EFDA–JET–CP(07)05/01

A. Murari, J. Vega, J.A. Alonso, E. De La Luna, J. Farthing, C. Hidalgo,
G.A. Rattà, J. Svensson, G. Vagliasindi and JET EFDA contributors

New Techniques and Technologies for Information Retrieval and Knowledge Extraction from Nuclear Fusion Massive Databases

New Techniques and Technologies for Information Retrieval and Knowledge Extraction from Nuclear Fusion Massive Databases

A.Murari [1], J.Vega[2], J.A.Alonso[2], E.De La Luna[2], J.Farthing[3], C.Hidalgo[2], G.A.Rattà[2], J.Svensson[4], G.Vagliasindi[5] and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]Consorzio RFX-Associazione EURATOM ENEA per la Fusione, I-35127 Padova, Italy.
[2]Asociación EURATOM-CIEMAT para Fusión, CIEMAT, Madrid, Spain
[3]EURATOM/UKAEA Fusion Association, Culham Science Centre, Abingdon, UK
[4]Max-Planck-Institut für Plasmaphysik, Teilinstitut Greifswald, EURATOM Association, Wendelsteinstr.1, 17491 Greifswald, Germany
[5]Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi-Università degli Studi di Catania, 95125 Catania, Italy
** See annex of M.L. Watkins et al, "Overview of JET Results ", (Proc. 21 st IAEA Fusion Energy Conference, Chengdu, China (2006)).*

Preprint of Paper to be submitted for publication in Proceedings of the
IEEE International Symposium on Intelligent Signal Processing,
Alcala de Henares, Madrid, Spain
(3rd October 2007 - 5th October 2007)

ABSTRACT

Reactor relevant experiments for Magnetic Confinement Fusion, like JET, produce already tens of GBytes of data per shot and the next step device, ITER, is expected to require orders of magnitude more. Managing such vast quantities of data in an efficient way needs new techniques, ranging from signal storage and information retrieval to data analysis for physical interpretation. At JET significant efforts are being devoted to all the main issues. Lossless data compression is under development for both mono and bi-dimensional signals, together with new techniques and technologies for image processing (directional wavelets and Cellular Non-linear Networks). Structural pattern recognition has shown great potential for information retrieval. Statistical methods, like Bayesian inference and regression trees, are being systematically investigated, to extract the required knowledge from all the available measurements. Other Soft Computing techniques, like Fuzzy Logic and Artificial Neural Networks, are very powerful tools to handle the great complexity and uncertainties of present day and near future experiments.

1. THE PROBLEM OF INFORMATION RETRIEVAL AND DATA ANALYSIS IN MAGNETIC CONFINEMENT FUSION

Increasing the energetic content of the plasma and the length of the pulses are essential ingredients in the quest for Magnetic Confinement Fusion as an alternative energy source but they certainly pose significant scientific and technological issues. Modern devices have become big, complex non-linear systems, which require the acquisition of thousands of signals, derived from many measurements systems called diagnostics, for both the operation of the machines and the interpretation of the experiments. In the last campaigns, it has already been demonstrated that JET Joint Undertaking (JET) diagnostics can produce more than 10 Gbytes of data per shot and the volume of information is bound to increase in the next years. Managing in an integrated way massive databases, of the dimensions required by reactor class devices, like the next step machine ITER, presents problems ranging from data storage and information retrieval to data validation, modeling and real time processing.

At the level of the **first processing and storage** (see section 2), various methods for data compression without loss of information are being systematically applied to nuclear fusion typical signals. Bayesian statistics is a key element in the Integrated Data Modeling programme, whose objective consists of making systematic and statistically coherent use of all the information available from a diagnostic set. This approach provides for the first time a sound method to include physical information about the diagnostics in the evaluation of the error bars and to combine constructively the measurements of the same parameter performed by independent physical measurements. With regard to the analysis, **information retrieval** (see section 3), the problem of extracting useful information from massive databases, is a major time consuming activity for many scientists. Since visual inspection is a routine activity in plasma physics, a “pattern oriented” approach to data analysis is appropriate. Fusion plasmas, in addition to being very complex, are also often affected

by significant uncertainties, so it can be very difficult to obtain the required “knowledge” from the available signals, even after the relevant information has been retrieved from the data base. To help in the direction of deriving physical information from the signals and to cope with the high level of uncertainty in the data, several “soft computing” methods for **data mining and physical analysis** (see section 4) are being pursued. Fuzzy Logic, Regression Trees and Artificial Neural Networks are among the most systematically investigated approaches.

Real time control (see section 5) very often requires specific solutions, given the additional requirements in terms of speed and automatic processing. In the quickly growing field of image processing, the innovative approach of Cellular Nonlinear Networks has already found several applications, in particular for the interpretation of infrared camera data for safety applications.

2. DATA STORAGE AND INTEGRATED SIGNAL PROCESSING

As mentioned in the introduction, present day experiments tend to produce substantial amount of data. There are many reasons behind this general trend. First of all, since continuous operation would be very desirable in a fusion reactor, the length of the discharges is being systematically increased. Approaching ignition parameters requires also a more complete set of measurements to guarantee the integrity of machine operation. The necessity to understand more sophisticated physical phenomena also pushes systematically in the direction of performing more measurements, with higher temporal and spatial resolution. On the other hand for physical studies in particular, any loss of information would be strongly undesirable. To preserve the signals, in Magnetic Fusion delta techniques are the most widespread compression methods. They are based on storing not the digital code of the measurement absolute value but the digital code corresponding to the difference between a temporal sample and the previous one (since the name of delta methods). Since the statistical distribution of the differences between subsequent samples is much more constrained and typically peaked around zero, an appropriate choice of symbols can reduce the stored volume significantly. The simplest code alphabet would therefore be to encode the value 0 with the shortest possible number of bits and then increase the encoding length of the symbols as the delta absolute value increases (since big differences between subsequent samples are highly unlikely in the vast majority of signals). More sophisticated implementations of this principle are reported extensively in [1] for mono-dimensional signals, but the main principle remains very similar to the one just described. The same approach is being now developed for images and, to give an idea of its potential, it has been proved that a reduction of about 90% of the data volume can be routinely achieved.

In addition to improvements in storage efficiency, a systematic revision of first data processing is also under way at JET. Advanced signal processing methods based on dynamic identification of the system, like recursive filtering, are under development, particularly in the field of real time. Following a general trend in science and industry, increasingly more attention is being devoted to image processing. In particular, to identify fast phenomena and instabilities, the use of wavelets has proven particularly useful (see fig.1). New directional wavelets, for example, have been used to

determine the plasma flow, as seen in the edge of JET plasmas with a completely new fast visible camera. Again in the field of image processing, in addition to the problem of reducing the amount of information, the speed of computation is also a significant issue, particularly for machine protection and real time applications [2]. In this perspective the innovative method of Cellular Nonlinear Networks has already found several applications, in particular for the interpretation of infrared camera data [3] (see section V).

Even if diagnostics have received a significant impulse in the last years, particularly at JET, thermonuclear plasmas, being delicate ionised gases at tens or hundreds of million degrees, remain difficult to access for measurement. Large nuclear fusion experiments therefore present a remarkable interpretation problem, since the internal parameters, like temperature and purity, must be derived from quantities, like radiation and escaping particles, detected outside the plasma. This leads to complex processes of inference, which sometimes have to rely on sophisticated physics models and complex inversion algorithms, to provide a good estimate of the physical quantities of interest from the available external measurements. In this framework, another aspect of the data processing, which is undergoing a fundamental methodological revision in the field of fusion, is the extraction of physical information from measurements and its validation. To associate error bars to the diagnostic signals and to derive better estimates of the physical parameters needed to analyse the experiments, the general approach of 'Bayesian Probability Theory' is proving very powerful. Based on Bayes theorem, it constitutes now a solid and coherent methodology [4] that can be applied to a series of different problems, ranging from error bar determination to theory falsification. Up to now at JET the most successful efforts have concentrated on the problem of treating, in a unified way, all uncertainties in the measurements, from the statistical fluctuations to the systematic errors and model imperfections. First promising, even if preliminary, results have been obtained in the determination of various quantities, which are measured with independent diagnostic techniques, like the electron temperature, the electron density and the plasma purity (Zeff). At a more fundamental level and on a longer time scale, it is planned to apply Bayesian statistics at JET to attack in a systematic way the more general issue of diagnostic integration. The basic idea would consist of including the signals of a comprehensive set of diagnostics in a global Bayesian estimator and from this reach a more comprehensive understanding of the internal state of fusion plasmas, as inferred directly from the measurements, with minimum recourse to hypotheses and models. This programme will require a reformulation of the inference problem for each individual diagnostic as a Bayesian inference on a common unknown physics 'state'. At the core of this strategy is a new inversion method, Bayesian current tomography [5], to determine the internal magnetic topology directly from the magnetic measurements. This is certainly the first step in the process since the configuration of the magnetic fields provides the natural coordinate systems in which to interpret the data of all the other measurements. An example of the reconstructed magnetic fields at JET using the signals of the magnetic pickup coils, without any hypothesis about the equilibrium, is shown in fig.2.

3. INFORMATION RETRIEVAL WITH STRUCTURAL PATTERN RECOGNITION

In Nuclear Fusion, visual inspection of the data is a very common practice. Scientists not only look at the acquired waveforms during the first screening of the signals, for example to identify the most interesting discharges, but can also make recourse to inspection by eye in various other phases of the data analysis process. Irrespective of the stage of the investigation, this activity consists of identifying the relevant patterns in the time evolution of mono or bi-dimensional signals, since these patterns can be linked to the physical phenomena under study. Up to now the data of fusion machines have been organised on the basis of pulse number and time evolution and therefore the search for the relevant patterns has become a quite time consuming and labour intensive activity, now that massive databases have to be traversed. In recent periods at JET and other devices, like TJ-II, the new approach of "structural pattern recognition" has been successfully tested. In general, the methodology consists of extracting from the signals the features more relevant to the typology of phenomena under investigation. By definition of a suited distance in the space of these features, it is then possible to classify the various signals in similarity classes and identify the ones more similar to a certain model. In fusion research up to now mainly the methods of 'structural' pattern recognition are the ones more investigated. At JET, for example, two alternatives have been tested, the first for entire signals, covering the whole discharge, and the other for substructure inside signals. To select entire waveforms [6], the feature extraction is performed by applying a wavelet transform (using the Haar function). This allows reducing the dimensionality of the problem to a computationally acceptable level without a significant loss of information. The signals are then classified using various supervised clustering criteria in the space of the transforms. To speed up the search process, the signals are organised in a tree structure and the search process is carried out only in the relevant class. The second development in the field of structural pattern recognition has been aimed at finding similar patterns inside waveforms (instead of looking for completely similar signals). This problem has been addressed adopting the technique of primitives, which are reference patterns used to codify the signals [7]. One of the general purpose methods implemented at JET defines a discrete series of gradients, associating to each gradient value a specific letter. The signals are then divided in suitable time intervals and stored as strings, in which each letter is determined by the gradient of the waveform in that interval. The quest for a similar pattern is then reduced to a much faster search for identical or similar strings (depending on the level of similarity appropriate for the problem under investigation). An example of a typical application at JET and the relative performance is reported in fig.3, which illustrates the potential of the technique. It must be emphasised that the process of building the data base is completely transparent to the user, who in the end does not need to perform any computation and can simply select the desired signals or part of them with the cursor of a visual interface. These structural pattern recognition methods are an important step in the programme of storing data according to technical and scientific criteria, instead of time intervals and pulse numbers. It is worth mentioning that the same pattern recognition methods are being applied also to 2D dimensional signals, particularly to the images of infrared and visible cameras.

4. DATA MINING AND SOFT COMPUTING FOR PHYSICS STUDIES

Magnetically confined fusion plasmas of reactor relevance are complex systems, driven far from equilibrium by a series of additional heating schemes, which inject tens of megawatts into the discharge. Since the plasma is a conducting fluid (an ionized gas), control is achieved via external electromagnetic fields, generated in systems of coils carrying kAs of current. The phenomena taking place inside the plasma, and the interactions between the ionized gas and the surrounding magnetic and mechanical boundaries, can have a strong non-linear character. In this framework it is sometimes the case that the interpretation of the physics is problematic because of the uncertainties either in the available measurements, in the integration of the information or in the physical models.

The statistical approaches, based on Bayesian inference, used to validate diagnostic information have been briefly summarised in section 2. New methods are also being implemented at JET for data mining, which in our context can be described as the process of extracting hidden information from massive databases. Since in fusion plasmas the same phenomenon can affect a very high number of signals, the investigation of correlations between various quantities is an important and daunting task. This activity can become particularly important in the field of machine protection, because the simultaneous behaviour of various signals can provide essential information to prevent dangerous situations. A typical example is the problem of disruptions, sudden and difficult to predict losses of confinement, which can be very harmful even for the structural component of the machines, like the vacuum vessel. Disruptions can be caused by a vast spectrum of different causes, ranging from unexpected plasma effects to anomalous behaviour of the control systems. On the other hand, if the approach of a dangerous situation can be predicted sufficiently in advance, at least about 100 ms before the actual disruption occurs, remedial action can be undertaken and at least the most harmful consequences avoided. Understanding which are the most important signals to monitor for disruption prediction is therefore of great importance for the safe operation, particularly of big devices like JET and ITER. Given the fact that thermonuclear plasmas and their interactions with the environment are strongly non-linear, this correlation analysis must be performed with adequate tools. In JET the approach of regression trees has been quite successful and has been implemented using the Classification and Regression Trees (CART) [8] software package. This tool consists of a non-parametric statistical method, based on a decision tree to solve classification and regression problems. In the process of constructing the tree and providing the required correlations, the algorithm traverses the whole database and checks all possible input variables and all their values. The aim is to find the best variable to split the parent node into two child nodes with the best “purity”. In the case of binary variables, like being disruptive or not, the best achievable purity of the child nodes is considered the partition which maximises the separation between disruptive and not disruptive discharges, i.e. the one which puts the maximum of disruptive discharges in one child node and the maximum of non disruptive discharges in the other. This process is applied recursively starting from the root and building a complete tree using all the variables available. Several criteria have been developed for determining the splits and for optimising the tree to find the best compromise between fitting accuracy and generalisation capability

(for a more detailed description see [8]). This approach has been used in JET to determine, on a sound correlation basis, the relative importance of the most relevant variables for disruption prediction. The relevance of the various quantities is reported in table I, showing a strong dependence from the time interval separating the moment of prediction from the disruption. This information, together with Fuzzy logic methods, is being used at the moment to design better disruption predictors for JET (see later).

The uncertainties in the physical phenomena, to be studied when investigating thermonuclear plasmas, are not limited to the error bars in the measurements, due to the diagnostic limitations, or to the degree of correlation between the various parameters, but they are of a more fundamental nature. In some cases the process are so complex and the relevant quantities so poorly known, due to the limited accessibility of the plasma for measuring purposes, that algorithmic solutions to the investigated problems are not available or would require excessive computational time to be determined with fully adequate physical models. To overcome these difficulties "soft computing" approaches are now quite common in the fusion community. Particularly popular are Artificial Neural Networks (ANNs), which have a quite solid mathematical basis and can be implemented nowadays with tested and validated software packages. In various machines ANNs have been used extensively in the field of disruption prediction, with the long-term programme to obtain a general predictor, trained over a multi-machine database, for extrapolation to ITER. At JET ANNs have also been used to attack a series of other problems, like the determination in real time of the total radiated power and the plasma internal inductance. On a more fundamental level, some evidence is emerging that the traditional Boolean logic, at the basis of western logic and philosophy, could not always be the best conceptual tool to investigate at least some of the phenomena of relevance in a fusion device. A case in point, to use an example already mentioned, is the one of the disruptions. The multiplicity of disruption causes and the blurred borders between the safe and dangerous regions of the operational space have suggested the investigation of "Fuzzy Logic" as a possible tool for both understanding and prediction [9]. "Fuzzy Logic" is a coherent logical edifice, founded not on crisp sets, like Boolean logic, but on fuzzy sets. The concept of fuzzy set constitutes an alternative to the traditional notion of set membership and logic, that hinges on the so-called "Law of the Excluded Middle", which states that an element X must either be in set A or in set not-A. On the other hand, in many practical applications it is often necessary to handle phenomena characterised by unsharp boundaries. The case of disruptions is representative, since the boundaries, between the regions of safe and unsafe behaviour in the operational space, are far from clear cut and the transition covers a grey area. Pioneering work on the application of these fuzzy methods to disruption prediction has been carried out at JET. Various predictors, with about ten signals as inputs and almost forty fuzzy rules, have been devised. The results are very competitive with other alternatives but much more transparent in terms of the logical operations than black box approaches like ANNs [10].

5. REAL TIME TECHNIQUES AND TECHNOLOGIES

With the increase of the machine dimensions and the energetic content of the plasma, the need for

efficient and reliable real time strategies has been keenly felt in the last years. Nowadays at JET feedback schemes are routinely in operation to guarantee the machine integrity, covering various aspects from the control of the magnetic field topology to the gas injection and the thermal loads on the first wall. More ambitious approaches are also being actively investigated, whereby difficult plasma parameters, like the current profile or the emitted radiation, are being controlled to improve the plasma performance. In ITER, the discharges being much longer and with higher energetic content, the availability of reliable feedback schemes will be even more crucial.

From the point of view of signal processing, real time applications present a series of specific issues. In particular, computational speed and reliability assume a higher relevance than accuracy with respect to inter-shot analysis. These requirements can become quite demanding in case of image processing, which is normally heavy in terms of computational resources, or when sophisticated pre-processing is required to derive the physical quantity of interest from a measured quantity. In the field of image processing for real time applications, the approach of "Cellular Nonlinear Networks" is being actively pursued and a specific project is in place at JET to test the viability of this solution in a reactor relevant environment. Cellular Nonlinear Networks are first of all mathematical objects, whose abstract definition was originally introduced in 1988 [11]. They can be described as 1, 2 or 3 dimensional infinite arrays of non linear dynamic units called cells. These cells are located at the nodes of the array and interact locally, via weighted connections, within a finite radius r, which is called the neighbourhood of the cell and can be varied depending on the application. Various mathematical models have been developed to compute the core cell value on the basis of its input and the status of the neighbourhood cells. In the case of the most popular models, libraries of functions, called templates, have been developed and are available for implementing the basic operations on the cells. These templates allow performing fundamental steps, like thresholding or interpolation. The mathematical paradigm of the CNNs has been successfully applied to a wide spectrum of computational applications, ranging from differential equation solutions to modelling of complex biological systems. In the perspective of image processing, the most attractive features of CNNs are their local connectivity and the repetition of identical cells, which are particularly suited to implementation with VSLI technologies. Moreover, the inherent topological architecture of CNNs, based on an analog local core that permits very fast exchange of information between the cells, allows parallel processing of the cells, a significant advantage in real time applications when speed is at a premium. On the other hand, the analog nature of the original architecture does not constitute a serious drawback since it can be combined with digital components in the so called CNN-Universal Machine (CNN-UM). This solution, which provides the capability to program sequential operations and to store the intermediate results, implements a computational paradigm, which is as universal as a Turing Machine [12]. The hardware implementation adopted for the applications described in this paper is the ACE16K [13], the third generation of CNN chips, which consists of two layers of CMOS components. The first is an array of 128X 128 identical analog cells, designed to act as photo-receptors. The second layer contains

the analog processing unit (implementing the CNN processing paradigm) and a series of digital processing units (for storage, programming and other functions). The images can be acquired directly by the device or can be read by means of a 32-bit bidirectional data bus (this second alternative is the feature used in the applications described in this paper). The chip can perform up to 330 GOPS (Giga operations per second). In modern fusion devices, infrared (IR) thermography has become a common tool for surveillance and machine protection [2]. The first wall of the machines is monitored with very sophisticated IR cameras, to make sure that the plasma wall interactions remain limited to the appropriate locations and do not cause excessive overheating of the surfaces. The timely detection of hot spots, regions of the first wall where the temperature approaches dangerous levels, can therefore be a crucial factor in big devices and is expected to play a fundamental role in ITER. Since preserving the integrity of the plasma facing components will be one of the main issues presented also by the operation of JET with the future new Be wall, the CNN technology has been applied to the real time identification of hot spots. The images of JET wide angle infrared camera were used as input to the CNN chip. To cover the varied typologies of interesting phenomena, which can give rise to IR emission, two different types of algorithms have been devised. The first one, for the so called static detection, performs the analysis of a single frame at the time and is more suited to monitor fixed parts of the machine, like limiters and divertor. A second approach, called dynamic detection, is differential and concentrate on the difference between subsequent frames. An example of the results obtained during a typical JET discharge is shown in fig.4. The execution time is about 120 ms to process the whole 384x384 frame of the IR wide angle camera. On the other hand, for each single 128x128 subframe, the maximum unit that the present ACE16K chip can process, the execution time is of the order of 20 ms for both the static and the dynamic algorithm. These performances are considered more than adequate for the monitoring of thermal events, which evolve on much slower time scales.

CONCLUSIONS AND OUTLOOK

Present day fusion devices generate considerable amounts of data per shot and therefore their databases reach massive dimensions and cannot be managed with usual, simple approaches. Moreover, the conceptual main objective of fusion experiments remains the formulation of adequate models to understand the physics of thermonuclear plasmas. Therefore efficient ways have to be found not only to store the diagnostic signals but also to easily retrieve the required information and to reformulate it in adequate mathematical form. In addition to this, more sophisticated feedback control algorithms are being developed to guarantee safe operation and to achieve better performance. It is worth pointing out that all these issues will become significantly more severe in the next generation of devices, like ITER. In present day machines, and particularly in JET, a series of new tools are being developed to address practically all the aspects of managing massive databases. In the near future particular progress is expected in the field of Bayesian statistics, which should allow a much more integrated analysis of the diagnostics measurements. Its use should also become more

extensive for model selection and theory falsification. In addition to further progress in pattern recognition methods, particularly for images, finding hidden or not obvious correlations between signals will probably emerge as a very important application of data mining techniques. To support the formulation of physical models starting directly from the data, soft computing approaches, from Fuzzy Logic to Bayesian statistics, seem particularly promising and their application should become more widespread in the future.

REFERENCES

[1]. J.Vega et al "Encoding technique for high data compaction in data bases of fusion devices" Rev. Sci. Instrum., Vol. **67**, No. 12, December 1996

[2]. J.M. Travere et al "Advance Image Analysis of Infrared Quantitative Data Applied to Tokamak Survey and Safety: present and future" this conference

[3]. G. Vagliasindi et al "Application of Cellular Neural Network Methods to real time image analysis in plasma fusion" this conference

[4]. P. Gregory "Bayesian Logical Data Analysis for the Physical Sciences" Cambridge University Press

[5]. J. Svensson et al "Bayesian Data Analysis in Fusion Research" this conference

[6]. J. Vega et al "Data mining technique for fast retrieval of similar waveforms in Fusion Massive databases" to be published in Fusion Eng and Design

[7]. S. Dormido-Canto et al "Search and retrieval of plasma wave forms: Structural pattern recognition approach" Rev. Sci. Instrum., **77**, 10F514 (2006)

[8]. CART Tree-structured data Analysis www.salford-systems.com

[9]. L.A. Zadeh, "Fuzzy sets", Infor. Control, **8**, pp.338-353,1965

[10]. G. Vagliasindi et al. "Fuzzy Logic Approach to Disruption prediction at JET" to be published in IEEE Trans. on Plasma Science

[11]. L.O. Chua and L. Yang, "Cellular neural networks: Theory" IEEE Trans. Circuits Syst., vol. **35**, no. 10, pp. 1257–1272, Oct. 1988

[12]. K.R. Crounse and L.O. Chua, "The CNN Universal Machine is as Universal as a Turing Machine", IEEE Trans. On Circuits and Systems–I, vol. **43**, no. 4, April 1996

[13]. A. Rodríguez-V·zquez, el al. ,"ACE16K: The Third Generation of Mixed-Signal SIMD-CNN ACE Chip Toward VSoCs", IEEE Transaction on Circuits and Systems-I, vol.**51**, no.5, May 2004

[td – 440, td – 100]		[td – 200, td – 100]		[td – 320, td – 220]		[td – 440, td – 340]	
Variable	Importance	Variable	Importance	Variable	Importance	Variable	Importance
dW_{dia}/dt	100.0	dW_{dia}/dt	100.0	dl_i/dt	100.0	Ipla	100.0
dl_i/dt	82.88	dl_i/dt	50.46	dW_{dia}/dt	99.20	dl_i/dt	69.42
Ipla	70.88	$d\beta_p/dt$	37.01	Ipla	79.40	P_{net}	68.23
P_{net}	67.96	P_{net}	35.03	l_i	70.47	l_i	54.78
dq_{95}/dt	54.90	q_{95}	27.54	q_{95}	62.06	dq_{95}/dt	53.92
Dens	53.24	β_p	24.80	P_{net}	59.00	P_{inp}	49.58
$d\beta_p/dt$	52.96	l_i	24.37	β_p	57.05	dW_{dia}/dt	38.78
l_i	52.36	Loca	22.74	dq_{95}/dt	56.19	q_{95}	37.97
Loca	46.75	P_{inp}	20.49	$d\beta_p/dt$	38.68	Loca	37.39
P_{inp}	45.26	Ipla	16.36	P_{inp}	38.41	$d\beta_p/dt$	36.57
β_p	43.47	Dens	13.34	Dens	37.91	Dens	35.55
q_{95}	33.55	dq_{95}/dt	9.87	Loca	30.13	β_p	28.77

Table I: The relevance of various quantities for disruption prediction as calculated by CART. The columns refer to different time intervals from the actual occurrence of the disruption (as indicated in the first row).

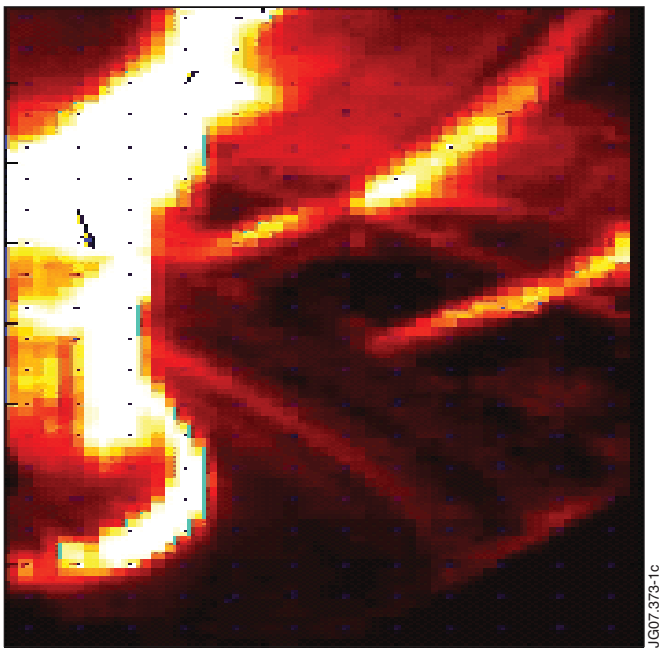


Figure1: Image of JET fast visible camera. The plasma flow, indicated by the arrows, has been obtained applying directional wavelets to the structures present in the frame.

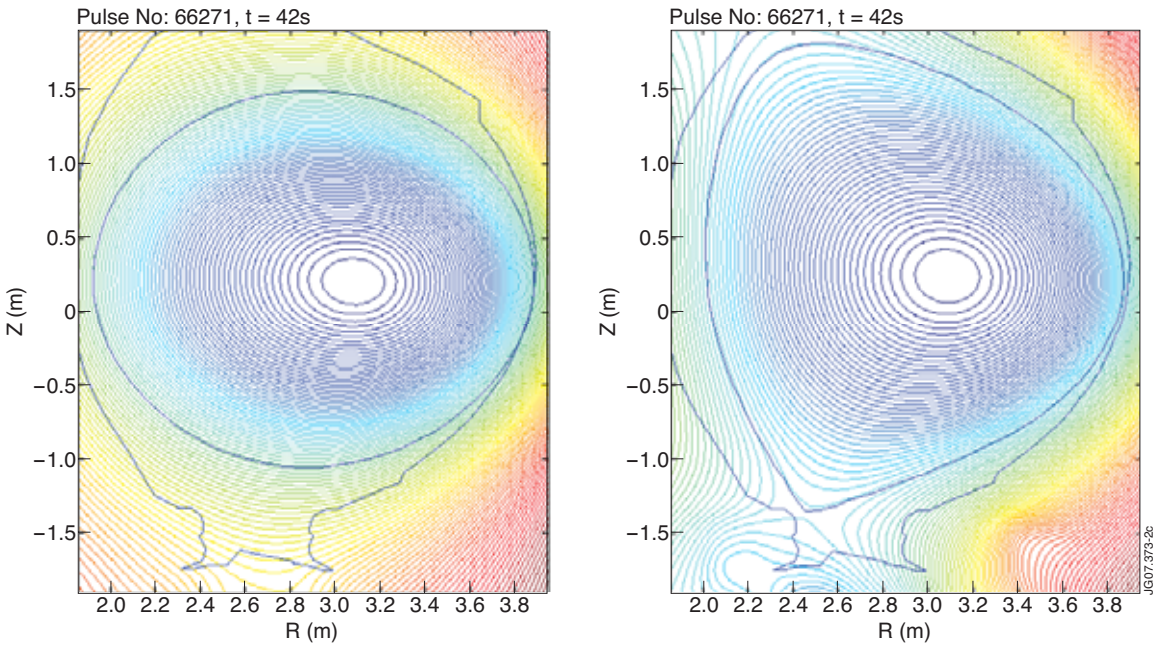
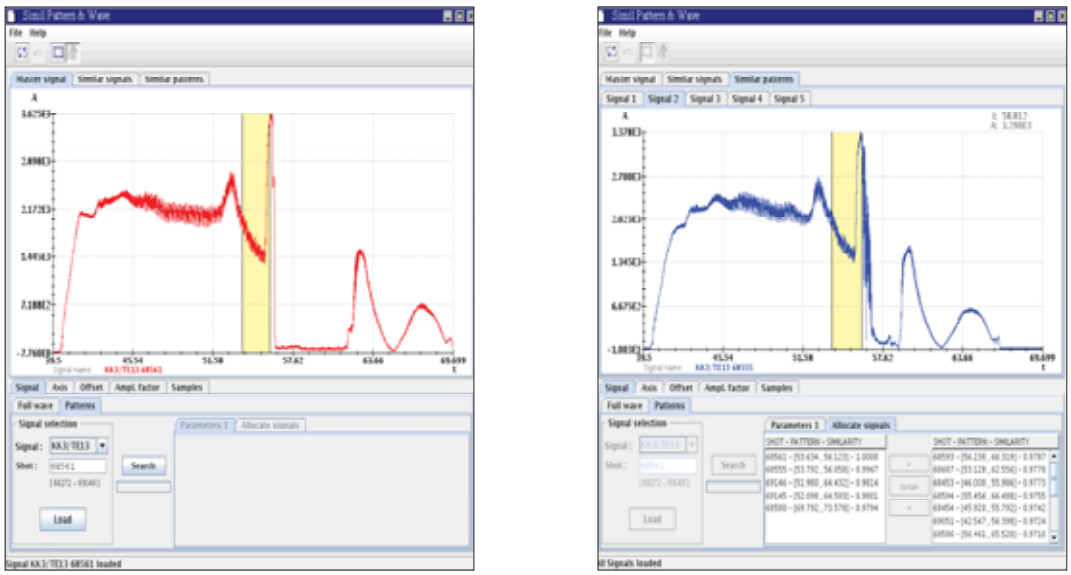


Figure 2: Example of magnetic topology determined with the Bayesian current tomography without any recourse to equilibrium hypotheses. The left end figure shows a limiter configuration and the right end one an X-point configuration.



Input: a pattern inside a waveform

Output: similar patterns ordered by similarity

	Shots (C17 campaign)	Storage in PPF (MBytes)	Additional storage (MBytes)	Search CPU time (JAC cluster)
EFIT/WDIA	706	1.23	0.83	300 ms
KK3/TE13	501	48.93	0.72	90 ms

JG07.373-3c

Figure 3: Example of pattern recognition using primitives applied to JET radiometer for Electron Cyclotron Radiation detection.

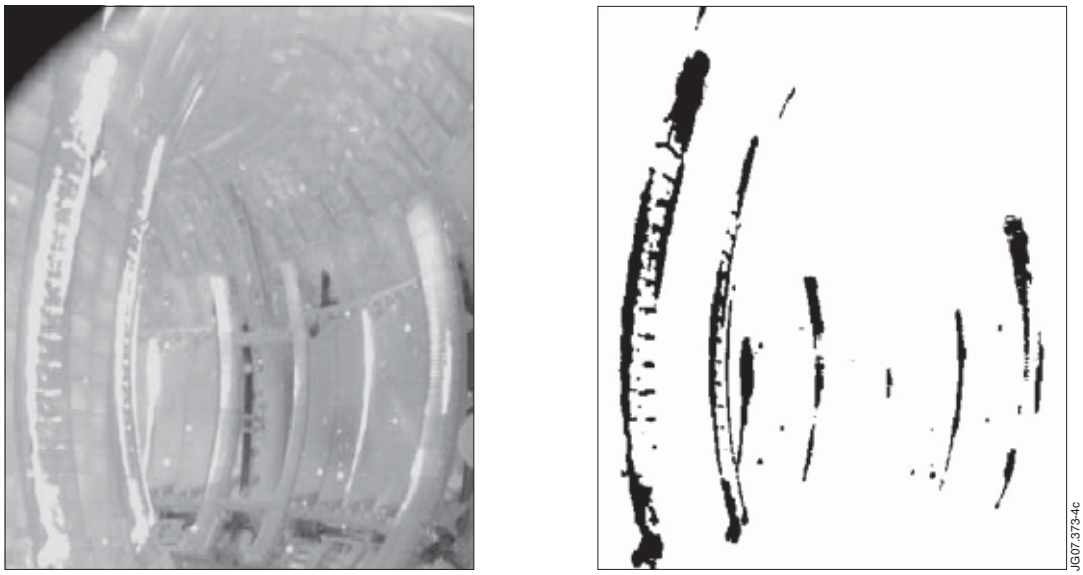


Figure 4: In the picture on the left an example of the different kind of hot-spots at JET. On the right the result of the static detection of hot-spots via CNN-based algorithms. The hottest point in the region of interest taken into account, limiter and RF antennas, are shown in black.