

A. Murari, J.Vega, G.De Arcas G.Vagliasindi
and JET EFDA Contributors*

New Information Processing Methods for Control in Magnetically Confinement Nuclear Fusion

“This document is intended for publication in the open literature. It is made available on the understanding that it may not be further circulated and extracts or references may not be published prior to publication of the original when applicable, or without the consent of the Publications Officer, EFDA, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK.”

“Enquiries about Copyright and reproduction should be addressed to the Publications Officer, EFDA, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK.”

New Information Processing Methods for Control in Magnetically Confinement Nuclear Fusion

A. Murari¹, J.Vega², G.De Arcas³, G.Vagliasindi⁴
and JET EFDA Contributors*

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

¹*Consorzio RFX – Associazione EURATOM ENEA per la Fusione, Corso Stati Uniti 4 Padova Italy*

²*Asociación EURATOM/CIEMAT para Fusión. Avda. Complutense, 22. 28040 Madrid, Spain*

³*Grupo de Investigación en Instrumentación y Acústica Aplicada. Universidad Politécnica de Madrid, Ctra. Valencia Km 7, 28031, Madrid, Spain*

⁴*Università degli Studi di Catania - Dip. di Ingegneria Elettrica, Elettronica e dei Sistemi, v.le A. Doria 6, 95125, Catania, Italy*

* See annex of M.L. Watkins et al, "Overview of JET Results",
(Proc. 21st IAEA Fusion Energy Conference, Chengdu, China (2006)).

Preprint of Paper to be submitted for publication in Proceedings of the
8th International FLINS Conference on Computer Intelligence in Decision and Control, Madrid, Spain.
(21st September 2008 - 24th September 2008)

ABSTRACT

Thermonuclear plasmas are complex and highly non-linear physical objects and therefore, in the most advanced present day devices for the study of magnetic confinement fusion, thousands of signals have to be acquired for each experiment, in order to progress with the understanding indispensable for the final reactor. On the other hand, the resulting massive databases, more than 40Tbytes in the case of the JET joint Undertaking, pose significant problems. In this paper, solutions to reduce the sheer amount of data by different compression techniques and adaptive sampling frequency architectures are presented. As an example of methods capable of providing significant help in the data analysis and real time control, a Classification and Regression Tree software is applied to the problem of regime identification, to discriminate in an automatic way whether the plasma is in the L or H confinement mode.

1. INTRODUCTION

All open living systems, from unicellular beings to complex societies, have to preserve their internal organization by processing matter and energy in order to delay the inevitable heat death. To survive they need to control the extraction of matter and its modifications, its transformation into energy and the elimination of the undesired by-products: waste and entropy. Their internal complex structure, indispensable to survival, can be successfully maintained only thanks to sophisticated information processing and carefully programmed interventions on the surrounding physical world. The higher the level of complexity of the planned tasks and/or the faster the rhythm of energy conversion, the more massive becomes the amount of information to be processed. These aspects are at the base of the dramatic surge of information technologies witnessed by the industrial revolution (increase in the speed of material processing) and the present communication age (increase in the complexity of society).

Magnetic confinement fusion devices are complex, non linear open systems, which have to be driven well away from equilibrium to produce energy in an economically competitive way. They therefore present all the issues typically faced by open, self sustaining systems, from fuelling to exhaust, from energy conversion to contamination and sudden degradation. In the last years, the need to improve performance has induced the scientific community to increase the amount of energy input into thermonuclear plasmas. Moreover, again to progress towards an economically viable reactor, new and more complex tasks have to be performed in real time. In order to control thermonuclear plasmas of reactor relevance, therefore, thousands of signals, derived from many measurements systems called diagnostics, are required. During the last campaigns, it has already been demonstrated that the 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 generation of devices like ITER. JET database contains already more than 40 Tbytes of data, which need to be analyzed with new methods if a sufficient level of understanding and control has to be achieved. Managing massive databases of these dimensions poses problems ranging from signal acquisition and data storage to validation and real time processing.

At the level of **acquisition and storage**, new methods are being developed to reduce the quantity of data but at the same time assuring the quality of the stored information (see section 2). Very efficient lossless compression algorithms can reduce significantly the amount of information to be stored and made available through the real time networks. Adaptive acquisition systems have been developed to optimise the sampling frequency depending on the nature of the phenomena to be investigated.

Since the traditional identification techniques, used in other fields to determine dynamical models of the systems under study, are not easily applicable, several “**soft computing**” techniques are being implemented. To help in the direction of deriving physical information from the signals and to cope with the high level of uncertainty in the data, various data mining and unsupervised learning methods (Classification and Regression Trees, unsupervised Clustering etc) are used to analyse the databases in an unbiased and systematic ways. As an example of these techniques, the potential of a plasma regime identifier based on Classification and Regression Trees (CART) is illustrated in section three for a series of JET experiments.

2. DATA ACQUISITION, DATA STORAGE AND FIRST SIGNAL PROCESSING FOR REAL TIME

The sheer amount of data to be collected in every discharge of present reactor relevant fusion devices has reached several Gbytes and therefore adequate measures have to be taken to make sure that useless signals are not stored. On the other hand, since fusion devices are still experimental machines, it is difficult to tell a priori which part of the signals can be discarded and which one contains the relevant information. To assess the basic quality of the measurements and perform a first real time screening and reduction of the available information, a new adaptive acquisition system based in the ITMS platform [1] is being implemented at JET, which allows various real time activities such as adjusting the sampling rate to the bandwidth of the signals. In general Data Acquisition (DAQ) systems are used in many scientific applications as passive systems to acquire large amounts of data during an experiment for later offline data analysis and processing. The problem is that not all the data contains relevant information, but this cannot be detected previously because the system is based on a post-processing architecture. By adopting an advanced intelligent instrument approach [2], the information content of the signals can be determined during the acquisition, thus optimizing the overall system performance. One possible strategy consists of data acquisition systems with self adaptive sampling rate capabilities, which allow reducing the amount of data acquired without losing information by selecting the sample rate of the acquisition depending on the amount of information carried by the acquired signal. This can be accomplished by using commercially available DAQ boards and software decimation as shown in Figure 1.

The input signals are acquired continuously at the expected maximum sample rate by means of a standard data acquisition board and transferred to a memory buffer in the host processor. Data is then read from the memory buffer in blocks and for each block an intelligent decimation algorithm (IDA) is applied. The decimation algorithm has to establish the signal bandwidth for each block in order to choose the optimum sample rate for that block, and from there the decimation factor to be

used, thus the term “intelligent” decimation. The first step to determine the signal’s bandwidth is to obtain its spectra using the Fast Fourier Transform (FFT). The problem now is how to determine the signal bandwidth. In situations where the spectra can have different shapes the implementation of such an algorithm is not straightforward. If these shapes are known beforehand, it is possible to implement an IDA based on classification methods following a scheme such as the one shown in Figure 2. In this application, typical of JET correlation reflectometry, there are three possible types of spectra: in case A the acquired signal can be decimated to optimize the acquisition since the relevant information is only in the lower frequencies; in case B no decimation can be done as there is information in the whole frequency range; and in case C there is no information in the data, so it can be fully discarded.

Another alternative, to reduce the burden on the storage and communication systems, is based on lossless data compression algorithms. Delta techniques are the most widespread compression methods in fusion today since in general the sampling rate is much shorter than the time constant of the signals. These methods therefore store not the digital code of the measurement absolute value but an integer corresponding to the difference between a temporal sample and the previous one (from this process comes the name of delta methods). Since the statistical distribution of the difference between subsequent samples of a signal is typically peaked around zero, an appropriate choice of symbols can reduce the stored volume significantly. For example the shortest code alphabet can be used to encode the value 0 and then progressively longer symbols can be adopted as the delta absolute value increases (since bigger differences between subsequent samples are increasingly unlikely in the vast majority of cases). More sophisticated implementations of this principle are reported in detail in [3] for mono-dimensional signals, but the main principle remains very similar to the one just described and reduction of 79/80 % of the data are commonly achieved. The same approach is being now developed for images and it has been proved that a reduction of about 90% of the data volume can be obtained for both visible and infrared cameras at JET.

3. “SOFT COMPUTING” TECHNIQUES FOR CONTROL: CLASSIFICATION AND REGRESSION TREES

Fusion plasmas are very complex objects difficult to control with traditional methods. Adequate mathematical formulation of energy transport, particle transport and various instabilities are not available and therefore model-based approaches are impossible to apply in a direct way. On the other hand, identification using input-output data, gathered in open loop perturbative experiments, has proved to be very time consuming. Moreover, the machine conditions are not completely under control and therefore minor changes in the plasma parameters, like the impurity content, can very often jeopardise the previous identification and require additional open loop experiments. These difficulties have motivated the investigation of alternative soft computing approaches to attack difficult problems for which algorithmic solutions and traditional identification techniques are not practical. For brevity sake, in the following the analysis will focus on a regime identification method based on Classification and Regression Trees (CART). In today thermonuclear fusion devices it is

possible to distinguish distinct types of plasma confinement regimes, which present different performance in terms of confinement time. Among them, the so-called H-mode discovered in ASDEX (Axially Symmetric Divertor EXperiment) Upgrade [4] is particularly relevant, because its energy confinement time can be about a 100% higher than the normal L-regime. Discriminating among these two operation modes could therefore represent a useful feature for an efficient control of an experiment. On the other hand present day control schemes are static, i.e. the plasma is expected to reach a certain confinement state at a certain time “a priori”, on the basis of the pre-programmed discharge parameters. The real time control strategies are therefore also based on the status assumed to be reached by the plasma at the time of the feedback. Since this can result in far from optimal control strategies, various methods, from Support Vector Machines to Fuzzy Logic, have been studied to determine in real time the plasma confinement regime. In this paper the results of Classification and Regression Trees (CART) for the identification of confinement regimes in JET are reported. The main strength of this approach is its non biased and nonlinear character. The algorithm traverses the whole database and builds a tree, using the values of the available signals which produce the best separation of the discharges in the L and H mode categories. Initially a total of 15 signals were selected among the data available in real time from JET’s diagnostics and provided as inputs to build a classification tree. Five of these signals proved to have the highest information content and were retained. They are the normalised beta, the density and the Fourier transform of three Da lines (three different lines of sight in the divertor). Using these signals the tree produced by CART was also used for the classification of an independent set of discharges. The success rate is very high as shown in table I. The number of correct and wrong classifications is reported in the last two columns of the table. The percentage of correct identifications for both the L and H mode regimes is also shown. The results are quite encouraging. They are also robust to noise on the acquired signals. A more detailed description of this approach to regime identification can be found in [5].

REFERENCES

- [1]. E. Barrera et al. “PXI-based Architecture for Real Time Data Acquisition and Distributed Dynamic Data Processing” IEEE TNS 53-3 (2006) 923-926
- [2]. M. Ruiz et al. “Distributed real time data processing architecture for the TJ-II data acquisition system” Rev. Sci. Instr. **75**-10 (2004) 4261-4264.
- [3]. 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.
- [4]. F. Wagner et al. “Regime of improved confinement and high beta in neutral beam heated divertor discharges in the Asdex tokamak” Physics Review Letters **49**, 1408 (1982).
- [5]. G. Vagliasindi et al this Conference

Actual Class	Total Time Slices	% Correct	L	H
L	3.599	88.22	3.175	424
H	7.811	97.80	172	7.639
Total:	11.410,00			
% Correct:		94.78		

Table I: Automatic classification of the confinement regime with CART.

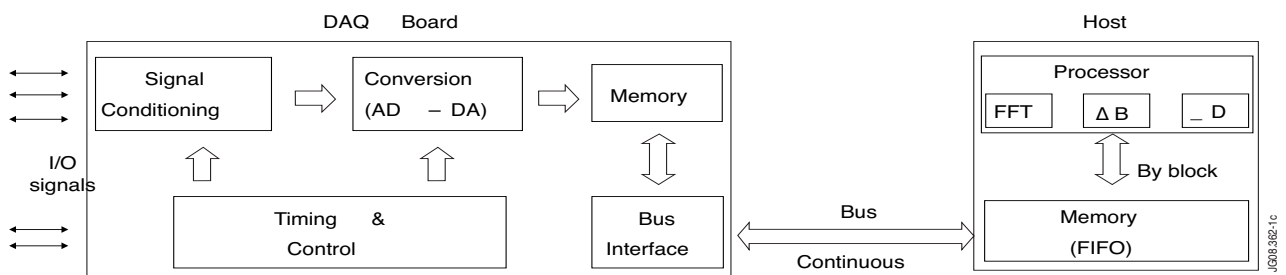


Figure 1: Hardware and software architecture of a data acquisition system with self-adaptive sampling rate capabilities.

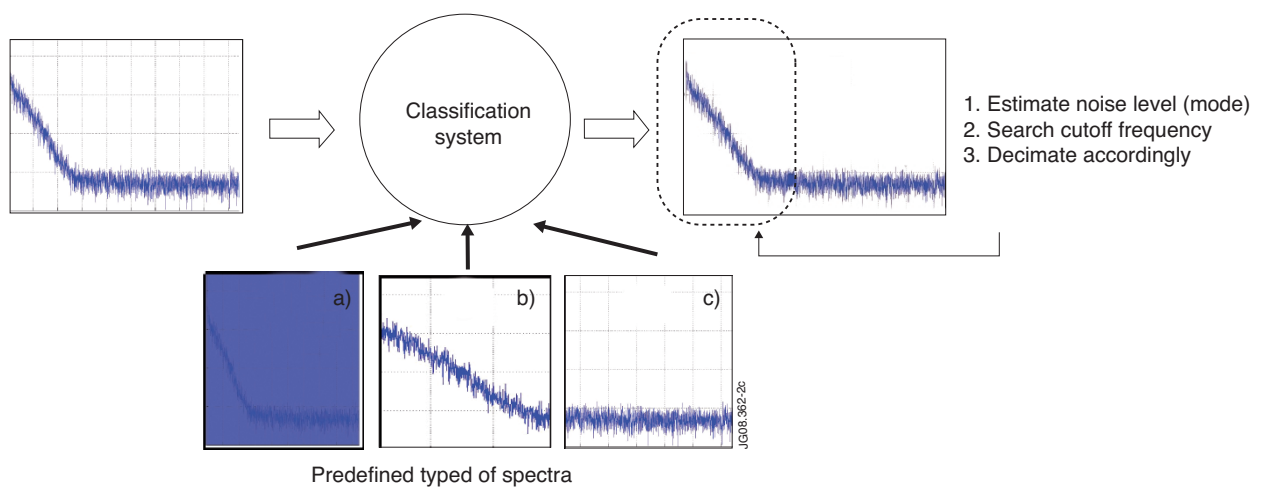


Figure 2: Intelligent decimation algorithm based on the classification of the spectra.