl. Sci.* **12** 6798

[85] Rossi R. *et al* 2023 A systematic investigation of radiation collapse for disruption avoidance and prevention on JET tokamak *Matter Radiat. Extremes* **8** 046903

[86] Wyss I., Murari A., Spolladore L., Peluso E., Gelfusa M., Gaudio P. and Rossi R. (on-behalf-of-JET-contributors) 2023 Comparison of a fast low spatial resolution inversion method and peaking factors for the detection of anomalous radiation patterns and disruption prediction *Fusion Eng. Des.* **193** 113625

[87] Boozer A.H. 2012 Theory of tokamak disruptions *Phys. Plasmas* **19** 058101

[88] Vega J. *et al* 2022 Disruption prediction with artificial intelligence techniques in tokamak plasmas *Nat. Phys.* **18** 741–50

[89] Wenninger R. *et al* Power handling and plasma protection aspects that affect the design of the DEMO divertor and first wall *Proc. 2016 26th IAEA Fusion Energy Conf. (Kyoto, Japan, 17–22 October 2016)* FIP /P7-14 (available at: https://nucleus.iaea.org/sites/fusionportal/Shared%20Documents/FEC%202016/fec2016-preprints/preprint0322.pdf)

[90] Murari A., Peluso E., Lungaroni M., Rossi R. and Gelfusa M. 2020 Investigating the physics of tokamak global stability with interpretable machine learning tools *Appl. Sci.* **10** 6683

[91] Rossi R. *et al* 2023 A systematic investigation of radiation collapse for disruption avoidance and prevention on JET tokamak *Matter Radiat. Extrem.* **8** 046903

[92] Murari A., Lungaroni M., Gelfusa M., Peluso E. and Vega J. (JET Contributors) 2019 Adaptive learning for disruption prediction in non-stationary conditions *Nucl. Fusion* **59** 086037

[93] Murari A., Rossi R., Lungaroni M., Baruzzo M. and Gelfusa M. 2021 Stacking of predictors for the automatic classification of disruption types to optimize the control logic *Nucl. Fusion* **61** 036027

[94] Murari A., Rossi R., Peluso E., Lungaroni M., Gaudio P., Gelfusa M., Ratta G. and Vega J. 2020 On the transfer of adaptive predictors between different devices for both mitigation and prevention of disruptions *Nucl. Fusion* **60** 056003

[95] Craciunescu T. and Murari A. 2023 Detection of changes in the dynamics of thermonuclear plasmas to improve the prediction of disruptions *Nonlinear Dyn.* **111** 1–15

[96] Kamada Y., Di Pietro E., Hanada M., Barabaschi P., Ide S., Davis S., Yoshida M., Giruzzi G. and Sozzi C. the JT-60SA Integrated Project Team 2022 (JT-60SA Integrated Project Team) *Nucl. Fusion* **62** 042002

[97] Vega J., Dormido-Canto S., Ramírez R., Farias G., Murari A. and Gadariya D. (JET Contributors) Real-time disruption prediction in multi-dimensional spaces with privileged information not available at execution time *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27320/)

[98] Vega J., Murari A., Dormido-Canto S., Hernández F., Cruz T., Gadariya D. and Rattá G.A. 2020 A linear equation based on signal increments to predict disruptive behaviours and the time to disruption on JET *Nucl. Fusion* **60** 026001
[99] Vega J., Dormido-Canto S., Castro R., Fernández J.D. and Murari A. 2024 Real-time disruption prediction in multi-dimensional spaces leveraging diagnostic information not available at execution time *Nucl. Fusion* **64** 046010
[100] Vega J., Dormido-Canto S., López J.M., Murari A., Ramírez J.M., Moreno R., Ruiz M., Alves D. and Felton R. (JET-EFDA Contributors) 2013 Results of the JET real-time disruption predictor in the ITER-like wall campaigns *Fusion Eng. Des.* **88** 1228–31
[101] Snipes J.A., Campbell D.J., Haynes P.S., Hender T.C., Hugon M., Lomas P.J., Lopes Cardozo N.J., Nave M.F.F. and Schüller F.C. 1988 Large amplitude quasi-stationary MHD modes in JET *Nucl. Fusion* **28** 1085–97
[102] Nave M.F.F. and Wesson J.A. 1990 Mode locking in tokamaks *Nucl. Fusion* **30** 2575–83
[103] Vapnik V. and Vashist A. 2009 A new learning paradigm: learning using privileged information *Neural Net.* **22** 544–57
[104] Ruiz M. *et al* Real-time implementation of intelligent data processing applications: gamma/neutron discrimination and hot spot identification *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27334/)
[105] Astrain M. *et al* 2021 Real-time implementation of the neutron/gamma discrimination in an FPGA-based DAQ MTCA platform using a convolutional neural network *IEEE Trans. Nucl. Sci.* **68** 2173–8
[106] Esquembri S., Nieto J., Carpeno A., Ruiz M., Astrain M., Costa V. and de Gracia A. 2021 Application of heterogeneous computing techniques for the development of an image-based hot spot detection system using MTCA *IEEE Trans. Nucl. Sci.* **68** 2151–8
[107] Alhage J. *et al* A comparative study of event detection methods in fusion devices with an application to edge-localized modes *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27313/)
[108] Tan Y. *et al* 2023 A novel method to find jumps in waveforms *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27322/)
[109] Puig A. *et al* Detection of thermal events using machine learning for the feedback control of thermal loads in Wendelstein 7-X *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27335/)
[110] Caputo L. *et al* Predictive maintenance in fusion devices with an application to the ohmic heating circuit at JET *5th TM on Fusion Data Processing, Validation and Analysis* ((*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27324/)
[111] Zang L. Phase tracking with Hilbert transform and nonlinear wave-wave coupling analysis on HL-2A and Heliotron J *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27336/)
[112] Ahmed S. *et al* Analysis of intermittent data time series from the far scrape-off layer in Alcator C-Mod at high Greenwald fractions *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27315/)
[113] Ahmed S., Erik Garcia O., Kuang A.Q., LaBombard B., Terry L.J. and Theodorsen A. 2023 Strongly intermittent far scrape-off layer fluctuations in Alcator C-Mod plasmas close to the empirical discharge density limit *Plasma Phys. Control. Fusion* **65** 105008
[114] Carli S. Sensitivity-based uncertainty quantification for plasma edge codes: status and challenges (available at: https://conferences.iaea.org/event/346/contributions/27317/)
[115] Wiesen S. *et al* 2015 The new SOLPS-ITER code package *J. Nucl. Mater.* **463** 480–4
[116] Carli S., Hascoët L., Dekeyser W. and Blommaert M. 2023 Algorithmic Differentiation for adjoint sensitivity calculation in plasma edge codes *J. Comput. Phys.* **491** 112403
[117] Carli S., Dekeyser W., Blommaert M., Coosemans R., Van Uytven W. and Baelmans M. 2021 Bayesian maximum a posteriori-estimation of κ turbulence model parameters using algorithmic differentiation in SOLPS-ITER *Contrib. Plasma Phys.* **62** e202100184
[118] Dekeyser W., Blommaert M., Ghoos K., Horsten N., Boerner P., Samaey G. and Baelmans M. 2018 Divertor design through adjoint approaches and efficient code simulation strategies *Contrib. Plasma Phys.* **58** 643–51
[119] Horsten N., Carli S. and Dekeyser W. 2024 Sensitivity calculation for Monte Carlo particle simulations of neutrals in the plasma edge *Contrib. Plasma Phys.* **64** e202300138
[120] Blommaert M., Reiter D. and Baelmans M. 2017 An efficient methodology to analyze plasma edge model parameter sensitivities *Nucl. Mater. Energy* **12** 1049–54
[121] Pinches S. Adoption and validation of IMAS data (available at: https://conferences.iaea.org/event/346/contributions/27332/)
[122] Fluery L. *et al* WEST plasma reconstruction chain and IMAS related tools *SOFT 2020*
[123] Kwon G. and Lee H. 2024 Multi-scale recurrent transformer model for predicting KSTAR PF superconducting coil temperature *Plasma Phys. Control. Fusion* **66** 055011
[124] Liu Y. Data analysis of quasi-two-dimensional nonlinear interactions in avalanche like phenomena in HL-2A plasmas *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27326/)
[125] Hall J. Confinement scaling with machine size in the updated ITPA global H-Mode confinement database *5th TM on Fusion Data Processing, Validation and Analysis* (*Ghent, Belgium, 12–15 June 2023*) (Ghent University) (available at: https://conferences.iaea.org/event/346/contributions/27325/)