CHAPTER 11
DATA VALIDATION, ANALYSIS, AND APPLICATIONS FOR FUSION PLASMAS

S. A. ARSHAD,[a,b]* J. G. CORDEY,[b] D. C. McDONALD,[b] J. FARTHING,[b] E. JOFFRIN,[c]
M. VON HELLERMANN,[d] C. M. ROACH,[b] and J. SVENSSON[e]

[a]*EFDA-CSU, Culham Science Centre, Abingdon, Oxfordshire OX14 3DB, United Kingdom*
[b]*EURATOM/UKAEA Fusion Association, Culham Science Centre, Abingdon, Oxfordshire OX14 3DB, United Kingdom*
[c]*Association EURATOM/CEA sur la fusion, F-13108 St. Paul Lez Durance, France*
[d]*FOM Instituut voor Plasmafysica, EURATOM Association, TEC, Postbus 1207, NL-3430 BE, Nieuwegein, The Netherlands*
[e]*IPP Greifswald, Max-Planck-Institut für Plasmaphysik, Wendelsteinstr. 1, D-17491 Greifswald, Germany*

Received September 7, 2006
Accepted for Publication November 3, 2007

Principal techniques and trends in the validation and analysis of data in magnetic fusion research are described and examples of applications are given. Well-established methods to obtain key physical quantities are outlined, as well as newer techniques employing integrated analysis of multiple diagnostics to improve quality and extract additional information from the data. Plasma control, confinement scaling, and transport studies, including model validation and development, are presented as important examples of applications of validated data. Finally, aspects essential to successful operation of future devices, which bring challenges due to a harsher environment for diagnostics, increased real-time requirements, and a geographically more distributed user community, are highlighted.

KEYWORDS: *data validation and analysis, plasma control, confinement scaling and transport modeling*

Contents—Chapter 11

*E-mail: shakeib.arshad@jet.efda.org

I. INTRODUCTION

In magnetic fusion experiments large amounts of data need to be handled in order to interpret experiments and to operate the experimental device in a safe and flexible way. In order to ensure that the quality of data remains high at all times, it is becoming common practice to implement formal procedures for data validation. These procedures are aimed at early detection of erroneous data, e.g., due to a hardware fault, and checking whether required physical parameters can be determined with adequate accuracy using the available information. Principle uses of validated data include plasma control and physics

studies, which include the development and validation of models of the plasma.

Data-validation procedures range from relatively simple consistency checks on raw data through to searching for anomalies between sophisticated data analyses and models of the physical processes in the experiment. At the basic level, data validation starts by comparing individual channels of raw data within each diagnostic system. In the next stage, consistency between different diagnostics is checked, e.g., by comparing the same physical parameters obtained from different diagnostics. At the more sophisticated level, input parameters to models of physical processes in the experiment are obtained from the measurements and the outputs of the models are compared with observations to check for consistency.

In this chapter the main procedures typically involved in data validation are described. Owing to the background of the authors, European machines (and JET in particular) are used as examples, since these are most familiar to the authors. In addition, the review focuses on the tokamak device. However, the procedures described are representative of the approach to data validation across all mainstream magnetic fusion research, including stellerators and reverse-field pinches.

Data analysis in magnetic fusion research is performed using one of two alternative approaches. The first, and most widely used, approach is sequential data analysis, in which physical parameters are calculated from data provided by individual diagnostics and then combined to yield a unified model of the plasma. A second, relatively new, approach employs integrated data analysis (IDA), in which a single model of the plasma is fitted to all of the available diagnostic data. In this chapter both of these analysis techniques will be discussed separately.

In Sec. II an overview of data validation for the sequential data analysis approach is given, covering the motivation, process, and infrastructure requirements. The processes described in Sec. II relate to raw and analyzed diagnostic data, and the comparison of these data with predictions of models. This is indicated in Fig. 1, a diagram of the typical data flow sequence in magnetic fusion research.

Section III provides examples of how principal physical parameters are derived from diagnostic data, for use in higher-level validation tests. It should be noted that Sec. III is not intended to be a comprehensive survey of analysis in magnetic fusion, which is beyond the scope of this chapter. However, a number of physical parameters of importance in fusion research appear in the text; a description of these is provided in Table I.

In Sec. IV data analysis and validation in terms of IDA, and in particular using Bayesian techniques, are presented. Modeling and data analysis are combined in this approach, encompassing a large part of the data flow sequence, as indicated in Fig. 1.

In Sec. V some important uses of validated data are presented, with discussions of plasma control and physics studies. The discussion on plasma control touches on some aspects of real-time validation, a crucial ingredient of safe and efficient operation of the experimental device. A wide range of physics studies make use of diagnostic data in magnetic fusion research. Two particularly

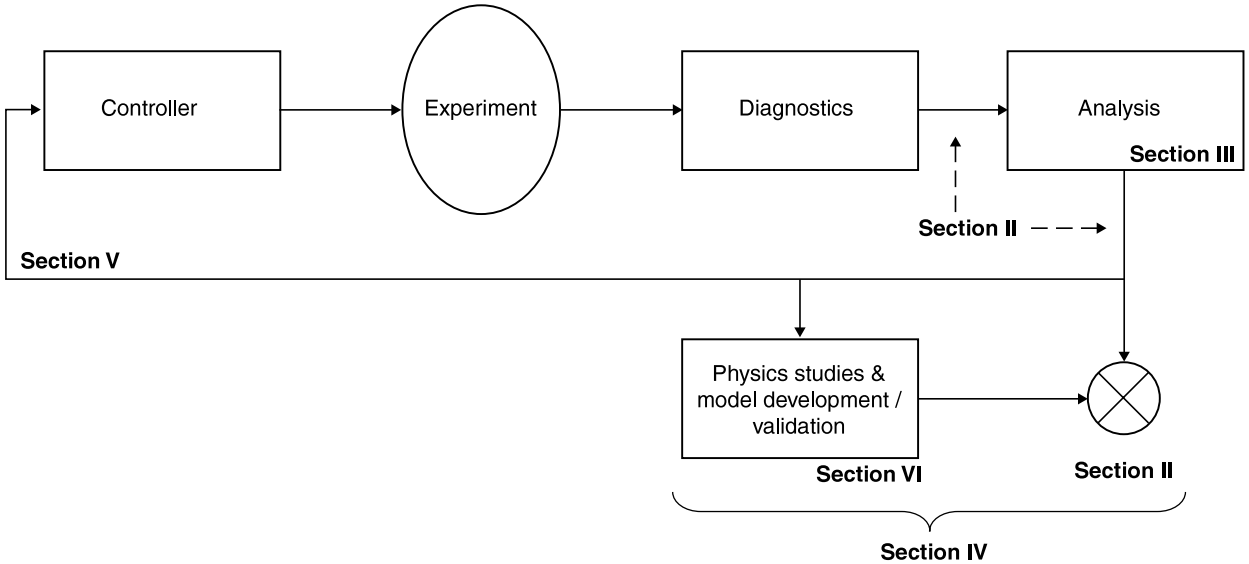

Fig. 1. Typical data flow sequence in magnetic fusion research. The circle containing a cross represents comparison of model outputs with analyzed diagnostic data. Sections of this chapter, relevant to different parts of the flow sequence, are marked in bold.

TABLE I

Steps in the Determination of Key Physical Parameters

Description	Determined Physical Parameter	Required Calculations and Measurements
	All	Derivation of magnetic flux surfaces
Plasma kinetic energy ($W_{thermal}$)	Electron and impurity densities $n_e(\rho)$, $n_z(\rho)$ Electron and ion temperatures $T_e(\rho)$, $T_i(\rho)$ Toroidal angular frequency $\Omega(\rho)$	Cross calibration of Thomson scattering and far infrared; deduction of hydrogen density n_i from electrons and impurity ions
Fast particle energy (W_{fast})	$n_e(\rho)$, $n_z(\rho)$, $T_e(\rho)$, $T_i(\rho)$ Neutral beam power and energy (P_{beam}, E_{beam})	Calculation of beam source rate S from beam stopping processes along entire beam path through plasma
Plasma effective ion charge from visible spectroscopy ($\langle Z_{eff}\rangle_{vis.}$)	$n_e(\rho)$, $T_e(\rho)$ Continuum radiation I_{cont}	Mapping of n_e and T_e along line of sight, absolute calibration of continuum radiation, modeling of gaunt factor G_{ff}
Plasma effective ion charge from charge exchange ($Z_{eff}(\rho)$)	$n_e(\rho), \Phi_z(\rho)$	Evaluation of charge exchange intensity Φ_z, local beam density $n_b(\rho)$, modeling of atomic emission rates $\langle\sigma v\rangle_{cx}$
Neutron yield (Y)	$n_e(\rho), n_z(\rho), T_i(\rho)$, $\Omega(\rho)$, P_{beam}, E_{beam}, f_{beam}	Modeling of neutron rates, source rate from beam data and attenuation vector (see Ref. 35)

important examples of physics studies are discussed in Sec. VI: confinement scaling studies, which develop empirical scaling laws to describe the global confinement properties of devices, and local transport studies, which develop and validate more detailed local models for the radial transport of heat and particles. Both approaches require validated data from existing experiments, and the resulting models have an important role in predicting the performance of future devices. Section VI describes the use of international multimachine databases, which have been gathered to develop and validate both empirical scaling models of global plasma confinement and physics-based transport models that describe the detailed evolution of the plasma profiles. In this approach the models are tested by comparing their predictions with validated experimental data. In principle, robust models of the future could also be used as a tool for data validation, by comparing model predictions with measurements.

Finally, in Sec. VII a discussion is presented of the outlook and challenges for data validation in ITER, the next major international fusion device, which will be characterized by more reactor-like plasmas with high fusion reaction rates sustained for long durations, compared with contemporary devices.

II. DATA VALIDATION

II.A. Motivation

Plasma diagnosis is a complex field encompassing physics effects, diagnostics (including optical, mechanical, and electronic systems), computational hardware, and data-processing software (including physics models of plasma processes). An unfortunate consequence of this complexity is the increased possibility of faults resulting in corruption of the final data. To counter this, data-validation processes are used to enable such faults to be identified early so that remedial action can be taken. Experience has shown that this can be very effective in enhancing the efficiency and productivity in magnetic fusion research. This will be illustrated below with a few simple examples.

The possible faults may be grouped in two classes: system failures, where the diagnostic or computational hardware does not function as expected, and faulty physics models, where the physics models of the diagnostic and plasma processes used in the data reduction or analysis are inappropriate.

System failures can result in increased noise, systematic drifts, highly unpredictable behavior, or even a complete loss of data. Examples of causes are an overheated component in the data collection system, a broken sensor in a diagnostic, or a window in the optical path of a diagnostic that becomes coated with an impurity. Such faults are often easy to identify, as they usually result in large abnormalities in the data produced. They are also usually clearly associated with a particular diagnostic or system. However, system failures are often difficult to compensate for, and the required information may not be retrievable. Nevertheless, identification of such a fault enables the underlying malfunction to be repaired, thereby rendering the diagnostic operational again for future use.

Physics models are usually separated according to whether they are used in the data reduction method, in which the desired physical quantities are calculated from the raw diagnostic measurements (for example, the physics relating the width of an observed spectral line to a temperature), or are related directly to the aims of an experiment (as is the case for a transport model). The identification of faults in the data-reduction method is usually considered part of the data-validation process. Investigation of the models related to the aims of the experiment itself is usually seen as the physics study and referred to as the physics model validation or simply model validation. However, data and model validation are increasingly being treated together, removing the need for this somewhat arbitrary distinction (see Sec. VI.B). Faults in the data-reduction method include methods that lack important physics, that are taken outside their range of validity, or that contain errors. Examples would be a data-reduction method for a passive spectroscopy diagnostic, which accounts only for bremsstrahlung radiation, ignoring an important emission line, or a data-reduction method for electron cyclotron emission (ECE) based on black-body emission applied to a plasma with an optically thin edge region. An example of a fault in the physics model would be an incorrect model for thermal transport. The effects of faulty data-reduction methods or physics models are usually more subtle than those of system failures, but if found they can be removed by reprocessing the data retrospectively in cases where the validity of the model can be extended to the relevant conditions. This often involves including additional physics in the data-reduction method or physics model. For example, if a window in the optical path of a diagnostic system becomes coated with some material, a modified transmission coefficient of the window could be included in the data-reduction method for the diagnostic to eliminate the effects of the coating.

II.B. Process

The aim of data validation is to assess the quality of diagnostic data by assessing how well a unified picture of the plasma conditions can be produced from the data. As it exists today, data validation usually takes place at different levels involving increasingly sophisticated physics models and, in general, increasing amounts of human intervention. The use of such a hierarchy enables many faults to be detected and removed at an early stage, ensuring that they do not mask the faults that can only be detected at the higher levels. This subsection outlines this hierarchy and offers examples at a variety of different levels within it.

Most data-validation systems begin with a very low-level assessment. Low-level data-validation systems typically check that data exist and lie within an acceptable range, and they possibly perform some rudimentary consistency checks between data from different diagnostics, making related measurements. Such low-level data validation is usually aimed at identifying system failures. The checks are often straightforward and can be performed automatically. An example of this is the data analysis monitoring (DAM) system in DIII-D (Ref. 1). The DAM system runs between discharges, after a set of data analysis codes have been executed. It verifies the data from a wide range of diagnostics, testing for the existence and basic quality of data. By testing for discrepancies between the values of physical parameters (such as the neutron rate) estimated from independent measurements, the DAM system also extends into higher-level data validation. The results are displayed on a single page after each discharge, to quickly alert the operating team to any problems. Similar systems exist on most currently operating magnetic fusion devices.

A simple example of low-level data validation of some JET data is shown in Fig. 2. Here, time traces of the line-integrated density along a single chord, obtained from two different techniques (interferometry and polarimetry[2]) are compared (for a full description see Ref. 3). In the case of the interferometry system, the line-integrated density is proportional to the measured phase shift of a laser beam after passing through the plasma. Since phase changes of integer multiples of 2π are indistinguishable (corresponding to density changes or "fringe jumps" of integer multiples of $1.12 \times 10^{19}\ \mathrm{m}^{-3}$ for the JET system), an ambiguity results in the measurement. An incorrectly inferred density, because of this ambiguity, is an example of a fault in the data-reduction method. The problem is routinely circumvented by taking the line-averaged density to be zero before the plasma is formed and constraining it to vary smoothly throughout the discharge. However, any loss of the signal or a limitation in the response time of the diagnostic can result in fringe jumps in the resulting line-integrated density. For the discharge of Fig. 2, a channel of the interferometry diagnostic was found to be inconsistent with an alternative measurement derived from polarimetry. The estimate of line-averaged density derived from polarimetry is not subject to fringe jumps, but in JET it has larger uncertainties in its measurement than interferometry and its analysis requires knowledge of the magnetic field along

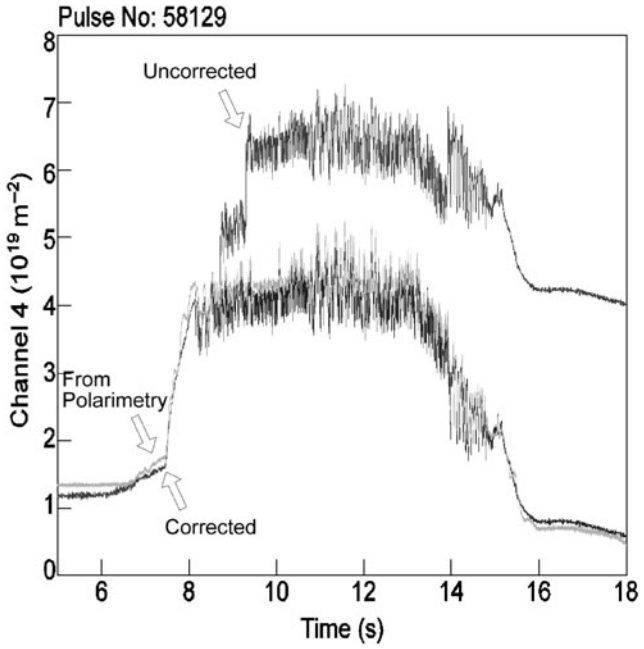

Fig. 2. Time traces of the line-integrated density from a line of sight of the JET interferometry system, derived from phase-shift measurements (before and after correction) and from polarimetry measurements for discharge #58129. The random errors on the interferometry measurement are $\sim 3 \times 10^{17}$ m^{-2}. (*Figure reproduced, with permission, from Ref. 3.*)

the line of sight, which introduces further uncertainties. By comparing the two independent measures of line-integrated density, the interferometry measurement was reanalyzed and a more consistent reconstruction was found. These three traces are shown in Fig. 2. Low-level validation techniques are highly effective for real-time validation, where system failures usually result in large anomalies in the data. Their relative simplicity renders them well suited to real-time implementation where high-speed processing is a priority, while providing an adequate level of validation for operational requirements.

A higher level of data validation is achieved by including an equilibrium reconstruction in the analysis. Although more sophisticated models are used in special cases, for conventional tokamaks this is usually the Grad-Shafranov model (see, e.g., Refs. 4 and 5), which assumes toroidal symmetry, low plasma rotation, and equipartition. Codes commonly used in this respect are EFIT (Ref. 6), CLISTE (Ref. 7), CREATE (Ref. 8), XLOC (Ref. 9), and ESC (Ref. 10). It should be noted, however, that versions of these codes are available that include the effects of toroidal asymmetry (see, e.g., Ref. 11) and toroidal rotation (see, e.g., Ref. 12). Data from the available core diagnostics are mapped onto the derived surfaces of constant magnetic flux in the plasma. This procedure allows measurements of the same physical quantity obtained from different diagnostics to be cross-checked. Coupled to this is the validation of the equilibrium reconstruction itself. By assuming the plasma to be approximately thermalized, or by taking a simplified model for nonthermal particles, the mapped density, effective ion charge (Z-effective or Z_{eff}), and temperature measurements can be combined to produce an estimate of the total plasma energy. This can then be compared for consistency with independent estimates of the total energy derived from the diamagnetic loop and the magnetic equilibrium. This procedure is performed automatically, or with fairly minimal human intervention, on several machines with, for example, the Chain2 system[13] in JET; the Charge Exchange Analysis Package (CHEAP) system (see Sec. III) in TEXTOR, Tore-Supra, and JET; and the TPROF system in Tore Supra.

In higher-level data validation, where an equilibrium reconstruction is used, the quality of the equilibrium reconstruction itself can affect the outcome significantly. An example is shown in Figs. 3 and 4. Figure 3 shows the "safety factor" (magnetic field line pitch, represented as

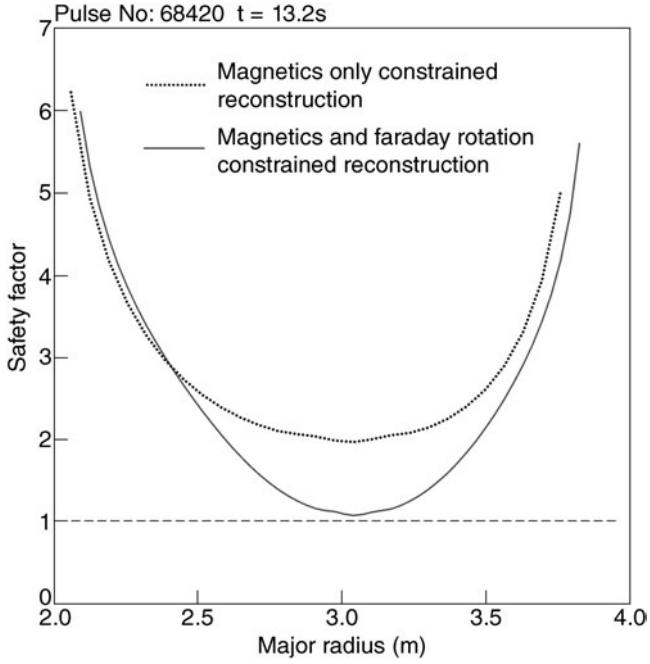

Fig. 3. Safety factor profiles reconstructed by EFIT constrained by magnetics alone (dotted line) and with magnetics and Faraday rotation constraints (solid line). The vertical axis represents the safety factor on the plasma midplane. The horizontal axis represents the major radius in meters. The random error in the minimum safety factor value, estimated from time point to time point random variation, is at least 0.3. The reconstructions are for JET discharge #68420 at 13.2 s.

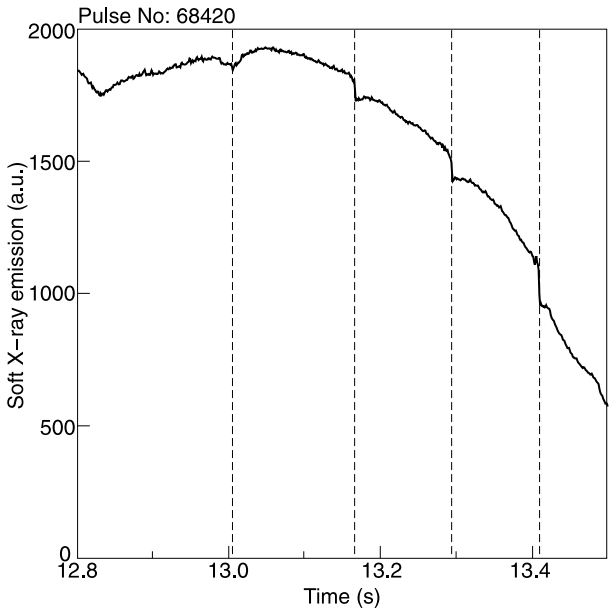

Fig. 4. Time trace showing the soft X-ray emission on a line of sight passing through the plasma core from the same discharge as in Fig. 3. Sawtooth crashes (dashed lines) can be identified with the periodic falls in this signal indicating the presence of a region of the plasma with safety factor of unity. The random error on the measurements is <1% at all time points.

q profiles for JET pulse #68420 at ~13.2 s into the discharge. The profile shown with a dotted line was calculated by the EFIT, which was constrained only by measurements of the magnetic field strength obtained from a set of inductive sensor coils (magnetics). This profile lies above unity everywhere in the plasma. However, as can be seen from Fig. 4, measurements with a soft X-ray diagnostic[14] clearly show the presence of sawtooth oscillations,[15] which are known to be associated with the existence of a region in the plasma where the safety factor is close to or below unity. Generally, for JET plasmas EFIT reconstructions constrained by magnetics alone are consistent with the existence, or otherwise, of sawtooth oscillations. However, disparities can exist for certain discharges in which the current profile is unconventional, as in the example shown in Figs. 3 and 4 where active shaping of the current profile was being studied.[16] The problem is believed to be due to the fact that, as only a few degrees of freedom can be permitted in equilibria constrained by magnetic measurements alone, there is insufficient freedom in the equilibrium reconstruction to reconstruct unconventional cases. This is an example of a faulty data-reduction method. In this case, the data consistency was improved by further constraining the equilibrium reconstruction with information on the internal magnetic field from Faraday rotation measurements made with the polarimetry diagnostic. These enabled reconstructions to be run with more degrees of freedom and resulted in the reconstruction shown by the solid line in Fig. 3. The equilibrium reconstructed with the additional polarimetry constraint remains above unity everywhere but is significantly lower and closer to unity near the center of the plasma. In this sense, the more constrained profile is more consistent with the observed sawtooth oscillations. In discharges with unconventional current profiles, such equilibrium reconstructions are typically more consistent with the existence, or otherwise, of sawtooth activity. However, this procedure is only applied when the reconstruction constrained by magnetics alone shows problems, because of the large manual effort required to validate Faraday rotation measurements before they can be used. This will now be illustrated with an example of a data-validation effort using equilibrium reconstruction to map measurements of the same physical quantity (the electron temperature T_e in this case) from different diagnostics (the ECE system and the Thomson scattering system) onto one another for comparison. The EFIT code, constrained by magnetics alone, was used to map electron temperature profiles from a Michelson interferometer ECE diagnostic[17] and a LIDAR Thomson scattering diagnostic[18] onto the same radial grid. The peak electron temperatures measured by the two diagnostics for a database of discharges with both ohmic and auxiliary heating are shown in Fig. 5. Error bars are not shown on Fig. 5, but the ECE statistical error is estimated as ~10%, and the LIDAR statistical error is estimated as ~7%. Agreement is relatively good for discharges with core electron temperatures below 5 keV, but above this the ECE estimate is systematically and significantly higher than the LIDAR measurement by up to 20%. These cases correspond to discharges with strong auxiliary heating (an example of the measured temperature profiles of such a discharge is given in Fig. 6). The study found that this discrepancy might be the result of non-Maxwellian electron energy distribution functions in the energy range $\sim k_B T_e$, where k_B is Boltzmann's constant. For further information on this study, see Ref. 19. An "oblique ECE" diagnostic is proposed for installation in JET (Ref. 20), which would allow the hypothesis of non-Maxwellian distribution functions to be tested. If this was found to be the case, the observed discrepancy would be seen as being due to a faulty physics model, and the electrons in the core of these plasmas would have to be treated as having a non-Maxwellian distribution function. It should be noted that the core region affected is relatively small, with minor radius <20% of the total minor radius in all cases.

Figure 7 shows another example for JET (for a full discussion, see Ref. 21). Measurements of ion and electron densities, Z_{eff}, and ion and electron temperatures have been mapped onto magnetic flux surfaces reconstructed with the EFIT code. The resulting profiles of these parameters were used in the PION ion cyclotron

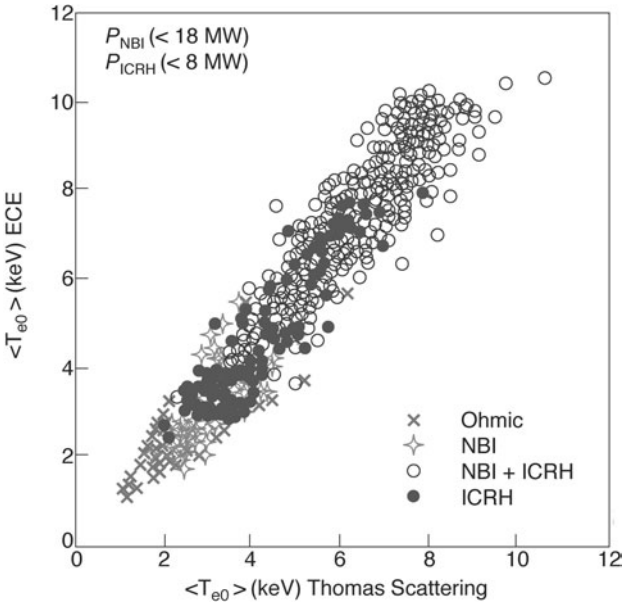

Fig. 5. Comparison between central temperatures measured by Thomson scattering (LIDAR) and ECE (Michelson interferometer, second harmonic, X mode), for plasmas heated purely ohmically, with additional ICRH only, with additional NBI only, or with additional ICRH and NBI. These discharges have a range of toroidal fields from 2 to 3.4 T. The ECE statistical error is estimated as ~10%, and the LIDAR statistical error is estimated as ~7%. (*Figure reproduced, with permission, from Ref. 19.*)

resonance heating (ICRH) modeling code[22] to reconstruct fusion reaction rates, including the contributions of nonthermal particles. Comparison of an early reconstruction (dotted curve) with the measured reaction rate (solid curve) showed an inconsistency. Investigation found that this inconsistency was reduced when the redistribution of the high-energy resonant ions by the sawteeth, assuming that they followed the reconnecting magnetic field lines during the sawtooth crash,[23] was taken into account in the underlying physics model (broken curve). Thus, the original discrepancy is an example of a faulty physics model.

The highest level of data validation involves assessing the consistency of density and temperature profile measurements and global measurements, such as neutron emission rates and diamagnetic energy. This requires modeling of neutral particle source rates and fast particle distributions. At the same time, the equilibrium is solved self-consistently with the available pressure profile measurements. This is commonly performed by the TRANSP code,[24–26] which runs on several current machines, but also with alternative codes such as CRONOS (Ref. 27). This process provides the most complete test of data consistency available at present. Inconsistencies found in such checks may arise from system failures or faulty physics models of the diagnostic or plasma processes. By tracing the inconsistencies back through the processing chain, it is usually possible to identify which of these is the cause. Although it involves well-established codes, the process requires heavy manual intervention with profiles for the main physical parameters fed into the code from different diagnostics, or combinations of diagnostics, and is continued until the most consistent unified representation of the plasma is constructed. Consequently, such high-level validation is usually restricted to a subset of plasma pulses deemed to be scientifically the most interesting. Figure 8 shows an example for JET (for a full discussion, see Ref. 28). The TRANSP code was used to produce a self-consistent reconstruction of the plasma in the manner discussed above. The diffusion and convection profiles for the transport of trace amounts of tritium were parameterized, and the predicted time history of the signals seen by the 19 channels of the JET neutron profile monitor was calculated. The least-squares method was then used to quantify how close the forward-modeled predicted diagnostic signals were to the measured data. In such a method, the time history of the trace tritium density profiles resulting from the diffusion and convection profiles is calculated, and from

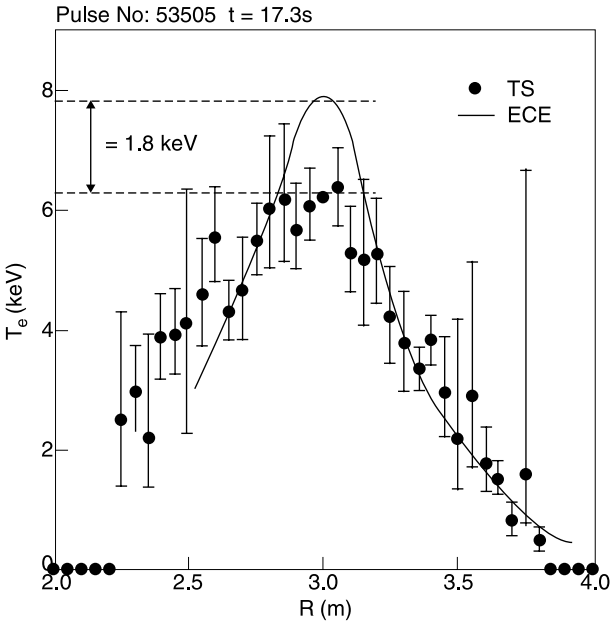

Fig. 6. Electron temperature profiles on the plasma midplane measured by Thomson scattering (LIDAR) and ECE (Michelson interferometer, second harmonic, X mode) for a plasma with ICRF and NBI heating (P_{ICRH} = 5 MW, P_{NBI} = 3 MW), showing the disagreement existing between the two measurements in the plasma core. The horizontal axis represents the plasma major radius. (*Figure reproduced, with permission, from Ref. 19.*)

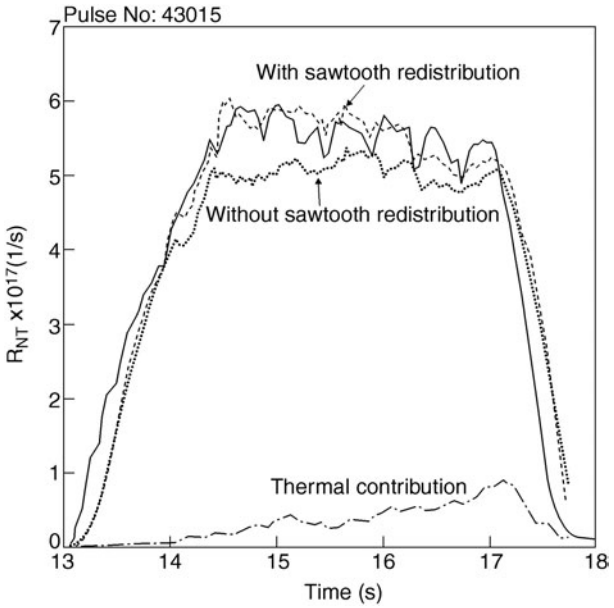

Fig. 7. Measured (solid curve) and simulated deuterium-tritium fusion reaction rates for JET discharge #43015 which was ion cyclotron resonant heated at the deuterium minority fundamental resonance. Simulation results are shown both with (broken curve) and without (dotted curve) sawtooth redistribution. The uncertainty in the calculated neutron yield is ~10%. (*Figure reproduced, with permission, from Ref. 21.*)

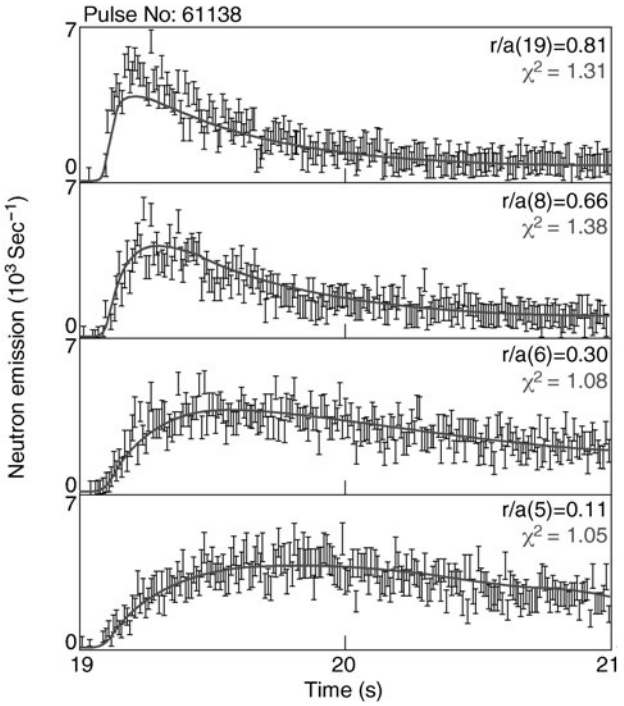

Fig. 8. Neutron emission measured along four lines of sight of a neutron profile monitor and calculated along the same lines of sight in TRANSP (solid curves) for JET pulse #61138. The total normalized chi-squared for all 19 neutron profile monitor channels and the total neutron yield is $\chi^2_N = 1.48$. (*Figure reproduced, with permission, from Ref. 28.*)

this the predicted time history of the signals seen by the 19 channels of the JET neutron profile monitor was forward-modeled. A fitting algorithm was then used to minimize the discrepancy between the modeled signals and those actually measured. Relatively large initial inconsistencies between the modeled and the measured neutron yields highlighted the need for recalibration of the neutron cameras[29] and an improvement of the transport model used.[28] The fit achieved after these were performed is shown in Fig. 8. The modeled (solid line) and measured (points) signals can be seen to agree well, but an analysis of the underlying errors revealed a normalized chi-squared for the fit to all 19 channels of the neutron profile monitor together with the total neutron yield of 1.48. This is significantly greater than the normalized chi-squared of 1.00 ± 0.02 that would be expected from random errors, indicating that some inconsistency in the data or the physics model still remains.

Further examples of this high-level validation are given in Sec. III once the methodology for deriving the physical parameters required in such a study has been discussed in more detail.

II.C. Infrastructure Requirements

The need for effective data validation for a fusion machine places a number of requirements on the data-processing infrastructure. An important aspect in this respect is traceability, that is, the ability to be able to recreate a given piece of processed data from the original data. Traceability requires that the appropriate version of the data and codes used to manipulate the data can be identified in sequence. The key to this is an efficient code and data management system. The data that must be stored can be conveniently classified as

1. *raw data*: the lowest-level data produced by the diagnostics
2. *metadata*: low-level data that are required in addition to the raw data to fully describe the status of the experiment
3. *processed data*: data derived from the raw data and metadata that have been reduced to the required physical parameters.

Raw data are usually in uncalibrated form and in measured units such as volts. An example would be the timing of gamma-ray detection events. Raw data are often preprocessed to reduce the amount of data that is stored, e.g., to reduce the signal from a gamma-ray spectrometer to counts. From the data-validation viewpoint, at this

stage such preprocessing has the disadvantage that it may cause important information to be lost and should therefore be kept to a minimum. Data reduction is better done at the data-processing stage because if mistakes are made, the data reduction can be repeated without any loss of information. However, eliminating preprocessing altogether can impose prohibitive demands on storage capacity in the computing infrastructure, and so a reasonable compromise needs to be found. An important feature of raw data is that, once generated, they are unalterable. Magnetic fusion experiments generally have effective systems in place for storing their raw data, with most large machines having a centralized repository where all raw data may be accessed. The MDSplus system[30] is used at several laboratories for this purpose, but most machines have developed their own database system for this purpose.

Metadata include calibration coefficients, diagnostic settings, data related to the performance of a diagnostic (such as the temperature of a detector or details of the computational hardware), and software employed. Examples would be the absolute calibration of a Thomson scattering system, the spatial position of the chords of an interferometric system, the operating system used by the data collection system, or the version number of a code used in the analysis. In conjunction with the raw data, it provides a comprehensive description of the experiment. It is therefore important that metadata should be stored in a well-defined database, although this practice has not been fully adopted. On existing experiments, metadata tend to be less coherently stored than raw and processed data, often existing in several different databases, hard-coded into analysis software, or, in some cases, in hard-copy documents. In part this is because the definition of metadata is somewhat open-ended. One procedure to encourage proper storage of metadata is to construct analysis codes that are machine-independent and thus require the metadata to be read in explicitly, as input.

Processed data include any data produced from the raw data and metadata, either automatically or with some manual intervention. Examples are the calibrated and Fourier-transformed data from an array of sensor coils measuring fluctuations of the magnetic field, the calculated error bar for an ion temperature measurement taken from a charge exchange diagnostic, or a transport coefficient derived from an interpretative plasma analysis. The MDSplus system may also be used to store processed data, although, again, the majority of machines have developed their own database system for this. For efficient data validation such database systems must permit storage of different processed versions of the same raw signal, the ability to identify the preferred version, flags to identify whether data have been validated, and the storage of errors. The process can be further optimized by using database systems incorporating pattern recognition tools to streamline search and retrieval processes.[31] In JET, the SAS system[32] is used extensively for statistical analysis of large numbers of discharges. The efficiency of this analysis is greatly enhanced by using a separate subsampled database of processed information, to minimize redundant data retrieval/processing.

In addition to data management, good code management is also required. Two important elements here are traceability (the ability to identify codes used to produce given processed data) and the presence of a task management system, to ensure that processed data are produced in the right sequence. Traceability is achieved by date-stamping all processed data and using a version control system for the processing codes. The MDSplus system has been used at some facilities to provide task management, but the majority of facilities use systems developed in-house for this purpose. In JET, a first analysis of diagnostic data is performed between pulses by a suite of analysis codes, referred to as steps, whose dependence on raw data and processed data from other steps is recorded in a database. A task management system refers to this database to determine the order in which steps should be launched such that the inputs required by each step are available at launch. If any reprocessing of data becomes necessary, e.g., because the calibration of a diagnostic was incorrect in the metadata the first time, the task manager refers to the same database to execute all of the steps that depend directly or indirectly on the changed metadata, in sequence. At JET, the suite of first-analysis codes, the task management system, and the supporting database are collectively referred to as the "Chain1 System."[33]

Finally, the human resources involved in data validation have to be coordinated. The process involves diagnosticians, software developers, experimentalists, and modelers and is therefore a community exercise. On most contemporary machines an individual is appointed to be responsible for the data quality from each diagnostic. This individual is responsible for correcting faults and performing low-level data validation. Experimentalists contact the relevant responsible individuals when queries over data consistency arise, and the responsible individuals investigate the underlying problems. Such investigations can involve physics studies and dedicated experiments in their own right, with results reported to and discussed at internal scientific meetings attended by the data-validation community. On larger machines, a formal system for clearing scientific publications usually assumes ultimate responsibility for adjudicating on data validation.

III. DERIVATION OF KEY PHYSICAL PLASMA PARAMETERS

In magnetic fusion research key physical parameters often cannot be measured directly and must be derived from related measurements. Examples of such parameters include the energy confinement time (the characteristic time for the plasma to cool down if all external heating is switched off) or the total thermal energy content of the plasma. The derivation involves the use of physics

models of diagnostics and plasma processes. As noted in Sec. II, faults in such physics models can give rise to inconsistencies in processed data. The data-validation process aims at identifying these inconsistencies and underlying faults so that corrective action can be taken. In this section the physics models behind some important physical parameters are described and examples of the approach to validating the results are presented. The interdependence of measured physical parameters in a fusion plasma has led to a joint approach of exploiting a range of different diagnostic techniques simultaneously. One of the main ingredients in assessing data consistency is the mapping of comprehensive experimental data onto a common radial and temporal grid. This implies that the main physical quantities are constants on a surface of constant magnetic flux, and moreover, that poloidal and toroidal symmetry can be assumed. Another principle is the assumption of local charge neutrality, which enables the deduction of the bulk ion density from measured electron and impurity ion densities. This is an example of how mapping data from different diagnostics, in this case for the bulk ion density from a charge exchange recombination spectroscopy (CXRS) system and the electron density from a Thomson scattering system, onto a common grid can yield an important new measurement.

To demonstrate the principles of data analysis combining data from several diagnostics, a few examples taken from the JET experiment are presented below.

III.A. Kinetic Plasma Energy and Global Confinement Time

The kinetic plasma energy is derived from electron and ion pressure profiles as measured by electron and ion diagnostics. The kinetic plasma energy content is usually benchmarked against the energy derived from a measurement of the diamagnetic flux made with a poloidal magnetic loop.[34] Here, the reconstruction of the kinetic plasma energy is presented and the magnetic measurements are referenced for comparison only. Key input diagnostics for the kinetic energy reconstruction are, for electron density and temperature, a Thomson scattering system, a far-infrared laser interferometer, an ECE system, and for the ion temperature, ion densities, and plasma rotation, a CXRS system. The total ion pressure includes contributions from thermal and nonthermal ions as created, for example, by high-power neutral beam heating (see Refs. 35 and 36).

For the derivation of a fast beam population, determination of the local source rate of fast ions is a critical issue. This may be deduced experimentally from beam emission spectroscopy (see Refs. 37 and 38). High-quality electron density profiles are needed, as well as a full coverage of the main impurity ion concentrations. The attenuation of neutral beams increases exponentially along the path through the plasma. Hence, errors of electron density may lead to substantial errors in the deduction of local beam strengths. In view of the importance of absolute electron densities, the approach used at JET in the CHEAP system is examined below.

The CHEAP electron density profile is based on two electron density diagnostics. In the first case, line-integrated electron densities are provided by a far-infrared interferometer. The density data derived are directly related to phase shifts between probing and reference far-infrared hydrogen-cyanide (HCN) laser beams. The absolute calibration of this system is found to have high stability. In the second case, a highly resolved profile is measured by a (LIDAR) Thomson scattering diagnostic. In this case the absolute calibration tends to drift in the course of operation because of contamination of observation optics. After initial absolute calibration this diagnostic is periodically cross-calibrated against the interferometer. This is achieved by calculating, in a first step, a symmetrized electron density profile with magnetic flux indices (see Fig. 9) as coordinate. The second step is the mapping of the path of the interferometric observation channel onto its corresponding magnetic flux surfaces and the calculation of the equivalent density line-integral using local LIDAR data. The mapping process requires data from the EFIT equilibrium code. An

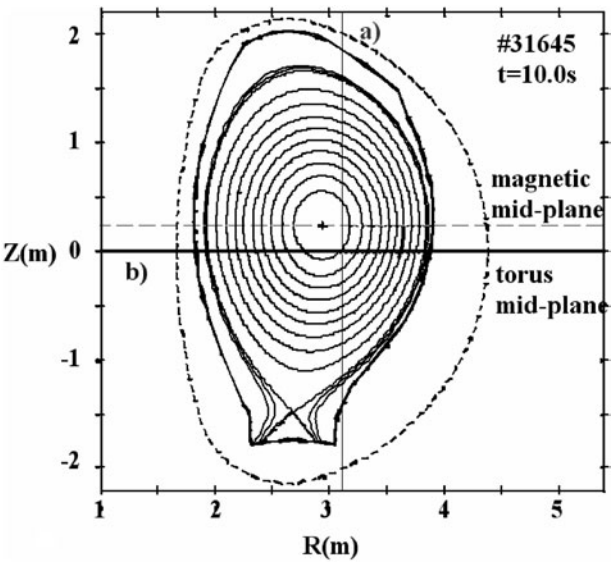

Fig. 9. Schematic layout of cross calibration of radially resolved LIDAR electron density diagnostic and line-of-sight integrated interferometric density data. The radial LIDAR profile (horizontal line-of-sight "b") is symmetrized on magnetic flux coordinates. The line-density equivalent of the HCN data is calculated along the vertical line-of-sight "a". Note, neither line of sight passes through magnetic axis.

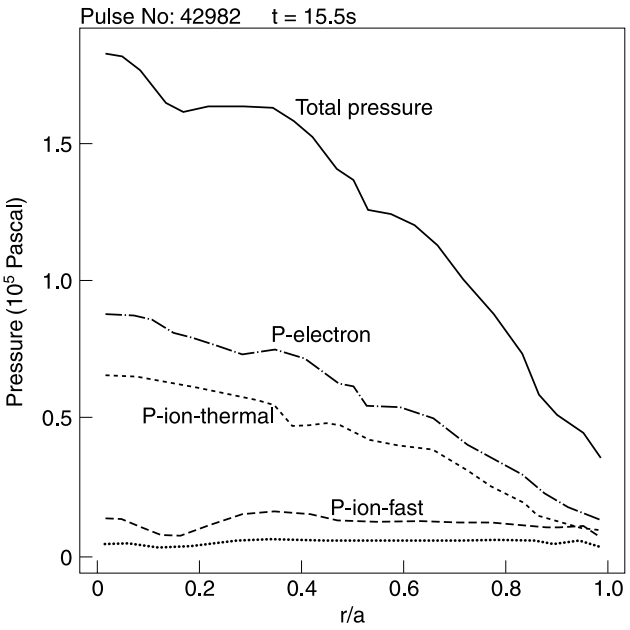

Fig. 10. Particle pressure reconstructed from electron and ion data. The fast particle contribution is calculated from beam target interaction and beam source rates deduced from beam stopping calculation (CHEAP). The random uncertainty is ~10% in the electron pressure and ~18% in the ion pressure.

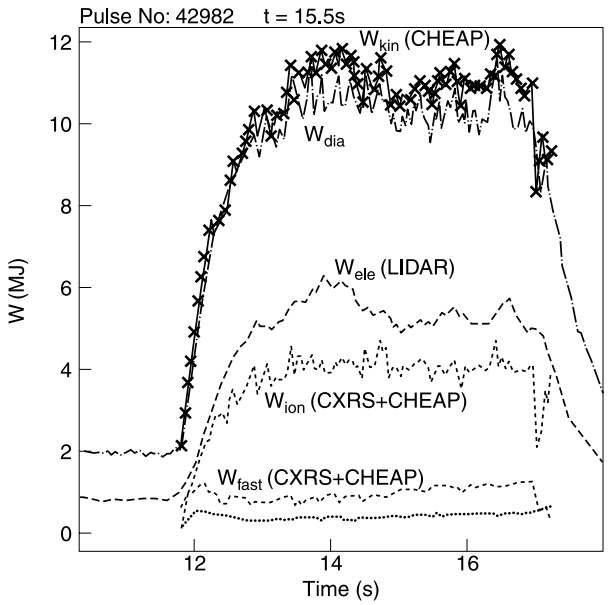

Fig. 11. Reconstruction of kinetic energy from electron and ion diagnostics and comparison with diamagnetic energy as deduced from magnetic loop measurements in a hot ion H-mode fueled with deuterium and tritium. Also shown is the fast particle energy content. The uncertainty in the reconstructed stored energy is typically ~14%.

example that illustrates the reconstruction of the kinetic energy of a "hot-ion–mode" pulse fueled with deuterium and tritium (#42982) is shown in Fig. 10 (radial pressure profiles). Time traces of the reconstructed energy signals are shown in Fig. 11 for this case and in Fig. 12 for the case of a standard high-confinement mode (H-mode). Uncertainties in such data are typically at a level of ~14% (see Sec. VI.A). The latter case has been added as a cross-reference for the two examples of neutron yield reconstructions in the case of a thermal-thermal and beam-thermal dominant neutron-production case (see Sec. III.C). For comparison, Figs. 11 and 12 also show the plasma energy calculated with the diamagnetic method. Note that the diamagnetic energy includes the contribution from fast particles, whereas the kinetic energy does not. For consistency checks this contribution, calculated, for example, from the Fokker-Planck equation, must first be added to the kinetic energy, as has been done for the example in Fig. 11.

The global thermal energy confinement time is defined as the ratio of plasma thermal energy to the power loss, i.e., it is the characteristic time over which the plasma cools if all heating is switched off. This parameter is particularly important because it represents the overall quality of plasma confinement. In contemporary machines the fusion power produced is usually negligible compared to the power input, so that the power loss can

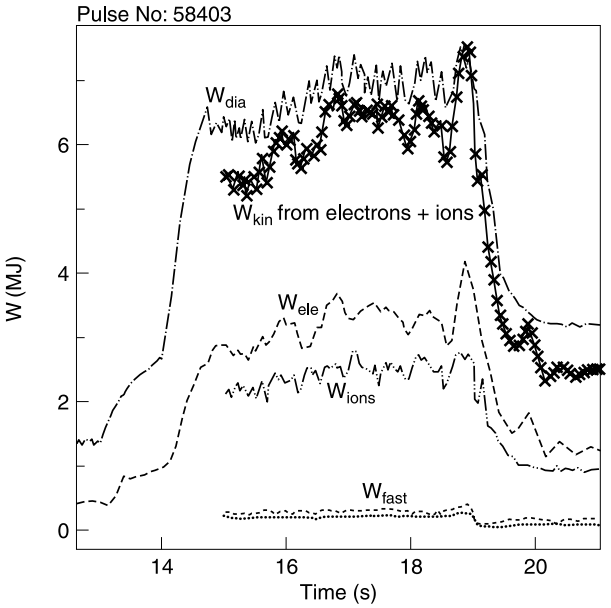

Fig. 12. Reconstruction of kinetic energy from electron and ion diagnostics and comparison with diamagnetic energy as deduced from loop measurements in the case of a standard H-mode scenario. Also shown is the fast particle energy content from two neutral beam injectors. The uncertainty in the reconstructed stored energy is typically ~14%.

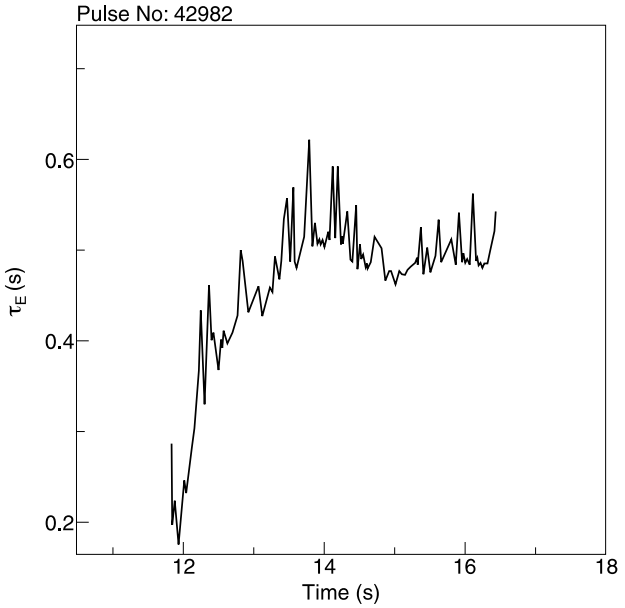

Fig. 13. Confinement time obtained from the plasma kinetic energy and total power input. The uncertainty in the plasma energy (typically ~14%) is the main source of uncertainty in the confinement time.

be replaced with the power input less the rate of change of plasma thermal energy:

$$\tau_E = W_{th}/(P_{in} - dW_{th}/dt) \ ,$$

where

τ_E = confinement time
W_{th} = plasma thermal stored energy
P_{in} = power input.

The calculation is routinely performed using the total power input together with either the kinetic energy or the diamagnetic energy (after subtraction of fast particle energy). Time traces of the global thermal energy confinement time obtained from these two derivations of thermal energy are shown in Fig. 13.

III.B. Effective Ion Charge and Bulk Ion Density

The value of the effective ion charge Z_{eff} plays an import role in the characterization of fusion plasmas. Although a number of different diagnostic techniques for measuring Z_{eff} exist, here spectroscopic measurements are discussed (i.e., line-of-sight averaged values of Z_{eff} as deduced from the measurement of visible bremsstrahlung) and compared to localized measurements of impurity ion densities as provided by CXRS. In the context of this chapter, the role of Z_{eff} is noteworthy as a means to establish consistency checks between different spectroscopic techniques and measurements of radiated power by bolometry. Using the plasma as a common radiation source accessible to all spectroscopic diagnostics, Z_{eff} provides a natural tool for cross-calibration procedures (see Ref. 39). The cross calibration is performed by comparing the line-of-sight integrated measurement with the localized Z_{eff} data reconstructed from CXRS impurity ion concentrations (see Table I), mapped onto the same line of sight.

Local effective ion charge values are obtained from Abel inversion of line-integrated measurements of bremsstrahlung emission (see, e.g., Ref. 40). For a number of reasons, the inversion technique may lead to doubtful results. For example, hollow Z_{eff} profiles and hollow continuum emissivity profiles (which occur in H-mode plasmas) are difficult to reconstruct from noisy experimental data. In other cases, a background level of diffuse reflection of continuum radiation from the wall of the vacuum vessel may lead to an apparent enhancement of Z_{eff} toward the edge channels. To counter this, it is common practice to constrain the result of the inversion by introducing parameterized functions to represent realistic emissivity profiles.

In the CXRS method several boundary conditions need to be satisfied in order to achieve a sensible reconstruction of Z_{eff}. The main low-Z impurities need to be assessed, and therefore a number of instruments need to be operated simultaneously to measure the dominant CX spectra. In the case of JET, intrinsic impurities such as carbon and beryllium originating from the plasma wall, as well as externally injected (seeded) impurities such as argon, neon, nitrogen, or helium, need to be measured. In some cases even several ionization stages of the same impurity need to be considered. Figures 14 and 15 show the contributions $Z_z(Z_z - 1)c_z$ of Ar^{+18} and Ar^{+16} to Z_{eff} in an experiment with argon seeding. Here, c_z is the impurity concentration (ratio of the impurity density to electron density) for species Z. Z_{eff} is obtained from these contributions from the formula $Z_{eff}^{CX}(\rho) = 1 + \sum_{z=2}^{z=18} Z_z(Z_z - 1)\, c_z(\rho)$, where ρ is the normalized minor radius r/a. In this example argon is seen to be a significant contributor to Z_{eff}. Similar arguments apply to the reconstruction of the bulk ion density or the plasma dilution factor:

$$\frac{n_d}{n_e} = 1 - \sum_{z=2}^{z=18} Z_z \frac{n_z}{n_e} \ .$$

In each case a comprehensive decomposition into the contributing impurity ion densities is needed. In most cases, the CXRS reconstruction of Z_{eff} appears to be systematically below the values deduced from visible bremsstrahlung (examples are shown in Figs. 16 and 17). The reasons for this are not well understood and are the subject of ongoing investigations. Although the error margins in these measurements are substantial (estimated errors are ~15% for Z_{eff} from visible bremsstrahlung and 20% for Z_{eff} from CXRS), their measurement has been shown to lead to reasonable global data consistency for the energy content and neutron yield.

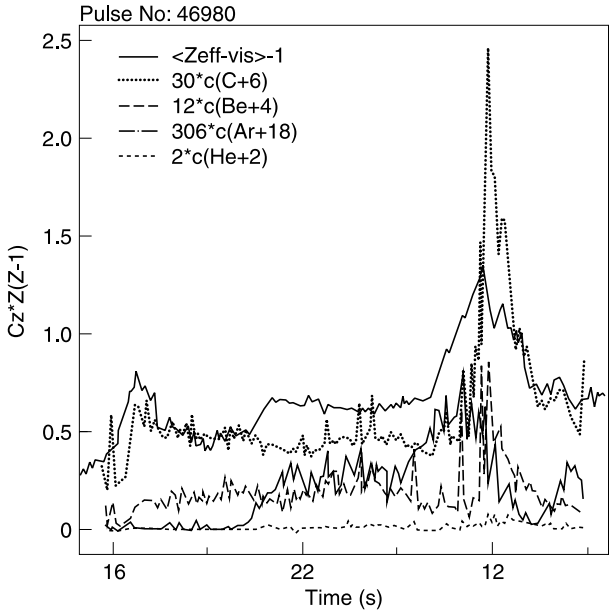

Fig. 14. Impurity ion contributions $Z_z(Z_z - 1)c_z$ to Z_{eff} in an argon-seeded plasma. Each ion species Z is characterized by its charge Z_z and concentration c_z. Charge exchange data are from single chord measurements at $R = 3.1$ m. Contributing ions are He^{+2}, Be^{+4}, C^{+6}, and A^{r+18}. The random uncertainty is ~15% in the bremsstrahlung measurement and ~20% in the charge exchange measurement.

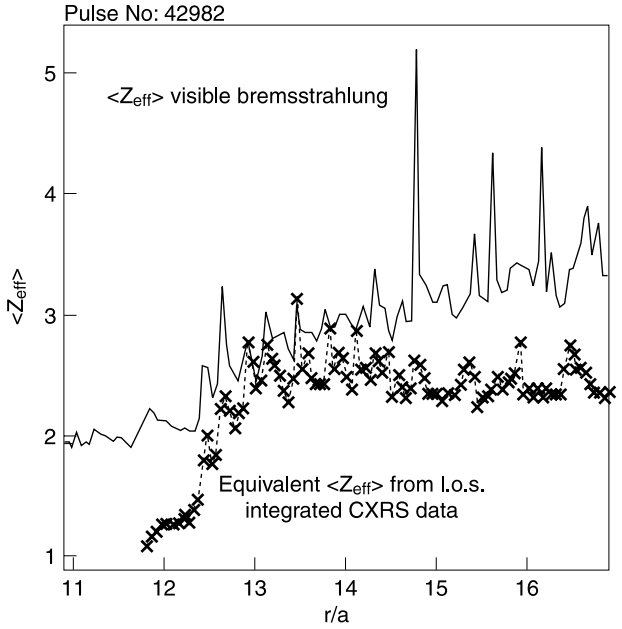

Fig. 16. Comparison of line-averaged Z_{eff} from visible bremsstrahlung and its equivalent charge exchange line integral reconstructed from contributions of C^{+6}, B^{e+4}, and He^{+2}. The random uncertainty is ~15% in the bremsstrahlung measurement and ~20% in the charge exchange measurement.

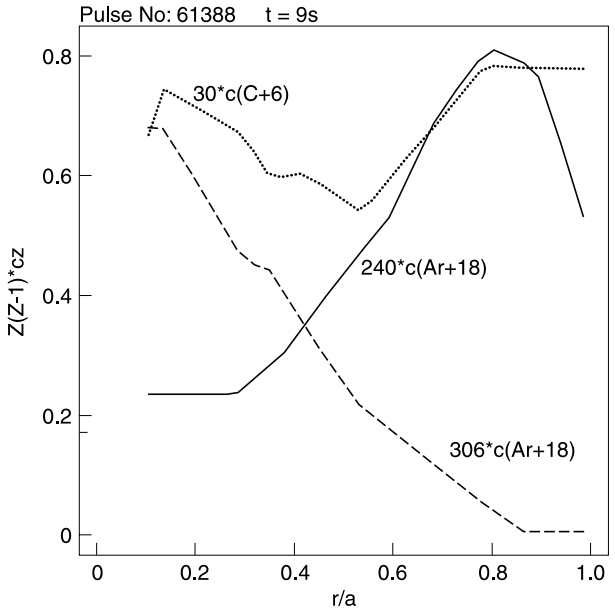

Fig. 15. Contributions of C^{+6} and two ionization stages of seeded Ar to Z_{eff}. The traces represent the individual contributions $Z_z(Z_z - 1)c_z$ from each ion z with charge Z_z.

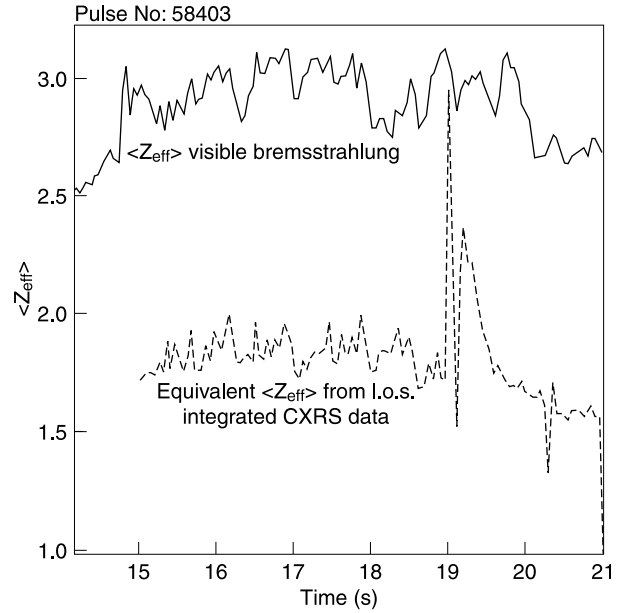

Fig. 17. Comparison of line-averaged Z_{eff} from visible bremsstrahlung and its equivalent charge exchange line-integral reconstructed in this example from contributions of C^{+6} only. The random uncertainty is ~15% in the bremsstrahlung measurement and ~20% in the charge exchange measurement.

III.C. Thermal and Nonthermal Neutron Yields

The total neutron yield is a key performance indicator for hot fusion plasmas, and for this reason it is important to test the quality of measured plasma data that are used to predict this. In the following, two dominant neutron-production mechanisms are considered in a hot-ion high-performance plasma at JET. The first is the thermal-thermal neutron-producing reaction either between deuterons or between deuterons and tritons. The second is the beam-thermal neutron-producing reaction, which describes the interaction between the deutrons injected in neutral beam heating and the thermal deuteron background target. In the case of JET D-T experiments, a fast triton population is created by using tritium-seeded neutral beams.

The thermal-thermal neutron yield, for example, is a strong function of the local ion temperature,[41] and high accuracies (<5%) are needed in the measurement to achieve acceptable accuracy in the predictions. To obtain the beam-thermal neutron yield, time-dependent Fokker-Planck equations have to be solved (see Ref. 35). Examples of the reconstruction of neutron yields from thermal-thermal and beam-thermal components are given in Figs. 18 and 19. Neutron yields can also be obtained more directly with time-resolved neutron counters or spectrometers (see, e.g., Ref. 42). Spectrometers, based on a magnetic proton recoil technique, time-of-flight measurements, or scintillation pulse-shape and pulse-height analysis, are installed in JET (Refs. 29 and 43). Cross-checks of inferred and measured neutron yields constitute a valuable part of the consistency tests of plasma data.

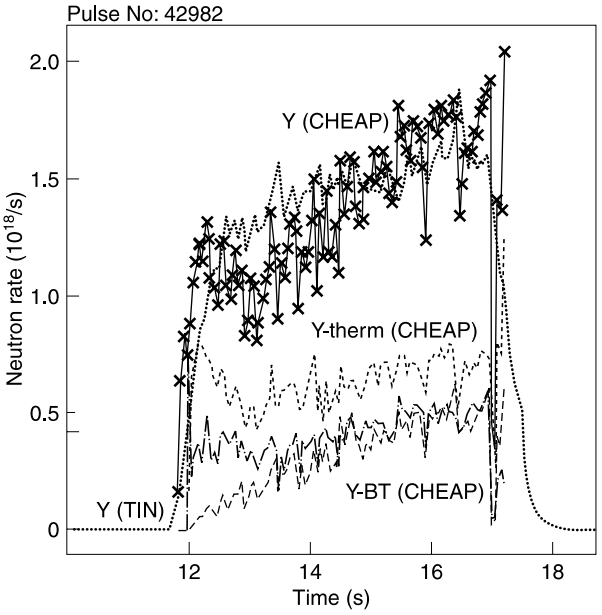

Fig. 18. Time traces of the measured neutron yield [Y(TIN)] and the reconstructed neutron yield from the CHEAP code [Y(CHEAP)] for a D-T plasma. Also shown are the individual contributions from the thermal-thermal [Y-therm(CHEAP)] and beam-thermal [Y-BT(CHEAP)] reactions. The two beam-thermal signals represent the two injection boxes in octant 4 and 8, respectively.

III.D. Discussion

Different physical parameters require different measurements and calculations. The definitions of key physical parameters and the main steps needed in their determination are summarized in Table I. Among the measurements of the kinetic plasma energy, the effective ion charge, and the neutron yield, the kinetic energy appears to be the most robust, and independent magnetic measurements can be well reproduced. Not surprisingly, the reconstruction of the total neutron yield with its nonlinear dependencies is the most challenging. In the case of dominant thermal-thermal neutron production (e.g., in hot-ion–mode plasmas), the accuracy of ion temperature profiles appears to be adequate for a reasonable prediction. In such cases the quality of the measured ion density profile is less critical, since high-performance plasmas with high neutron yields are typically characterized by bulk ion densities that correspond closely to electron densities, which are generally measured with good quality. Although the ion composition and the relative contribution by a number of low-Z impurities is adequately established from CXRS data, an

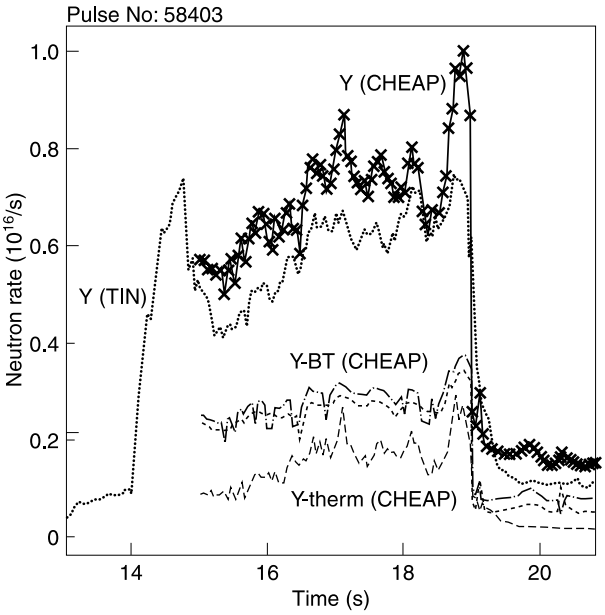

Fig. 19. Reconstruction of neutron yield from thermal-thermal and beam-thermal reactions in a 3-MA H-mode plasma. In this case, beam-thermal neutrons are the dominant rate. The overprediction of the total yield is probably due to unaccounted losses of fast particles.

ongoing challenge persists in resolving nontrivial deviations between Z_{eff} as deduced from bremsstrahlung and Z_{eff} derived by the summation of several impurity ion concentrations.

IV. INTEGRATED DATA ANALYSIS

Most analysis of raw data in fusion experiments is currently done independently for each diagnostic. While this approach gives valuable information on individual physical parameters, the analysis procedure does not fully exploit the fact that diagnostics measure different and partially overlapping aspects of one single underlying plasma "state" of densities, temperatures, currents, etc. The approach can be compared with trying to fit spectral lines by using different fits for different parts of a spectrum and then afterward trying to make the different parts fit together, instead of analyzing the whole spectrum with a single model. The formulation of the data analysis problem in terms of unified models, using data from a number of diagnostics taken together, is the philosophy behind IDA (Fig. 20). This method can substantially improve the accuracy of inferred physical parameters. For example, an electron density profile simultaneously fitted to data from both a multichannel interferometry diagnostic and a separate Thomson scattering diagnostic would give better resolved profiles than using either diagnostic on its own. There are two main approaches to such integrated schemes, as described below.

In the first approach, a large least-squares or maximum-likelihood optimization is performed to find the values of the physical parameters that best match the raw or analyzed data from several diagnostics simultaneously. Examples of this can be found in Refs. 28 and 44. The advantage of this approach is that it is comparatively easy to implement and uses standard analysis tools. The main disadvantage is that the approach does not lend itself readily to rigorous treatment of systematic uncertainties or poorly specified statistical errors, often biasing the results too much in favor of certain diagnostics. This is usually dealt with by applying weighting factors to adjust the relative importance of data from different diagnostics in the fitting procedure. The weighting factors can affect the results significantly, but they often lack a rigorous basis.

In the second approach, described later in this section, so-called Bayesian probability theory[45,46] is used. Bayesian probability theory has received a lot of interest in recent decades, since it provides a unified way of handling all types of uncertainties (systematic, statistical, model, etc.), and is at the same time a generic method for solving so-called "inverse problems" (curve fitting, tomographic inversions, etc.). It is based on the usage of probabilities and probability distributions to represent all of the available knowledge on model parameters. It can be used to specify, if necessary, detailed statistical models of the diagnostics themselves,[47,48] where systematic uncertainties are explicitly described by probability distributions over diagnostic parameters whose values are not fully known, such as calibration factors, viewing geometry, instrument functions, etc. If this is done carefully, it allows data from multiple diagnostics to be combined in a rigorous way, eliminating the need for weighting factors, since the causes behind the biases that necessitated the weighting factors are dealt with directly, and the influences of the systematic uncertainties on the physical parameters can then be accounted for. Also, in such an approach, employing probability theory, the information on the unknown physical parameters is represented as probability distributions. The multidimensional probability distribution for the unknown physical parameters is, in this scheme, updated for each new diagnostic datum by exactly the information this datum contributes to the joint set of unknowns. Each datum is therefore allowed to update the information on

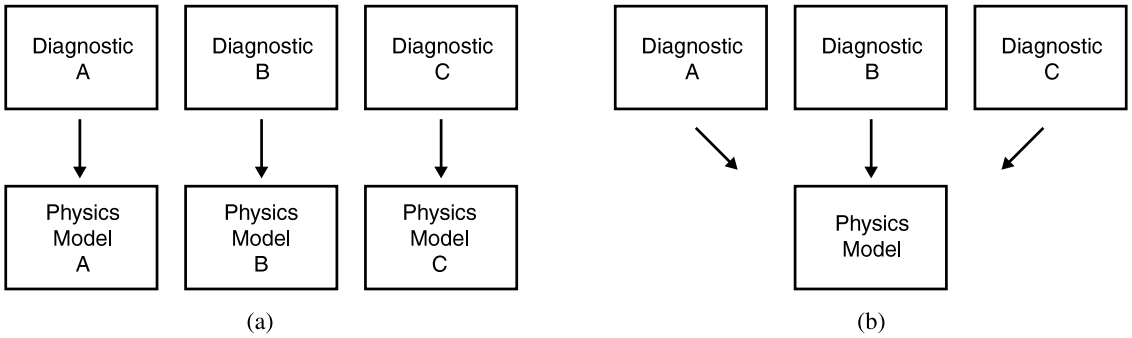

Fig. 20. (a) Approach in conventional diagnostic analysis, where each diagnostic module infers physical parameters using its own data, and (b) an integrated approach, where physical parameters are inferred jointly from multiple, heterogeneous diagnostics.

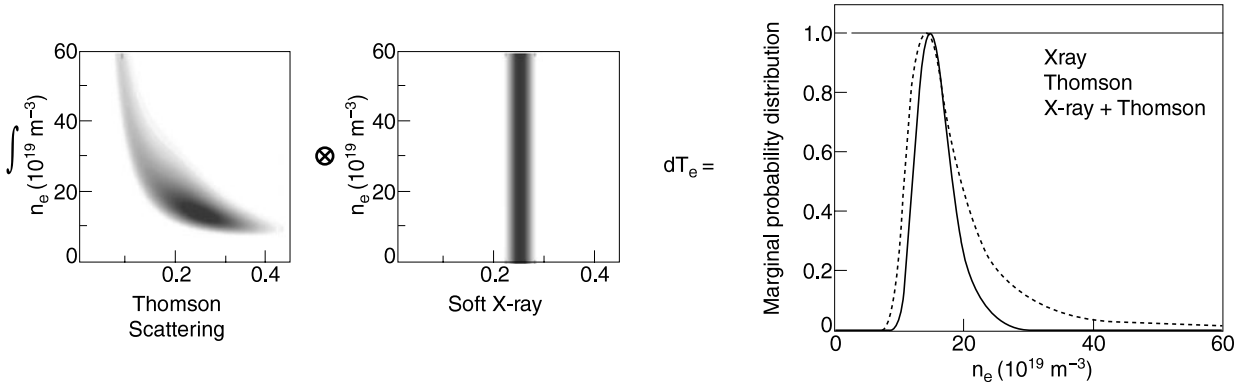

Fig. 21. Example of results from a Bayesian integrated analysis. In this example, information on electron temperature and electron density, obtained from a Thomson scattering diagnostic, is combined with an estimate of the electron temperature only, obtained from a soft X-ray diagnostic. The temperature is obtained from the soft X-ray diagnostic through the use of two narrow spectral filters, exploiting the wavelength dependence of Bremsstrahlung radiation. (*Figure reproduced, with permission, from Ref. 49.*)

all physical parameters simultaneously, and after all data have been used, the final so-called posterior probability distribution fully describes the joint uncertainty of the inferred physics quantities. This process is illustrated in Fig. 21, where the probability distribution over electron density and temperature from a Thomson scattering diagnostic is adjusted by separate information on electron temperature to yield an updated probability distribution that shows a much lower uncertainty.[49] Note that the uncertainty of electron density has been reduced as well as the uncertainty on the electron temperature, even though the extra information added was information about the electron temperature only. This is an effect of the joint model: temperature and density are correlated in the left graph, so extra information on one quantity simultaneously gives information on the other, increasing the overall accuracy.

A large-scale Bayesian integrated model has been created for the W7-AS stellarator,[50] in which up to four different diagnostics were combined. In this model, density and temperature were assumed constant on three-dimensional stellarator flux surfaces, and the flux surface topology was itself part of the model. The flux surface geometry is usually determined by an equilibrium calculation using magnetic and/or profile data, and is therefore an inferred quantity that itself has uncertainties. Since physics models are usually expressed in flux coordinate systems and most diagnostics need to be mapped to flux surfaces (see also Sec. III), a large integrated system should infer flux surface topology self-consistently with other quantities, such as density and temperature profiles, that is, there should be a joint model for magnetic topology and the physics quantities mapped to the inferred magnetic coordinate system. Figure 22 shows flux surfaces and pressure profile that were self-consistently inferred from interferometry, charge exchange ion temperatures, Thomson scattering, and diamagnetic energy

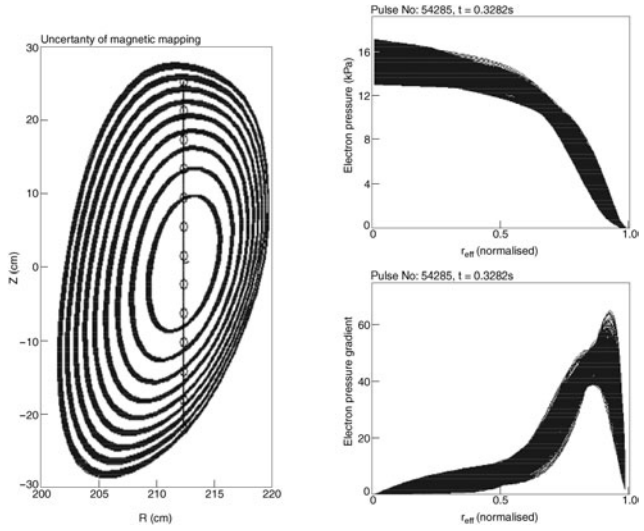

Fig. 22. Quantities inferred from a Bayesian integrated model for the W7-AS stellarator. The width of lines corresponds to the uncertainty of the reconstruction. (a) Reconstructed flux surfaces, (b) simultaneously inferred electron pressure profiles, and (c) the gradient of electron pressure profile. (*Figure reproduced, with permission, from Ref. 51.*)

measurements.[51] The posterior uncertainties correspond to the width of the lines in Fig. 22 and represent all uncertainties from the full system, including uncertainties in the magnetic coordinate system itself. In this example, inference of profiles, magnetic topology, and mapping of the profiles to the magnetic coordinate system are done in one unified, internally consistent, process. Also note that the resulting (marginalized) posterior over physical parameters is very easy to interpret. Since

the information gained on the physical parameters is expressed as a probability distribution, it is possible to answer questions such as "What is the probability that, given the model and all data used, the electron pressure gradient at normalized minor radius $r_{eff} = 0.8$ is higher than some specific value, taking into account all uncertainties in the joint system?"

There is certainly a great deal of potential in these methods, but it remains to be seen whether it will be possible to apply such unified models routinely on a large scale, utilizing large amounts of heterogeneous data. A major challenge with this approach lies in its complexity; each diagnostic is usually a very complicated system on its own, often showing complex relationships to both underlying parameterized physics models and internal diagnostic parameters that could be sources of systematic uncertainties. To remodel each such diagnostic using a probability theory approach and, furthermore, combine it with other equally complex instruments is a formidable task. On the other hand, if successful, it would allow full utilization of all available data, increased accuracy of inferred physical parameters, and higher levels of internal consistency.

V. PLASMA CONTROL

In plasma operation aiming at high fusion gain (ratio of fusion power production to total power input) steady-state physics issues will impact strongly on device operational constraints.[52,53] Active plasma control in such operation will require the use of a wide range of physical parameters obtained from real-time measurements and analysis, together with appropriate actuators.[54,55] These include measurements obtained with magnetic diagnostics (such as the total plasma current, core and divertor plasma geometry, vertical position, etc.), kinetic profiles (density and temperature), and core and divertor radiation. Since steady-state operation will require the optimization of the bootstrap current (a self-generated contribution to plasma current, driven by pressure gradients, which could reduce or eliminate the need for inductive current drive, which implies pulsed operation), the pressure profile (electron and ion density and temperature profiles) will have to be actively controlled at particular flux surfaces. This could also be required in order to minimize particle and heat transport or to avoid potential macroscopic instabilities. Control of some macroscopic instabilities (such as the neoclassical tearing mode or toroidal Alfvèn eigenmodes) may also require fine tailoring of the q profile. In addition, the extension of the pulse length to 500 or 1000 s will bring more challenges to plasma control diagnostics in terms of reliability, robustness, and resilience to neutron flux. Specific diagnostics will be required to survey plasma-facing surface components, which will receive power flows of the order of 5 to 10 MW/m^2. Real-time control of the surface temperature of plasma-facing components, the neutral pressure, and radiation will therefore be

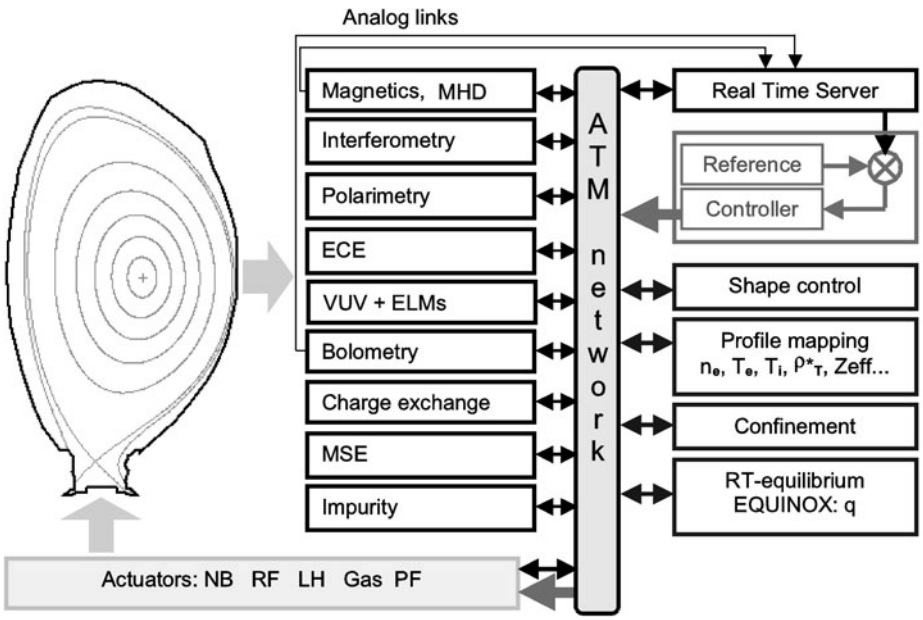

Fig. 23. Real-time measurement and control system developed at JET. This system comprises real-time diagnostic [left of asynchronous transfer mode (ATM) network], real-time processes, the real-time signal server, and the central controller (right of ATM network) where the gains are set up for a feedback control experiment. All of these systems are connected to the ATM communication network. (*Figure reproduced, with permission, from Ref. 56.*)

essential. In a reactor it will also be essential to measure the fusion power and the alpha power in real time, in order to achieve and sustain the required fusion gain.

Several fusion devices (tokamaks and alternative machines such as stellerators or compact tokamaks) have recently developed comprehensive real-time measurement networks capable of providing most of the data required for the control of a steady-state discharge. JET (Ref. 56) (Fig. 23), Tore Supra (Ref. 57), JT-60U (Ref. 58), and DIII-D (Ref. 59), in particular, have implemented integrated real-time diagnostic data acquisition and processing using fast communication networks. In the first instance, to implement feedback loops it is essential to measure or calculate the control parameters in real time. In addition, other plasma parameters relevant for control are also calculated online by validated and dedicated codes. As an example, magnetic diagnostics produce direct measurements of flux and field at given locations in the vacuum chamber. However, for the control of the plasma shape and for machine protection purposes, it is often more convenient for plasma operation to control the distances between the plasma and the plasma-facing components. To obtain these distances the flux and field measurements are used as input to real-time plasma boundary reconstruction codes.

Validation of real-time data for both magnetic and profile control raises an acute problem since the data must be validated in real time before being used for control. This is achieved using real-time alarm generation, which may, for example, check that raw data are within a reasonable range and are consistent with similar measurements or with a database containing data from previous operation under similar conditions. In some cases a complete model of the diagnostic chain may be available for comparison with measurements in real time. These low-level validation techniques (see Sec. II) are usually highly effective in detecting system failures, which usually result in large anomalies in the data. In addition, diagnostics need to be built to high reliability to minimize interruptions to operation.

In Secs. V.A, V.B, and V.C, we present the range of diagnostics commonly used in real-time control across three different areas: magnetic control, kinetic control, and instability control. Real-time calculations associated with these diagnostics and the validation procedure for real-time data are discussed and specific examples given.

V.A. Real-Time Diagnostic Validation for Magnetic Control

The magnetic diagnostic is commonly used in tokamaks for the control of the plasma shape, position, and current. Plasma scenario requirements of shaped configurations (high triangularity and elongation) and high reliability have resulted in the installation of a large number of magnetic sensors in the vacuum vessel for redundancy.[60,61] The number of poloidal and radial field sensors (i.e., pickup coils and flux loops) is therefore generally much higher, by a factor of 3 to 4, than the number of control points. This ensures robutsness of the control in case a sensor fails and also helps improve the accuracy of the measured parameters.[62] Real-time validation of these signals, e.g., by continuous comparison of similar sensors, can be applied to detect faulty signals and trigger corrective action, such as shutting down the machine or substituting faulty signals in the processing chain with alternative signals.

Magnetic control has the essential function of maintaining the plasma shape, position, and current in the presence of internal plasma instabilities [such as sawteeth, minor disruptions, or edge-localized modes (ELMs)]. In addition, the plasma is inherently vertically unstable, and this instability is controlled using data from magnetic diagnostics. To achieve these goals in practice, the controllers use real-time calculations such as the distance between the plasma boundary and the first wall at specified positions around the cross section, or the vertical speed of the plasma current centroid. These quantities are controlled by a real-time feedback system using external poloidal field coils as actuators. Recently, significant advances have been made in the control of highly shaped plasma configurations in the presence of strong variations of the plasma pressure and internal inductance.[62] Before being implemented for control, real-time control schemes were developed using electromagnetic models which had been validated against extensive past experimental data.[63] Various engineering limits of the research device, such as power loading limits or poloidal field coil current limits, are also sometimes included. A systematic real-time check of the data file structures is commonly performed before operation to check that the system is behaving normally before the discharge is launched. More sophisticated consistency checks to detect faults in individual sensors during operation are not commonly implemented on contemporary machines.

Until recently, the underlying basis of magnetic plasma shape calculation in magnetic devices was the Grad-Shafranov ideal magnetohydrodynamic plasma equilibrium equation applied to the vacuum region between the plasma and the vacuum vessel (no current) and using magnetic measurements as input. From these real-time computations, the loop voltage and Shafranov current moments can also be inferred. It is therefore possible to determine the normalized plasma pressure β_p ($\beta = 2\mu_0 p/B^2$, where p is the plasma pressure and B is the magnetic field, in this case, the poloidal field as indicated by the subscript p) and the internal inductance when the plasma is elongated.[a] For a circular plasma, β_p can also be determined using diamagnetic loops. They provide essential volume-averaged parameters for the control of plasma

[a]Note that the accuracy of this procedure is such that it yields little information on the core current profile.

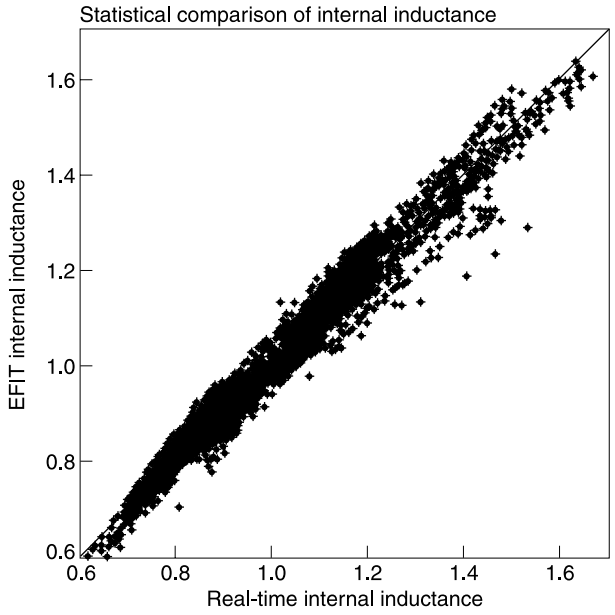

Fig. 24. Validation of the real-time data internal inductance inferred from magnetic boundary code against the EFIT data for a very large database of pulses. (*Figure reproduced, with permission, from Ref. 143.*)

performance. These real-time data are usually validated against the magnetic equilibrium database (see Fig. 24). Numerous examples are available in DIII-D (Ref. 64), JT-60U (Ref. 65), and JET (Ref. 66) where the control of the plasma performance, using the stored energy or normalized plasma pressure, is achieved through modulation of auxiliary power input.

More recently, efforts have been undertaken to develop a full real-time magnetic equilibrium reconstruction capability. Real-time Grad-Shafranov solvers have been implemented on several devices such as JET, Tore Supra (EQUINOX code[67]), and DIII-D (EFIT code[68]) to compute the complete equilibrium using magnetics alone as input. Generally, these codes were developed and are optimized to produce an updated equilibrium every 10 to 20 ms, which is sufficient for digital plasma control systems using equilibrium data. These processes are usually validated against the existing off-line equilibrium code to ensure that their behavior is robust in the vast majority of plasma configurations that the device has to run.

However, the development of more advanced operation, demanding a higher degree of control over the plasma profile, has motivated the integration of other internal flux measurements [such as infrared interferopolarimetry and/or motional Stark effect (MSE) measurements] to compute real-time current density and q profiles with enhanced accuracy. This challenging objective requires improved reliability in the magnetics as well as the other diagnostics.

V.B. Real-Time Diagnostic Validation for Kinetic Profile Control

Until recently, tokamak control operated with a clear distinction between magnetic and kinetic control functions in terms of the sensors and actuators employed. However, with the development of more challenging scenarios characterized by reversed magnetic "shear" ($dq/dr < 0$) profiles associated with transport barriers, combined magnetic and kinetic control has emerged as an essential tool to address research priorities.

The goal of kinetic control is primarily to control the performance of the plasma core, including fusion power output, energy and particle losses, and radiated power. Kinetic data are characterized by electron density, ion and electron temperature, impurity content, and current density profiles across the entire plasma cross section. Ultimately, kinetic real-time data are also expected to yield the real-time fusion power source profile. In contrast to magnetic control, the physics governing kinetic control is not well established.[69] In particular, heat and particle transport physics is still not understood or modeled well enough to simulate and validate the kinetic data and to design a nonlinear controller.[70]

All contemporary major tokamak experiments have developed a wide range of real-time diagnostics for kinetic control. The ECE diagnostic produces real-time electron temperature profiles on a routine basis in a large number of devices.[71] Challenging real-time hardware and analysis software have been developed for spectroscopic diagnostics such as CXRS (Ref. 72) and MSE (Ref. 73). Both of these diagnostics are now producing the ion temperature, the toroidal plasma velocity profiles, and the pitch angles in several devices. Visible spectrometry real-time analysis has also been set up to diagnose hydrogenic species (H/D/T). On the software side, off-line analysis has been used as a benchmark for the validation of the real-time analysis. The raw data undergo some basic checks (check for negative values, zero values, etc.) before being processed. However, there are also important technical issues that real-time data analysis has to take into account such as the presence, or otherwise, of the diagnostic neutral beam for charge exchange and MSE, or the knowledge of the toroidal field strength for the processing of ECE real-time data.

With these developments, a major milestone has been reached in the architecture of integrated real-time control of steady-state scenarios in contemporary devices. However, in a device such as ITER, the effects of neutrons, gamma rays, high levels of fusion power, and large quantities of alpha heating will bring new challenges for diagnostics and control systems, and advances will be required to allow their use in a routinely operating plasma control system of a reactor.

The requirements for profile control have also resulted in the implementation of real-time q profile measurements. In JET and Tore Supra, real-time infrared

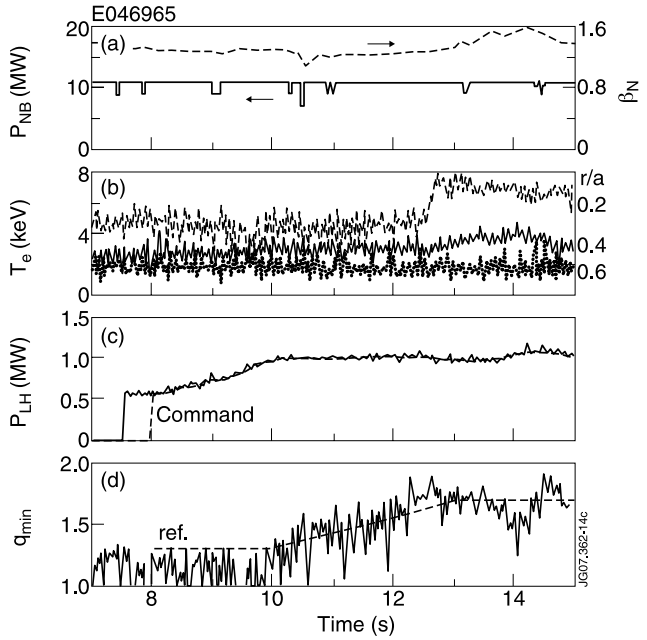

Fig. 25. Waveforms in the discharge with real-time current profile control. (a) Neutral beam injection power (P_{NB}) and normalized beta (β_N), (b) electron temperature, (c) lower hybrid power (injection: solid line and command: dashed line) and (d) q_{min} (real-time estimation: solid line and reference: dashed line). (*Figure reproduced, with permission, from Ref. 76.*)

polarimetry is used for feedback control of the q profile.[72] The q and density profiles are reconstructed in real-time by Abel inversion. The data are also used to constrain the real-time equilibrium reconstruction. In Tore Supra the hard X-ray diagnostic is also used to determine the real-time lower-hybrid deposition profile for current profile control.[74] For the computation of the current and q profiles, DIII-D and JET used a newly developed Grad-Shafranov solver with internal flux measurements from infrared polarimetry or MSE measurements. This facility is now routinely used in DIII-D for feedback control of the q profile in the current ramp-up.[75] JT-60U has also carried out feedback control experiments of the q profile using lower hybrid current drive as actuator[76] (Fig. 25). Though essential for a potential steady-state scenario and also for more accurate mapping of kinetic profiles onto flux coordinates, these real-time parameter data are not always stable because of inconsistencies within the raw input data. For the determination of the equilibrium and q profiles in real-time, it is important to combine data from several different (and sometimes inconsistent) sources such as magnetics, MSE, polarimetry, or kinetic pressure. This type of processing is not sufficiently developed to be routinely available. In addition, some effects, such as the contribution to the total pressure of fast particles generated by neutral beam injection (NBI) or ion cyclotron heating, are not yet accounted for.

V.C. Real-Time Sensor Issues for Limit Avoidance Schemes

In addition to optimizing performance, real-time control systems play a central role in maintaining the plasma discharge within allowed operational limits (particularly in view of instabilities that can lead to disruptions, radiation events, impurity influx, etc.) and the control of the power load on plasma-facing components. This function is to be separated from machine protection that leads to plasma shutdown. To identify the instability amplitude and location, real-time control systems often use sophisticated algorithms validated against extensive operational experience. Several machines, such as DIII-D (Ref. 59), and studies[55] have already shown that such systems will be mandatory in the integrated control system of a reactor.

For these reasons real-time measurements of plasma internal radiation (e.g., total radiated power and metallic impurity line radiation) have been developed. Metallic spectral lines (Ni, Fe, or Cu) are currently used to protect radio-frequency (rf) antennas from arcing on several devices. Plasma radiation from bolometers is also used to avoid the radiation limit at high plasma density using the radiation fraction as a sensor and gas injection as an actuator.[77]

In addition, real-time infrared imaging has proven to be an essential tool in plasma control and avoidance of the surfaces directly interacting with the plasma. For instance, it has been used to identify hot spots and their temperature on plasma components and antennas and to act on input power or plasma position.[78] This is achieved by careful analysis of the source of the hot spot (fast electrons generated by the lower hybrid launcher, rf sheath effect in front of ICRH antenna, etc.). The most offending areas are identified on the infrared image, and their temperature evolution is monitored in real time and linked with the appropriate control action.

Specific algorithms and measurements for real-time detection of internal plasma instabilities (magnetic coil combinations, ECE mode tracking, etc.) have also been implemented in several devices. For instance, the real-time stabilization of neoclassical tearing modes has been achieved in DIII-D using tracking of the 3/2 rational surface inferred from a real-time magnetic equilibrium reconstruction.[79] The stabilization and real-time detection of the $n = 1$ ideal kink mode is also crucial to many steady-state tokamak scenarios with high normalized pressure and high bootstrap current fraction. In the presence of a resistive wall, the kink mode becomes a slowly growing resistive wall mode (RWM), which can lead to a disruption. In DIII-D, RWMs have been effectively suppressed using magnetic error field coils as the actuator[80] (Fig. 26). It is also recognized that the ELM instability, which causes energy and particles from the edge of the plasma to be expelled transiently, will have to be mitigated to avoid damage to plasma-facing materials in

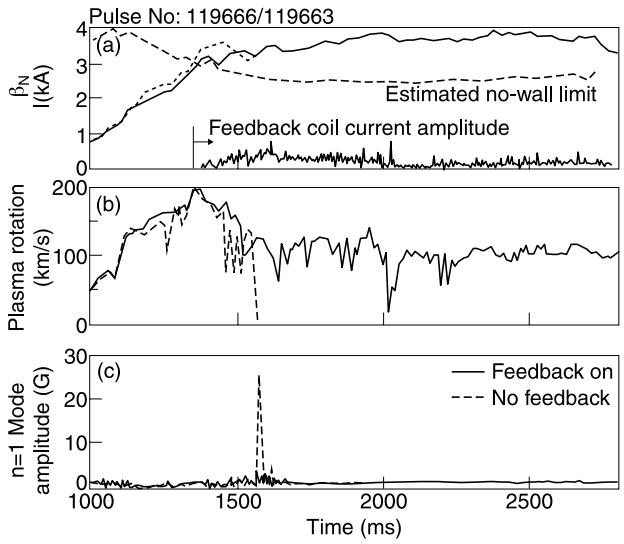

Fig. 26. Active wall mode stabilization using 12 internal coils in $n = 1$ symmetry for operation above the no wall limit in DIII-D for about 200 growth time. (*Figure reproduced, with permission, from Ref. 144.*)

steady-state discharges. Several techniques, such as edge resonant magnetic perturbation in JET (Ref. 81) and DIII-D (Ref. 82) or pellet pacing[83] in ASDEX Upgrade, have been successfully applied to mitigate ELM amplitude. The power load induced by ELMs can be recorded by infrared cameras. ELM detectors using visible spectroscopy are already implemented in a few devices, but more physics studies are necessary to quantify the effectiveness of actuators for the control of ELMs.

Finally, real-time disruption avoidance schemes are also under study in many devices. Significant effort is being directed at identifying the causes of disruptions or obtaining early warning of disruptions using the neural network technique,[84] which takes real-time plasma parameters as input (such as plasma current, internal inductance, radiated power, etc.). Although the latter technique looks attractive, it is difficult to extrapolate it to machines in which data are not included in the database. At present, the more promising approach appears to be identifying and acting on the causes of disruptions.

In the operation of contemporary machines, a key hardware issue for sensors for limit avoidance or plasma control is their reliability. This situation can be ameliorated with stricter qualification and commissioning requirements for diagnostics and data-processing chains for use in these applications. Adequate sensor redundancy together with long-pulse operation will bring new challenges of managing higher volumes of information. Going forward, the validation of real-time data for control is therefore likely to depend increasingly on sophisticated artificial intelligence and fault detection techniques.

VI. PHYSICS STUDIES

Physics studies in magnetic fusion research constitute a wide field, aimed at predicting and understanding plasma behavior in future devices. In turn, this guides the detailed designs of future devices and allows tools that address the control requirements of future devices to be developed and tested on contemporary machines. As two important examples of physics studies, confinement and transport studies are outlined in Sec. VI.A and VI.B. A more detailed discussion of these subjects can be found in Ref. 85.

Confinement studies aim at predicting the confinement time (see also Sec. III.A) on future devices by empirical scaling of data from a range of machines. The scalings are found to depend on a number of machine parameters, such as size and magnetic field strength, and thus play a central role in defining the machine parameters of future devices. Transport studies aim at obtaining a deeper and more detailed understanding of the processes governing plasma confinement and involve the development of transport models as well as their application to fusion plasmas. Validated models of heat and particle transport are essential to make predictive simulations of the fusion performance in future devices such as ITER. Such models are under continuous development, and considerable challenges need to be addressed before models achieve the required precision. The model development process includes model validation as an integral part, whereby strengths and weaknesses of models are identified by consistency checks against validated data and other models to direct the model development process.

VI.A. Global Confinement Studies

Empirical scaling studies attempt to find the relationship between measured physical quantities by statistically fitting parameterized formulas to experimental databases. The resulting scalings can be used to study physics trends in the data, to compare consistency of the experimental data with known physics models, or for the extrapolation of results to regimes outside of the existing dataset. The methods used for fitting the parameters to the dataset and testing the consistency of the resulting fits can also be applied to more complicated physics models, such as those discussed in Sec. VI.B. However, the greater simplicity of the databases and formulas used in scaling studies has meant that it has been practical to use more advanced statistical methods in these cases. Empirical scaling studies are thus discussed here since they provide both an important area of data application in themselves and illustrate statistical methods that are common to almost all comparison of measured data with empirical or theoretically derived physics models.

One important class of such studies aims at identifying scaling laws for the energy confinement time, a measure of the thermal insulation of the tokamak (see also

Sec. III.A). Two well-established modes of tokamak operation are the L-mode (low-confinement mode) and the H-mode. L-mode and H-mode confinement databases[86,87] containing data from many different sizes of machines were first assembled in 1989. Since then, data have been added to both databases so that now they both contain over 10 000 observations from some 20 different devices. Standard datasets of typically around 3000 observations are extracted, containing the highest-quality data from each device, and these are used in the data analyses.

An ordinary least-squares regression analysis has been completed for both the L-mode[86] and H-mode[88] databases. These expressions are as follows:

L-mode ITERL-97P:

$$\tau_\varepsilon = 0.023 I^{0.96} B^{0.03} R^{1.83} n^{0.40} \varepsilon^{0.06} \kappa^{0.64} M^{0.2} P^{-0.73} \quad (1)$$

and

H-mode IPB98(y,2):

$$\tau_\varepsilon = 0.0562 I^{0.93} B^{0.15} R^{1.97} n^{0.41} \varepsilon^{0.58} \kappa^{0.78} M^{0.19} P^{-0.69} \quad , \quad (2)$$

where

I = current (MA)
B = toroidal field (T)
n = density (10^{19} m^{-3})
P = power (MW)
R = major radius (m)
ε = inverse aspect ratio
κ = elongation
M = isotope mass.

These expressions are used extensively to predict the performance of the ITER device and in the design of future power plants.

Although the dependencies in these scalings agree well with those found in experimental studies on individual machines, for some time now there has been concern over the accuracy of the precise dependency of these expressions. Some single-scan studies, where one of the variables listed above is varied while the rest are kept fixed, do not reproduce the dependence precisely with respect to the power P and density n. For the H-mode scaling, this prompted a statistical review of the underlying database.[89,90] The scatter of the data about the fit is ~17% (Ref. 91) and this was found to be significantly above the measurement error in the fitted parameter, thermal energy, of ~14%. This implies that there is missing physics in the scaling and/or that some important source of errors has been ignored. Further, it was found that for both databases the condition is rather poor with respect to three of the variables: the elongation κ, the mass M, and the safety factor (in cylindrical geometry) q_{cyl}. This means that the independent variation of these variables is not significantly greater than their estimated errors. Ordinary least-squares fits to such poorly conditioned databases may result in biased results.[92]

Two different techniques have been used to remove this biasing. The first, which has been used on the H-mode database, is to reduce the number of variables used in the regression to five (I, n, R, ε, and P) and take narrow ranges in the other variables (κ, M, and q_{cyl}). This effectively removes variables known to be highly uncertain from the fit. The cost is that the fit no longer describes the dependence of confinement on the removed variables, and so its applicability is reduced. This dataset is found to be well conditioned with respect to the five variables, and the ordinary least-squares regression has the following form:

$$\tau_\varepsilon = 0.096 I^{1.06} R^{1.78} n^{0.39} \varepsilon^{0.56} P^{-0.61} \quad . \quad (3)$$

This form does indeed have a weaker dependence on power, but is still not as weak as that found in the single-scan experiments. This approach was not viable for the L-mode database because a suitable range for restricting the elongation variable to produce a well-conditioned dataset could not be found.

The second technique, which has been used on both the L-mode[93] and H-mode[90] databases, is an "errors-in-variables" method.[92] Errors-in-variables methods are less susceptible to the biasing that results from poorly conditioned datasets. However, these methods are numerically more complicated and also require reliable estimates of the errors in all of the regressor variables (I, B, n, etc.) as well as the errors in the measurements of τ_ε. Compared with the ordinary least-squares regressions, errors-in-variables fits to the H-mode and L-mode databases had a significantly different scaling of energy confinement with respect to the power P and density n. For the H-mode database, the resulting scaling was

$$\tau_E = 0.0150 \cdot I^{0.75} B^{0.32} R^{1.23} n^{0.35} a^{-1.53} A^{1.14} M^{0.06} P^{-0.61} \quad , \quad (4)$$

where A is the plasma cross-sectional area (m^2) and a is the minor radius (m). In comparison with the ordinary least-squares fit of Eq. (2), it can be seen that the scaling with power has become less negative and the scaling with density has become less positive. As for the method where the number of variables was reduced, the fits with the errors-in-variables method have been shown to be more consistent with the single-scan experiments than the ordinary least-squares regressions.[90]

When the H-mode scalings from the three methods of Eqs. (2), (3), or (4) are applied to the ITER [baseline point[53] ($\beta_N \equiv 1.8$)], it is found that the predicted energy confinement time is almost identical. However, at higher

β_N, which may be accessed in future fusion power stations, the recently derived expressions of Eqs. (3) and (4) give higher values for τ_E.

VI.B. Transport Studies

One of the main goals of tokamak plasma measurements is to develop our understanding of plasma transport and to capture this knowledge in transport models. Validated plasma data are essential to test such models. In this section the approach to validating transport models is described. The ultimate goal of this activity is to produce a transport model which is sufficiently robust to make reliable predictions for new experiments and future devices. Present-day transport models can be used to make such predictions (e.g., for the confinement time in ITER) and provide valuable indicators, but their accuracy and range of validity are limited. As these models become more mature, it is conceivable that their predictions could also be used directly in data validation.

VI.B.1. Data Required to Test Transport Models

Tokamak transport models provide, for each plasma species s, the radial plasma heat flux q_s and radial particle flux Γ_s as functions of the flux surface label coordinate. This information can be combined with radial profiles of the heat source P_s and particle source S_s to determine the time evolution of the plasma density n_s and temperature T_s profiles by solving the one-dimensional heat q_s and particle diffusion Γ_s equations:

$$\frac{\partial}{\partial t} W_s + \frac{1}{V'} \frac{\partial}{\partial \rho} V' \langle |\nabla \rho| \rangle q_s = P_s$$

and

$$\frac{\partial}{\partial t} n_s + \frac{1}{V'} \frac{\partial}{\partial \rho} V' \langle |\nabla \rho| \rangle \Gamma_s = S_s \ ,$$

where W_s is the energy density of species s ($W_s = 3 n_s T_s / 2$), ρ is any flux surface label (ρ is commonly taken as the square root of the normalized toroidal flux), V is the plasma volume inside a given flux surface, primes denote derivatives with respect to ρ, and angle brackets denote flux surface averaged quantities.

Transport models, which are discussed in Sec. VI.B.2, can be tested using experimental data in two ways. First, given the equilibrium metric coefficients V' and $\langle |\nabla \rho| \rangle$, heat source $P_s(\rho, t)$ and particle source $S_s(\rho, t)$, and initial conditions for the profiles $n_s(\rho, 0)$ and $T_s(\rho, 0)$ from experimental measurements, the transport equations can be solved using the transport model's prescription for the heat and particle fluxes. The modeled temperature and density profiles may then be directly compared with the measured profile evolution. In a second approach, the heat and particle fluxes are estimated directly from experiment and are compared directly with the transport model predictions of the fluxes. Transport coefficients are sensitive quantities that can vary dramatically over short time and length scales, and it is usually more convenient to compare temperature and density profiles, which are more resilient because they respond to time and spatial averages of the more transient fluxes.

Obtaining the data that is required for the testing of tokamak transport models from experiments is challenging, and uncertainties may be large and highly correlated. Heat and particle sources are especially difficult to determine and usually require modeling themselves. Experiments may have multiple diagnostics to measure the same physical quantity, and statistical techniques (e.g., Bayesian methods[94]) can be used to help to maximize data consistency in the interpretation of a range of experimental measurements. Computational tools [e.g., tokamak plasma transport simulation codes such as ASTRA (Ref. 95), CRONOS (Ref. 27), JETTO (Ref. 96), and TRANSP (Refs. 24 and 25)] are used to perform power and particle balance calculations and enable comparisons between the fluxes derived from transport models with those measured experimentally. Equivalently, many of these codes can also be used to apply the transport models to predict the plasma profiles, which can then be directly compared with experimental profile measurements. Such codes are also used to optimize data consistency; that is, to find values for the physical quantities being measured that are most consistent with the full range of experimental measurements. In solving the transport equations (1) for the plasma profiles, transport codes require the source and sink terms, which may either be prescribed or modeled inside the transport code itself, and the boundary conditions, e.g., the edge temperature profile.

VI.B.2. Transport Models

Neoclassical transport theory has yielded formulas for the plasma fluxes that arise from collisional processes in the absence of turbulent fluctuations, but the experimentally measured fluxes are normally considerably larger than neoclassical theory's predictions, owing to the important role of turbulent fluctuations in the plasma equilibrium. One-dimensional transport models that predict the flux surface averaged radial fluxes of heat and particles as functions of minor radius have been developed to model the anomalously large transport that is observed, and a number of such transport models have been reviewed in Ref. 85. Many of these transport models are guided by first-principles physics considerations, such as models that are based on mixing-length estimates of the anomalous transport expected to arise from particular classes of microinstabilities (e.g., ion temperature gradient driven drift waves, trapped electron modes, microtearing modes, etc.). Other transport models have been obtained rather more empirically by fitting experimental transport observations from various confinement

regimes and devices. Detailed and up-to-date descriptions of many transport models are given in Ref. 85.

VI.B.3. ITER International Multitokamak Confinement Profile Database

The ITER confinement profile database[97] started to be accumulated in the mid-1990s (Ref. 98), in order to test transport models more extensively than had ever previously been attempted. The main aims of this exercise were to discriminate between the ability of the models to describe present-day devices and to exploit this knowledge to improve our confidence in predictions of the performance of next-step fusion devices such as ITER. Initially, 11 of the world's leading tokamak experiments contributed analyzed data from approximately 200 discharges. The database provided a convenient way for physicists to access data from a wide range of different tokamaks, facilitating the study of important tokamak confinement issues across machines. The collected data contained the physical quantities, which are essential for the testing of one-dimensional transport models, and constituted a broad and representative set of reference discharges with a range of different heating methods and from various confinement regimes (H-mode, L-mode, etc.). The database was made available to the public in late 1998 and has been described in detail in Ref. 97.

A collaboration of transport modelers set out to solve the transport equations (1) with an agreed standard set of assumptions (e.g., identical boundary conditions, sources, sinks, etc.), and to provide stand-alone program modules for each of the various transport flux models that were to be tested. This allowed the transport codes and their numerical schemes to be benchmarked and each transport flux model to be tested rigorously. It was a particular priority to test each transport flux model using transport codes that were independent of the model developer. A standard set of discharges were selected from the profile database, and these were well diagnosed and in steady confinement regimes considered to be of most relevance to ITER operation (L-mode and ELMy H-mode). Discharges were only selected if the plasma profiles evolved to conditions that were sufficiently close to steady state, and a single steady-state time slice was selected from each discharge. Heat transport equations were solved to obtain the corresponding steady-state temperature profiles. The density, heat source, and heat sink profiles were prescribed and taken from the best data and models that were available to each experiment. The edge boundary condition was chosen so that the modeled temperatures matched the experimental values at $\rho = 0.9$ (since the models are not expected to reproduce the behavior of the plasma edge). Figures of merit were devised to quantify the discrepancy between the modeled and experimental profiles; in particular, the incremental stored energy W_{inc} is defined as

$$W_{inc} = \frac{3}{2}\int [n_e(\rho)(T_e(\rho) - T_e(0.9)) + n_i(\rho)(T_i(\rho) - T_i(0.9))]\, dV \ .$$

The root-mean-square error ΔR in the incremental stored energy averaged over the N discharges is defined as

$$\Delta R = \sqrt{\frac{1}{N}\sum\left(\frac{W_{inc}^{sim}}{W_{inc}^{\exp}} - 1\right)^2} \ ,$$

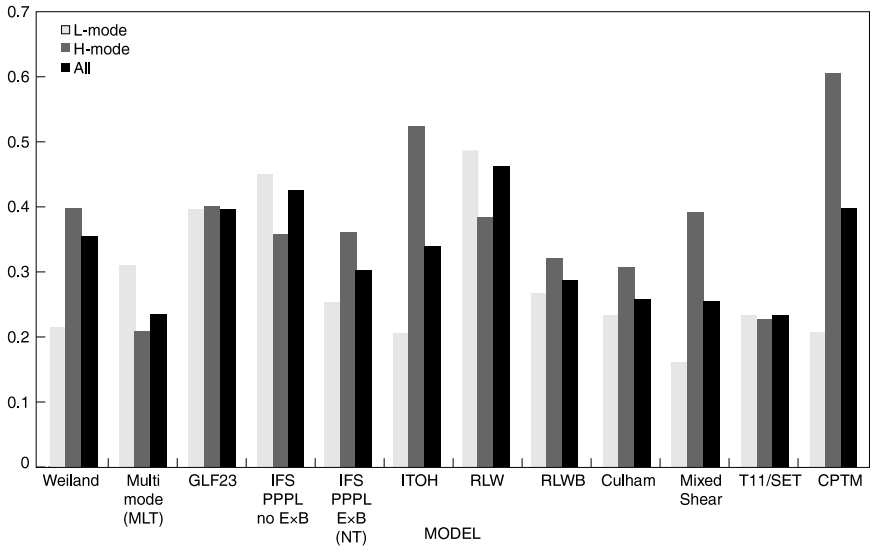

Fig. 27. Root-mean-square error in the incremental stored energy W_{inc} simulated by twelve transport models for the 55 discharges, which have measured ion temperature profiles. *(Figure reproduced, with permission, from Ref. 97.)*

where W_{inc}^{sim} and W_{inc}^{exp} are the simulated and experimental values, respectively, of the incremental stored energy. Figure 27, from Ref. 97, plots ΔR for each of the tested transport models and shows poor differentiation between the various transport models. This result was disappointing; it remains unclear to what degree the transport models themselves are in error, and whether the assumptions (metrics, sources, etc.) were determined with sufficient accuracy that a perfect transport model would show better agreement with the data (note also that, as discussed in Sec. VI.A, the uncertainty in experimental stored energy measurements is estimated at ~14%). More sophisticated statistical analysis of the various uncertainties, their correlations, and error propagation are probably required in order to answer such challenging questions more definitively.

Despite poor discrimination between models, the model testing exercise was an extremely valuable one that led to improvements in models and codes and uncovered a number of key issues for next-step devices. Many of the physics-based transport models predict a sharp rise in anomalous transport above a critical value of R/L_T [where R is the major radius and $L_T = T/(dT/dr)$ is the temperature gradient scale length], and the anomalous transport fluxes are zero below the critical value. These so-called stiff models predict that the stored energy is highly sensitive to the edge temperature and that core confinement improves substantially if the pedestal temperature can be increased. Improving our understanding of edge physics is presently a high-priority area of tokamak research, not only directly through the importance of edge phenomena such as ELMs but also indirectly through the edge influence on the core.

The ITER profile database continues to serve the fusion community and is presently held and maintained at the Culham Laboratory in the United Kingdom, under the International Tokamak Physics Activity framework; it is accessible via the Internet (see Ref. 99). Documentation is available to describe the various tools for easily accessing the database and interfacing it to physics codes. The version of the profile database which was used in the modeling exercise was made available to the public in 1998, and since that date approximately 100 new discharges have been added, including world record fusion power discharges from D-T plasma operation in JET and TFTR; discharges with internal transport barriers (ITBs) from JET, JT60-U, and DIII-D; high-performance hybrid scenario discharges from DIII-D and JT60U; H-mode parameter scans from AUG; low aspect ratio plasmas from the spherical tokamak MAST; and electron heated discharges from FTU, T10, and Tore Supra. In addition to providing convenient access to data from a range of experimental devices, the profile database is being used by modelers to store their predictive calculations of ITER plasmas in a standardized format, which is valuable because it allows key modeling assumptions to be scrutinized in detail by the broader modeling community.

An observation that has attracted much attention in recent years is that tokamak confinement can improve significantly by tailoring the q profile, with the formation of an ITB deep within the plasma. Improving our understanding of these phenomena may help the achievement of high-performance burning plasmas in next-step devices. Study of the physics underlying ITBs via international multimachine comparisons has been facilitated using the profile database.[100] In particular, this analysis has reported two important findings. First, in conditions where ITBs can be generated, a threshold input power P_{th} is required to generate the ITB, and this is reduced with negative magnetic shear s, where $s = r/q\ dq/dr$. Second, in ITBs the growth rates of microinstabilities lie close to the radial equilibrium flow shear,[b] which can suppress the drives for turbulent fluctuations. Gyrokinetic calculations find reduced growth rates at negative magnetic shear, so that reduced power thresholds at negative magnetic shear appear to be associated with reduced drives for microturbulence. High-pressure gradients in ITB plasmas increase the parameter $\alpha = -q^2\beta R/p\ dp/dr$. Gyrokinetic calculations have demonstrated that high α has an important stabilizing influence on microstability, and comparisons with data confirm that this theoretical model is indeed important in experimental ITB plasmas.[101] Describing profile evolution in ITB plasmas is challenging in present-day transport models, and several models have been modified in order to attempt to address this.

Another important goal in progress toward a fusion reactor is demonstration of steady-state operation. Discharge scenarios are being sought where pulse lengths can be extended by reducing the requirement for inductive current (driven by transformer action and hence limited in duration by the flux swing capacity). This can be achieved either by applying alternative auxiliary current drive mechanisms or by increasing the self-driven bootstrap current. Advanced-tokamak and hybrid scenarios, where the plasma current driven by the solenoid is small compared to the noninductively driven current, have been developed on a number of machines over the last few years in order to work toward this goal. The plasma current is usually reduced in these scenarios (thereby reducing confinement), but it has been found experimentally that with careful tuning of the plasma current profile the energy confinement can be optimized. Considerable effort has been invested in trying to understand the influences on plasma confinement in these regimes. A number of such discharges from a range of machines have been added to the profile database to make them accessible to modelers for detailed transport modeling and microstability analysis. The comparison of state-of-the-art transport modeling simulations with data from experiments with hybrid plasmas has been reported

[b] The equilibrium radial electric field generates $E \times B$ flows that lie in the flux surfaces perpendicular to the magnetic field. The radial equilibrium flow shear is the rate of change of this flow with respect to minor radius.

in Refs. 102 and 103, where projections have also been made to assess the viability of the hybrid scenario in ITER. The quality of the match between simulations and data was mixed, and projections to ITER using a range of plausible transport models predicted significant scatter in the fusion performance parameter Q, $4 < Q < 13$, where Q is the ratio of fusion power to total plasma heating power. The projection of current drive performance to ITER is another very important area of study, and the status of modeling in this area has been presented in Ref. 104.

The increasing power of computation is making challenging nonlinear gyrokinetic microturbulence simulations with physically reasonable assumptions increasingly feasible. Gyrokinetic codes are advancing rapidly and are able to exploit modern supercomputers effectively. So far, no single calculation has yet included all of the following: global plasma geometry, full electron and ion physics, the full magnetic perturbation, particle collisions, particle and heat sources, and sinks. Nevertheless, progress is rapid and these calculations are increasingly true to nature, which makes it increasingly appropriate to compare them with experiment. We can hope to see considerable improvement in our understanding of tokamak plasma confinement from more routine comparisons of these first-principles models of turbulent plasma transport with experiment. Parallel advances in diagnostics will also permit detailed comparisons between the experimental fluctuations and those predicted by these models.

VII. CONSIDERATIONS FOR NEXT-STEP DEVICES

Contemporary magnetic fusion experiments have reached a scale and level of complexity such that the systems and processes developed and operational experience gained for the validation and analysis of data are directly relevant to next-generation burning plasma experiments such as ITER. There are, nevertheless, several unique features of ITER that will have important implications for data validation and analysis. Notably, in ITER access to diagnostic hardware will be reduced (or entirely precluded in some cases), diagnostics will be operated in the much harsher environment of strong neutron and gamma-ray fluxes, there will be high alpha particle pressure and strong alpha heating, and long steady-state discharges will become the norm.[105,106] ITER will also produce larger data volumes and will serve a larger and more disparate community of scientists.

Reduced access to diagnostic hardware will mean that diagnosing and correcting system failures will become more difficult. Known faults in data will tend to persist. To mitigate against this, more redundancy and reliability in diagnostics will be required, and the use of efficient data analysis techniques such as IDA will be important, in order to extract the maximum output information from available data. The effects of the environment on individual diagnostic components and systems need to be understood, and this may highlight the need for new analysis tools to allow identification and compensation of the prompt and long-term effects of neutron and gamma irradiation.

Steady-state long-pulse operation brings new challenges for plasma control, with the need for more real-time diagnostics. Increased use of intelligent tools may be necessary to manage the processing load associated with the validation of real-time data and to allow timely corrective action when faults arise. Furthermore, diagnostic redundancy to allow such corrective action may have to be increased, in view of the reduced access to reinstate faulty diagnostics.

Finally, improved infrastructure will be needed to allow the higher data volumes to be analyzed by a more geographically disparate community. With rapid advances in remote collaboration and the adoption of grid computing by several fusion facilities, progress in this direction is well underway. There have been major developments in this area, and these are outlined in detail in the Appendix.

VIII. SUMMARY

Data validation in magnetic fusion research aims at early identification of faulty data, which may arise from system failures or faulty physics models, in order that timely corrective action may be taken to minimize the impact of the fault on the experiment and physics conclusions that are drawn from it. System failures can result in loss of data, and in such cases, corrective action is required to eliminate the problem for the future. By contrast, the effects of faulty physics models are in general reversible, if the model can be corrected. The process of data validation involves checks of consistency in information obtained from different diagnostics, and may range from rudimentary (often fully automatic) to highly complex (requiring extensive manual intervention).

A powerful method to assess and improve the consistency of plasma data is via comparison of measurements of the same physical quantity obtained from different diagnostics. This approach is applicable to the measurement of a range of plasma parameters such as the ion temperature, plasma impurities, and neutron yields. Most of the mainstream plasma diagnostics stand to benefit when this approach is systematically used in measuring key plasma parameters.

As an alternative to the conventional approach of checking measurements from different diagnostics for consistency, IDA is an approach in which heterogeneous information from different diagnostics is analyzed together, in a Bayesian probabilistic framework, yielding a "superfit" to all available information. The method typically makes use of physics interdependencies between

different measurements, together with models of diagnostics. With respect to conventional analysis, this approach increases the reliability of results, can yield new information, and includes comprehensive error analysis. The quality enhancement by IDA will be especially important for next-generation fusion devices, where diagnostic data could be degraded in the harsher environment and where access will be reduced.

Two categories of data usage in fusion research can be defined: plasma control and physics studies (confinement and transport studies, as important examples), involving the validation of physics models. A prerequisite for safe and efficient exploitation of magnetic fusion devices is the provision of validated real-time data for control of the magnetic configuration, plasma profiles, and first-wall loading. The requirement is to detect and eliminate faults in data quickly enough to avoid any adverse effects on the plasma. To achieve this, faults may be detected by comparing similar measurements or by reference to past data, and any faulty measurements found can then be removed from the processing chain in real time. Redundancy must be built into control diagnostics to allow operation to continue uninterrupted when a faulty measurement is excluded. In recent years, this general approach has been applied to increasingly sophisticated measurements in view of the rapidly advancing field of plasma control, which has evolved from magnetic configuration control to the control of plasma profiles, plasma instabilities and first-wall loading.

An important example of physics studies is the development of scaling laws for the global energy confinement. Scaling laws for tokamaks have been obtained from regression analysis of a large database containing observations from some 20 different machines. These are the basis for performance predictions for future tokamaks. Refinement of these laws is ongoing, with recent improvements in the regression technique indicating that in larger tokamaks performance at high β_N may be more favorable than previously predicted. A similar approach is taken to obtain confinement scaling laws for stellerators.

Another important example of physics studies is the validation of transport models. First-principles models of turbulent plasma transport represent our best understanding of plasma confinement and constitute an additional tool to predict performance on future fusion devices. Model validation aims at comparing predictions made by various models with experimental data in order to identify the most realistic models. To facilitate this activity, a multitokamak profile database has been set up to make use of the best available experimental data, and a standardized approach has been developed for code benchmarking. Although detailed and quantitative discrimination between models is found to be poor at present, many advances have been made in improving the qualitative understanding of the key physics elements that influence transport, and there is cause for optimism for the future. Increasing computational power is allowing more realistic assumptions to be brought into the models and more routine comparisons with experiment. In addition, advances in plasma diagnostics and data analysis techniques are expected to allow more detailed comparisons with experiment.

Data validation and analysis is a community exercise, involving diagnostic physicists, software developers, and operational physicists. As a result of the growing international character of fusion research, the community working on any one device is frequently distributed in several countries. To enable this community to work together effectively, a range of tools for remote collaboration have been developed, providing teleconferencing, access to documentation, data and computing power, and remote monitoring of diagnostics during experiments. Expanded use of such facilities is foreseen in the future, together with enhancements in the collaboration tools to improve the quality of interaction and to cater for growing data volumes and computing power requirements of fusion research.

Looking to the future, with the growing scale and complexity of contemporary fusion devices, much of the experience being gained in data validation is of direct relevance to next-step devices. Nevertheless, ITER will introduce new challenges due to reduced access to diagnostics, a harsher environment for diagnostic components, longer plasma pulses, and larger data volumes. It is encouraging that the magnetic fusion community has already begun to address many of these challenges as well.

APPENDIX

REMOTE COLLABORATION IN MAGNETIC FUSION RESEARCH

In recent years the growth in the international nature of fusion research has resulted in an increased reliance on remote participation technology. With growing use of shared facilities in the future, this trend is likely to continue.

The basic techniques of remote computer access (RCA) and remote data access (RDA) coupled with remote access to documentation and teleconferencing facilities allow dispersed scientists to analyze data, monitor diagnostics, and collaborate on experiments in real time as if they were physically at the same experimental site. All the techniques rely on adequate network connectivity and must be integrated with remote and local information technology (IT) security polices.

A.I. REMOTE COMPUTER ACCESS

The RCA technique provides a secure remote computer connection to a site network to allow remote users

to run applications on the site computers as if they were physically present. Examples of the applications that are commonly run in this way are analysis and predictive codes, control room tools to monitor experimental progress, pulse preparation tools, and office computing.

The degree of security required depends on the IT security policies at the remote and local sites. Techniques used include strong authentication using a personal identification number and one-time passcode associated with a hardware or software token, virtual private networks[107] (VPNs), VPNs combined with secure shell[108] (SSH), and SSH combined with public key infrastructure X.509 certificates.[109] RCA typically provides a connection to an off-line part of a segregated site network, allowing access to experimental data and read-only access to plant monitoring systems. Further security constraints must be satisfied before write access to the plant control networks is provided. This is analogous to someone working in an office at the site from where they would not be able to alter plant settings without additional security measures. Examples of the RCA methods used include Windows servers running CITRIX metaframe,[110] virtual network computing[111] (VNC) running on Windows or Unix, and SSH (Ref. 108) running on Unix. The key feature of all of these is the availability of client software for all major computer platforms.

A.II. REMOTE DATA ACCESS

The RDA technique provides access to fusion data (servers) at remote locations across a network (usually the Internet) allowing remote users to run display and analysis applications (clients) on their own computer networks. Usually data are returned at the signal rather than the file level. All of the knowledge of the data structures and low-level data access routines is hidden in the server and only a small amount of code is required in the client application, thus simplifying installation and management.

The de facto standard for RDA is MDSplus (Ref. 112), which was developed in the early 1990s by MIT, LANL, and RFX Padova. From Ref. 113: "MDSplus allows all data from an experiment or simulation code to be stored into a single, self-descriptive, hierarchical structure." Some fusion experiments (e.g., Alcator C-Mod, RFX, NSTX, TCV, and KSTAR) use, or plan to use, MDSplus as their primary data acquisition and archive system and hence can use the RDA features of MDSplus directly. Others (e.g., JET, ASDEX Upgrade, DIII-D, and Tore Supra) have interfaced MDSplus on top of their existing raw and processed data archive systems. Interfaces have been written for most popular programming languages, including FORTRAN, C, perl, python, IDL, Matlab, and Scilab. Most fusion data analysis and data display programs are now able to read data from MDSplus servers, and most fusion experiments provide MDSplus server access to their data. This symmetry, sometimes combined with RCA, allows physicists to use their favorite tools no matter where they are located or which data they wish to access.

Control of access to the MDSplus servers residing on the site networks depends on the site IT security policies. Read-only MDSplus access is typically granted to a list of Internet-protocol (IP) names defined as the "fusion community". Further restrictions are put in place for write access. For example, a version of MDSplus using Globus,[114] which requires the client application to supply an X.509 certificate,[109] is available. RCA and RDA may be seen as complementary techniques for remote data analysis. The choice of which one to use depends on the availability of software tools at the different sites, the quantities of data that will be transferred, the network capacity, the site IT security policies, and also on personal preference.

A.III. REMOTE ACCESS TO DOCUMENTATION

Web technology is now used by all experiments to provide descriptions of diagnostics, descriptions of raw and processed signals, etc. RCA can be used to access intranets locally. Access across the Internet is controlled by the IT security policies in place at the various sites. At JET, three tiers of web servers are in place providing (a) general public access, (b) access to the fusion community as defined by a list of IP addresses, and (c) local intranet access, which is only available remotely using RCA (Ref. 115). More sophisticated solutions based on setting up trust relationships between federations of sites using PAPI software[116] are being developed.[117]

A.IV. TELECONFERENCING

Most fusion laboratories have now installed one or more of the IP-based videoconferencing systems [H.323 (Ref. 118), Virtual Room Videoconferencing System[119] (VRVS), and Access Grid[120] (AG)] to cater for meeting types ranging from one-on-one desktop meetings through medium-sized working meetings and remote control room communication to formal large-scale shared presentations.

A family of standards approved by the International Telecommunications Union,[121] H.323 defines how audiovisual conferencing data are transmitted across networks. Several hardware and software vendors support the protocol. VRVS is an Internet-based videoconferencing system developed by Caltech and originally intended for the CERN computing collaboration. It is based on the concept of virtual rooms where participants gather as if in the same physical room. Some dedicated fusion rooms have been created. The AG system is a set of tools including multimedia large-format displays, presentation, and interactive environments. Interoperability between these systems is possible but can be problematic. Ideally, all fusion laboratories would adopt a single standard. Today, the most likely candidate for fusion is H.323 (Ref. 122).

For desktop sharing, and hence transmission of electronic presentations, there is a general consensus to use VNC. VNC is embedded in the AG software and is widely used by sites broadcasting meetings based on H.323 and VRVS (Refs. 123 through 126).

A.V. PARTICIPATING IN EXPERIMENTS REMOTELY

The ability to follow, and participate in, experiments from remote locations is provided by combining the techniques of RCA, RDA, remote access to documentation, and teleconferencing.

Existing control room tools can be used via RCA and could be used to set plant parameters if local rules allowed access to the plant control networks. Most fusion experiments distinguish between control room physics and engineering roles, with the engineers responsible for machine safety. When this is the case, the location of the physicists is not important, provided that adequate teleconferencing connections to the control room staff are available.[127] Remote control of diagnostic settings via web applications[128] has been provided in some cases. Some experiments broadcast control room screens in read-only mode on the Internet using VNC technology,[125] and there is also a general trend to make these tools and interactive data display and analysis tools available across the Internet via web browsers, thus avoiding the requirement for a remote computer log-on; see, for example, Refs. 129, 130, and 131.

The RCA technique can also be used to allow remote software maintenance of diagnostic data acquisition systems, which is a requirement for systems where the maintenance responsibility rests with the institute that installed the diagnostic.

The U.S. FusionGrid[132] is deploying more elaborate technologies, including the use of AG for complex communication and large shared display walls for data sharing.[123,127]

A.VI. GRID TECHNOLOGY

The driving force behind the emergence of the GRID (Ref. 133) has been high-energy physics (HEP), especially the LHC experiment, which is predicted to collect 15 petabytes of raw data annually soon after it comes online in 2007. The analysis and storage of such large data volumes pose new challenges for the computing infrastructure. The LHC Computing Project aims "to build and maintain a data storage and analysis infrastructure for the entire HEP community that will use the LHC" (Ref. 134). This project is now part of a wider project called Enabling GRIDS for E-SciencE[135] (EGEE). EGEE integrates GRID infrastructure and applications from many scientific disciplines, including fusion.[136]

Experimental data volumes in fusion are much more modest than in HEP. JET raw data volumes have followed Moore's Law for more than twenty years, roughly doubling every two years.[137,138] A simple extrapolation from JET to ITER predicts <10 petabytes of raw data will be collected per year when ITER starts operating. With the increase in disk and tape densities expected over the next 10 yr, it is probable that the ITER raw and processed data could be stored at the ITER site without resorting to the multitier transfer strategy adopted by LHC, although the methods developed for the GRID might be used to make multiple geographically separated copies of the data for reasons of data security. At present, relatively small local clusters are adequate for the usual intershot analysis of fusion experimental data, and this is also likely to be the case for ITER.

The U.S. FusionGrid has made larger more-complex codes like TRANSP available as network services using Globus[139] and has also tried running reduced TRANSP runs between pulses in the same way.[140] The benefit of this approach is that the codes only have to be supported at one location. This trend is likely to continue.

Large-scale simulation codes, however, are predicted to require the resources available via GRIDs in the near future. The European Fusion Development Agreement Integrated Tokamak Modelling (ITM) Task Force[141] predicts $\sim$1000 teraFLOPS (floating-point operations per second) will be required to simulate 1 min of an ITER burning plasma.[142]

The use of the GRID in fusion research is in its infancy. Peak computing demands like those predicted for the ITM could be met by pooling the ITER partner computing resources into an international fusion GRID.

ACKNOWLEDGMENTS

The authors wish to thank the many workers who have contributed to work presented in this paper. The authors are particularly grateful to the following people for their contributions and helpful comments during the writing of this paper: M. Brix, V. Dose, H. Dreier, E. de la Luna, L.-G. Eriksson, Georg Kühner, A. Murari, R. Preuss, U. von Toussaint, L. Zabeo, and K.-D. Zastrow.

REFERENCES

1. S. FLANAGAN et al., *Rev. Sci. Instrum.*, **75**, 2043 (2004).

2. G. BRAITHWAITE et al., *Rev. Sci. Instrum.*, **60**, 2825 (1989).

3. A. MURARI et al., *Rev. Sci. Instrum.*, **77**, 10F529 (2006).

4. H. GRAD and H. RUBIN, "Hydromagnetic Equilibria and Force-Free Fields," *Proc. 2nd United Nations Int. Conf. Peaceful Uses of Atomic Energy*, Geneva, Switzerland, September 1–13, 1958, Vol. 31, p. 190, Columbia University Press, New York (1959).

5. V. D. SHAFRANOV, *Zh. Ehksp. Teor. Fiz.*, **33**, 710 (1957) and *Sov. Phys.—JETP*, **6**, 545 (1958).

6. L. L. LAO et al., *Nucl. Fusion*, **25**, 1611 (1985).

7. P. J. MCCARTHY, *Phys. Plasmas*, **6**, 3554 (1999).

8. R. ALBANESE and F. VILLONE, *Nucl. Fusion*, **38**, 723 (1998).

9. D. P. O'BRIEN et al., *Nucl. Fusion*, **33**, 467 (1993).

10. L. E. ZAKHAROV and A. PLETZER, *Phys. Plasmas*, **6**, 4693 (1999).

11. W. ZWINGMANN et al., "Equilibrium Reconstruction of Tokamak Discharges Variation," *Europhysics Conference Abstracts: 32nd Conf. Controlled Fusion and Plasma Physics*, Tarragona, Spain, June 27–July 1, 2005, Vol. 29C, P-2.044, European Physical Society (2005).

12. S. A. SABBAGH et al., *Nucl. Fusion*, **46**, 635 (2006).

13. S. D. BAILEY et al., *Chain2 User's and Developer's Manual*, JET Data Note, JDN/T1(02)003, Joint European Torus (2002).

14. B. ALPER et al., *Rev. Sci. Instrum*., **68**, 778 (1997).

15. J. WESSON, *Tokamaks*, Oxford University Press, London (2004).

16. J. MAILLOUX et al., "Development of ITB Plasmas at High β_N and high δ in JET," *Europhysics Conference Abstracts: 34th Conf. Controlled Fusion and Plasma Physics*, Warsaw, Poland, July 2–6, 2007 (to be published).

17. D. V. BARTLETT et al., in *Proc. 9th Joint Workshop Electron Cyclotron Emission and Electron Cyclotron Resonance Heating*, Borrego Springs, California, January 23–26, 1995, p. 511, World Scientific Publishing, Singapore (1995).

18. H. SALZMANN et al., *Rev. Sci. Instrum*., **59**, 1451 (1988).

19. E. DE LA LUNA et al., *Rev. Sci. Instrum*., **74**, 1414 (2003).

20. C. SOZZI et al., "First Measurements of the Oblique ECE System at JET," *Europhysics Conference Abstracts: 34th Conf. Controlled Fusion and Plasma Physics*, Warsaw, Poland, July 2–6, 2007 (to be published).

21. L.-G. ERIKSSON et al., *Nucl. Fusion*, **39**, 337 (1999).

22. L.-G. ERIKSSON et al., *Nucl. Fusion*, **33**, 1037 (1993).

23. D. ANDERSON et al., *Nucl. Fusion*, **34**, 217 (1994).

24. R. J. HAWRYLUK, "An Empirical Approach to Tokamak Transport," *Physics of Plasmas Close to Thermonuclear Conditions*, B. COPPI et al., Eds., Vol. 1, pp. 19–46, European Commission, Brussels (1980).

25. R. J. GOLDSTON et al., "New Techniques for Calculating Heat and Particle Source Rates due to Neutral Beam Injection in Axisymmetric Tokamaks," *J. Comput. Phys.*, **43**, 61 (1981).

26. R. V. BUDNY et al., *Nucl. Fusion*, **35**, 1497 (1995).

27. V. BASIUK et al., "Simulations of Steady-State Scenarios for Tore Supra Using the CRONOS Code," *Nucl. Fusion*, **43**, 822 (2003).

28. K. D. ZASTROW et al., *Plasma Phys. Control. Fusion*, **46**, B255 (2002).

29. S. POPOVICHEV et al., "Performance of Neutron Measurements during Trace Tritium Experiments on JET," *Europhysics Conference Abstracts: 31st Conf. Controlled Fusion and Plasma Physics*, London, United Kingdom, June 28–July 2, 2004, Vol. 28G, P-5.173, European Physical Society (2004).

30. T. W. FREDIAN and J. A. STILLERMAN, *Fusion Eng. Des.*, **81**, 1963 (2006); see also http://www.mdsplus.org.

31. S. DORMIDO-CANTO et al., "Search and Retrieval of Plasma Wave Forms: Structural Pattern Recognition Approach," *Rev. Sci. Instrum.*, **77**, 10F514 (2006).

32. SAS OnlineDoc, version 8; available on the Internet at http://v8doc.sas.com/sashtml (1999).

33. J. P. CHRISTIANSEN, "Integrated Analysis of Data from JET," *J. Comput. Phys.*, **73**, 85 (1987).

34. G. TONETTI et al., "Measurement of the Energy Content of the JET Tokamak Plasma with a Diamagnetic Loop," *Rev. Sci. Instrum.*, **57**, *8*, 2087 (1986).

35. W. G. F. CORE and K-D. ZASTROW, "Efficient Modelling of Time Dependent Beam Target Interactions in NBI Heated JET Plasmas," JET-R(96)01, Joint European Torus (1996).

36. K.-D. ZASTROW et al., "Transfer Rates of Toroidal Angular Momentum during Neutral Beam Injection," *Nucl. Fusion*, **38**, 257 (1998).

37. W. MANDL et al., "Beam Emission Spectroscopy as a Comprehensive Plasma Diagnostic Tool," *Plasma Phys. Control. Fusion*, **35**, 1373 (1993).

38. M. G. VON HELLERMANN et al., "Recent Progress in Beam Emission and Charge Exchange Spectroscopy," *Proc. Int. Conf. Advanced Diagnostics for Magnetic and Inertial Fusion*, Varenna, Italy, September 3–7, 2001, P. STOTT et al., Eds., Plenum Publishers, New York (2002).

39. G. MATTHEWS et al., "Scaling Radiative Plasmas to ITER," *J. Nucl. Mater.*, **241–243**, 450 (1997).

40. N. SCHOON et al., Technical Report No. 92, Laboratory for Plasma Physics of the Belgian Royal Military Academy (1990).

41. H. S. BOSCH and G. M. HALE, *Nucl. Fusion*, **32**, 611 (1992).

42. O. N. JARVIS et al., *Fusion Technol.*, **20**, 265 (1991).

43. A. GIACOMELLI, "Advanced Neutron Diagnostics for JET and ITER Fusion Experiments," *Nucl. Fusion*, **45**, *1191*, 12019 (2005).

44. C. GIROUD et al., "Method for Experimental Determination of Z Dependence of Impurity Transport on JET," *Nucl. Fusion*, **47**, 313 (2007).

45. D. S. SIVIA, *Data Analysis—A Bayesian Tutorial*, Clarendon Press, Oxford (2006).

46. A. GELMAN et al., *Bayesian Data Analysis*, Chapman and Hall, London (2003).

47. R. FISCHER et al., "Thomson Scattering Analysis with the Bayesian Probability Theory," *Plasma Phys. Control. Fusion*, **44**, 1501 (2002).

48. M. KRYCHOWIAK et al., "Bayesian Analysis of the Effective Charge from Spectroscopic Bremsstrahlung Measurement in Fusion Plasmas," *J. Appl. Phys.*, **96**, 4784 (2004).

49. R. FISCHER et al., "Bayesian Modelling of Fusion Diagnostics," *Plasma Phys. Control. Fusion*, **45**, 1095 (2003).

50. J. SVENSSON et al., "An Integrated Data Analysis Mode for the W7-AS Stellarator," *Europhysics Conference Abstracts: 30th Conf. Controlled Fusion and Plasma Physics*, St. Petersburg, Russia, July 7–11, 2003, European Physical Society (2003).

51. J. SVENSSON et al., "Integrating Diagnostic Data Analysis for W7-AS Using Bayesian Graphical Models," *Rev. Sci. Instrum.*, **75**, 4219 (2004).

52. J. WESLE et al., "Operation and Control of ITER Plasmas," *Nucl. Fusion*, **40**, *3Y*, 485 (1998).

53. B. GREEN, *Plasma Phys. Control. Fusion*, **45**, 687 (2003).

54. J. WESLEY et al., "Plasma Control Requirements for ITER," *Fusion Technol.*, **32**, 495 (1997).

55. J. LISTER et al., *Nucl. Fusion*, **40**, *6*, 1167 (2000).

56. E. JOFFRIN et al., *Plasma Phys. Control. Fusion*, **45**, A445 (2003).

57. D. VAN HOUTTE et al., presented at 23rd Symp. Fusion Technology, Fondazione Cini, Venice, Italy, September 20–24, 2004.

58. T. OIKAWA and JT-60 TEAM, *Nucl. Fusion*, **40**, 1125 (2000).

59. D. HUMPHREYS et al., *Proc. 21st Fusion Energy Conf.*, Chengdu, China, October 16–21, 2006, IT-2.6, International Atomic Energy Agency (2006).

60. P. MOREAU et al., "A Magnetic Diagnostic on Tore Supra," *Rev. Sci. Instrum.*, **74**, *10*, 4324 (2003).

61. A. CENEDESE et al., *Fusion Eng. Des.*, **74**, *1–4*, 825 (2005).

62. P. SPUIG et al., *Fusion Eng. Des.*, **66–68**, 953 (2003).

63. R. ALBANESE et al., "Plasma Response Models for Current, Shape and Position Control in JET," *Fusion Eng. Des.*, **66–68**, 715 (2003).

64. R. ALBANESE et al., "Design, Implementation and Test of the XSC Extreme Shape Controller in JET," presented at 23rd Symp. Fusion Technology, Fondazione Cini, Venice, Italy, September 20–24, 2004.

65. J. B. LISTER et al., *Nucl. Fusion*, **42**, *6*, 708 (2002).

66. T. FUKUDA et al., "Active Feedback Control of Steady-State Improved Confinement Discharges in JT-60U," *Fusion Eng. Des.*, **46**, 337 (1999).

67. K. BOSAK et al., "Code for Real-Time Plasma Equilibrium Reconstruction," *Europhysics Conference Abstracts: 30th Conf. Controlled Fusion and Plasma Physics*, St. Petersburg, Russia, July 7–11, 2003, European Physical Society (2003).

68. J. FERRON et al., *Nucl. Fusion*, **40**, 1167 (2000).

69. D. GATES et al., *Nucl. Fusion*, **46**, 17 (2006).

70. T. TALA, *Nucl. Fusion*, **45**, *9*, 1027 (2005).

71. M. ZERBINI et al., *Proc. 12th Joint Workshop Electron Cyclotron Emission and Electron Cyclotron Resonance Heating*, Aix-en-Provence, France, May 13–16, 2002, World Scientific Publishing Company (2003).

72. P. HEESTERMAN et al., *Rev. Sci. Instrum.*, **74**, 1783 (2003).

73. D. ALVES et al., *Proc. 4th Technical Mtg. Control, Data Acquisition, and Remote Participation for Fusion Research*, San Diego, California July 21–23, 2003, International Atomic Energy Agency (2003).

74. D. MAZON et al., "Real-Time Tool Developments for Steady State Plasma Operation in Tore Supra," *Europhysics Conference Abstracts: 30th Conf. Controlled Fusion and Plasma Physics*, St. Petersburg, Russia, July 7–11, 2003, European Physical Society (2003).

75. J. FERRON et al., *Nucl. Fusion*, **46**, L13 (2006).

76. T. TAKENAGA et al., *Nucl. Fusion*, **47**, *10*, S668 (2007).

77. P. DUMORTIER, *Europhysics Conference Abstracts: 30th Conf. Controlled Fusion and Plasma Physics*, St. Petersburg, Russia, July 7–11, 2003, European Physical Society (2003).

78. P. MOREAU, "IR Imaging Analysis and Processing Tore Supra," *Proc. Conf. Decision and Control*, San Diego, California, December 13–15, 2006, Institute of Electrical and Electronics Engineers (2006).

79. R. PRATER et al., *Proc. 21st Fusion Energy Conf.*, Chengdu, China, October 16–21, 2006, EX-4.2, International Atomic Energy Agency (2006).

80. M. OKABAYASHI et al., *Phys. Plasma*, **9**, 1997 (2002).

81. Y. LIANG et al., "Active Control of Type-I Edge Localized Modes on JET," *Europhysics Conference Abstracts: 34th Conf. Controlled Fusion and Plasma Physics*, Warsaw, Poland, July 2–6, 2007 (to be published).

82. T. EVANS et al., *Nucl. Fusion*, **45**, *7*, 595 (2005).

83. P. LANG, *Nucl. Fusion*, **45**, 505 (2005).

84. M. K. ZEDDA et al., "Disruption Classification at JET with Neural Techniques," *Europhysics Conference Abstracts: 30th Conf. Controlled Fusion and Plasma Physics*, St. Petersburg, Russia, July 7–11, 2003, European Physical Society (2003).

85. ITER PHYSICS EXPERT GROUP ON CONFINEMENT AND TRANSPORT et al., "Chapter 2: Plasma Confinement and Transport," *Nucl. Fusion*, **39**, 2176 (1999); see also "Progress in the ITER Physics Basis," *Nucl. Fusion*, **47**, *6* (2007).

86. S. KAYE et al., *Nucl. Fusion*, **37**, 1303 (1997).

87. D. C. McDONALD et al., *Nucl. Fusion*, **47**, 147 (2007).

88. "ITER Physics Basis," *Nucl. Fusion*, **39**, 2403 (1999).

89. K. THOMSEN et al., in *Europhysics Conference Abstracts: 31st Conf. Controlled Fusion and Plasma Physics*, London, United Kingdom, June 28–July 2, 2004, Vol. 28G, P-5.173, European Physical Society (2004).

90. J. G. CORDEY et al., *Proc. 20th Fusion Energy Conf.*, Vilamoura, Portugal, November 1–6, 2004, IAEA-CN-116/IT-3.32, International Atomic Energy Agency (2006).

91. O. J. W. F. KARDAUN, *Plasma Phys. Control. Fusion*, **41**, 429 (1999).

92. W. A. FULLER, *Measurement Error Models*, Wiley-Interscience, New York (1987).

93. D. ELBEZE et al., *Europhysics Conference Abstracts: 32nd Conf. Controlled Fusion and Plasma Physics*, Tarragona, Spain, June 27–July 1, 2005, European Physical Society (2005).

94. A DINKLAGE et al., "Topics and Methods for Data Validation by Means of Bayesian Probability Theory," *Fusion Sci. Technol.*, **46**, 355 (2004).

95. G. V. PEREVERZEV and P. N. YUSHMANOV, "An Automatic System for Transport Analysis in a Tokamak," IPP Report 5/98, Institute for Plasma Physics (2002).

96. G. GENACCHI and A. TARONI, "JETTO: A Free Boundary Plasma Transport Code (Basic Version)," Report ENEA RT/TIB, ENEA (1988).

97. D. BOUCHER et al., "The International Multi-Tokamak Profile Database," *Nucl. Fusion*, **40**, 1955 (2000).

98. J. W. CONNOR and THE ITER CONFINEMENT DATABASE AND MODELLING GROUP, "Validation of 1-D Transport and Saw-Tooth Models for ITER," *Proc. 16th Int. Fusion Energy Conf.*, Montreal, Canada, October 7–11, 1996, Vol. 2, pp. 935–944, International Atomic Energy Agency (1997).

99. The International Multi-Tokamak Confinement Profile Database; available on the Internet at http://tokamak-profiledb.ukaea.org.uk (2006).

100. P. GOHIL et al., "Increased Understanding of the Dynamics and Transport in ITB Plasmas from Multi-Machine Comparisons," *Nucl. Fusion*, **43**, 708, (2003).

101. C. BOURDELLE et al., "Impact of the α Parameter on the Microstability of Internal Transport Barriers," *Nucl. Fusion*, **45**, 110 (2005).

102. J. E. KINSEY et al., "Transport Modelling and Gyrokinetic Analysis of Advanced High Performance Discharges," *Nucl. Fusion*, **45**, 450 (2005).

103. F. IMBEAUX et al., *Plasma Phys. Control. Fusion*, **47**, B179 (2005).

104. W. A. HOULBERG et al., "Integrated Modelling of the Current Profile in Steady-State and Hybrid ITER Scenarios," *Nucl. Fusion*, **45**, 1309 (2005).

105. A. J. H. DONNE and A. E. COSTLEY, "Key Issues in Diagnostics for Burning Plasma Experiments," *IEEE Trans. Plasma Sci.*, **32**, *1*, 177 (2004).

106. A. E. COSTLEY et al., "Technological Challenges of ITER Diagnostics," *Fusion Sci. Eng.*, **74**, 109 (2005).

107. "A Framework for IP Based Virtual Private Networks," The Internet Society; available on the Internet at http://tools.ietf.org/html/rfc2764 (2000).

108. "The Secure Shell (SSH) Protocol Architecture," The Internet Society; available on the Internet at http://tools.ietf.org/html/rfc4251 (2006).

109. "Recommendation X.509," International Telecommunications Union; available on the Internet at http://www.itu.int/rec/T-REC-X.509 (2007).

110. Citrix Systems homepage; available on the Internet at http://www.citrix.com (2007).

111. T. RICHARDSON et al., "Virtual Network Computing," *IEEE Internet Comput.*, **2**, *1*, 33 (1998); also see Real VNC homepage, available on the Internet at http://www.realvnc.com (2007).

112. J. A. STILLERMAN et al., *Rev. Sci. Instrum.*, **68**, 939 (1997).

113. "MDSplus Wiki"; available on the Internet at http://www.mdsplus.org (2007).

114. The Global Alliance homepage; available on the Internet at www.globus.org (2007).

115. W. SUTTROP et al., "Remote Participation at JET Task Force Work: User's Experience," *Fusion Eng. Des.*, **60**, 459 (2002).

116. "The PAPI AA Framework"; available on the Internet at http://papi.rediris.es (2007).

117. R. CASTRO, "PAPI Based Federation as a Test-Bed for a Common Security Infrastructure in EFDA Sites," *Proc. 6th Technical Mtg. Control, Data Acquisition and Remote Participation in Fusion Research*, Inuyama, Japan, June 4–8, 2007, International Atomic Energy Agency (2007).

118. "Recommendation H.323," International Telecommunications Union; available on the Internet at http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-H.323 (2007).

119. VRVS homepage; available on the Internet at http://www.vrvs.org (2007).

120. L. CHILDERS et al., "Access Grid: Immersive Group-to-Group Collaborative Visualization," *Proc. 4th Int. Immersive Projection Technology Workshop*, Ames, Iowa, June 19–20, 2000, Iowa State University (2000); also see AccessGrid homepage, available on the Internet at http://www.accessgrid.org (2007).

121. International Telecommunication Union homepage; available on the Internet at http:///www.itu.int (2007).

122. U. SCHWENN, "Status of Videoconferencing Standards," *Proc. 5th Technical Mtg. Control, Data Acquisition and Remote Participation for Fusion Research*, Budapest, Hungary, July 12–15, 2005, International Atomic Energy Agency (2005).

123. G. ABLA, "A Remote Control Room at DIII-D," *Proc. 6th Technical Mtg. Control, Data Acquisition and Remote Participation in Fusion Research*, Inuyama, Japan, June 4–8, 2007, International Atomic Energy Agency (2007).

124. EFDATV homepage; available on the Internet at http://efdatv.efda.org (2007).

125. http://jtv.jet.efda.org.

126. D. SCHISSEL et al., "Advances in Remote Participation for Fusion Experiments," *Fusion Eng. Des.*, **74**, 803 (2005).

127. D. SCHISSEL, "The Collaborative Tokamak Control Room," *Fusion Eng. Des.*, **81**, 2031 (2005).

128. J. VEGA et al., "Overview of the TJ-II Remote Participation System," *Fusion Eng. Des.*, **81**, 2045 (2005).

129. "NSTX Web Tools"; available on the Internet at http://nstx.pppl.gov/nstx/Software/WebTools (2007).

130. "jScope"; available on the Internet at http://www.mdsplus.org/mdsplus/jScope.html (2007).

131. E. EMOTO, "Development of Web Interfaces for Analysis Codes," *Proc. 6th Technical Mtg. Control, Data Acquisition and Remote Participation in Fusion Research*, Inuyama, Japan, June 4–8, 2007, International Atomic Energy Agency (2007).

132. FusionGrid homepage; available on the Internet at www.fusiongrid.org (2007).

133. I. FOSTER and C. KESSELMAN, *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publishers, Inc., San Francisco (1999).

134. LHC Computing Grid Project homepage; available on the Internet at http://lcg.web.cern.ch/LCG (2007).

135. EGEE homepage; available on the Internet at http://www.eu-egee.org (2007).

136. EGEE FUSION Virtual Organization homepage; available on the Internet at http://grid.bifi.unizar.es/egee/fusion-vo (2007).

137. J. W. FARTHING, "Twenty Years of Data Acquisition at JET," *Proc. 4th Technical Mtg. Control, Data Acquisition and Remote Participation in Fusion Research*, San Diego, California, July 21–23, 2003, International Atomic Energy Agency (2003).

138. J. FARTHING, "Control and Data Acquisition System Changes for the JET Enhanced Performance (JET-EP) Programme," *Proc. 5th Technical Mtg. Control, Data Acquisition and Remote Participation in Fusion Research*, Budapest, Hungary, July 12–15, 2005, International Atomic Energy Agency (2005).

139. J. R. BURRUSS, "Remote Computing Using the National Fusion Grid," *Fusion Eng. Des.*, **71**, 251 (2004).

140. J. R. BURRUSS, "Security on the U.S. Fusion Grid," *Fusion Eng. Des.*, **81**, 1949 (2005).

141. Integrated Tokamak Modelling Task Force homepage; available on the Internet at http://www.efda-taskforce-itm.org (2007).

142. B. GUILLERMINET, Personal Communication regarding A. BECOULET, US-JAPAN Workshop on Integrated Modelling (2005).

143. O. BARANA et al., *Fusion Eng. Des.*, **66–69**, 735 (2003).

144. M. OKABAYASHI et al., *Nucl. Fusion*, **45**, 1715 (2005).